

Análise de agrupamentos exploratória dos usuários do Programa Multidisciplinar de Tratamento do Tabagismo do HUAC, Campina Grande – PB

Exploratory cluster analysis of users of the Multidisciplinary Tobacco Treatment Program at HUAC, Campina Grande – PB

Análisis exploratorio de conglomerados de usuarios del Programa multidisciplinario de tratamiento del tabaco en HUAC, Campina Grande - PB

Recebido: 09/07/2020 | Revisado: 15/07/2020 | Aceito: 16/07/2020 | Publicado: 30/07/2020

Kleber Napoleão Nunes de Oliveira Barros

ORCID: <https://orcid.org/0000-0003-2515-3292>

Universidade Estadual da Paraíba, Brasil

E-mail: kleberbarros@cct.uepb.edu.br

Mácio Augusto de Albuquerque

ORCID: <https://orcid.org/0000-0002-0113-9130>

Universidade Estadual da Paraíba, Brasil

E-mail: marcioaa@uepb.edu.br

Amanda dos Santos Gomes

ORCID: <https://orcid.org/0000-0003-3456-3680>

Universidade Federal de Campina Grande, Brasil

E-mail: amanda.natalia.gomes@gmail.com

Deborah Rose Galvão Dantas

ORCID: <https://orcid.org/0000-0002-8561-054X>

Universidade Federal de Campina Grande, Brasil

E-mail: deborahdantas4@hotmail.com

Resumo

Considerando a relação entre os dados do consumo do tabagismo e as percepções das pessoas sobre os efeitos desse consumo, objetivou-se fornecer uma análise exploratória completa dos dados, utilizando técnicas de agrupamento, traçando perfis de participantes do Programa Multidisciplinar de tratamento do Tabagismo (PMTT). A partir das 117 variáveis de 714 usuários do PMTT produzidas do questionário de avaliação médica do paciente original, foram selecionadas 32 variáveis relevantes. Foram filtrados 495 usuários que continham informações nas 32 variáveis selecionadas, sendo 5 variáveis quantitativas, 22 qualitativas

nominais e 5 qualitativas ordinais. Para estes participantes, foi empregada análise de agrupamentos. O menor entrelaçamento médio foi utilizado para escolher a medida de dissimilaridade empregada conjuntamente com o método de Ward. Foram identificados três grupos distintos de tabagistas. Os resultados da análise agrupamento confirmam a presença de três grupos distintos de usuários do PMTT. Em acréscimo, o estudo indicou haver padrões diferenciados entre variáveis sexo, religião, renda e histórico de enfermidades, além de diferenças de perfis tabagistas. O trabalho apresenta considerações sobre a responsabilidade das pessoas na prevenção e redução das taxas de consumo entre a sua população. Os resultados obtidos traçam perfis de fumantes que possuem consideráveis tempos de tabagismo e altos consumos diários de cigarros. Tais perfis ressaltam a importância de políticas públicas abrangentes e articuladas, de preferência preventivas, apostando na educação e promoção da saúde ou no incremento da qualidade de vida.

Palavras-chave: Tabagismo; Promoção da saúde; Agrupamento.

Abstract

Considering the relationship between data on smoking consumption and people's perceptions of the effects of smoking, the objective was to provide a more complete exploratory analysis of the data, using grouping techniques, drawing profiles of participants in the Multidisciplinary Tobacco Treatment Program (PMTT). From the 117 variables of 714 PMTT users produced from the original patient's medical evaluation questionnaire, 32 variables relevant to this study were selected. 495 users were filtered which contained information on the 32 selected variables, being 5 quantitative, 22 nominal qualitative and 5 ordinals qualitative variables. For these participants, cluster analysis was used. The lowest mean interlacing was used to choose the dissimilarity measure used in conjunction with Ward's method. Three distinct groups of smokers were identified. The results of the cluster analysis confirm the presence of three distinct groups of PMTT users. In addition, the study indicated that there are different patterns between variables sex, religion, income, and history of illnesses, in addition to differences in smoking profiles. The work presents considerations about people's responsibility in preventing and reducing consumption rates among their population. The results obtained trace profiles of smokers who have considerable smoking times and high daily cigarette consumption. Such profiles underscore the importance of comprehensive and articulated public policies, preferably preventive, focusing on education and health promotion or increasing the quality of life.

Keywords: Smoking; Health promotion; Grouping.

