Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Luciana Schleder Gonçalves

Associação entre nutrição e determinantes sociais de saúde: people analytics para gestão da saúde

Curitiba
2024

Luciana Schleder Gonçalves

Associação entre nutrição e determinantes sociais de saúde: people analytics para gestão da saúde

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. César Augusto Taconeli

Curitiba
2024

Associação entre nutrição e determinantes sociais de saúde: people analytics para gestão da saúde

Association between Nutrition and Social Determinants of Health: People Analytics for Health Management

Luciana Schleder Gonçalves[1], César Augusto Taconeli[2]

[1]Aluna do programa de Especialização em Data Science & Big Data, lsgk@ufpr.br
[2]Professor do Departamento de Estatística - DEST/UFPR, taconeli@ufpr.br

A promoção da saúde requer uma abordagem abrangente que considere os determinantes de saúde, visando não apenas o tratamento de doenças, mas a criação de condições que permitam a todos alcançar seu máximo potencial de saúde, uma vez que têm uma relação direta com as disparidades de saúde entre diferentes grupos populacionais. OBJETIVO: Compreender os fatores sociodemográficos associados a classificações ordinais sobre Nutrição em uma base de dados do sistema MyStrengths MyHealth. MÉTODO: Regressão logística de chances proporcionais para variáveis categóricas ordinais. RESULTADOS: Foram obedecidos os pressupostos do método, como uma primeira abordagem à base de dados. Verificou-se relações entre variáveis relacionadas a raça, etnia, renda e nível educacional com a variável dependente (nutrição). CONCLUSÕES: A base de dados tem potencial de análise por diversas abordagens, futuros estudos incluem experimentos com variáveis de outros domínios e a realização de ajustes para aprimorar o modelo.
Palavras-chave: Ciência de dados, Análise de pessoas, Nutrição, Determinantes sociais de saúde, Gestão em saúde.

Health promotion requires a comprehensive approach that considers the determinants of health, aiming not only at the treatment of diseases but also at creating conditions that enable everyone to achieve their maximum health potential. This is directly related to health disparities among different population groups. OBJECTIVE: To understand the sociodemographic factors associated with ordinal classifications of Nutrition within the MyStrengths MyHealth database. METHOD: Proportional odds logistic regression for ordered categories. RESULTS: The assumptions of the method were met as an initial approach to the database. It was found that there are relationships between variables related to race, ethnicity, income, and educational level with the dependent variable (nutrition). CONCLUSIONS: The database has the potential for analysis through various approaches; future studies will include experiments with variables from other domains and making adjustments to enhance the model.
Keywords: Data science, People analytics, Nutrition, Social determinants of health, Health management.

1. Introdução

No Brasil, a busca por uma compreensão profunda do fenômeno da saúde, tanto individual quanto coletiva, historicamente vai além da simples análise de sintomas e doenças. Isso envolve uma visão holística que considera fatores interligados que influenciam o bem-estar. Entre esses fatores, destacam-se os condicionantes e determinantes de saúde, dois conceitos que desempenham papéis distintos, mas complementares, na construção de uma vida saudável.

Compreender as diferenças entre condicionantes e determinantes de saúde é fundamental para profissionais de saúde, formuladores de políticas e para a sociedade em geral. Enquanto os condicionantes enfatizam a importância das escolhas individuais e do autocuidado, os determinantes ressaltam a necessidade de abordar questões sistêmicas para criar ambientes mais saudáveis. A realização de ações de promoção da saúde requer uma abordagem abrangente que considere tanto os condicionantes quanto os determinantes, visando não apenas o tratamento de doenças, mas a criação de condições que permitam a todos alcançar seu máximo potencial de saúde. Uma questão importante para os profissionais de saúde é que os determinantes de saúde têm uma relação direta com as disparidades de saúde entre diferentes grupos populacionais. Comunidades marginalizadas, de baixa

renda ou com acesso limitado a recursos muitas vezes enfrentam determinantes adversos, resultando em condicionantes desfavoráveis. Essa conexão entre determinantes e condicionantes pode contribuir significativamente para a persistência das desigualdades de saúde [1].

Desigualdades de acesso a serviços de saúde vem sendo cada vez mais amplificadas, devido à vários motivos, o que pode levar à iniquidades, ou seja, às desigualdades consideradas injustas ou decorrentes de alguma forma de injustiça [1]. Ressalta-se a necessidade de abordagens colaborativas e globais para promover a saúde e o bem-estar de todos, considerando a diversidade das condições de vida da população. Investir em educação, sistemas de saúde resilientes, políticas de combate à pobreza e ações sustentáveis são passos cruciais para construir um futuro saudável e equitativo, tanto local quanto globalmente.

