



Universidade de Brasília
Departamento de Estatística

Influência das comorbidades na qualidade do sono dos residentes de Ouro Preto durante a pandemia da Covid-19: uma aplicação de análise estatística de mediação

André Fonsêca Ferreira

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2025

André Fonsêca Ferreira

Influência das comorbidades na qualidade do sono dos residentes de Ouro Preto durante a pandemia da Covid-19: uma aplicação de análise estatística de mediação

Orientador: Frederico Machado Almeida

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2025**

Agradecimentos

Agradeço, primeiramente, a Deus, criador de toda a vida e ciência. Àquele que sustentou a mim e a minha família desde sempre, dedico esse trabalho de conclusão de curso.

Agradeço à minha mãe, Ana Raquel, pela dedicação, carinho e abnegação, sem os quais eu não teria conquistado nada do que alcancei. À minha irmã, Talita, uma das minhas maiores incentivadoras e apoiadoras, que, por meio de palavras e ações, me deu forças para chegar até aqui. Ao meu pai, Anivaldo (in memoriam), que sempre me incentivou a estudar e batalhar pelos meus sonhos. À minha tia Eurípia que me apoiou toda vez que precisei, e cujo exemplo de sucesso e esforço guardarei para sempre.

Agradeço à minha noiva, Giovanna, que esteve comigo em diversos momentos da graduação, sempre me incentivando a persistir, e à sua família — Claiton, Janayna, Eduarda e Jonathan — que demonstraram afeto e cuidado para comigo, torcendo e orando pelo meu sucesso.

Aos meus irmãos da Primeira Igreja Presbiteriana de Taguatinga, os pastores Adelson, Adriano e Rafael, que, direta ou indiretamente, me incentivaram a persistir na graduação. Ao Paulo, Mariana e Mateus Atique, exemplos de que é possível se formar na UnB.

Aos meus amigos de faculdade, Pedro, Luciano, Victor, Iago, Luiz e João Pedro, que estudaram comigo para diversas provas e trabalhos.

Ao meu orientador, Professor Doutor Frederico Machado Almeida, pela paciência e dedicação na minha orientação.

Resumo

Em muitas situações da vida real somos confrontados com problemas em que o principal objetivo é estudar a relação entre as variáveis. Os métodos comumente utilizados na literatura estatística pressupõem que o efeito da variável explicada (ou dependente), é particionado em duas partes: uma que é devido a parte determinística do modelo (efeito da(s) variável(is) independente(s)), e outra devido a uma série de fatores não controlados pelo experimentador. Entretanto, em determinadas situações surge a necessidade de avaliar a magnitude da influência de determinados fatores na relação entre as variáveis. Desta forma, no presente trabalho estudamos a Análise de Mediação, uma abordagem estatística que nos permite compreender o quanto de uma relação entre uma (ou várias) características independentes na variável dependente, pode ser influenciada pela inclusão de uma ou mais variáveis intermediárias (mediadoras) na relação de causa e efeito. Diferentemente de uma relação puramente direta entre as variáveis, a mediação revela que parte do efeito do fator de exposição sobre a variável resposta ocorre de forma indireta, por meio da variável mediadora. Neste trabalho, essa técnica é aplicada a um conjunto de dados que busca explicar a relação entre comorbidades e qualidade do sono, considerando os sintomas de ansiedade como variável mediadora, no contexto da pandemia de COVID-19. O objetivo principal é, primeiramente, apresentar a análise de mediação e seus desdobramentos teóricos para, em seguida, aplicá-la ao contexto estudado e interpretar os resultados obtidos.

Palavras-chave: Análise de mediação, variável mediadora, comorbidades, qualidade do sono, ansiedade, COVID-19, KHB.

Abstract

In many real-life situations, we are faced with problems where the general objective is to study the relationship between variables. The methods commonly used in the statistical literature assume that the effect of the explained (or dependent) variable is divided into two parts: One due to the deterministic part of the model (effect of the independent variable(s)), and the other due to a set of factors not controlled by the experimenter. However, in certain situations, it is necessary to estimate the extent of the influence of certain factors on the relationship between variables. Therefore, in this present work, we study the Mediation Analysis, a statistical approach that allows us to understand how the relationship between one (or more) independent factors in the dependent variable may be modified by introducing one or more intermediated covariates (mediators covariates) in the causal relation. Unlike a purely direct relationship between the involved covariates, mediation reveals that part of the effect of the exposure factor on the response variable occurs indirectly, through the mediating factor. In this work, this technique is applied to a dataset aiming to explain the relationship between comorbidities and sleep quality, considering anxiety symptoms as a mediating variable in the context of the COVID-19 pandemic. The objective is to introduce the theoretical foundations of mediation analysis, explore its developments, and apply it in practice, analyzing the obtained results.

Keywords: Mediation analysis, mediator variable, non-communicable chronic diseases, sleep quality, anxiety, COVID-19, KHB method.

Lista de Tabelas

1	Modelo reduzido da 1ª aplicação (resultados)	25
2	Modelo de mediação da 1ª aplicação (resultados)	25
3	Modelo completo da 1ª aplicação (resultados).	26
4	Modelo reduzido da 2ª aplicação (resultados).	28
5	Modelo de mediação da 2ª aplicação (resultados)	28
6	Modelo de mediação da 2ª aplicação (resultados)	29
7	Modelo completo da 2ª aplicação (resultados)	29
8	Tabela de contingência do gênero e qualidade do sono	31
9	Tabela de contingência dos sintomas de ansiedade e qualidade do sono	32
10	Tabela do número de comorbidades e qualidade do sono	34
11	Modelo reduzido da 3ª aplicação (resultados)	34
12	Modelo de mediação da 3ª aplicação (resultados)	35
13	Modelo completo da 3ª aplicação (resultados)	35
14	Modelo reduzido da 4ª aplicação (resultados)	37
15	Modelo de mediação da 4ª aplicação (resultados)	37
16	Modelo reduzido da 4ª aplicação (resultados)	37
17	Dados 1ª aplicação.	42
18	Dados 2ª aplicação.	43

Lista de Figuras

1	Diagrama do efeito total.	16
2	Diagrama conceitual de um modelo de mediação simples.	17
3	Diagrama de mediação múltipla com k mediadores.	20
4	Fluxograma de análise de mediação.	24
5	Coeficientes do modelo com 1 mediador.	26
6	Coeficientes do modelo com 2 mediadores.	29
7	Qualidade do sono e ansiedade por gênero.	32
8	Qualidade do sono e a presença de sintomas de ansiedade	33
9	Qualidade do sono por presença de comorbidades	33
10	Coeficientes estimados do modelo do banco de dados da qualidade do sono	36

Sumário

1 Introdução	9
1.1 Organização do trabalho	11
2 Revisão da literatura	12
2.1 Análise de regressão linear	12
2.2 Regressão Logística	14
2.3 Análise de mediação simples	16
2.4 Análise de mediação múltipla	19
2.5 Método KHB	21
3 Resultados	23
3.1 Conjunto de dados	23
3.2 Aplicação 1: Modelo com 1 mediador	24
3.3 Aplicação 2: Modelo com 2 mediadores	27
3.4 Aplicação 3: Banco de dados da qualidade do sono	31
3.5 Aplicação 4: Decomposição do efeito total via método KHB	36
4 Conclusão	38
Referências	40

1 Introdução

A pandemia da COVID-19, que teve início em março de 2020 (Organização Mundial da Saúde, 2020), representou um fenômeno singular na história da humanidade. A confrontação com uma doença nova, ainda sem vacina, e as taxas crescentes de contágio geraram temores intensos. A título de exemplo, no Brasil foram registrados em 2020, 7 milhões de casos e cerca de 195 mil mortes BRASIL (2024). O receio de contaminação própria ou de entes queridos serem infectados pelo vírus resultou em sérios danos à saúde mental das pessoas que viveram essa experiência. Associado a esse medo de contaminação pela COVID-19, as medidas implementadas pelo governo para conter o coronavírus também impactaram significativamente a rotina diária de todos. Por exemplo, a interrupção de eventos festivos, encontros familiares e as atividades aglomerativas resultaram em uma diminuição acentuada do convívio social, afetando profundamente o bem-estar emocional. Adicionalmente, tais restrições tiveram um impacto econômico significativo, levando à perda de empregos para aqueles cuja renda dependia desses eventos, comprometendo assim a capacidade de prover para as famílias.

Um estudo realizado com uma amostra de indivíduos da comunidade acadêmica da Universidade Federal de Ciência da Saúde de Porto Alegre (UFCSPA), mostra como a pandemia da COVID-19 teve um impacto negativo na saúde mental dos estudantes e funcionários dessa instituição. Conforme Ribeiro et al. (2021) "a pandemia de COVID-19 exerceu efeitos na saúde mental da comunidade acadêmica, em que prevaleceram sentimentos de impotência, angústia, medo de perder familiares, amigos ou conhecidos, irritabilidade e tristeza [...]". Ademais, como resultado da pandemia, muitas pessoas foram desenvolvendo uma série de doenças como ansiedade, depressão, hipertensão, etc. Doenças essas que tiveram impacto significativo na qualidade do sono. O distanciamento social pode ser apontado como um dos fatores que influenciou negativamente a qualidade do sono entre os períodos antes e durante a pandemia, distúrbios do sono e a percepção subjetiva da qualidade do sono pioraram durante o período pandêmico (FERREIRA et al., 2021).

Por desempenhar um papel fundamental em várias funções biológicas importantes, como a consolidação da memória, a regulação da temperatura corporal e a restauração da energia, bem como na manutenção do metabolismo cerebral, qualquer interrupção no padrão de sono pode ter um impacto significativo na qualidade de vida dos indivíduos incluindo o: bem-estar físico, desempenho no trabalho ou na escola, funções cognitivas e interações sociais (CARDOSO et al., 2009).

Estudos apontam que distúrbios do sono podem acarretar em consequências adversas para os indivíduos, incluindo uma perda geral na qualidade de vida, disfunções no

sistema autônomo, uma redução no desempenho profissional ou acadêmico, um aumento na suscetibilidade a distúrbios psiquiátricos e uma diminuição na vigilância (CARDOSO et al., 2009), o que pode levar a acidentes. Pessoas que experimentam problemas no sono têm uma maior probabilidade de enfrentar problemas de saúde, uma expectativa de vida potencialmente menor e sinais prematuros de envelhecimento. Portanto, a importância de uma boa qualidade de sono não pode ser subestimada.

Vários estudos apontam que, além de terem sido responsáveis por cerca de 43 milhões de mortes em todo mundo, só no ano de 2021, a presença de doenças crônicas não transmissíveis (DCNTs) pode igualmente influenciar a qualidade do sono (WHO, 2014; MARRERO; ADASHI, 2015; FIGUEIREDO; CECCON; FIGUEIREDO, 2021). As DCNTs também conhecidas como comorbidades, são problemas de saúde, como por exemplo: a hipertensão, diabetes, câncer, obesidade, doenças cardiovasculares, entre outras. Uma abordagem mais detalhada pode ser igualmente encontrada em Lima et al. (2022). Este estudo também indica que problemas com o sono são mais frequentes à medida que o número de DCNTs declaradas aumenta. Algumas comorbidades, causam dores nos indivíduos, dificultando um bom sono, que por sua vez, provoca mais dores, criando assim um círculo vicioso. Dessa forma, surge a necessidade de investigar os reais efeitos das comorbidades na qualidade do sono e como esses efeitos podem ter sido agravados pela pandemia da COVID-19.