Resumen

Teniendo en cuenta la relación entre los datos sobre el consumo de tabaco y las percepciones de las personas sobre los efectos del hábito de fumar, el objetivo era proporcionar un análisis exploratorio más completo de los datos, utilizando técnicas de agrupación, dibujando perfiles de participantes en el Programa Multidisciplinario de Tratamiento del Tabaco (PMTT). De las 117 variables de 714 usuarios de PMTT producidas, se seleccionaron 32 variables relevantes para este estudio. Se filtraron 495 usuarios que contenían información sobre las 32 variables seleccionadas, siendo 5 variables cuantitativas, 22 cualitativas nominales y 5 cualitativas ordinales. Para estos participantes, se utilizó el análisis de conglomerados. El entrelazado medio más bajo se utilizó para elegir la medida de disimilitud utilizada junto con el método de Ward. Se identificaron tres grupos distintos de fumadores. Los resultados del análisis de conglomerados confirman la presencia de tres grupos distintos de usuarios de PMTT. Además, el estudio indicó que existen diferentes patrones entre las variables sexo, religión, ingresos e historial de enfermedades, además de las diferencias en los perfiles de tabaquismo. El trabajo presenta consideraciones sobre la responsabilidad de las personas para prevenir y reducir las tasas de consumo entre su población. Los resultados obtuvieron perfiles traza de fumadores que tienen tiempos considerables de fumar y un alto consumo diario de cigarrillos. Dichos perfiles destacan la importancia de políticas públicas integrales y articuladas, preferiblemente preventivas, centradas en la educación y la promoción de la salud o en el aumento de la calidad de vida.

Palabras clave: De fumar; Promoción de la salud; Agrupamiento.

1. Introdução

Iniciado em 2007, o Programa Multidisciplinar de Tratamento do Tabagismo (PMTT), vinculado ao Curso de Medicina do CCBS, atende por ano cerca de 400 usuários tabagistas do município de Campina Grande e circunvizinha. A equipe é multiprofissional, multidisciplinar e interdisciplinar, composta por docentes e discentes de diversos cursos da área da saúde da UFCG e de outras universidades parceiras.

Cada grupo de cerca de 30 novos usuários assistem a uma palestra inicial, que abrange os males associados ao tabagismo e explica todo o funcionamento do PMTT. Posteriormente, cada um deles é entrevistado e avaliado pelas equipes que integram o Programa, recebendo assistência multiprofissional. A partir de indicação médica, os usuários recebem o medicamento bupropiona, em quantidade suficiente para 15 dias, quando serão reavaliados

por todas as equipes, nos retornos em grupo. O tratamento dura 12 semanas (6 encontros denominados de retornos), sendo os usuários acompanhados e incentivados a participarem de todas as atividades. Os dados da entrevista são anotados no questionário de avaliação médica do Paciente. Os dados dos retornos são registrados na Ficha de Retorno.

O objetivo deste trabalho é reunir os indivíduos de uma amostra da população estudada por técnicas de análise de agrupamento. O resultado dos agrupamentos pode ser influenciado pela escolha da medida de similaridade ou dissimilaridade, bem como pela definição do número de grupos (Albuquerque, T. et al., 2016; Albuquerque, M. et al., 2016).

O agrupamento é realizado de forma a minimizar as diferenças homogêneas entre os indivíduos em estudo dentro do agrupamento, e maximizar as diferenças heterogêneas entre os indivíduos de agrupamentos diferentes (Albuquerque, M. et al., 2016).

Desta forma, neste trabalho se objetivou fornecer uma análise exploratória mais completa dos dados, a técnicas de agrupamento e de validação dos grupos, e aumentar o conhecimento que pode ser obtido com a aplicação de um conjunto de sentenças lógicas em análise de agrupamento.

2. Metodologia

2.1 Tratamento dos dados

Conforme Pereira et al. (2018), este estudo se apresenta como trabalho descritivo de ênfase no método quantitativo. Detalhes da metodologia aplicada podem ser encontrados em Albuquerque & Barros (2020).

A partir das 117 variáveis de 714 usuários do PMTT produzidas do questionário de avaliação médica do paciente original, foram selecionadas 32 variáveis relevantes para este estudo. A relevância das variáveis foi avaliada pelos critérios: variáveis não pertencentes se à seção informações pessoais; quantidade não superior 20% de dados faltantes; informações completas para ambos os sexos; correlação significativa pelo menos 10% com a variáveis número de cigarros fumados diários. Por último foram filtrados 495 usuários que continham informações nas 32 variáveis selecionadas, sendo 5 variáveis quantitativas, 22 qualitativas nominais e 5 qualitativas ordinais. Neste trabalho, estão sendo consideradas apenas as variáveis qualitativas.

2.2 Procedimentos de Análise dos Dados

Em uma primeira exploração dos dados, para agrupar indivíduos, é necessário a definição de uma medida de similaridade ou dissimilaridade. Com base nessas medidas os indivíduos similares são agrupados e os demais são colocados em grupos separados. Existem várias medidas que podem ser utilizadas como medidas de distâncias ou dissimilaridade entre elementos de uma matriz de dados (Cormack, 1971).

A distância mais utilizada, para dados em escala intervalar ou de razão, em análise de agrupamento é distância euclidiana (Cornish, 2007). Os atributos, por sua vez, devem ser transformados para a escala binária.