Entretanto, são poucos os métodos padronizados para medir a saúde integral das pessoas, incluindo a resiliência individual e comunitária [2]. Tem-se no Sistema Omaha, uma terminologia de saúde multidisciplinar que permite uma avaliação integral da pessoa em quatro domínios da saúde: Ambiental, Psicossocial, Fisiológico e Comportamentos Relacionados à Saúde. Consiste em três instrumentos padronizados: o Esquema de Classificação de Problemas, o Esquema de Intervenção e a Escala de Avaliação de Problemas para Resultados. Está mapeado dentro da Nomenclatura Sistemática de Medicina - Termos Clínicos (SNOMED CT) e é acessível em todas as disciplinas de saúde para permitir comunicação inclusiva e troca de dados [**?**] . Por sua vez, o sistema MyStrengths MyHealth (MSMH) foi desenvolvido para aprimorar a coleta de dados dos pacientes em serviços de saúde norte-americanos, e recentemente vem sendo traduzido para idiomas falados em diversos outros países, incluindo o português do Brasil. A terminologia nele incorporada é a do Sistema Omaha, porém numa linguagem mais simplificada, no intuito de que os próprios pacientes autorreferenciem dados da sua saúde, além dos biológicos, incluindo as dimensões física, emocional, social, ambiental e espiritual. A esta abordagem chama-se whole-person, ou voltada para cuidado integral, em que o foco está na pessoa como um todo. Assim, o MSMH organiza os organiza em domínios: My Living (minha vida); My Mind and Networks (minha mente e redes); My Body (meu corpo) e My Self-Care (meu autocuidado) [3] .

Já existe um corpo de evidências científicas que se relacionam à viabilidade de usar dados padronizados e anonimizados do sistema MSMH, relatados por pacientes com características diversas, para compreender o conceito de resiliência e da saúde integral, incluindo forças, desafios, e necessidades ([4] ; [6] ; [7] ; [8] ; [5] ; [2] . A maioria deles utiliza de ferramentas de estatística inferencial para tanto, e refletem a aplicação da ciência de dados na área da saúde, contribuindo para a compreensão dos fenômenos e para melhoria da tomada de decisões na área.

Sabe-se que a Ciência de Dados tem o compromisso de oportunizar soluções a partir do entendimento dos dados e informações relativos aos problemas concretos da área, ao aplicar o pensamento crítico e a análise para obter conhecimento. Nota-se que o foco da aplicação da Ciência de Dados na Saúde seja a ~interpretação e entendimento específico de dados do paciente, principalmente, oriundos de sua jornada de tratamento de saúde (...) e pode também ser um instrumento válido na pesquisa clínica para compreender e permitir a predição de comportamentos em seres vivos~ [9] (p.2250).

Neste contexto, dentre outras abordagens estatísticas, a abordagem de People Analytics - análise de pessoas, envolve o estudo dos comportamentos e características de pessoas ou grupos em relação a desfechos importantes para áreas de negócios ou gestão de organizações a partir de métodos que podem incluir os qualitativos e quantitativos A fim de prever esses desfechos, faz-se necessário modelá-los tanto para compreender o que os influencia quanto para potencialmente prevê-los no futuro [10] .

Salienta-se que modelar um desfecho com o objetivo principal de entender o que o influencia pode ser bastante diferente de modelar um desfecho com o objetivo principal de prever se ele ocorrerá no futuro [10] . No primeiro caso, o objetivo é aprofundar em um modelo para inferir como cada variável age sobre esse desfecho e além: saber quais variáveis são significativas ou não e quantificar sua 'explicabilidade' com base em nossas variáveis. No segundo caso, é menos importante aprofundar-se no modelo, considerando que não é preciso explicar o desfecho; mas ter confiança acerca da precisão da predição. Um modelo construído para entender um desfecho é chamado de modelo inferencial, sendo os mais conhecidos os modelos de regressão, os quais proporcionam uma ampla gama de medidas que contribuem para explicar a relação entre as variáveis de entrada e o desfecho de interesse. A questão de pesquisa para a qual se definiu um método científico para sua resolução pode ser definida

como: qual é a associação entre variáveis relacionadas aos determinantes sociais de saúde com os níveis autorreferidos de Nutrição na base de dados do sistema MyStrengths MyHealth (MSMH)?

2. Objetivo

Compreender os fatores sociodemográficos associados a classificações ordinais sobre Nutrição em uma base de dados do MSMH.