Uma forma de analisar a influência das DCNTs na qualidade do sono, é por meio da aplicação das técnicas estatísticas que permitem relacionar as variáveis intervenientes por meio de uma expressão matemática. A análise de regressão linear, provavelmente a primeira escolha quando se objetiva avaliar a relação entre as variáveis, tem sido amplamente utilizada na literatura estatística. Tal análise pressupõe que a variável dependente seja numérica, podendo as variáveis independentes serem quantitativas ou qualitativas. Já, no caso em que a variável dependente for binária, a regressão logística é a técnica mais indicada.

No presente trabalho, contudo, optou-se pela análise estatística de mediação, uma abordagem metodológica que, apesar de seu potencial, ainda é pouco explorada e conhecida. Essa escolha se justifica pelas vantagens oferecidas, incluindo uma compreensão mais aprofundada de como uma variável independente pode influenciar a variável dependente, considerando também os efeitos indiretos mediados por outras variáveis no modelo. A metodologia que apresentamos nesse trabalho permite captar relações complexas e oferece, não só uma compreensão mais detalhada das diferentes influências que a variável explicada pode sofrer, mas também, uma forma diferente de interpretar os coeficientes de regressão em termos de efeitos: indireto (ou efeito mediado), e o efeito direto (ou efeito não mediado).

Com o intuito de obter as respostas dos questionamentos anteriormente levanta-

dos, serão utilizadas ferramentas estatísticas para que a partir dos dados coletados nos residentes de Ouro Preto, uma cidade do estado de Minas Gerais, seja possível generalizar os resultados do estudo na população onde esses dados foram coletados. E quiçá, esperamos que os resultados desse estudo sejam replicados em outros estados brasileiros, por forma a auxiliar os tomadores de decisão no entendimento de que a influência das DCNTs na qualidade do sono pode, de alguma forma ser mediada por outros fatores de interesse, e conseqüentemente, sugerir medidas de saúde pública que busquem mitigar fatores que corroboram para uma piora da qualidade do sono.

Para obter as respostas de interesse, as principais contribuições do presente trabalho incluem: Investigar a influência das comorbidades na qualidade do sono dos residentes de Ouro Preto durante a pandemia da COVID-19. Tal avaliação será feita a partir de: (i) uma caracterização descritiva do banco de dados; (ii) análise das relações entre os fatores que resultam na piora da qualidade do sono; (iii) bem como uma profunda compressão de como as variáveis consideradas mediadoras na relação entre as DCNTs e qualidade de sono podem ser eficazes na decomposição do efeito total.

1.1 Organização do trabalho

O presente trabalho é composto por quatro capítulos. O capítulo 2 apresenta de forma exaustiva o referencial teórico, onde a definição e explicação detalhada da análise de mediação pode ser encontrada, nesse mesmo capítulo são exibidos 2 exemplos simulados de aplicação encontrados em MacKinnon (2012). No capítulo 3 apresentamos os resultados das análises estatísticas, baseados em diferentes aplicações. Por fim, no capítulo 4 são apresentadas as considerações finais do trabalho.

2 Revisão da literatura

2.1 Análise de regressão linear

Segundo Kutner, Nachtsheim e Neter (2004), um modelo de regressão é um meio de expressar a tendência da variável resposta Y_i em variar sistematicamente com uma ou mais variáveis preditoras x_1, x_2, \dots, x_p . Essa característica se deve ao fato de haver: uma distribuição de probabilidade de Y_i para cada nível de x_i , e as médias dessas distribuições variam de forma sistemática com x_i . Para estabelecer um primeiro modelo, considere a estrutura:

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad (2.1.1)$$

em que

- \mathbf{x}_i é um vetor linha das variáveis independentes;
- y_i é o valor observado da variável resposta no i -ésimo experimento;
- $f(x_i, \boldsymbol{\beta})$, denota a parte determinística do modelo;
- ε_i representa a parte aleatória, que não é controlada pelo pesquisador.

Observe igualmente que $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ denota o vetor p -dimensional dos coeficientes de regressão, e x_i denota um valor particular da variável independente. Note que, a depender de como a parte determinística do modelo em (2.1.1) for fixada, o modelo de regressão pode ser classificado em: (i) linear simples, se $f(x_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_i$, ou (ii) linear múltipla, caso $f(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i \boldsymbol{\beta}$, onde \mathbf{x}_i denota a i -ésima linha da matriz experimental $\mathbf{X}_{n \times (p+1)}$, que inclui uma coluna de 1's para acomodar o intercepto.

A versão particular do modelo em (2.1.1), também conhecida como modelo de regressão linear múltipla, é dada na sua forma matricial por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.1.2)$$

quando a matriz \mathbf{X} possui posto completo e com $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ denotando o vetor da variável resposta, $\mathbf{X} = [X_{ij}]$, com $i = 1, 2, \dots, n$ e $j = 0, 1, \dots, p$ e $\boldsymbol{\beta}$ são respectivamente, a matriz experimental e o vetor desconhecidos dos coeficientes de regressão, ambas citadas anteriormente. Por fim, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$ denota o vetor dos erros aleatórios, em que é suposto ser normalmente distribuídos, com vetor de média $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ e matriz de variância-covariância dada por $Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$, em que \mathbf{I}_n é a matriz identidade de

ordem n . Note que, a formulação da $Cov(\boldsymbol{\varepsilon})$ sugere claramente que os erros devem ser homoscedásticos (variância constante) e independentes entre si.

Consequentemente, o vetor aleatório \mathbf{Y} tem média $E\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$ e a matriz de variância-covariância dada por $Cov(\mathbf{Y}) = Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$.

Para encontrar estimadores com boas propriedades, é empregado o método dos mínimos quadrados (KUTNER; NACHTSHEIM; NETER, 2004) que considera para cada observação a soma dos quadrados dos desvios, denotada por Q e definida por

$$\begin{aligned} Q &= \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned} \quad (2.1.3)$$

Assim, os estimadores dos mínimos quadrados são os valores $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^\top$ que minimizam a soma do quadrado dos erros Q . A obtenção das estimativas é detalhada em Kutner, Nachtsheim e Neter (2004). O estimador $\hat{\boldsymbol{\beta}}$ é o melhor estimador linear e não viciado, ou seja, dentre todos os estimadores lineares em Y e não viciados para $\boldsymbol{\beta}$ este é o que possui menor variância pelo Teorema de Gauss-Markov (WOOLDRIDGE, 2023). Para testar a significância dos coeficientes da regressão, realiza-se um teste t . Este teste verifica se uma variável explicativa x_i influencia significativamente o comportamento médio da variável resposta, mantendo constantes as demais variáveis. As hipóteses do teste são

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_1 : \beta_j &\neq 0, \end{aligned}$$

em que $j = 1, \dots, p$, H_0 indica que x_i não tem nenhum efeito sobre o valor esperado de y_i , enquanto H_1 sugere que x_i possui um efeito significativo sobre as variações de Y_i . Uma abordagem detalhada sobre os procedimentos inferenciais, que incluem estimação pontual e intervalar pode ser igualmente encontrada em Faraway (2009), Hastie, Tibshirani e Friedman (2017), Montgomery, Peck e Vining (2021), entre outras referências.

A análise de regressão linear é uma ferramenta fundamental na análise de dados, empregada para descrever a relação entre a média de uma variável resposta e uma ou mais variáveis explicativas. Esse método pressupõe que a variável dependente seja numérica, podendo ser discreta ou contínua. No entanto, quando a variável resposta de interesse é binária ou dicotômica, é necessário recorrer a técnicas específicas de modelagem. Nesses casos, a regressão logística se apresenta como a abordagem mais utilizada.

2.2 Regressão Logística

Muitas variáveis resposta categóricas apresentam apenas duas categorias, como “sim” ou “não”, 0 ou 1, “falso” e “verdadeiro”, ou “sucesso” e “fracasso”. Considere uma variável Y_i cujos possíveis valores são $Y_i = 0$ (fracasso) e $Y_i = 1$ (sucesso). Nesse caso, dado o vetor de covariáveis \mathbf{x}_i , Y_i segue uma distribuição de Bernoulli, em que a probabilidade de sucesso é $P(Y_i = 1|\mathbf{x}_i) = \pi(\mathbf{x}_i)$, e a probabilidade de fracasso é $P(Y_i = 0|\mathbf{x}_i) = 1 - \pi(\mathbf{x}_i)$. A esperança de Y_i é dada por $E(Y_i) = \pi(\mathbf{x}_i)$. Para n observações independentes dessa variável, o número de sucessos segue uma distribuição binomial com parâmetros n e $\pi(\mathbf{x}_i)$ (AGRESTI, 2007).

A regressão logística pode ser vista como um caso particular dos modelos lineares generalizados que busca compreender o comportamento da variável resposta em função de mudanças nas variáveis explicativas \mathbf{x}_i . Assim, a probabilidade de Y_i é representada da seguinte forma: $\pi(\mathbf{x}_i)$, para indicar $E(Y|\mathbf{x}_i)$. Um possível modelo análogo ao da regressão linear utilizado para uma resposta binária é:

$$\pi(\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Esse modelo não está bem definido visto que probabilidades estão restritas ao intervalo $[0, 1]$ enquanto modelos lineares podem assumir valores em toda reta real, essa equação pode resultar, quando valores de x são muito grandes ou muito pequenos, em $\pi(x_i) < 0$ ou $\pi(x_i) > 1$. Então, se torna necessário uma modelagem mais complexa (AGRESTI, 2007). A abordagem utilizada para a modelagem no caso de respostas dicotômicas utiliza a distribuição logística cuja estrutura sistemática para a regressão logística é, segundo Hosmer e Lemeshow (2000):

$$\pi(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})}.$$

A transformação logito é de grande importância para o estudo da regressão logística, definida por

$$\text{logit}(\pi(\mathbf{x}_i)) = \log\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right) = \mathbf{x}_i \boldsymbol{\beta}.$$

Sob o modelo de regressão linear os parâmetros dos modelos são estimados através da técnica dos mínimos quadrados ordinários e os valores de $\hat{\beta}$ que minimizam a soma dos desvios quadrados servem como estimativa dos parâmetros. Entretanto, quando trata-se de variáveis resposta binárias, a técnica de mínimos quadrados ordinários não atende os mesmos pressupostos que atende em variáveis contínuas. Segue da independência dos Y_i 's que, o método de estimação que será aplicado no contexto da regressão logística é o método da máxima verossimilhança (HOSMER; LEMESHOW, 2000). Para definir o procedimento de estimação por máxima verossimilhança, considere que seja Y_i uma variável dicotômica e \mathbf{x}_i um vetor com os valores fixados para uma variável quantitativa

que explica Y_i , então pode-se afirmar que $\pi(\mathbf{x}_i) = P(Y_i = 1|\mathbf{x}_i)$ e que $1 - \pi(\mathbf{x}_i) = P(Y_i = 0|\mathbf{x}_i)$, a contribuição do par (\mathbf{x}_i, y_i) na função de probabilidade conjunta (ou função de verossimilhança) pode ser expressa por

$$L_i(\boldsymbol{\beta}) = \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}.$$

