

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ICEEx – Instituto de Ciências Exatas
Curso de Especialização em Estatística

**DEFINIÇÃO DE GRAU DE COMPLEXIDADE DE IDOSOS FRAGILIZADOS
CAPTADOS POR SERVIÇO DE MEDICINA PREVENTIVA DE OPERADORA
PRIVADA DE PLANOS DE SAÚDE POR MEIO DE ANÁLISE DE
AGRUPAMENTOS**

Camila Rachel Camargo Takemori

Belo Horizonte
Agosto, 2014

Camila Rachel Camargo Takemori

**DEFINIÇÃO DE GRAU DE COMPLEXIDADE DE IDOSOS FRAGILIZADOS
CAPTADOS POR SERVIÇO DE MEDICINA PREVENTIVA DE OPERADORA
PRIVADA DE PLANOS DE SAÚDE POR MEIO DE ANÁLISE DE
AGRUPAMENTOS**

Trabalho de conclusão do curso
de Especialização em
Estatística.

Orientadora: Profa. Dra. Sueli
Aparecida Mingoti.

Belo Horizonte

Agosto, 2014

RESUMO

Partindo-se de banco de dados de 481 idosos já identificados como fragilizados por serviço de medicina preventiva de operadora privada de planos de saúde no Brasil, foram aplicados métodos estatísticos de Análise de *Cluster*, utilizando-se uma série de indicadores de saúde, coletados em avaliação presencial domiciliar realizada por enfermeiro. O objetivo é particionar os idosos em grupos, a partir de variáveis que indicam o estado de saúde geral de cada um deles, para que sejam definidos níveis de complexidade, de maneira que os idosos de cada um dos grupos formados recebam um plano de cuidados adequado à necessidade, constituído por atendimentos domiciliares e telemonitoramentos, cuja frequência e quantidade sejam fundadas no princípio da equidade.

Palavras-chave: Análise de *Cluster*, Indicadores de Saúde, Idosos Fragilizados, Serviço de Medicina Preventiva.

ABSTRACT

Cluster Analysis was applied to the database of 481 elderly already identified previously as frail beneficiaries of private operator of health plans in Brazil, using a range of health indicators, listed in household face assessment by nurse. The goal is to find the partition of the elderly into groups based on variables that indicate the general health condition of each person, in order to define levels of complexity, so that the elderly of each formed groups could receive the appropriate care plan that they really need, consisting of homecare and telehomecare, whose frequency and quantity are founded on the principle of equity.

Keywords: Cluster Analysis, Health Indicators, Frailty Elderly, Preventive Medicine Service.

SUMÁRIO

1	INTRODUÇÃO.....	6
2	OBJETIVO.....	8
3	METODOLOGIA.....	8
4	BANCO DE DADOS.....	10
5	ANÁLISE ESTATÍSTICA DOS DADOS.....	15
5.1	Análise Descritiva.....	15
5.2	Análise de Associação entre Variáveis.....	18
5.3	Análise de Agrupamentos.....	20
5.3.1	Análise de Agrupamentos Hierárquico.....	21
5.3.2	Análise de Agrupamentos não Hierárquico – k-Médias.....	22
5.3.3	Análise de Agrupamentos não Hierárquico – k-Modas.....	22
5.4	Discussão dos Resultados.....	22
5.4.1	Análise de Agrupamentos desconsiderando-se a variável “Limitação de força em membros superiores”.....	29
5.4.2	Análise da variável quantitativa “Idade” no melhor agrupamento obtido com k=3.....	32
6	CONSIDERAÇÕES FINAIS.....	34
	REFERÊNCIAS BIBLIOGRÁFICAS.....	37
	ANEXO I – Fórmulas Matemáticas da Medida de Distância e dos Métodos de Agrupamentos Utilizados.....	39
1.	Relação entre distância Euclidiana ao quadrado e coeficiente de dissimilaridade para dados binários.....	39
2.	Descrição dos Métodos de Agrupamentos utilizados na monografia ...	40
2.1.	Método de Ligação Simples (<i>Single Linkage</i>).....	40
2.2.	Método de Ligação Completa (<i>Complete Linkage</i>).....	40
2.3.	Método de Ligação das Médias (<i>Average Linkage</i>).....	40
2.4.	Método de ligação de Ward.....	41
2.5.	Método das k-Médias.....	41
2.6.	Método k-Modas.....	42
	ANEXO II – Tabelas Contendo os Resultados dos Testes estatísticos Qui-quadrado Apresentados na Seção 5.2.....	44
	ANEXO III – Distribuição de Frequências, Vetor de Modas e Dissimilaridade por <i>Cluster</i> para Cada Valor de k e Método de Agrupamento Utilizados.....	47