Na escala binária, o número 1 indica a presença da característica e 0, a ausência. Variáveis que possuam maior coincidência de presenças e ou ausências das características serão mais semelhantes entre si. Para cada par de indivíduos são contabilizadas as coincidências/discordâncias com a seguinte notação:

a: número de coincidências do tipo 1 – 1 para cada par de características;

b: número de discordâncias do tipo 1 – 0 para cada par de características;

c: número de discordâncias do tipo 0 – 1 para cada par de características;

d: número de coincidências do tipo 0 – 0 para cada par de características;

$p = a + b + c + d$: número total de características binários estudados.

Observe na Tabela 1, um resumo de algumas das medidas de similaridade utilizadas neste trabalho (Gower & Legendre, 1986). Um coeficiente de similaridade permite concordância negativa quando são levadas em consideração concordâncias do tipo 0–0, isto é, d é incluído no coeficiente. A terceira coluna da Tabela 1 informa se o coeficiente de concordância negativa, o d, está incluída ou não no respectivo coeficiente de similaridade. Também são exibidos os intervalos de variação na quarta coluna.

Tabela 1: Coeficientes de similaridade analisados (Gower & Legendre, 1986).

Coeficientes	Expressão da Similaridade	Concordância Intervalo Negativa	
D1: <u>Jaccard</u>	$s_1 = a/(a+b+c)$	Excluída	[0;1]
D2: <u>Concordância simples</u>	$s_2 = (a+d)/(a+b+c+d)$	Incluída	[0;1]
D3: <u>Sokal e Sneath I</u>	$s_3 = a/(a+2(b+c))$	Excluída	[0;1]
D4: <u>Rogers e Tanimoto</u>	$s_4 = (a+d)/(a + 2(b + c) + d)$	Incluída	[0;1]
D5: <u>Dice ou Sorensen</u>	$s_5 = 2a/(2a + b + c)$	Excluída	[0;1]
D6: <u>Hamann</u>	$s_6 = (a - b - c + d)/(a + b + c + d)$	Incluída	[-1;1]
D7: <u>Ochiai</u>	$s_7 = a/\sqrt{(a + b)(a + c)}$	Excluída	[0;1]
D8: <u>Sokal e Sneath II</u>	$s_8 = ad/\sqrt{(a + b)(a + c)(d + b)(d + c)}$	Incluída	[0;1]
D9: <u>Phi de Pearson</u>	$s_9 = (ad - bc)/\sqrt{(a + b)(a + c)(d + b)(d + c)}$	Incluída	[-1;1]
D10: <u>Gower e Legendre</u>	$s_{10} = a/(a+b+c+d)$	Excluída	[0;1]

Fonte: Autores.

A técnica de agrupamento utilizado foi o hierárquico aglomerativo, isto é, cada novo grupo formado contém os grupos das etapas anteriores (Johnson & Wichern, 2002; Albuquerque & Barros, 2020). Nos métodos aglomerativos mais utilizados (Kaufman & Rousseeuw, 2009; El-Hamdouchi & Willett, 1989), método do vizinho mais próximo (single linkage), método do vizinho mais distante (complete linkage), método da distância média (UPGMA), método do centroide (UPGMC), método de Ward (UPGMC). As técnicas aglomerativas permitem uma representação visual conhecida como dendrograma. Quando dois dendrogramas são justapostos e as terminações ligadas, gera-se um emaranhamento ou entrelaçamento que pode ser mensurado. Quanto menor o entrelaçamento, maior a concordância entre duas medidas de dissimilaridade e esta ideia pode ser utilizada para escolher entre as medidas da Tabela 1. Para definir o número de grupos de um agrupamento utilizamos o método do gráfico do cotovelo (*elbow method*).

Para escolher entre as medidas da Tabela 1 quanto menor o entrelaçamento, maior a concordância entre duas medidas de dissimilaridade.