Como as observações são independentes, segue que,

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^n L_i(\boldsymbol{\beta}) = \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right)^{y_i} \left(1 - \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right)^{1-y_i} \\ &= \prod_{i=1}^n \exp(\mathbf{x}_i \boldsymbol{\beta})^{y_i} (1 + \exp(\mathbf{x}_i \boldsymbol{\beta}))^{-y_i} (1 + \exp(\mathbf{x}_i \boldsymbol{\beta}))^{y_i-1} \end{aligned}$$

Para facilitar as operações matemáticas é aplicado o logaritmo:

$$\begin{aligned} \ell(\boldsymbol{\beta}) &= \log [L(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \log [\pi(\mathbf{x}_i)] + (1 - y_i) \log [1 - \pi(\mathbf{x}_i)]\}, \\ &= \sum_{i=1}^n \{y_i \mathbf{x}_i \boldsymbol{\beta} - \log [1 + \exp(\mathbf{x}_i \boldsymbol{\beta})]\}. \end{aligned} \quad (2.2.1)$$

Portanto, o estimador de máxima verossimilhança $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^\top$ de $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ é obtido maximizando a equação (2.2.1). Tal como apresentado no modelo normal (modelo clássico), a significância dos coeficientes de regressão pode ser feita por meio do teste de Wald e intervalos de confiança. Para tal, as hipóteses de interesse são:

$H_0 : \beta_j = 0$ a probabilidade de sucesso independe do valor que x_i assume

$H_1 : \beta_j \neq 0$ a probabilidade de sucesso é influenciada pelo valor de x_i ,

com x_i denotando elementos de \mathbf{x}_i . A estatística do teste é dada por $z = \sqrt{\hat{\beta}_j^2 / \text{ep}(\hat{\beta}_j)}^2$, onde $\text{ep}(\hat{\beta}_j)$ denota o erro padrão de $\hat{\beta}_j$.

Para compreender melhor as relações entre múltiplas variáveis, é fundamental investigar se o efeito de uma variável preditora x_i influencia direta ou indiretamente o comportamento de uma variável resposta Y_i . É importante considerar como a inclusão de uma variável mediadora no modelo pode alterar esse efeito. Para isso, utiliza-se a análise de mediação, que permite identificar e quantificar o papel que uma variável mediadora desempenha na relação entre x_i e Y_i .

2.3 Análise de mediação simples

A análise estatística de mediação é um método utilizado para compreender como uma variável independente x_i transmite seu efeito para a variável resposta Y_i (MACKINNON, 2012). Na teoria de análise de mediação é considerada uma variável intermediária M_i , chamada de mediadora. A variável mediadora explica como e porque uma variável independente influencia a variável dependente.

Uma vantagem dessa técnica sobre os demais métodos que objetivam estudar a relação entre as variáveis é que, a análise de mediação fornece uma compreensão mais completa e detalhada das relações entre tais variáveis, identificando os processos intermediários que conectam as variáveis independentes às dependentes. A influência de x_i em Y_i , sem intermediação, é conhecida como relação do efeito total e um dos objetivos da análise de mediação é decompor tal efeito.

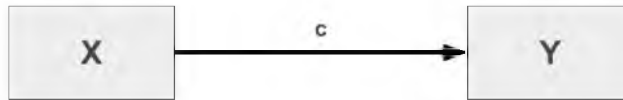


Figura 1: Diagrama do efeito total.

Supondo que Y_i é uma variável quantitativa, então o efeito total, representado pela Figura 1 e denotado por c (que é o efeito a ser decomposto), quantifica a diferença estimada em Y_i , para cada variação unitária de x_i . Portanto, quando as variáveis são quantitativas, a seguinte equação pode representar o efeito total:

$$Y_i = i_Y + cX_i + e_Y,$$

onde i_Y denota o intercepto no modelo reduzido (ou modelo do efeito total), e e_Y o termo de erro, que segundo a notação apresentada no Capítulo 2.1, supõem-se ter uma distribuição normal centrada em 0 e variância constante. Entretanto, o efeito total mede a alteração média do Y_i , para cada variação unitária de x_i . Matematicamente, tem-se

$$c = [\hat{Y}_i|(X = x_i + 1)] - [\hat{Y}_i|(X = x_i)],$$

em que \hat{Y}_i é o valor ajustado para Y .

Ademais, na presença de uma variável mediadora, o efeito total é decomposto da seguinte maneira:

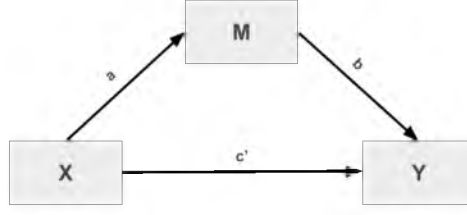


Figura 2: Diagrama conceitual de um modelo de mediação simples.

Um exemplo de um gráfico de mediação é apresentado na Figura 2. Nessa figura observa-se que há dois caminhos que ligam X_i à variável resposta Y_i . O caminho que conecta X_i diretamente à Y_i , é denominado *efeito direto*, representado por c' . Já, o caminho que conecta X_i à variável resposta Y_i passando pelo mediador M_i é chamado de *efeito indireto* (HAYES, 2017). Observe que na Figura 2, há duas variáveis que são influenciadas pelas demais que são (Y_i e M_i), quando ambas são contínuas, podem ser modeladas pelas seguintes equações:

$$M_i = i_M + aX_i + \epsilon_M,$$

$$Y_i = i_Y^* + c'X_i + bM_i + \epsilon_Y,$$

onde i_M e i_Y^* são os interceptos, ϵ_M e ϵ_Y são os erros aleatórios nos ajustes de M_i e Y_i , respectivamente e a , b e c' são os coeficientes de regressão (HAYES, 2017). Mais especificamente, temos que,

- a : quantifica a mudança de M_i em a unidades para cada variação unitária de X_i , ou seja,

$$a = [\hat{M}_i|(X = x_i + 1)] - [\hat{M}_i|(X = x_i)]$$

em que, \hat{M}_i é o valor ajustado de M_i ;

- b : quantifica a mudança em b unidades de Y_i , para cada variação unitária de M_i , mantendo-se constante o fator de exposição X_i , isto é,

$$b = [\hat{Y}_i|(M = m_i + 1, X = x_i)] - [\hat{Y}_i|(M = m_i, X = x_i)].$$

O efeito indireto de X_i em Y_i através de M_i é representado pelo produto ab . Esta quantidade é interpretada da seguinte maneira: para cada variação de uma unidade em X_i , será estimada uma diferença de ab unidades em Y_i , como resultado do efeito de X_i em M_i , e do efeito de M_i em Y_i . Esses parâmetros podem ser estimados utilizando o método dos mínimos quadrados ordinários para obter \hat{ab} (MACKINNON, 2012). O efeito direto é a influência que X_i tem em Y_i sem passar pelo mediador M_i , ou seja, é a parcela do efeito total que é transmitido somente à variável resposta Y_i . O coeficiente c' denota o

efeito direto, uma interpretação é que, para cada variação unitária de X_i , mantendo M_i constante, é provocada uma mudança de c' unidades em Y_i . Isto é,

$$c' = [\hat{Y}|(X = x_i + 1, M = m_i)] - [\hat{Y}|(X = x_i, M = m_i)].$$

O sinal de c' determina como o efeito direto impacta o ajuste da variável resposta. Um coeficiente $c' > 0$ significa que para cada aumento de uma unidade em X_i , o valor ajustado \hat{Y}_i aumentará, enquanto o efeito direto negativo significa que, para valores maiores de X_i , serão ajustados valores menores de Y_i . No caso em que Y_i e M_i são variáveis contínuas o efeito total é expresso como a soma dos efeitos direto e indireto:

$$c = c' + ab,$$

e portanto o efeito indireto ab é também encontrado pela diferença entre o efeito total e o efeito indireto: $c - c'$, que pode ser estimado utilizando os mínimos quadrados ordinários, obtendo $\hat{c} - \hat{c}'$ (MACKINNON, 2012).

Para compreender o quanto do efeito total é mediado por M_i , é calculado o seguinte percentual:

$$\frac{ab}{c} \times 100.$$

Outro aspecto que é importante ressaltar é que há uma diferença entre um fator de confusão e uma variável mediadora. Segundo MacKinnon (2012), o fator de confusão explica a relação entre variáveis porque está relacionado a ambos, X_i e Y_i , mas não faz parte do processo de mediação causal. Já o mediador explica a relação entre X_i e Y_i porque o efeito de X_i é transmitido a Y_i por meio de M_i .

Como dito anteriormente, tanto \hat{ab} , quanto $\hat{c} - \hat{c}'$ fornecem uma estimativa do efeito mediado. Para calcular o intervalo de confiança desses estimadores, é necessário definir o seu erro padrão. Os erros padrões de ambas as equações podem ser usados para a construção de limites superiores e inferiores do intervalo de confiança. Primeiro para \hat{ab} , a forma do erro padrão mais utilizada é $s_{\hat{ab}}$ (MACKINNON, 2012), definida por

$$s = \sqrt{\hat{a}^2 s_b^2 + \hat{b}^2 s_a^2}, \quad (2.3.1)$$

onde, s_a^2 e s_b^2 correspondem a variância de \hat{a} e \hat{b} , respectivamente. Para calcular o erro padrão de $\hat{c} - \hat{c}'$, ou seja, o erro padrão da diferença entre dois coeficientes de regressão é utilizado:

$$s_{\hat{c}-\hat{c}'} = \sqrt{s_{\hat{c}}^2 + s_{\hat{c}'}^2 - 2rs_{\hat{c}}s_{\hat{c}'}}, \quad (2.3.2)$$

em que $2rs_{\hat{c}}s_{\hat{c}'}$ é a covariância entre \hat{c} e \hat{c}' . Em geral, tanto o erro padrão para \hat{ab} , quanto

para $\hat{c}-\hat{c}'$ são bastante similares, mas por simplicidade a equação (2.3.1) é a mais utilizada (MACKINNON, 2012).

O intervalo com $\gamma = 95\%$ de confiança para o efeito mediado é calculado utilizando:

$$\text{Limite inferior (LI)} = \hat{ab} - z_{\alpha} s_{\hat{ab}},$$

$$\text{Limite superior (LS)} = \hat{ab} + z_{\alpha} s_{\hat{ab}}.$$

Para testar se o efeito mediado é estatisticamente significativo, é possível calcular a razão entre a estimativa do efeito mediado e o erro padrão e comparar com os valores tabulados da distribuição normal padrão. Em seguida compara-se o resultado da razão, em módulo, caso ultrapasse o quantil para o valor escolhido de γ , é possível afirmar que o efeito é significativo. Outro método que é usado para testar é verificar se o valor 0 está dentro do intervalo de confiança, caso não esteja o efeito mediado é significativo (MACKINNON, 2012).