1 INTRODUÇÃO

A empresa “A”¹ é uma empresa que provê a organizações de saúde em âmbito nacional serviços de gestão do cuidado através do acesso ao conhecimento, geração de informação, utilização de tecnologias de comunicação e informações adequadas. A metodologia de gestão do cuidado comercializada por essa empresa propõe um conjunto organizado de práticas assistenciais à carteira de pacientes acompanhados em programas de promoção à saúde e prevenção de agravos, para que a equipe de saúde possa atender aos pacientes de acordo com a necessidade apresentada pelos mesmos.

Verifica-se no Brasil atualmente o fenômeno chamado de transição demográfica, que constitui no aumento da expectativa de vida e progressivo envelhecimento da população. O aumento no número de idosos implica também no aumento da prevalência de condições crônicas, superando o número de doenças infecto-contagiosas, o que é conhecido por transição epidemiológica (ARAÚJO, 2012). Devido a esse fato explica-se o interesse em realização de ações de medicina preventiva voltada à população de idosos.

O programa de medicina preventiva a ser considerado no presente trabalho será o programa de atenção a idosos fragilizados, cujo perfil de pacientes identifica-se como população portadora de várias doenças crônicas, consumo de grande volume de medicamentos de uso contínuo e portadora de diferentes graus de dependência na realização de atividades básicas da vida diária. Esse conjunto de condições clínicas constituinte da síndrome de fragilidade no idoso indica condição de alto risco para quedas, hospitalização e re-hospitalização, incapacidade, institucionalização e morte (FRIED *et al.*, 2001). O objetivo de realização de um programa de medicina preventiva para esse público com forte tendência de crescimento e acúmulo de condições crônicas reside no fato de estabelecer melhoria da qualidade de vida e prevenir complicações de agravos já instalados na população de idosos, o que traz como consequência a redução dos custos em saúde.

¹ Por questões éticas, o nome da empresa foi preservado.

Para o efetivo acompanhamento e atendimento racional dos idosos fragilizados é apropriado que seja realizada a estratificação dos mesmos em níveis de complexidade, de acordo com a gravidade do quadro clínico apresentado. Devido ao fato de a condição de idoso fragilizado caracterizar-se como uma síndrome, resultante de várias características clínicas, torna-se complexa e muitas vezes incorreta a definição de graus de complexidade para esse perfil de pacientes, ao serem utilizadas técnicas empíricas. A utilização de técnicas estatísticas de agrupamento para definição dos graus de complexidade visa à partição da população de idosos fragilizados em grupos que apresentem condições semelhantes, para o efetivo atendimento das reais necessidades dos participantes do programa, bem como ao uso racional dos recursos humanos disponíveis nos serviços de medicina preventiva.

Não existe na literatura critérios pré-estabelecidos para a definição desses graus de complexidade voltados à população de fragilizados, uma vez que o conceito de fragilidade em idosos ainda é um conceito relativamente recente (TEIXEIRA, 2008). Atualmente, a atribuição dos graus de complexidade aos pacientes não é um serviço oferecido pela empresa “A” aos seus clientes como parte integrante do projeto, embora a classificação em níveis de complexidade seja ponto chave para definição do plano de cuidados ao qual o paciente será direcionado, e a empresa “A” proponha, em sua metodologia, três planos de cuidados distintos², de acordo com a gravidade em que o paciente se encontra, conforme Tabela 1.

Tabela 1 – Frequência de consultas e telemonitoramentos preconizados pela empresa A por nível de complexidade

Nível de complexidade	Consultas	Telemonitoramentos
Alta complexidade	semanal	semanal
Complexidade intermediária	semestral	mensal
Baixa complexidade	anual	semestral

A definição dos níveis de complexidade constitui ponto de atenção do projeto de implantação de gestão do cuidado, por ser uma atividade a ser realizada pelo cliente, de maneira não orientada e sistematizada.

² A empresa “A” sugere essa definição aos clientes no plano de projeto, porém alguns ajustes podem existir em função da capacidade operacional disponível.

2 OBJETIVO

Definir grau de complexidade de idosos de operadora privada de planos de saúde participantes de programa de atenção a idosos fragilizados, para que sejam definidas estratégias de ação a cada um dos grupos obtidos de acordo com o respectivo grau de complexidade atribuído. Para isso serão utilizados métodos estatísticos de Análise de Agrupamentos Hierárquicos e não Hierárquicos.