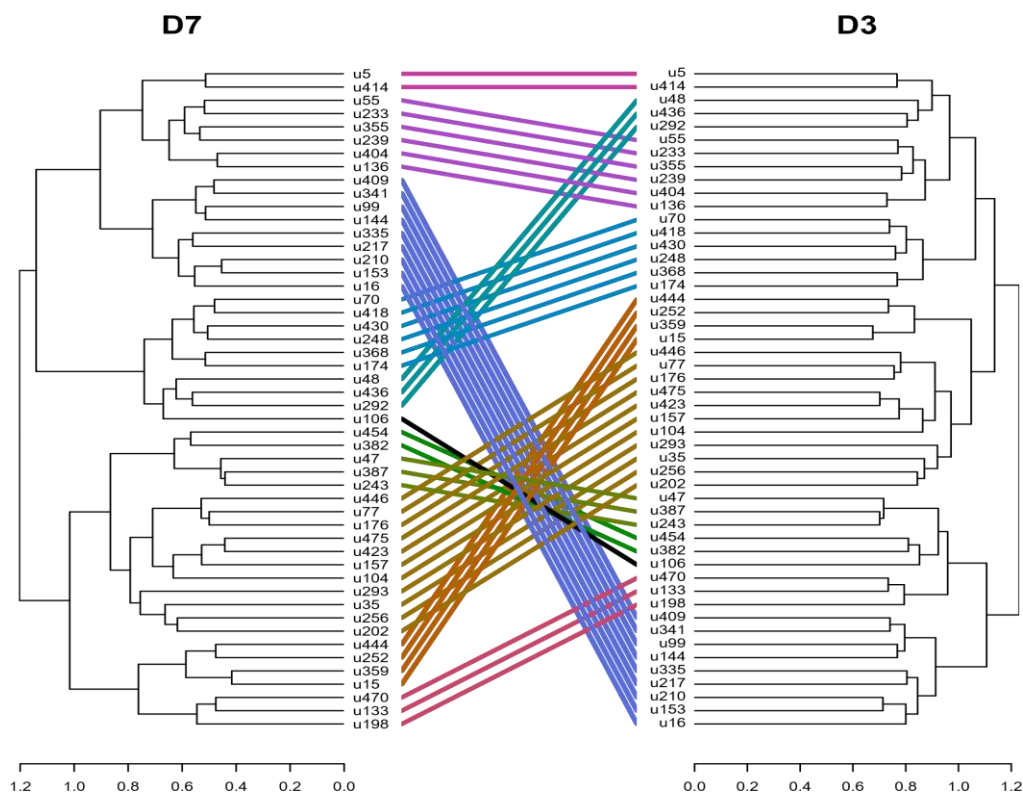
3. Resultados

Todas as análises deste capítulo foram realizadas com o software R (R Core Team, 2014). Os dendrogramas e medidas cruzados foram construídos a partir do pacote dendextended. Para a realização da análise de agrupamentos hierárquicos, foram selecionadas 27 variáveis qualitativas nominais e ordinais. As 22 variáveis nominais geraram 98 variáveis binárias, enquanto as 5 variáveis ordinais produziram 30 variáveis binárias. No total, a matriz de entrada para a análise de agrupamentos teve 495 linhas por 128 colunas. A partir deste ponto, escolheu-se o método de Ward para agrupar os indivíduos, uma vez que este método

leva a grupos com tamanhos mais homogêneos.

As medidas de entrelaçamento para 45 combinações possíveis de 2 dendrogramas para as 10 medidas de dissimilaridade obtidas a partir dos coeficientes de similaridades foram computadas. Na Figura 1 é ilustrado o cruzamento entre os dendrogramas para as medidas de dissimilaridade de Ochiai (D7) e de Sokal e Sneath (D3) a partir de 50 usuários selecionados aleatoriamente dos 495 usuários em estudo. Utilizando os 495 usuários houve um cruzamento de 22% entre os dois dendrogramas, o que indica que eles concordam em parte. Diversos usuários que estão juntos em D7 continuam próximos em D3, porém agrupados em outra posição.

Figura 1: Dendrogramas das dissimilaridades de Ochiai (D7) e de Sokal e Sneath (D3).



Fonte: Autores.

Observa-se na Tabela 2 estão as proporções de entrelaçamentos entre todas as combinações de dendrogramas para os 495 usuários. As medidas de entrelaçamento menores que 30% estão destacadas. A maior discordância acontece entre os coeficientes de Jaccard (D1) e de Sokal & Sneath (D3), enquanto não há entrelaçamento entre os coeficientes de concordância simples Sokal e Michener (D2) e de Hamann (D6).

Tabela 2: Entrelaçamento para os dendrogramas das dissimilaridades D1 a D10.

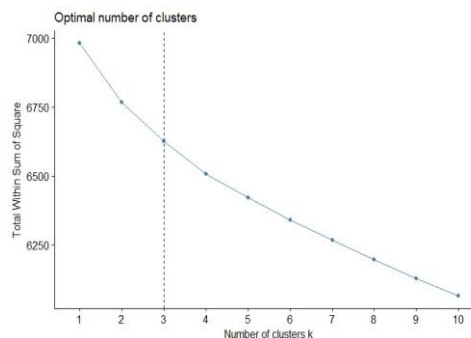
	D1	D2	D3	D4	D5	D6	D7	D8	D9
D2	0,282								
D3	0,532	0,403							
D4	0,338	0,286	0,253						
D5	0,292	0,278	0,383	0,341					
D6	0,248	0,000	0,314	0,205	0,357				
D7	0,259	0,280	0,221	0,232	0,248	0,310			
D8	0,289	0,355	0,355	0,286	0,240	0,250	0,285		
D9	0,360	0,368	0,358	0,295	0,310	0,371	0,313	0,275	
D10	0,306	0,381	0,473	0,359	0,226	0,371	0,338	0,265	0,354

Fonte: Autores.

Para escolher a melhor medida de dissimilaridade foram calculados os cruzamentos medianos obtidos com base na Tabela 2. O menor valor acontece para o coeficiente de Ochiai (D7) com entrelaçamento mediano igual a 0,269. Assim, devido ao seu grau de concordância com as demais medidas, escolheu-se este coeficiente D7 para as inferências.