Assim como na análise de regressão usual, os modelos de mediação necessitam atender certos pressupostos, dentre os quais, serão citados 4, sendo eles (MACKINNON, 2012):

- **Forma funcional correta:** Cada equação assume que há uma relação linear entre as variáveis, onde a mudança em 1 unidade na variável preditora afetará proporcionalmente a variável resposta. Adicionalmente, as variáveis possuem um relacionamento aditivo, ou seja, não interagem entre si.
- **Todas as influencias estão expressas:** Não há outras variáveis importantes e influentes que não estejam declaradas no modelo.
- **Medidas acuradas:** X_i , M_i e Y_i são medidas confiáveis e válidas.
- **Erros comportados:** Os erros possuem variância constante (Homoscedasticidade), não são correlacionados com as variáveis preditoras e são independentes entre si.

2.4 Análise de mediação múltipla

Há situações em que mais de uma variável mediadora é incluída na análise, segue diagrama conceitual apresentado na figura 3, retirado de Hayes (2017).

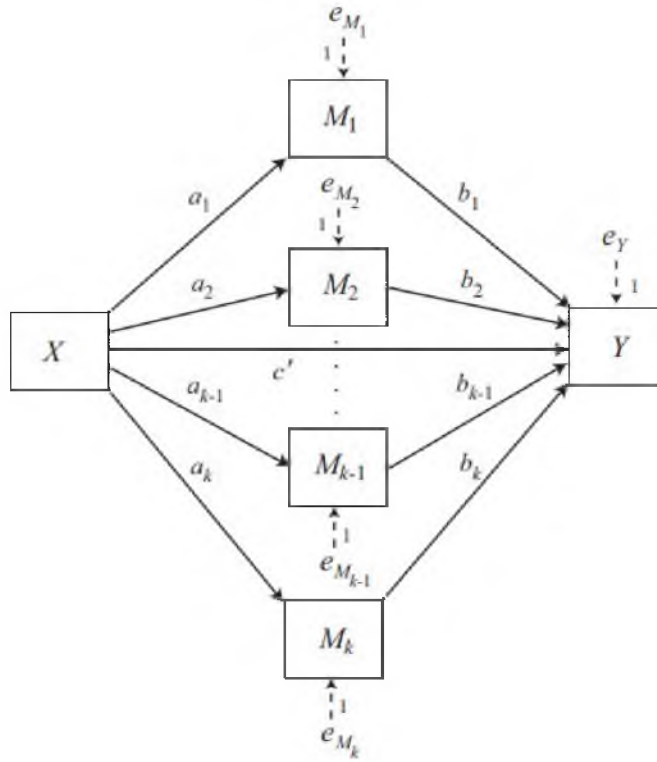


Figura 3: Diagrama de mediação múltipla com k mediadores.

As equações dos modelos com múltiplos mediadores são definidas por

$$M_j = i_{M_j} + a_j X_i + e_{M_j} \quad \text{para cada } j = 1 \text{ a } k,$$

$$Y_i = i'_Y + c' X_i + \sum_{j=1}^k b_j M_j + \epsilon'_Y.$$

Neste caso o efeito de X_i possui $k+1$ direções para influenciar Y , sendo um passando direto de X_i para Y_i e outros k através de cada um dos k mediadores: M_1 ($X \rightarrow M_1 \rightarrow Y$), outro através de M_2 ($X \rightarrow M_2 \rightarrow Y$), e assim por diante, até M_k ($X \rightarrow M_k \rightarrow Y$) (HAYES, 2017). O efeito mediado total é expresso por:

$$\text{Efeito mediado total} = \sum_{j=1}^k a_j b_j.$$

Analogamente ao caso com apenas 1 variável mediadora o efeito também pode ser obtido através da diferença entre o efeito total e o efeito direto:

$$\sum_{i=1}^k a_i b_i = c - c'.$$

Para obter o erro padrão de cada variável mediadora, segue similarmente ao caso de um

só mediador. Para o primeiro efeito mediado segue que

$$s_{\hat{a}_1\hat{b}_1} = \sqrt{\hat{a}_1^2 s_{b_1}^2 + \hat{b}_1^2 s_{a_1}^2}.$$

Também o erro padrão do efeito mediado total, no caso de 2 variáveis mediadoras pode ser expresso por:

$$s_{\hat{a}_1\hat{b}_1 + \hat{a}_2\hat{b}_2} = \sqrt{s_{\hat{a}_1\hat{b}_1}^2 + s_{\hat{a}_2\hat{b}_2}^2 + 2\hat{a}_1\hat{a}_2 s_{\hat{b}_1\hat{b}_2}},$$

sendo $s_{\hat{b}_1\hat{b}_2}$ a covariância das estimativas. Ou ainda:

$$s_{\hat{c}-\hat{c}'} = \sqrt{s_{\hat{c}}^2 + s_{\hat{c}'}^2 - 2r s_{\hat{c}} s_{\hat{c}'}}$$

Para testar se os coeficientes são significativos e também obter os valores para os intervalos de confiança basta proceder de forma análoga ao que foi apresentado na análise de mediação simples.

As técnicas e equações acima abordadas são aplicáveis somente quando a variável resposta Y_i e os mediadores M_i são contínuos. Para o caso de variáveis discretas utiliza-se outros métodos, em particular o método KHB (KOHLEH; KARLSON; HOLM, 2011) que é capaz de decompor o efeito total em relações de variáveis de quaisquer naturezas.

2.5 Método KHB

O método KHB é uma técnica que retorna o grau com que uma variável de controle M_i , media ou explica a relação entre X_i e Y_i bem como a decomposição do efeito total (KOHLEH; KARLSON; HOLM, 2011).

No caso da regressão linear clássica, a decomposição do efeito total em efeito direto e indireto é bem intuitiva o que não acontece nos casos em que a relação entre a variável resposta e o conjunto de variáveis explicativas não é linear. O contexto que será abordado é a comparação de coeficientes em modelos cuja resposta é binária.

Seja o modelo de mediação completo:

$$Y_i = i_Y + c'X_i + bM_i + \epsilon_Y,$$

e o modelo reduzido:

$$Y_i = i_Y + cX_i + e_Y.$$

Para a decomposição do efeito considera-se o efeito indireto como a diferença do

efeito total pelo efeito direto: $ab = c - c'$. Ao considerar o caso de variáveis independentes dicotômicas, é possível utilizar uma estimativa para esses coeficientes (KOHLEK; KARLSON; HOLM, 2011):

$$\hat{c} = \frac{c}{\sigma_c}, \quad \text{e} \quad \hat{c}' = \frac{c'}{\sigma_{c'}},$$

ambos σ_c e $\sigma_{c'}$ são parâmetros de escala. Os parâmetros c , c' , σ_c e $\sigma_{c'}$ não podem ser estimados, apenas as suas razões. Logo, temos:

$$\hat{c} - \hat{c}' = \frac{c}{\sigma_c} - \frac{c'}{\sigma_{c'}}.$$

Como a diferença entre \hat{c} e \hat{c}' combina mediação e reescalonamento, a questão é como obter uma estimativa do efeito indireto que não seja distorcida por diferenças nas escalas. O Método KHB soluciona essa questão extraindo toda a informação de M_i que não está em X_i . Isso é feito utilizando os resíduos ordinários da regressão de M_i em X_i :

$$R_i = M_i - (\beta_0 + \beta_1 x_i),$$

em que β_0 e β_1 são os coeficientes da regressão. R_i substitui M_i e o modelo reduzido fica:

$$Y_i^* = \hat{i}_Y + \tilde{c}X_i, \text{ e}$$

$$\tilde{c} - \hat{c}' = \frac{\tilde{c}}{\tilde{\sigma}_c} - \frac{c'}{\sigma_{c'}} = \frac{c - c'}{\sigma_c},$$

pois $\tilde{\sigma}_c = \sigma_{c'}$ (KOHLEK; KARLSON; HOLM, 2011).

Para testar a significância dos coeficientes é realizado o seguinte teste de hipótese:

$$H_0 : \tilde{c} - \hat{c}' = 0 \text{ o efeito indireto não é significativo,}$$

$$H_1 : \tilde{c} - \hat{c}' \neq 0 \text{ o efeito indireto é significativo,}$$

a estatística do teste é:

$$z = \frac{\sqrt{N}(b_R - b_F)}{\sqrt{\mathbf{a}\Sigma_{\mathbf{a}}}} \sim N(0, 1),$$

em que b_R é o valor do coeficiente de X_i no modelo reduzido (modelo sem o mediador) e b_F , o coeficiente de X_i no modelo completo (modelo com mediador).

Para mais detalhes ver Kohler, Karlson e Holm (2011).

3 Resultados

3.1 Conjunto de dados

Os dados que serão usados no presente trabalho foram obtidos por meio de um inquérito sorológico envolvendo uma pesquisa de amostragem complexa de base populacional, intitulado “COVID-Inconfidentes: Vigilância epidemiológica da covid-19 na região dos Inconfidentes/MG”, realizado pela Universidade Federal de Ouro Preto (UFOP). Os dados analisados foram coletados de outubro a dezembro de 2020, em duas cidades de médio porte da região centro-sul de Minas Gerais totalizando 1511 indivíduos que responderam à pesquisa, maiores detalhes podem ser encontrados em Junior et al. (2023).

O desfecho de interesse no presente estudo é a qualidade do sono. Sua construção é feita usando como referência, uma versão brasileira do Pittsburgh Sleep Quality Index (PSQI) traduzida pelos autores em Bertolazi et al. (2011). Em suma, uma seção do questionário supracitado é dedicada exclusivamente para obter informações que sejam úteis para a construção do índice PSQI. Em geral, o instrumento é composto por 19 perguntas categorizadas em sete componentes com pontuações variando de 0 a 3. Desta forma, cada participante vai pontuando em como ele classifica a sua qualidade de sono no período em análise. Feito isso, o desfecho de interesse (Y) será: $Y = 1$ se $PSQI \geq 5$ (Má qualidade do sono) e $Y = 0$ se $PSQI < 5$ (boa qualidade do sono). Observe que, PSQI é a soma dos escores obtidos para cada respondente. Nisso, a classificação de boa qualidade de sono é atribuída aos respondentes com uma soma menor a 5. De igual forma, tal classificação pode ser feita incluindo uma terceira categoria no desfecho de interesse (nível moderado), veja Wang et al. (2020) para mais informações.

A variável de exposição (comorbidades) foi obtida com base no auto-relato do entrevistado, conforme a sua situação clínica no período de avaliação. A classe das comorbidades avaliadas incluem: diabetes, asma, problemas de pressão, doenças pulmonares, transtornos de ansiedade, doenças cardíacas ou da tireoide. A variável de exposição (DCNT) denota a quantidade de comorbidades que um indivíduo relatou.