3 METODOLOGIA

A Análise de Agrupamentos, também conhecida como Análise de *Cluster* ou de Conglomerados, é um método estatístico multivariado utilizado com o propósito de separar elementos amostrais em grupos, de maneira que os elementos alocados em um mesmo grupo possuam maior similaridade possível entre si e maior dissimilaridade com os elementos constituintes de grupos diferentes, em relação às mesmas variáveis consideradas (MINGOTI, 2005).

O grau de semelhança entre os grupos para a definição e composição dos mesmos é obtido por meio do cálculo das distâncias matemáticas entre os vetores de dados dos elementos amostrais quando as variáveis usadas na comparação são quantitativas, ou de coeficientes de similaridades no caso de variáveis categóricas (MINGOTI, 2005).

Os métodos de agrupamentos são divididos em duas classes: a constituída pelos algoritmos Hierárquicos e aquela constituída pelos não Hierárquicos. Conforme Everitt (2011), Johnson e Wichern (2007) e Mingoti (2005) (apud MATOS, 2007), “os algoritmos Hierárquicos diferem dos não Hierárquicos basicamente em dois aspectos: no agrupamento Hierárquico, uma vez que observações ou grupos de observações foram combinados, eles não se separarão.” O segundo aspecto, ainda segundo a autora, consiste no fato de que o tipo não Hierárquico permite que esses elementos sejam trocados de um grupo para outro, durante as etapas constituintes do algoritmo. Outra diferença entre os procedimentos reside na escolha do número de grupos. Para o primeiro tipo de método, uma forma de se estimar o número de grupos da partição é identificar pontos em que exista decréscimo acentuado na similaridade dos

conglomerados unidos (MINGOTI, 2005); e para o segundo, esse número deve ser definido previamente à análise. Quando comparado com o método Hierárquico, o método não Hierárquico é computacionalmente mais rápido porque não é necessário calcular e armazenar, durante o processamento, a matriz de similaridade (BUSSAB, 1990).

Os métodos Hierárquicos podem ainda ser categorizados em aglomerativos e divisivos (HAIR *et al.*, 2005). O algoritmo aglomerativo se inicia com a quantidade de grupos igual ao número de elementos a serem agrupados ($k = n$) e termina com todos eles em um único grupo ($k = 1$). O método divisivo, ao contrário, começa com todos os objetos unidos e que são separados até que cada objeto esteja em um único grupo (MATOS, 2007). O que caracteriza os processos aglomerativos é que a reunião de dois grupos numa certa etapa do algoritmo é o resultado da união de dois grupos da etapa anterior, caracterizando a propriedade de hierarquia. Existem vários critérios usados para definir a distância (ou similaridade) entre grupos. Cada critério gera um método de agrupamento. Em cada etapa do algoritmo de agrupamento procura-se fundir grupos que tenham a menor distância entre si, ou a maior similaridade entre si (HAIR *et al.*, 2005). Nos métodos Hierárquicos, o número de grupos (k) não precisa ser pré-especificado. O método pode ser usado como uma ferramenta para estimação do número de grupos da partição a partir da análise estatística do histórico de agrupamento.

Nos métodos não Hierárquicos o número de grupos da partição deve estar previamente determinado. Sendo k o número de grupos desejado, a primeira etapa é a seleção das sementes de inicialização do algoritmo de agrupamento (k centróides iniciais ou pontos de partida para agrupamentos). No primeiro passo do algoritmo, os elementos amostrais são comparados às sementes, gerando o primeiro agrupamento. As sementes são então atualizadas de acordo com a partição obtida e os elementos amostrais são novamente comparados um a um com as novas sementes. Esse processo se repete e é interrompido quando se obtém a melhor partição final, de acordo com o critério estabelecido (HAIR *et al.*, 2005). Um dos métodos não Hierárquicos mais conhecidos é o k -Médias (HAIR *et al.*, 2005), apropriado para ser utilizado com variáveis numéricas, pois em cada passo do algoritmo cada elemento da amostra é comparado com o vetor de médias de cada um dos k grupos formados na partição do passo anterior, sendo

alocado no grupo cuja distância seja a menor possível, repetindo-se esse procedimento até que não haja mais nenhuma realocação a ser feita de elementos nos grupos formados. Nesse método a atualização das sementes em cada passo do algoritmo é feita através do cálculo do vetor de médias amostral de cada grupo formado na partição.