O método do cotovelo foi empregado para a determinação da quantidade de grupos a partir das 27 variáveis nominais e ordinais e 495 usuários. Ele indicou que se deve escolher três grupos, conforme Figura 2.

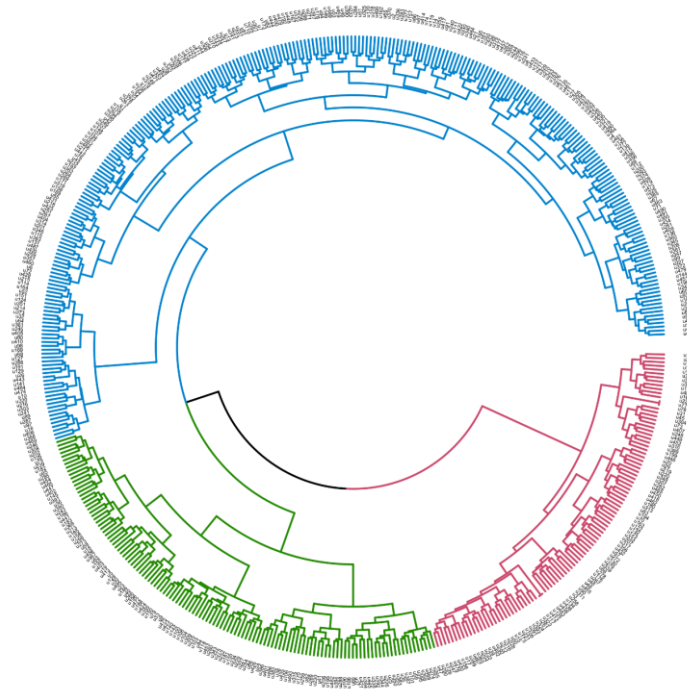
Figura 2: Determinação do número de grupos a serem utilizados.



Fonte: Autores.

A partir dos 495 usuários foi construído o dendrograma utilizando coeficiente de Ochiai, o método de Ward com três grupos, resultando na Figura 3. O procedimento apontou para uma divisão de 102 indivíduos para o primeiro grupo (G1, em vermelho), 122 usuários para o segundo grupo (G2, em verde) e 271 indivíduos no terceiro grupo (G3, em azul). Conhecendo os indivíduos é possível construir um perfil para cada grupo.

Figura 3: Dendrograma para os 495 usuários selecionados do PMTT.



Fonte: Autores.

As principais características sociodemográficas e relacionáveis ao tabagismo dos participantes dado estudo são apresentadas na Tabela 3. As entradas da tabela são divididas pelos grupos encontrados pelo dendrograma. Para as variáveis contínuas foram informadas as medianas (Md) e os intervalos quantílicos de confiança do tipo: [q2,5%; q97,5%]. Para as variáveis discretas são exibidas as quantidades de indivíduos com as características (n) e as respectivas porcentagens (%). As maiores quantidades e proporções entre os grupos são destacadas na Tabela 3. Características sociodemográficas e relacionáveis ao tabagismo dos participantes dado estudo do PMTT.

Tabela 3: Informações sociodemográficas e relacionáveis ao tabagismo da amostra.