A variável sintomas de ansiedade constitui um fator de mediação (JUNIOR et al., 2023). Esse fator é dicotômico, e seus níveis foram obtidos usando como base um índice de classificação frequentemente usado pelos pesquisadores da área. Para finalizar, outras covariáveis que podem ter alguma influência sobre a qualidade do sono são: gênero (1: se feminino, 0: masculino); idade (1: para 18 a 34 anos, 2: para 35 a 59 anos e 3: maior que 60 anos como o nível de referência); estado civil (1: se casado, 0: caso contrário); educação (1: se 0 ao 5º ano, 2: 6 ao 9º ano, 3: maior ou igual ao 10º ano é o nível de referência); renda familiar (1: se < 2, 2: 2 a 4 e 3: maior ou igual que 4 salários mínimos

é o nível de referência); ocupação (1: se trabalhador e 0: caso contrário). Todo o processo de análise de dados será feito usando o Software R e o STATA.

3.2 Aplicação 1: Modelo com 1 mediador

Os dados para o exemplo a seguir são de um estudo hipotético sobre o consumo de água, em que o efeito de um estímulo sobre uma resposta é mediado pelo organismo do indivíduo. O estímulo foi a temperatura, o organismo são os indivíduos estudados e a sua percepção de sede, e a resposta é o consumo de água. O objetivo deste estudo é verificar se a exposição a temperaturas mais altas, em locais como submarinos e estações espaciais, aumenta o consumo de água. Os dados para este estudo estão na Tabela 17 (Apêndice). Id representa cada indivíduo, total de 50 indivíduos; X , a temperatura (em Fahrenheit); M , um indicador de sede informado pelo indivíduo, com valores de 1 a 5; e Y , o número de decilitros de água consumidos.

Em Kim (2016) é criado um fluxograma baseado nos passos descritos por Baron e Kenny (1986) que servirá como um guia para a reprodução do estudo:

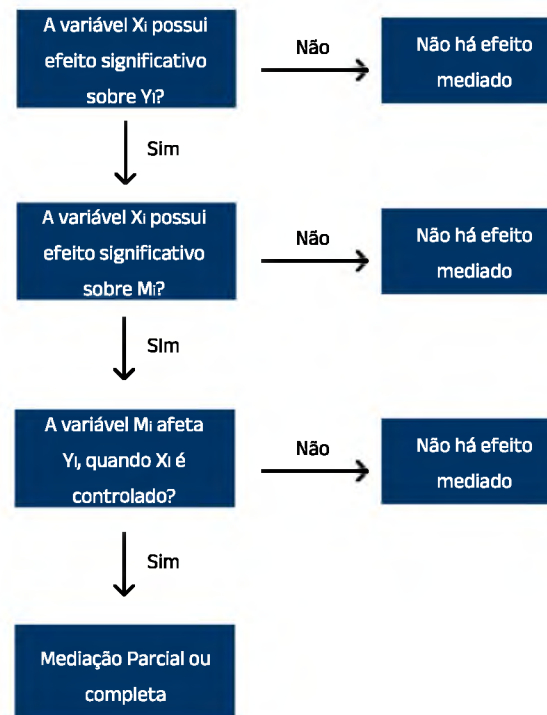


Figura 4: Fluxograma de análise de mediação.

O 1º passo é verificar se X_i possui efeito significativo sobre Y_i , após utilizar a

função lm do R, as estimativas dos coeficientes são:

Tabela 1: Modelo reduzido da 1ª aplicação (resultados)

	Estimativa	Erro Padrão	Estatística do teste	Valor-p
i_Y	-22,269	9,370	-2,377	0,021
c	0,363	0,133	2,721	0,009

o modelo incluindo as estimativas dos coeficientes é:

$$Y_i = i_Y + cX_i + e_Y,$$

$$Y_i = -22,269 + 0,363X_i.$$

Observando o Valor-p, conclui-se a um nível de significância de 5% que a variável X_i é uma boa preditora para Y_i , ou seja, para cada mudança de uma unidade em temperatura há um aumento de 0,363 vezes no consumo de água.

De acordo com o 2º passo deve-se testar se ao aumentar ou diminuir a temperatura do ambiente, a percepção de sede dos indivíduos também é alterada proporcionalmente.

Tabela 2: Modelo de mediação da 1ª aplicação (resultados)

	Estimativa	Erro Padrão	Estatística do teste	Valor-p
i_M	-20,702	8,588	-2,410	0,019
a	0,338	0,122	2,767	0,008

inserindo as estimativas dos coeficientes, temos:

$$M_i = i_M + aX_i + e_M,$$

$$M_i = -20,702 + 0,338X_i.$$

Segundo a tabela 2, é possível concluir que X_i também possui efeito significativo sobre M_i . Então, pode-se inferir que para cada aumento de 1º Fahrenheit na temperatura, a percepção de sede irá subir aproximadamente 0,34 vezes.

3º passo, verificar se X_i e M_i , são variáveis que influenciam significativamente Y_i .

Tabela 3: Modelo completo da 1ª aplicação (resultados).

	Estimativa	Erro Padrão	Estatística do teste	Valor-p
i_Y	-12,932	9,129	-1,417	0,163
c'	0,210	0,132	1,591	0,118
b	0,451	0,145	3,113	0,003

o modelo com os coeficientes estimados é:

$$Y_i = i_Y + c'X_i + bM_i + e_Y,$$

$$Y_i = -12,932 + 0,210X_i + 0,451M_i.$$

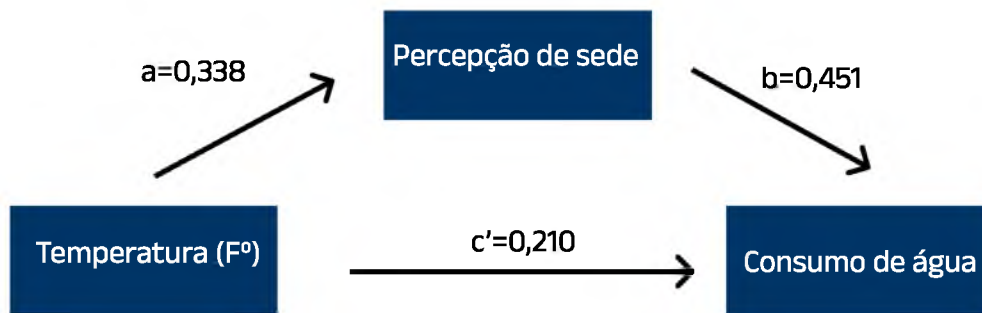


Figura 5: Coeficientes do modelo com 1 mediador.

Segundo os resultados acima, a um nível de significância $\alpha = 5\%$, a variável X_i não possui efeito significativo quando o mediador M_i é inserido no modelo linear, como também o intercepto, isso significa que uma parte expressiva do efeito de X_i é mediado por M_i , ou seja, o efeito da temperatura é mediado pelo organismo do indivíduo. Para saber o quanto do efeito foi mediado, considere fazer a proporção:

$$\frac{ab}{c} \times 100,$$

$$\frac{0,338 \times 0,451}{0,363} \times 100,$$

$$\frac{0,153}{0,363} \times 100,$$

Proporção do efeito mediado = 42%.

O aumento de 1 grau Fahrenheit influencia a percepção de sede que após sofrer o efeito do acréscimo de temperatura o transmite para o consumo de água do indivíduo. 42% do efeito do aumento de temperatura é transmitido pelo organismo do participante

do estudo. Vale ressaltar que o valor do coeficiente para o efeito mediado pode ser obtido de duas maneiras: $\hat{a} \times \hat{b} = 0,338 \times 0,451 = \hat{c} - \hat{c}' = 0,363 - 0,210 = 0,153$.

Prosseguindo ao cálculo do intervalo de confiança para o efeito mediado, será utilizado a expressão (2.3.1) que denota o erro padrão:

$$\begin{aligned} s &= \sqrt{\hat{a}^2 s_b^2 + \hat{b}^2 s_a^2}, \\ s &= \sqrt{(0,338)^2(0,145)^2 + (0,451)^2(0,122)^2}, \\ s &= 0,073. \end{aligned}$$

Calculando os limites do intervalo, para 95% de confiança:

$$\begin{aligned} \text{LI} &= 0,153 - 1,96(0,073) = 0,01, \\ \text{LS} &= 0,153 + 1,96(0,073) = 0,3, \\ 0,01 &\leq ab \leq 0,3. \end{aligned}$$

Portanto, o valor do coeficiente do efeito mediado em 95% das possíveis amostras da população estudada estará no intervalo de 0,01 e 0,3.

Com o intuito de verificar se o efeito mediado é estatisticamente significativo será feita a razão entre a estimativa $\hat{a} \times \hat{b}$ e o seu erro padrão:

$$\begin{aligned} &\frac{\hat{ab}}{\text{Erro padrão}}, \\ &\frac{0,338 \times 0,451}{0,073}, \\ &\frac{0,153}{0,073} = 2,095. \end{aligned}$$

Como o resultado da razão é maior que 1,96 e o valor 0 não está dentro do intervalo de confiança pode-se concluir que o efeito mediado é significativo.

3.3 Aplicação 2: Modelo com 2 mediadores

Novamente será apresentado um exemplo para elucidação, desta vez um caso em que há duas variáveis mediadoras, encontrado em MacKinnon (2012). A situação hipotética a ser estudada é a de como a expectativa dos professores sobre cada aluno e a dificuldade dos materiais utilizados podem influenciar o desempenho dos estudantes. Um total de 40 estudantes e 40 professores forneceram os dados. Cada estudante recebeu um escore de suas habilidades aleatoriamente, e a expectativa do professor baseada nesse

escore é a variável X_i . Supõe-se que a alta expectativa de um professor proporcione um clima social mais agradável e também que ao ter grandes expectativas para os alunos os professores tendem a escolher materiais maiores e mais desafiadores.

A resposta Y_i é a pontuação obtida em um teste de habilidades básicas aplicado no final do semestre, as variáveis M_1 e M_2 , correspondem a uma média de percepção do clima social dentro da sala de aula e uma média da percepção de volume e dificuldade do material utilizado, respectivamente. A motivação da pesquisa é verificar através da análise de mediação se um ambiente incentivador, e o nível de quantidade e dificuldade do material utilizado influenciam o desempenho estudantil. Os dados estão apresentados na Tabela 18.

Seguindo os passos descritos na Figura 4, serão analisados cada um dos modelos.

Tabela 4: Modelo reduzido da 2ª aplicação (resultados).

	Estimativa	Erro Padrão	Estatística do teste	Valor-p
i_Y	34,727	8,924	3,891	4×10^{-4}
c	0,707	0,173	4,081	2×10^{-4}

o modelo com as estimativas dos coeficientes é

$$Y_i = i_Y + cX_i + e_Y,$$

$$Y_i = 34,727 + 0,707X_i.$$

A variável X_i , expectativa do professor com base no escore de habilidades dos alunos, possui efeito significativo sobre o desempenho dos estudantes. Para cada aumento unitário na expectativa do professor, o desempenho dos estudantes aumenta cerca de 0,71 vezes.

Tabela 5: Modelo de mediação da 2ª aplicação (resultados)

	Estimativa	Erro Padrão	Estatística do teste	Valor-p
i_{M_1}	7,097	8,128	0,873	0,388
a_1	0,840	0,158	5,319	4×10^{-6}

o modelo então, é:

$$M_1 = i_{M_1} + a_1X_i + e_{M_1},$$

$$M_1 = 7,097 + 0,840X_i.$$

A variável X_i também possui influência significativa em M_1 , para cada aumento de 1 unidade da expectativa do professor a percepção de agradabilidade do ambiente é aumentada

0,84 vezes.