3. Materiais e métodos

Foi realizado um experimento estatístico fundamentado em People Analytics utilizando regressão logística de chances proporcionais para categorias ordenadas [10] . Frequentemente, cientistas de dados tratam o resultado como uma variável contínua e realizam regressão linear simples, o que pode levar a inferências extremamente imprecisas [10] . Dentre as outras possibilidades [10] ; [11] , tem-se modelo de odds proporcionais; modelo de odds proporcionais parciais; modelo de razão contínua modelo estereótipo. A decisão pela aplicação da regressão logística de chances proporcionais para categorias ordenadas justifica-se que esta opção pode ser considerada quando há mais de duas categorias de resultados que possuem uma ordem. Um pressuposto subjacente importante é que nenhuma variável de entrada tem um efeito desproporcional num nível específico da variável de resultado - isso é conhecido como suposição de chances proporcionais [10] . A sequência das análises deve incluir: verificar se a variável dependente é ordinal; independência das observações; ausência de multicolinearidade; construção do modelo ordinal; pressuposto das chances proporcionais; análise do modelo; resumo do modelo; p value; obtenção das odds ratio [12] ; [13].

4. Resultados

Este capítulo apresenta os resultados dos experimentos realizados para responder a cada um dos pressupostos do modelo de regressão logística de chances proporcionais (MOP) para categorias ordenadas. Para tanto, foi utilizado o software RStudio e seus pacotes.

3.1 Verificar se a variável dependente é ordinal e Independência das observações A partir da análise exploratória da base de dados, verificou-se que as variáveis presentes na base de dados são divididas em dois grupos:

3.1.1 Variáveis Demográficas (sócio-econômicas) Compreendem os conceitos de idade, gênero, raça, etnia, estado civil, escolaridade, renda, imigrante e morador de área rural.

3.1.2 Variáveis de conceitos de saúde Compreendem 42 variáveis referentes a conceitos de saúde que relacionam-se às forças dos indivíduos, representadas em quatro domínios: minha vida, minha mente e relacionamentos, meu corpo e meu auto-cuidado: - Minha Vida: Renda; Limpeza; Casa; Seguro em casa e no trabalho - Minha Mente e Redes: Conexões; Socialização; Mudança de papel; Relacionamentos; Espiritualidade ou fé (Espiritualidade); Luto ou perda; Emoções; Sexualidade; Cuidando; Negligência; Abuso; Crescimento e desenvolvimento (Crescimento) - Meu corpo: Audição; Visão; Fala e linguagem (Fala) 3.4 Saúde Bucal (Boca); Pensamento; Dor; Consciência; Pele; Movimentação; Respiração; Circulação; Digestão; 3.14 Intestino; Rins e bexiga (Urinário); Saúde Reprodutiva (Reprodutiva); Gravidez; Pós-Parto (Puerpério); Infecções - Meu Auto-Cuidado: Nutrição; Sono; Exercício (Atividade Física); Uso de substâncias (Substâncias); Planejamento Familiar (Planejamento); Cuidado à Saúde (Assistência); Medicações Estas são ordinais, são valoradas de 1 a 5 (muito ruim, ruim, regular, bom, ótimo). Quando não houve resposta foi atribuído o rótulo “NA” = “missing”. Ainda, escores 1 e 2 são considerados fraquezas e 4 e 5 forças, ou resiliência. O GRÁFICO 1 apresenta as frequências relativas dos dados referentes ao domínio Meu Autocuidado, do MSMH, que contempla a variável Nutrição

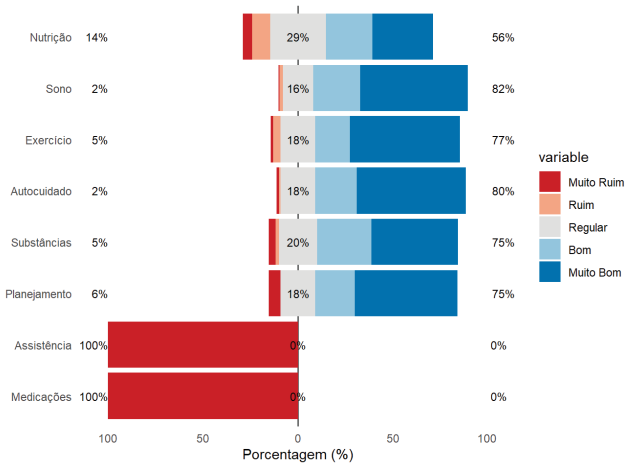

Figura 1: Dados do domínio Meu Autocuidado.