Variável	G1		G2		G3	
	Md [q _{2,5%} ; q _{97,5%}]		Md [q _{2,5%} ; q _{97,5%}]		Md [q _{2,5%} ; q _{97,5%}]	
Idade	50,9 [23,6; 72,6]		52,4 [30,0; 77,0]		54,8 [25,3; 77,3]	
Tempo de Tabagismo	30,5 [3,5; 58,0]		34,5 [14,0; 61,0]		35,0 [6,0; 62,3]	
Cigarros/dia	20 [5,5; 44,8]		20 [3,0; 60,0]		20 [3,0; 41,0]	
Xicaras de café/dia	4,5 [0; 20]		5,0 [0; 20]		3 [0; 20]	
	n	(%)	n	(%)	n	(%)
Sexo:						
Feminino	37	36,3	63	51,6	241	88,9
Masculino	65	63,7	59	48,4	30	11,1
Religião:						
Católica	55	53,9	99	81,1	197	72,7
Evangélica	16	15,7	19	15,6	54	19,9
Outra/não possui	31	30,4	4	3,3	20	7,4
Cor:						
Preta	8	7,8	18	14,8	23	8,5
Parda	60	58,8	72	59,0	152	56,1
Branca	33	32,4	30	24,6	89	32,8
Amarela/Indígena	1	1,0	2	1,6	7	2,6
Estado civil:						
Casado	58	56,9	56	45,9	103	38,0
Solteiro	29	28,4	37	30,3	87	32,1
Outro	15	14,7	29	23,8	81	29,9
Atividade Profissional:						
CLT	29	28,4	22	18,0	30	11,1
Autônomo	23	22,5	32	26,2	59	21,8
Servidor público	17	16,7	9	7,4	22	8,1
Dona de Casa	3	2,9	24	19,7	57	21,0
Desempregado	7	6,9	7	5,7	0	0,0
Aposentado	16	15,7	20	16,4	50	18,5
Outro	23	22,5	8	6,6	53	19,6
Escolaridade						
0 - 5 anos	8	7,8	42	34,4	88	32,5
6 - 12 anos	40	39,2	57	46,7	120	44,3
>= 13 anos	54	52,9	23	18,9	63	23,2
Renda familiar mensal (R\$)						
0 - 1000	7	6,9	64	52,5	146	53,9
1001 - 2500	53	52,0	46	37,7	94	34,7
>= 2501	42	41,2	12	9,8	31	11,4
Consumo após palestra:						
Aumentou	4	3,9	7	5,7	11	4,1
Diminuiu/Parou	45	44,1	75	61,5	123	45,4
Sem alteração	53	52,0	40	32,8	137	50,6
Mora com fumante:						
Não	75	73,5	88	72,1	183	67,5
Sim	27	26,5	34	27,9	88	32,5
Influenciado ao tabagismo por:						
Curiosidade	18	17,6	19	15,6	38	14,0
Mãe	4	3,9	9	7,4	19	7,0
Pai	9	8,8	12	9,8	40	14,8
Amigos/Colegas	43	42,2	53	43,4	107	39,5
Demais pessoas/motivos	28	27,5	19	15,6	67	24,7
Drogas ilícitas:						
Usuário	3	2,9	3	2,5	4	1,5
Ex-usuário	14	13,7	17	13,9	13	4,8
Nunca usou	85	83,3	102	83,6	254	93,7
Comorbidades:						
hipertensão arterial sistólica	20	19,6	29	23,8	105	38,7
Diabetes mellitus	1	1,0	16	13,1	28	10,3
Artrite	11	10,8	8	6,6	89	32,8
Gastrite	40	39,2	30	24,6	116	42,8
Doença na tireoide	1	1,0	5	4,1	27	10,0
Exercício Físico (vzs/semana)						
0 - 1	80	78,4	35	28,7	227	83,8
2 - 3	8	7,8	15	12,3	22	8,1
4 - 5	5	4,9	35	28,7	12	4,4
>= 5	9	8,8	37	30,3	10	3,7

Fonte: Autores.

O conglomerado G1 é formado pelos usuários mais jovens com idade mediana 50,9 anos. O tempo de tabagismo também é o menor 30,5 anos. A quantidade mediana de cigarros fumados por dia para o grupo foi 20, como os demais agrupamentos, porém com intervalo de confiança mais restrito, o que indica uma menor variabilidade. Os homens são maioria neste grupo (36,3%). A categoria apresenta a maior porcentagem de pessoas que não são católicas ou evangélicas (30,4%). Apenas 7,8% das pessoas deste grupo se identificaram como negras, enquanto 58,8% informaram a cor parda e 32,4% se identificaram como brancas. Os casados formam 56,9% dos participantes desse grupo (maior proporção entre os grupos). G1 também apresenta as maiores porcentagens de pessoas que informaram ter carteira assinada (28,4%). Os autônomos e os servidores públicos formam, respectivamente, 22,5% e 16,7% das atividades profissionais informadas em G1. Proporcionalmente, este grupo apresenta a maior renda: 42% declararam ganhar entre R\$ 1001,00 e R\$ 2500,00, enquanto 52% informaram ganhar R\$ 2500,00 ou mais; e a maior escolaridade: 59,2% afirmaram possuir 13 ou mais anos de estudo. 52% dos entrevistados não tiveram alteração do consumo de cigarros após a palestra. Os menores índice de pessoas que convivem com fumantes se encontram neste grupo (26,5%). 42,2% foram levados ao tabagismo por amigos ou colegas, enquanto 17,6% foram induzidos pela curiosidade. Dentre os participantes do grupo, 2,9% se identificaram usuários de drogas. As menores porcentagens de hipertensos e diabéticos se encontram em G1: 19,6% e 1%, respectivamente. 78,4% afirmaram não praticar exercício físico ou praticar apenas uma vez na semana.

No grupo G2, a idade mediana foi 52,4 anos. O tempo mediano de tabagismo foi 34,5 anos. Subtraindo as duas quantidades acima, verifica-se que neste grupo, os participantes iniciaram o hábito tabagista medianamente mais cedo (aos 17,9 anos). Apesar dos participantes também utilizarem, na mediana, 20 cigarros diários, há uma maior variabilidade nas respostas com respostas variando desde $q_{2,5\%} = 3$ a $q_{97,5\%} = 60$ cigarros diários. O agrupamento G2 ingere mais café (mediana de 5 xícaras). O grupo é uniforme quanto ao sexo: 51,6% mulheres e 48,4% homens. A religião predominante é a católica: 81,1%. As maiores porcentagens de autodeclarados pardos e negros se encontram nesta classe: 14,8% e 59%, respectivamente. Os casados e os solteiros somam 45,9% e 30,3%, respectivamente. As atividades profissionais mais informadas foram: autônomo (26,2%), CLT (18%) e dona de casa (19,7%). As menores escolaridades foram informadas pelas pessoas do grupo G2: 34,4% têm entre 0 e 5 anos de estudo, já 46,7% possuem 6 a 12 anos escolares. 52,5% dos usuários do grupo possuem renda familiar mensal entre zero e mil reais. Após a palestra, as pessoas deste grupo reportaram ambas as maiores porcentagem de aumento (5,7%) e de diminuição