Tabela 6: Modelo de mediação da 2ª aplicação (resultados)

	Estimativa	Erro Padrão	Estatística do teste	Valor-p
i_{M_2}	40,710	7,514	5,418	$3,580 \times 10^{-6}$
a_2	0,221	0,146	1,52	0,137

as estimativas dos coeficientes também podem ser expressas da seguinte maneira:

$$M_2 = i_{M_2} + a_2 X_i + e_{M_2},$$

$$M_2 = 40,710 + 0,221 X_i.$$

Analisando o modelo contendo a variável M_2 é constatado que para um nível de significância igual a 5%, X_i não é um bom preditor, contudo ainda vale a pena investigar se ao inserir os dois mediadores as estimativas dos parâmetros sofrem alguma alteração.

Tabela 7: Modelo completo da 2ª aplicação (resultados)

	Estimativa	Erro Padrão	Estatística do teste	Valor-p
i_Y	9,123	10,478	0,871	0,389
c'	0,112	0,207	0,541	0,592
b_1	0,570	0,157	3,629	8×10^{-4}
b_2	0,530	0,170	3,123	0,003

o modelo completo com as estimativas dos coeficientes é:

$$Y_i = i_Y + c' X_i + b_1 M_1 + b_2 M_2 + e_Y$$

$$Y_i = 9,123 + 0,112 X_i + 0,570 M_1 + 0,530 M_2$$

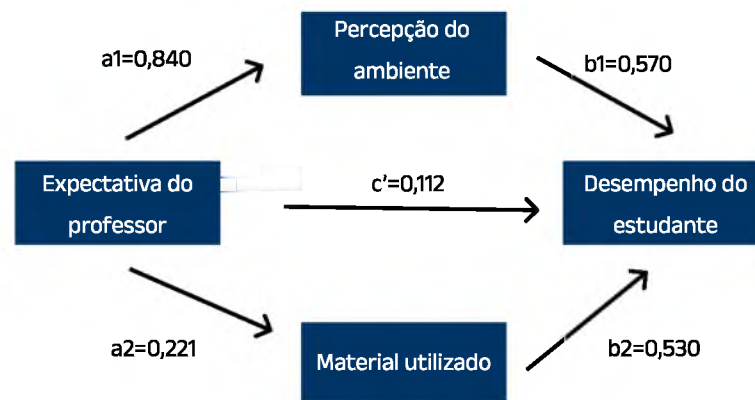


Figura 6: Coeficientes do modelo com 2 mediadores.

Os resultados acima mostram que tanto o intercepto quanto c' não possuem efeito significativo no modelo quando incluídos ambos os mediadores, porém estes últimos são significativos, o que indica que o efeito da expectativa do professor é mediado pela percepção de clima e a escolha por materiais mais difíceis e maiores, que por sua vez possuem influência na nota do estudante no teste final.

O cálculo do percentual do efeito que foi mediado é feito da seguinte maneira:

$$\frac{\frac{a_1b_1 + a_2b_2}{c}}{0,840 \times 0,570 + 0,221 \times 0,530} \times 100,$$

$$\frac{0,595}{0,707} \times 100,$$

Proporção do efeito mediado = 84%.

Os resultados acima mostram que 84% do efeito da expectativa do professor no desempenho do aluno é mediado pela percepção do clima e a dificuldade do material utilizado. Para conferência será calculado o efeito mediado através da diferença entre o efeito total e o efeito direto:

$$\sum_{i=1}^2 a_i b_i = c - c',$$

$$\sum_{i=1}^2 a_i b_i = 0,707 - 0,112,$$

$$\sum_{i=1}^2 a_i b_i = 0,595.$$

Agora serão construídos os intervalos de confiança para o efeito mediado. Os erros padrão para cada uma das variáveis mediadoras são:

$$s_{\hat{a}_1 \hat{b}_1} = \sqrt{\hat{a}_1^2 s_{b_1}^2 + \hat{b}_1^2 s_{a_1}^2},$$

$$s_{\hat{a}_1 \hat{b}_1} = \sqrt{(0,840)^2 (0,157)^2 + (0,570)^2 (0,158)^2},$$

$$s_{\hat{a}_1 \hat{b}_1} = 0,160,$$

e

$$s_{\hat{a}_2 \hat{b}_2} = \sqrt{\hat{a}_2^2 s_{b_2}^2 + \hat{b}_2^2 s_{a_2}^2},$$

$$s_{\hat{a}_2 \hat{b}_2} = \sqrt{(0,222)^2 (0,170)^2 + (0,530)^2 (0,146)^2},$$

$$s_{\hat{a}_2 \hat{b}_2} = 0,086.$$

Logo, o erro padrão para o efeito mediado total é:

$$\begin{aligned}
s_{\hat{a}_1\hat{b}_1+\hat{a}_2\hat{b}_2} &= \sqrt{s_{\hat{a}_1\hat{b}_1}^2 + s_{\hat{a}_2\hat{b}_2}^2 + 2\hat{a}_1\hat{a}_2s_{\hat{b}_1\hat{b}_2}}, \\
s &= \sqrt{(0,160)^2(0,086)^2 + 2(0,840)(0,222)(0,086)}, \\
s &= 0.189,
\end{aligned}$$

e os intervalos de confiança inferiores e superiores, respectivamente são:

$$\begin{aligned}
\text{LI} &= 0,595 - 1,96(0,189) = 0,225, \\
\text{LS} &= 0,595 + 1,96(0,189) = 0,966, \\
0,225 &\leq ab \leq 0,966.
\end{aligned}$$

Como o valor 0 não se encontra no intervalo construído com um nível de confiança de 95% pode-se considerar o efeito mediado como estatisticamente significativo.

3.4 Aplicação 3: Banco de dados da qualidade do sono

A seguir, apresento os resultados obtidos com a aplicação das técnicas previamente mencionadas no principal banco de dados deste estudo. Primeiramente, realizo uma análise descritiva dos dados. Em seguida, utilizo a regressão logística no contexto da análise de mediação. Como a maioria das variáveis do banco de dados é qualitativa, a apresentação dos resultados incluirá tabelas de contingência e gráficos representando percentuais.

Tabela 8: Tabela de contingência do gênero e qualidade do sono

Gênero	Sono		Total
	Bom	Ruim	
Masculino	395	552	947
Feminino	306	258	564
Total	701	810	1511

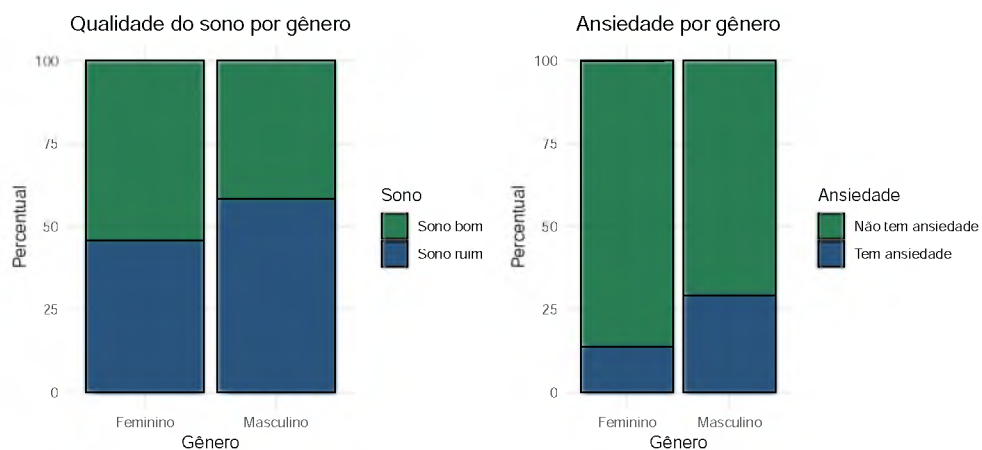


Figura 7: Qualidade do sono e ansiedade por gênero.

Com base na Figura 7, é possível perceber como a maioria dos homens relataram um sono ruim (58%). Já com as mulheres, o contrário, a maioria (54%) relatou ter o sono bom. Também é relevante investigar a possível associação entre o gênero dos entrevistados e a presença de sintomas de ansiedade.

A representação gráfica da Figura 7 evidencia uma prevalência relativamente baixa de relatos de ansiedade. Notavelmente, observa-se uma tendência que, em sua maioria, os entrevistados do sexo masculino relatam essa condição. Essa observação suscita a questão intrigante da possível associação entre a presença de ansiedade e a qualidade do sono.

Tabela 9: Tabela de contingência dos sintomas de ansiedade e qualidade do sono

Sintomas de ansiedade?	Sono		Total
	Bom	Ruim	
Não	635	522	1157
Sim	66	288	354
Total	701	810	1511

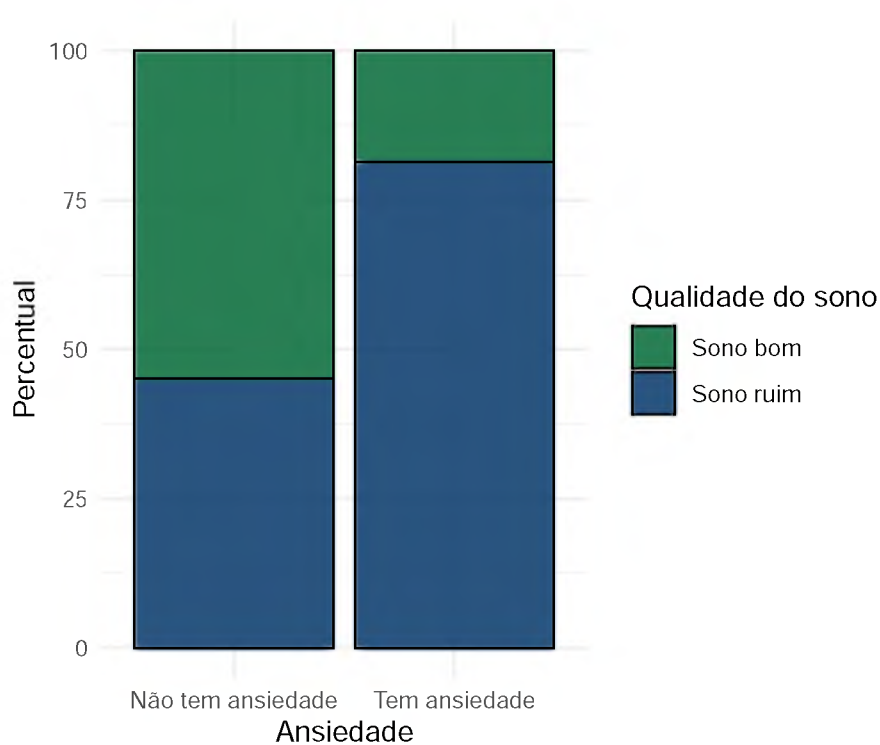


Figura 8: Qualidade do sono e a presença de sintomas de ansiedade

Tanto a Tabela 9, quanto a Figura 8 apontam para uma prevalência de uma pior qualidade do sono quando o indivíduo possui sintomas de ansiedade, indicando uma associação entre ansiedade e piora na qualidade do sono. A seguir será feita uma análise descritiva da possível associação entre qualidade do sono e a presença de comorbidades.