Ainda, testes de correlação de cada variável socio demográfica com dados dos domínios foram realizados. Para este experimento, optou-se por escolher como

desfecho a variável Nutrição, devido a maior frequência de escores "muito ruim"e "ruim", que referem-se à "necessidades"no MSMH, e também pelo interesse clínico e epidemiológico em compreender associações entre variáveis que possam influenciá-la [11]. Verificou-se relações entre variáveis relacionadas a raça, etnia, renda e nível educacional com a variável dependente (nutrição).

3.2 Pressuposto do Modelo das odds proporcionais (MOP) "No MOP são considerados (k – 1) pontos de corte das categorias sendo que o j-ésimo (j=1,..., k-1) ponto de corte é baseado na comparação de chances acumuladas. O termo j varia para cada uma das k categorias e cada não depende do índice j, implicando que a relação entre e Y é independente da categoria. Logo, o modelo possui uma suposição de odds proporcionais acerca dos (k-1) pontos de corte, também chamada de suposição de regressão paralela, que é assumida para cada covariável incluída no modelo. Essa suposição deve ser testada para cada caso separadamente e no modelo final, utilizando, por exemplo, o teste escore"(HOSMER, 2000 apud [11]).

3.3 Análise Gráfica da variável dependente: predição de chances À título de exemplificação, foram elaborados gráficos que apresentam a chance em atribuir o um conceito no domínio Nutrição de acordo com cada uma das variáveis sociodemográficas da base de dados. A Figura 2 exemplifica em relação a idade e estado civil; e a Figura 3, em relação a idade e escolaridade.

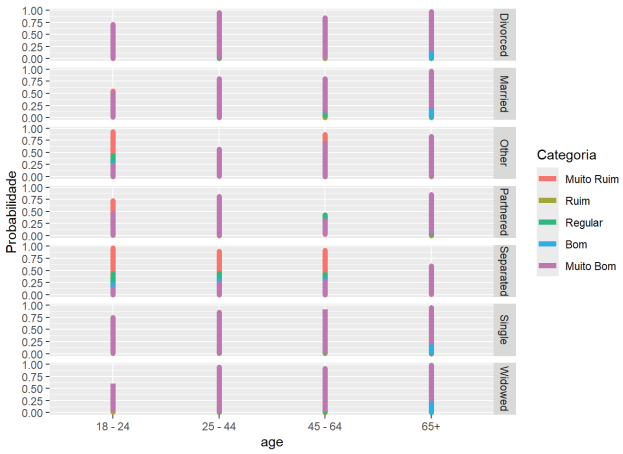

Figura 2: Predição da chance de atribuir conceito à variável Nutrição, segundo a idade e estado civil.

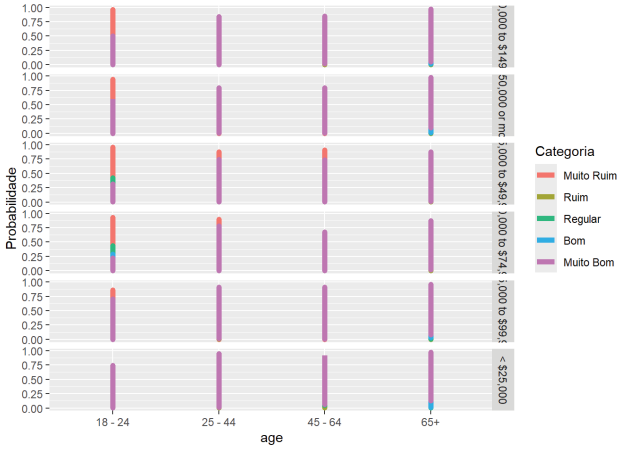

Figura 3: Predição da chance de atribuir conceito à variável Nutrição, segundo a idade e renda.

5. Conclusão

Este experimento contemplou uma primeira abordagem de análise dos dados referentes ao sistema MyStrengths MyHealth (MSMH), em que se verificou relações entre variáveis ligadas a determinantes sociais de saúde e o domínio Nutrição, que teve destaque na base de dados com escores baixos. Os mesmos experimentos realizados serão realizados para os demais domínios do sistema de informação, para nortear a tomada de decisão relacionada a saúde populacional, com base nos determinantes sociais de saúde, e não somente à luz de sintomas físicos ou biológicos. Nota-se a potencialidade para explorar ajustes e adequações a outros modelos de regressão ordinal, até mesmo de regressão linear, modelo linear generalizado, ou de análise multivariada fatorial ou de componentes principais ou agrupamentos. Futuros estudos incluem experimentos com bases de dados oriundas do sistema MSMH já traduzido para o português do Brasil, que já está sendo aplicado para reconhecimento da associação de variáveis ligadas aos determinantes sociais de saúde de modo a contribuir para o planejamento de ações de promoção de saúde pacientes brasileiros, em operadoras de planos de saúde.