(61,5%) do consumo de cigarros. 27,9% do grupo G2 relataram morar com outros fumantes. As maiores influencias provenientes do exemplo da mãe (7,4%) e de amigos/colegas (43,4%) se deram no grupamento. Ex-usuários de drogas compõem 13,9% do grupo. A maior porcentagem de diabéticos (13,1%) e uma grande proporção de hipertensos (23,8%) pertencem a G2. Um menor percentual de sedentário foi encontrado neste grupo: 12,3% afirmaram praticar atividades físicas 2 a 3 vezes na semana, 28,7% declararam que realizam atividade física de 4 a 5 vezes e 30,3% informaram praticar exercícios físicos 5 ou mais vezes na semana.

O grupo G3 é o mais idoso: mediana de 54,8 anos. O tempo mediano de tabagismo do grupo foi 35 anos. Eles consomem Md = 20 cigarros ao dia. Os indivíduos de G3 consomem menos café: Md = 3 xícaras por dia. Nota-se que a classe é majoritariamente composta por mulheres (88,9%). A maioria dos evangélicos (19,9%) estão agrupados em G3. Os que se declararam pardos e brancos formam 56,1% e 32,8%, respectivamente. A maioria dos solteiros está em G3 (32,1%). As donas de casa (21%) e os aposentados (18,5%) se destacam no grupo. Além disso, o algoritmo juntou todos os desempregados nos demais grupos. A escolaridade é intermediária neste grupo. A renda familiar é a mais baixa: 53,9% recebem entre zero e mil reais; 34,7% recebem de R\$ 1001,00 a R\$ 2500,00. Após a palestra 45,4% afirmaram que houve diminuição no consumo. Já 50,6% informaram que não houve alteração. Os usuários de G3 apresenta a maior taxa de convívio com outros fumantes (32,5%). O conglomerado apresenta a maior influência do pai, quando comparado com os demais grupos: 14,8%. A absoluta maioria daqueles que afirmaram nunca terem utilizado drogas (93,7%) na vida foram cadastrados no grupo. G3 ainda registra os maiores percentuais da maioria das comorbidades analisadas: hipertensão (38,7%); artrite (32,8%); gastrite (42,8%); e problema de tireoide (10%). Além disso, tem os maiores números absolutos de diabetes mellitus (28 participantes). Finalmente, G3 apresenta o menor percentual de exercitação: 83,8% realizam atividade física, no máximo, uma vez por semana.

4. Considerações Finais

Foram empregadas técnicas de análise de agrupamentos hierárquicos, a partir das quais foi possível classificar os usuários em três grupos com características diferentes. A investigação das principais diferenças entre os grupos formados, revelou que

- o perfil do grupo G1 é majoritariamente de homens, mais novos, com menor tempo de

tabagismo, casados, servidores públicos, com maior escolaridade e renda, indiferentes a palestra, não moram com outros fumantes e menores índices de comorbidades;

- G2 é composto amplamente por pessoas de ambos os sexos, de idade e tempo tabagista intermediários, maior variabilidade no consumo de cigarro, católicos, pardos ou pretos, autônomos, menos anos de estudo, renda familiar intermediária, maior sensibilidade à palestra, foram mais influenciados por amigos ou colegas para iniciar o tabagismo, com maiores índices de diabetes e maior atividade física;
- O perfil G3 é sobretudo formado por mulheres, de maior idade e tempo de tabagismo, autônomas, donas de casa ou aposentadas, escolaridade intermediária, menor renda familiar, não usuárias de droga, com maiores relatos de comorbidades e menor nível de atividade física.

Caram et al. (2009) relatam estudo relativo de pessoas que procuram programa de tratamento. As idades dos participantes, próximas dos 50 anos, são concordantes com este estudo. A maior procura ocorre por mulheres, embora este trabalho foi capaz de estratificar os gêneros, colocando a maioria dos homens num grupo, outro grupo com paridade dos sexos e um terceiro com predominância de mulheres.

Existem influências específicas do gênero feminino que contribuem para o início do tabagismo, como o fato de a mulher associar mais comumente o ato de fumar com estados de ânimo negativos (depressão, ansiedade, estresse); uma maior carga de trabalho atribuída à mulher na sociedade, como a necessidade imposta de conciliação do trabalho fora de casa e o cuidado do lar e dos filhos; uma maior sensibilidade e expectativa em relação à vida, o que leva a frustrações e estados de ânimo negativos e o fator peso corporal, principalmente em mulheres (Otero, 2008). O desbalanceamento na procura (Caram et al. 2009) demonstra um maior inconformismo com a situação entre as mulheres.