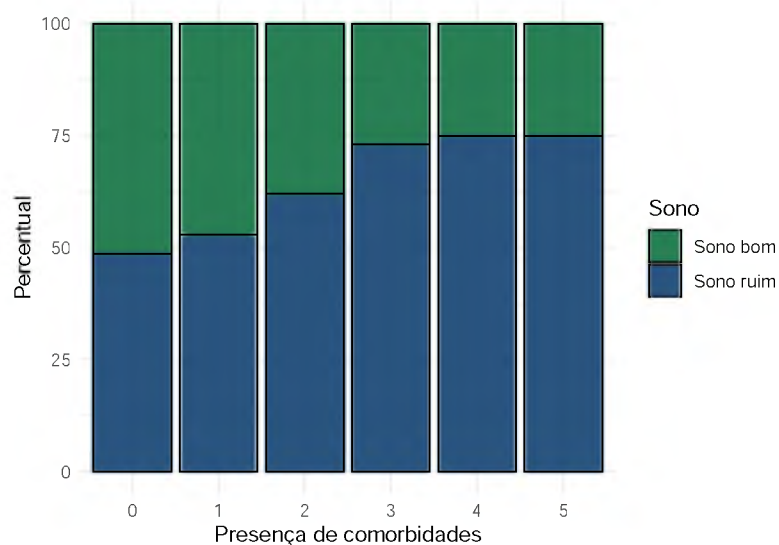


Figura 9: Qualidade do sono por presença de comorbidades

Tabela 10: Tabela do número de comorbidades e qualidade do sono

Número de comorbidades	Sono		Total
	Bom	Ruim	
0	386	365	751
1	205	230	435
2	80	131	211
3	21	57	78
4	7	21	28
5	2	6	8
Total	701	810	1511

A Figura 9, revela um padrão interessante pois a medida em que o número de comorbidades relatadas aumenta, o número de pessoas com uma má qualidade do sono também cresce, confirmando que o banco de dados do presente projeto aparenta possuir o mesmo padrão do estudo Lima et al. (2022). A maior parte dos indivíduos com pelo menos uma comorbidade relataram sono ruim, ou seja, há indícios que as comorbidades afetam de alguma maneira a qualidade do sono.

Com o intuito de utilizar a análise de mediação para o banco de dados em questão, faz-se necessário uma alteração na metodologia pois todos os métodos previamente abordados possuem aplicação somente em dados cuja variável dependente é quantitativa, portanto será feita a seguinte adaptação: os passos descritos na Figura 4 vão ser obedecidos mas a modelagem será feita através da regressão logística.

Primeiramente é necessário verificar se há uma relação entre ter doenças crônicas e a qualidade do sono dos indivíduos e se quanto maior for o número de comorbidades maiores são as chances de se ter uma má qualidade do sono, rodando o modelo os seguintes resultados foram obtidos:

Tabela 11: Modelo reduzido da 3ª aplicação (resultados)

	Estimativa	Razão de chances	Erro padrão da estimativa	Estatística do teste	Valor-p
i_Y	-0,088	0,920	0,066	-1,326	0,185
c	0,290	1,340	0,053	5,456	$4,88 \times 10^{-8}$

o modelo com as estimativas dos coeficientes é

$$\pi(x_i) = \frac{\exp(i_Y + cx_i)}{1 + \exp(i_Y + cx_i)},$$

$$\hat{\pi}(x_i) = \frac{\exp(-0,088 + 0,290x_i)}{1 + \exp(-0,088 + 0,290x_i)}.$$

É possível inferir que, à medida que o número de comorbidades aumenta em uma unidade, há um incremento na chance de apresentar um sono ruim. Esse incremento é de 34%, ou seja, para cada nova comorbidade relatada a chance de se ter um sono ruim aumenta 34%.

Para o 2º passo será feita a regressão de modo a verificar se a variável mediadora que no caso é o nível de ansiedade é influenciada pelo número de comorbidades.

Tabela 12: Modelo de mediação da 3ª aplicação (resultados)

	Estimativa	Razão de chances	Erro padrão da estimativa	Estatística do teste	Valor-p
i_M	-1,387	0,250	0,080	-17,214	2×10^{-16}
a	0,230	1,260	0,055	4,126	$3,680 \times 10^{-5}$

os coeficientes estimados são:

$$\pi(x_i) = \frac{\exp(i_M + ax_i)}{1 + \exp(i_M + ax_i)},$$

$$\hat{\pi}(x_i) = \frac{\exp(-1,387 + 0,230x_i)}{1 + \exp(-1,387 + 0,230x_i)}.$$

Quando se modela a variável ansiedade pelo número de comorbidades é possível identificar como a presença dessas doenças crônicas pode gerar ansiedade, pois para cada aumento de 1 unidade no relato de comorbidades a chance de se ter ansiedade aumenta 26%.

Tabela 13: Modelo completo da 3ª aplicação (resultados)

	Estimativa	Razão de chances	Erro padrão da estimativa	Estatística do teste	Valor-p
i_Y	-0,380	0,680	0,073	-5,300	$1,170 \times 10^{-7}$
c'	0,250	1,280	0,055	4,500	$6,870 \times 10^{-6}$
b	1,630	5,100	0,150	10,890	2×10^{-5}

o modelo completo é:

$$\pi(x_i) = \frac{\exp(i_Y + c'x_i + bm_i)}{1 + \exp(i_Y + c'x_i + bm_i)},$$

$$\hat{\pi}(x_i) = \frac{\exp(-0,380 + 0,250x_i + 1,630m_i)}{1 + \exp(-0,380 + 0,250x_i + 1,630m_i)}.$$

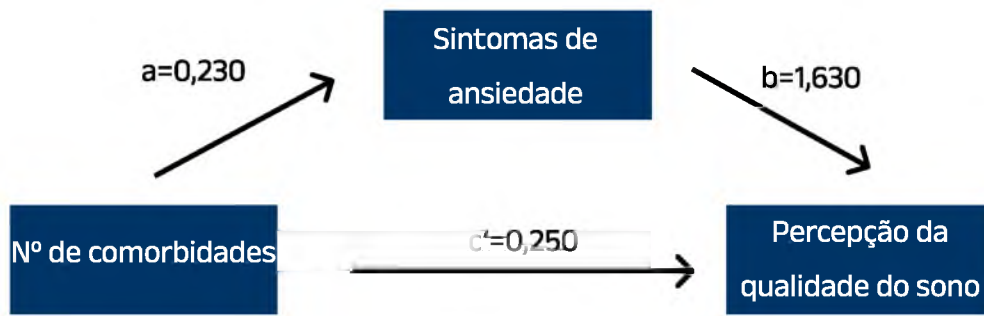


Figura 10: Coeficientes estimados do modelo do banco de dados da qualidade do sono

Os resultados acima mostram que o efeito que o número de comorbidades possui sobre a percepção de qualidade do sono é mediado pela presença ou não de sintomas de ansiedade. Para verificar o quanto desse efeito é mediado faz-se necessário a decomposição do efeito total em efeito indireto e efeito direto. Conforme a Tabela 11, o efeito total, denotado por c possui estimativa igual a 0,290, o efeito indireto é $a \times b$, ou seja, $0,230 \times 1,630 = 0,375$ e o efeito indireto, c' é 0,250, portanto observa-se que o efeito total não pode ser decomposto como foi feito previamente pois quando soma-se o efeito direto e o efeito indireto o resultado diverge daquele do efeito total, pois:

$$\begin{aligned}
 \text{Efeito total} &= \text{Efeito indireto} + \text{Efeito direto}, \\
 \text{Efeito total} &= 0,375 + 0,250, \\
 \text{Efeito total} &= 0,625, \\
 0,290 &\neq 0,625.
 \end{aligned}$$

Por não ser possível decompor o efeito total quando a variável de interesse é binária é necessário utilizar o método KHB, para assim conseguir eficientemente aplicar a análise de mediação em um contexto de respostas binárias.

3.5 Aplicação 4: Decomposição do efeito total via método KHB

O método KHB foi originalmente implantado no software STATA, portanto as seguintes aplicações serão feitas nessa plataforma. O primeiro resultado obtido no STATA corresponde ao modelo reduzido, aquele em que há apenas a variável de exposição que denota o efeito total:

Tabela 14: Modelo reduzido da 4ª aplicação (resultados)

	Estimativa	Razão de chances	Erro padrão da estimativa	Valor-p
Efeito total	0,270	1,310	1,251	0,027

Agora serão computados os valores do modelo completo (modelo contendo o mediador).

Tabela 15: Modelo de mediação da 4ª aplicação (resultados)

	Estimativa	Razão de chances	Erro padrão da estimativa	Valor-p
Efeito direto	0,130	1,140	1,220	0,286
Efeito indireto	0,140	1,150	0,410	0,001

Utilizando o método KHB, é possível decompor o efeito total, pois ele corresponde à soma do efeito direto (0,130) e do efeito indireto (0,140). Dessa forma:

$$0,130 + 0,140 = 0,270 = \text{Efeito total.}$$

De modo a facilitar a interpretação das estimativas dos coeficientes os autores Kohler, Karlson e Holm (2011) propuseram uma medida chamada razão de confusão que expressa a razão entre o efeito direto e o indireto. Segue tabela com esses resultados.

Tabela 16: Modelo reduzido da 4ª aplicação (resultados)

	Razão de confusão	Percentual mediado
Nº de comorbidades	2,074	51,800

O efeito total é 2,07 vezes maior que o efeito direto, evidenciando que uma parcela significativa do efeito é mediada. Conforme apresentado na segunda coluna da Tabela 16, o efeito indireto corresponde a aproximadamente 52% do efeito total. Isso sugere que o número de comorbidades influencia a presença de sintomas de ansiedade, os quais, por sua vez, exercem um impacto relevante na qualidade do sono dos entrevistados.

4 Conclusão

De acordo com os resultados obtidos, é possível concluir que há uma influência significativa do número de comorbidades na qualidade do sono. Conforme apresentado na Tabela 14, à medida que o número de doenças crônicas não transmissíveis aumenta em uma unidade, a probabilidade de o paciente relatar uma má qualidade do sono aumenta em aproximadamente 31%. Esse resultado está em consonância com a análise visual realizada na Figura 9, na qual se observa que 75% dos indivíduos com 5 comorbidades foram classificados, de acordo com a versão brasileira do PSQI (Pittsburgh Sleep Quality Index), como tendo uma má qualidade do sono.