6. Agradecimentos

À UFPR e aos docentes do curso de especialização em Data Science e Big Data. Ao Centro de Informática em Enfermagem da Universidade da Escola de Enfermagem da Universidade de Minnesotta. À Kelly Regina Durski Pinheiro e demais integrantes do grupo de pesquisa MSMH Brasil.

Referências

[1] ALBUQUERQUE, GSC de; SILVA, MJ S e. Sobre a saúde, os determinantes da saúde e a determinação social da saúde. Saúde Debate, Rio de Janeiro, v. 38, n. 103, p. 953-965, dez. 2014.

[2] AUSTIN RR, MATHIASON MA, LINDQUIST RA, MCMAHON SK, PIECZKIEWICZ DS, MONSEN KA. Understanding Women's Cardiovascular Health Using MyStrengths+MyHealth: A Patient-Generated Data Visualization Study of Strengths, Challenges, and Needs Differences. J Nurs Scholarsh. 2021 Sep;53(5):634-642. doi: 10.1111/jnu.12674. Epub 2021 May 16.

[3] MONSEN KA, AUSTIN RR, JONES RC, BRINK D, MATHIASON MA, EDER M. Incorporating a Whole Person Perspective in Consumer Generated Data: Social Determinants, Resilience, and Hidden Patterns. Comput Inform Nurs. 2021a Apr 8;39(8):402-410. doi: 10.1097/CIN.0000000000000730.

[4] GAO G, KERR MJ, MONSEN KA. Feasibility of Describing Wellbeing and Strengths at the Community Level Utilizing the Omaha System. Stud Health Technol Inform. 2016;225:1062-3.

[5] AUSTIN RR, LOZADA E, SECGINLI S, WERNER A, MONSEN KA, MARTIN KS. An International Research Collaborative to Examine Global Health Resilience Using the MyStrengths+MyHealth Application. Stud Health Technol Inform. 2022a Jun 6;290:1128-1129. doi: 10.3233/SHTI220301.

[6] MONSEN KA, AUSTIN RR, GOPARAJU B, CLARENCE JONES R, MATHIASON MA, PIRSCH A, EDER M. Exploring Large Community and Clinically-Generated Datasets to Understand Resilience Before and During the COVID-19 Pandemic. J Nurs Scholarsh. 2021 May;53(3):262-269. doi: 10.1111/jnu.12634.

[7] AUSTIN, RR.; MATHIASON, MA; LU, S; LINDQUIST, RA; MCMAHON, SK; PIECZKIEWICZ, DS; MONSEN, KA. Toward Clinical Adoption of Standardized mHealth Solutions: The Feasibility of Using MyStrengths+MyHealth Consumer-Generated Health Data for Knowledge Discovery. CIN: Computers, Informatics, Nursing 40(2):p 71-79, February 2022b. | DOI: 10.1097/CIN.0000000000000862

[8] RAJAMANI S, AUSTIN R, GEIGER-SIMPSON E, JANTRAPORN R, PARK S, MONSEN KA. Understanding Whole-Person Health and Resilience During the COVID-19 Pandemic and Beyond: A Cross-sectional and Descriptive Correlation Study. JMIR Nurs. 2022 May 16;5(1):e38063. doi: 10.2196/38063.

[9] VALERIO NETTO, A. Ciência de dados em saúde: contribuições e tendências para Aplicações. Rev. Saúde.Com 2021; 17(3):2249 –2253. DOI 10.22481/rsc.v17i3.6290 Disponível em https://periodicos2.uesb.br/index.php/rsc/article/view/6290/6171 01 jul 2024.

[10] MCNULTY K. Handbook of Regression Modeling in People Analytics. Chapman; Hall/CRC. Disponível em https://peopleanalytics-regression-book.org/index.html Acesso em 01 jul 2024.

[11] ABREU MNS; SIQUEIRA AL; CAIAFFA WT. Regressão logística ordinal em estudos epidemiológicos. Rev Saúde Pública 2009;43(1): 183-94.

[12] PERES F. Regressão logística ordinal (Parte 1). 2022a. Disponível em https://www.youtube.com/watch?v=wZLlL173uVc Acesso em 01 jul 2024.

[13] PERES F. Regressão logística ordinal (Parte 2). 2022b. Disponível em https://www.youtube.com/watch?v=9OYGpMIg0uo Acesso em 01 jul 2024.