Relacionamento entre alto grau de tabagismo e menor poder aquisitivo entre pessoas idosas também foi reportado por Senger et al. (2011).

Um estudo com adultos que iniciaram o tabagismo ainda na infância demonstrou que a iniciação ao tabaco foi fortemente associada à normalidade sociocultural e familiar do ato de fumar; identificação e aprendizado com figuras parentais fumantes; condições adversas de vida, incluindo trabalho infantil e poucas atividades lúdicas; e escassez de informações (Calsavara & Fontanella, 2007). Ocorrendo principalmente em espaços de socialização, fora do espectro familiar (Panaino, Soares & Campos, 2014). Traços deste perfil também são identificados aqui, onde parte dos participantes iniciaram o vício em torno dos 20 anos,

tiveram exemplo dos familiares mais próximos ou de amizades como gatilho para se tornarem fumantes.

Comorbidades estão sempre associadas em pessoas que procuram tratamento do tabagismo (Caram et al. 2009 & Bush et al., 2016). Além das comorbidades o algoritmo foi capaz de separar o grupo com maiores índices de doenças e relacionar com níveis diferenciados de atividade física.

Fica como sugestão para trabalhos futuros, analisar o impacto conjunto que os grupos encontrados e das variáveis analisadas provocam na probabilidade de cessação do tabagismo e do abandono do tratamento dos usuários do programa, por meio de modelo de regressão logístico.

Referências

Albuquerque, M. A., & Barros, K. N. N. O. (2020). *Introdução à Análise de Agrupamento: teoria e prática com aplicações em R*. [e-book]. Campina Grande. Ed. EDUEPB. Disponível em: <http://eduepb.uepb.edu.br/download/introducao-a-analise-de-agrupamento-teoria-e-pratica-com-aplicacoes-em-r/?wpdmdl=997&masterkey=5e97904980fc9>

Albuquerque, M. A., Barros, K. N. N. O., Gouveia, J. F., & Ferreira, R. L. C. (2016). Determination and validation of group numbers in a cluster analysis: A case study applied to forestry science. *Acta Scientiarum. Technology*, 38(3), 339-344.

Albuquerque, T. M., Araujo, G. A. B., Caminha, B. L., Albuquerque, M. L., & Albuquerque, M. A. (2016). Measures of association in epidemiological studies: smoking mothers and low birth weight children in the city of Campina Grande–PB. *Acta Scientiarum. Health Sciences*, 38(2), 179-84.

Bush, T., Lovejoy, J. C., Deprey, M., & Carpenter, K. M. (2016). The effect of tobacco cessation on weight gain, obesity, and diabetes risk. *Obesity*, 24(9), 1834-41.

Calsavara, T.V.S., & Fontanella, B.J.B. (2007). Uso de tabaco iniciado na infância: relatos de adultos em tratamento. *Jornal Brasileiro de Psiquiatria*, 56(4), 252-9.

Caram, L. M. D. O., Ferrari, R., Tanni, S. E., Coelho, L. S., Godoy, I. D., Martin, R. D. S., & Godoy, I. D. (2009). Perfil de fumantes atendidos em serviço público para tratamento do tabagismo. *Jornal Brasileiro de Pneumologia*, 35(10), 980-5.

Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society: Series A (General)*, 134(3), 321-53.

Cornish, R. (2007). *Statistics: cluster analysis*. Mathematics Learning Support Centre, 1-5.

El-Hamdouchi, A., & Willett, P. (1989). Comparison of hierarchic agglomerative clustering methods for document retrieval. *The Computer Journal*, 32(3), 220-7.

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of classification*, 3(1), 5-48.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

Otero, M. (2004). Tabaquismo en la mujer: consideraciones especiales. *Trastornos adictivos*, 6(2), 113-124.

Panaino, E. F., Soares, C. B., & Campos, C. M. S. (2014). Context of the beginning of tobacco use in different social groups. *Revista Latino-Americana de Enfermagem*, 22(3), 379-385.

Pereira, A. S., Shitsuka, D. M., Parreira, F. J., & Shitsuka, R. (2018). *Metodologia da pesquisa científica*. [e-book]. Santa Maria. Ed. UAB/NTE/UFMS. Disponível em: https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Viena, Austria. Disponível em: <http://www.R-project.org/>

Senger, A. E. V., Ely, L. S., Gandolfi, T., Schneider, R. H., Gomes, I., & Carli, G. A. (2011). Alcoolismo e tabagismo em idosos: relação com ingestão alimentar e aspectos socioeconômicos. *Revista Brasileira de Geriatria e Gerontologia*, 14(4), 713-719.

Porcentagem de contribuição de cada autor no manuscrito

Kleber Napoleão Nunes de Oliveira Barros – 30%

Mácio Augusto de Albuquerque – 30%

Amanda dos Santos Gomes – 20%

Deborah Rose Galvão Dantas – 20%