A utilização da técnica de análise de mediação foi fundamental para compreender a relação entre o número de comorbidades e a qualidade do sono, indo além do que uma análise de regressão tradicional poderia oferecer. Enquanto a regressão linear ou logística convencional permite identificar associações diretas entre variáveis, ela não é capaz de desvendar os processos intermediários que explicam como ou por que uma variável influencia outra. No presente estudo, a análise de mediação revelou que 51% do efeito do número de comorbidades sobre a qualidade do sono é mediado pela presença de sintomas de ansiedade. Esse detalhamento é crucial, pois demonstra que as comorbidades não afetam a qualidade do sono apenas de forma direta, mas também indiretamente, por meio de um fator psicológico.

Além disso, a análise de mediação permitiu quantificar a parcela do efeito total que é mediada pela ansiedade, o que seria impossível com uma abordagem de regressão usual. Essa distinção entre efeitos diretos e indiretos é essencial para o planejamento de intervenções mais eficazes. Por exemplo, os resultados sugerem que intervenções focadas apenas no manejo das comorbidades físicas podem não ser suficientes para melhorar a qualidade do sono, já que uma parte significativa do problema está relacionada à ansiedade. Portanto, a análise de mediação forneceu uma compreensão detalhada e valiosa, indicando a necessidade de abordagens integradas que considerem tanto os aspectos físicos quanto psicológicos da saúde.

Outra vantagem da análise de mediação é sua capacidade de identificar relações complexas e muitas vezes implícitas, o que é particularmente relevante em contextos como o da pandemia de COVID-19. Nesse cenário, fatores como isolamento social, incertezas e medos exacerbados podem amplificar sintomas de ansiedade, que, por sua vez, impactam negativamente a qualidade do sono. A regressão tradicional, ao não considerar esses caminhos indiretos, poderia subestimar ou até mesmo ignorar a importância de variáveis mediadoras, como a ansiedade, levando a conclusões incompletas ou equivocadas.

Em resumo, a análise de mediação mostrou-se superior à regressão usual por

permitir uma compreensão mais profunda e detalhada das relações entre as variáveis, destacando mecanismos indiretos que são essenciais para a elaboração de estratégias de intervenção mais eficazes. Essa técnica não apenas enriqueceu a interpretação dos dados, mas também forneceu subsídios para práticas clínicas mais holísticas e direcionadas, especialmente em populações com múltiplas comorbidades e em contextos de crise, como o da pandemia.

O uso da técnica de análise de mediação não deve ser aplicado de forma automática ou independentemente do contexto do estudo. É essencial um planejamento prévio bem estruturado, fundamentado em uma base teórica sólida e em hipóteses claras, para garantir que a mediação faça sentido no cenário investigado. Embora os caminhos analíticos possam cumprir os requisitos estatísticos estabelecidos, como a significância dos coeficientes, isso não implica, por si só, uma relação causal. A significância estatística deve ser interpretada com cautela, pois ela não substitui a necessidade de uma compreensão profunda do contexto e das relações entre as variáveis. Antes de aplicar a mediação, é imprescindível realizar uma avaliação criteriosa do fenômeno em estudo, definindo de forma clara o que se considera variável explicativa, variável resposta e, principalmente, o papel da variável mediadora. Somente com uma abordagem teórica e metodológica bem alinhada é possível evitar conclusões equivocadas e garantir que a análise de mediação contribua de maneira significativa para a compreensão das relações investigadas. De forma geral os objetivos propostos foram atingidos, pois as metodologias aplicadas possibilitaram compreender a real influência das comorbidades na qualidade do sono dos residentes de Ouro Preto durante a pandemia da COVID-19, contribuindo tanto para o entendimento de uma técnica pouco conhecida e utilizada quanto para o reconhecimento de um problema grave que ocorreu no contexto da pandemia.

Referências

- AGRESTI, A. **An Introduction to Categorical Data Analysis**. 2. ed. Hoboken, New Jersey: John Wiley Sons, Inc., 2007.
- BARON, R. M.; KENNY, D. A. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. **Journal of Personality and Social Psychology**, v. 51, n. 6, p. 1173–1182, 1986. Disponível em: <https://doi.org/10.1037//0022-3514.51.6.1173>. Acesso em: 25 nov. 2024.
- BERTOLAZI, A. N. et al. Validation of the brazilian portuguese version of the pittsburgh sleep quality index. **Sleep Medicine**, v. 12, n. 1, p. 70–75, 2011. ISSN 1389-9457. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389945710003801>. Acesso em: 25 nov. 2024.
- BRASIL. **Painel Covid-19**. [S.l.], 2024. Disponível em: <https://infoms.saude.gov.br/extensions/covid-19.html/covid-19.html.html>. Acesso em: 31 jul. 2024.
- CARDOSO, H. C. et al. Avaliação da qualidade do sono em estudantes de medicina. **Revista Brasileira de Educação Médica**, v. 33, n. 3, p. 349–355, jul 2009. ISSN 0100-5502. Disponível em: <https://doi.org/10.1590/S0100-55022009000300005>. Acesso em: 25 nov. 2024.
- FARAWAY, J. J. **Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models**. [S.l.]: Springer Nature BV, 2009.
- FERREIRA, C. R. T. et al. Qualidade do sono durante o distanciamento domiciliar na pandemia da covid-19 na amazônia ocidental. **Revista de Crescimento e Desenvolvimento Humano**, v. 31, p. 458 – 464, 12 2021. ISSN 0104-1282. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0104-12822021000300010&nrm=iso. Acesso em: 25 nov. 2024.
- FIGUEIREDO, A. E. B.; CECCON, R. F.; FIGUEIREDO, J. H. C. Chronic non-communicable diseases and their implications in the life of dependent elderly people. **Ciencia & saude coletiva**, SciELO Brasil, v. 26, p. 77–88, 2021.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Springer, 2017.
- HAYES, A. F. **Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach**. 5. ed. New York, NY: Guilford Press, 2017.
- HOSMER, D.; LEMESHOW, S. **Applied logistic regression**. 2. ed. Hoboken, New Jersey: John Wiley Sons, Inc., 2000.
- JUNIOR, L. A. A. de M. et al. Chronic non-communicable diseases and poor sleep quality during the covid-19 pandemic: an analysis of mediation by anxiety and depression symptoms. **Journal of Psychiatric Research**, 2023. Paper submitted.

- KIM, B. Introduction to mediation analysis. **StatLab Articles**, 2016. Disponível em: <https://library.virginia.edu/data/articles/introduction-to-mediation-analysis>. Acesso em: 25 nov. 2024.
- KOHLER, U.; KARLSON, K. B.; HOLM, A. Comparing coefficients of nested nonlinear probability models. **The Stata Journal**, v. 11, n. 3, p. 420–438, 2011. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/1536867X1101100306>. Acesso em: 25 nov. 2024.
- KUTNER, M.; NACHTSHEIM, C.; NETER, J. **Applied Linear Regression Models**. 4. ed. New York, NY: McGraw-Hill/Irwin, 2004.
- LIMA, M. G. et al. Associação do autorrelato de problemas no sono com morbidades e multimorbidades segundo sexo: Pesquisa nacional de saúde 2019. **Epidemiologia e Serviços de Saúde**, v. 31, n. spe1, p. e2021386, 2022. Disponível em: <https://www.scielo.br/j/ress/a/xzZD8QSRbvNs7FCTFqYkdpw/abstract/?lang=pt>. Acesso em: 27 nov. 2024.
- MACKINNON, D. **Introduction to Statistical Mediation Analysis**. New York, NY: Routledge, 2012.
- MARRERO, S.; ADASHI, E. Y. Noncommunicable diseases. In: THIEME MEDICAL PUBLISHERS. **Seminars in reproductive medicine**. [S.l.], 2015. v. 33, n. 01, p. 035–040.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. [S.l.]: John Wiley & Sons, 2021.
- Organização Mundial da Saúde. **Considerações sobre saúde mental e aspectos psicossociais durante o surto de COVID-19, 18 de março de 2020**. Genebra: Organização Mundial da Saúde, 2020. Acesso em: 07 dez. 2023, (Organização Mundial da Saúde, 2020, p. xi). Disponível em: <https://iris.who.int/bitstream/handle/10665/331490/WHO-2019-nCoV-MentalHealth-2020.1-eng.pdf>.
- RIBEIRO, L. da S. et al. Efeitos da pandemia de covid-19 na saúde mental de uma comunidade acadêmica. **Acta Paulista de Enfermagem**, v. 34, 2021. Disponível em: <https://doi.org/10.37689/acta-ape/2021AO03423>. Acesso em 26 nov. 2024.
- WANG, P. et al. Prevalence and associated factors of poor sleep quality among chinese older adults living in a rural area: a population-based study. **Aging Clinical and Experimental Research**, v. 32, n. 1, p. 125–131, 2020. Disponível em: <https://acta-ape.org/en/article/covid-19-pandemic-effects-on-the-mental-health-of-an-academic-community/>. Acesso em 26 nov. 2024.
- WHO, G. S. Global status report on noncommunicable diseases 2010. Geneva (Switzerland) WHO, 2014.
- WOOLDRIDGE, J. M. **Introdução à econometria: uma abordagem moderna**. 7. ed. Boston, MA: Cengage Learning, 2023.

Apêndice

I#	X	M	Y	I#	X	M	Y
1	70	4	3	26	70	3	4
2	71	4	3	27	70	2	3
3	69	1	3	28	69	3	4
4	70	1	3	29	69	4	3
5	71	3	3	30	70	3	3
6	70	4	2	31	71	2	1
7	69	3	3	32	70	1	3
8	70	5	5	33	70	2	5
9	70	4	4	34	70	2	1
10	72	5	4	35	71	4	3
11	71	2	2	36	68	2	1
12	71	3	4	37	72	4	3
13	70	5	5	38	69	3	2
14	71	4	5	39	70	3	3
15	71	4	5	40	68	3	2
16	70	2	2	41	68	3	3
17	70	4	4	42	70	4	3
18	69	3	5	43	71	4	4
19	72	3	4	44	69	2	2
20	71	3	3	45	69	3	3
21	71	2	4	46	71	3	4
22	72	3	5	47	71	4	4
23	67	1	2	48	71	3	2
24	71	4	4	49	72	4	5
25	71	3	2	50	70	2	2

Tabela 17: Dados 1ª aplicação.

I#	X	M1	M2	Y	I#	X	M1	M2	Y
1	51	41	54	59	21	53	69	44	84
2	40	34	51	60	22	53	67	40	82
3	55	42	53	60	23	40	49	45	74
4	35	22	56	61	24	34	40	37	62
5	47	34	45	47	25	32	40	49	54
6	58	52	79	84	26	56	60	51	81
7	56	57	55	69	27	55	46	65	89
8	53	49	55	85	28	51	58	54	83
9	38	42	46	75	29	50	53	56	75
10	73	80	48	87	30	45	61	52	72
11	57	42	65	85	31	63	42	40	63
12	54	62	55	73	32	46	39	51	69
13	68	54	55	77	33	60	62	53	66
14	46	41	62	50	34	48	41	56	72
15	48	44	43	58	35	46	40	46	68
16	56	54	54	69	36	50	51	52	73
17	67	73	61	99	37	49	51	55	69
18	47	61	38	64	38	35	39	46	46
19	60	59	42	65	39	50	44	46	70
20	54	51	55	68	40	47	40	68	76

Tabela 18: Dados 2ª aplicação.