É comum usar também o método k-Médias para situações em que as variáveis são categóricas ordinais, ou então categóricas binárias (MATOS, 2007). Embora o método k-Médias possa ser usado para agrupar dados categóricos binários, existem métodos de agrupamentos que são específicos para variáveis categóricas. Um dos métodos não Hierárquicos mais conhecido e que foi utilizado nessa monografia é o k-Modas proposto por Huang em 1997, e que apresenta três modificações em relação ao algoritmo das k-médias: uso de discordância simples como medida de dissimilaridade, modas em vez de médias, e um método baseado em frequências para atualizar essas modas em cada passo do algoritmo de agrupamento.

Para implementação dos métodos não Hierárquicos é necessário ter-se as k sementes de inicialização do algoritmo. Existem vários métodos propostos para a escolha das sementes (Mingoti, 2005; Everitt, 2011). Dois métodos muito utilizados são o de escolha aleatória, isto é, sorteia-se k elementos aleatoriamente do banco de dados e mantém-se os vetores de observações desses elementos como as sementes; ou vetores de médias (ou modas) provenientes de partições obtidas por métodos Hierárquicos para o valor de k pré-fixado. A escolha aleatória é feita automaticamente em alguns *softwares*, podendo ser realizada uma única vez ou várias vezes.

Detalhes matemáticos referentes aos métodos de Análise de Agrupamentos podem ser verificados no Anexo I.

4 BANCO DE DADOS

A empresa “A” disponibiliza aos clientes, como parte integrante da metodologia de gestão de cuidados comercializada, instrumentos de avaliação clínica embarcados em *software* próprio, para serem aplicados por equipe multiprofissional, durante o atendimento do paciente. Foram selecionadas para a

análise apresentada nessa monografia algumas variáveis do instrumento direcionado ao perfil de idosos fragilizados, consideradas como críticas para definição do grau de complexidade de cada um dos pacientes, com vistas à realização de plano piloto de Análise de Agrupamentos a ser proposta.

As variáveis originais selecionadas constituintes do banco de dados, com respectivos tipos e categorias, podem ser visualizadas na Tabela 2.

Tabela 2 – Descrição, tipo e categorias associadas às variáveis selecionadas no banco de dados utilizado

Variável	Tipo	Categorias
Idade	numérica	-
Úlcera de pressão	categórica	sim / não
cuidados paliativos	categórica	sim / não
Sintoma sistema circulatório	categórica	sim / não
Apresenta dispneia	categórica	quando descansa / atividade física mínima / atividade física moderada / quando anda mais que uma quadra ou sobe escadas / nunca
Uso de cigarros	categórica	sim / não
Paciente acamado	categórica	sim / não
Limitação de força	categórica	membro inferior direito / membro inferior esquerdo / membro superior direito / membro superior direito / não
Distúrbio de marcha	categórica	sim / não
Classificação de risco nutricional (DUARTE e DIOGO, 2000)	categórica	alto risco nutricional / risco nutricional moderado / bom

Para a realização da análise proposta, algumas variáveis foram modificadas, com a finalidade de se tornarem binárias, sem prejuízo do conteúdo, assumindo a seguinte formatação final:

- **Idade** – foi transformada em **Idade categorizada**, com os estratos: ‘menor que 80 anos’ e ‘80 anos ou mais’. O ponto de corte definido em 80 anos

está de acordo com o proposto pelos autores Moraes *et al.* (2008), Camarano (2002) e Chaimowicz (1997), entre outros, e também com os resultados do teste estatístico qui-quadrado realizado entre a variável idade categorizada e demais variáveis, conforme descrito no Anexo II.

- **Úlcera de pressão** – manteve-se conforme original. A úlcera de pressão (ou úlcera por pressão) caracteriza-se por lesão ocasionada por pressão não aliviada, cisalhamento ou fricção, que ocasione comprometimento cutâneo, geralmente em região de proeminências ósseas, podendo resultar em morte tecidual (MEDEIROS *et al.*, 2009).

- **Cuidados paliativos** – manteve-se conforme original. Por cuidados paliativos entende-se a condição de fragilidade e vulnerabilidade extrema, na qual o paciente é portador de doença que ameaça a vida, e os cuidados direcionados a esse tipo de paciente devem promover prevenção e alívio do sofrimento, controle da dor, além de tratamento de sintomas físicos, sociais, psicológicos e espirituais, conforme definido pela Organização Mundial da Saúde (WHO, 2002).

- **Sintoma sistema circulatório** – manteve-se conforme original. Os sintomas do sistema circulatório mais comuns são: dor no peito, dor na nuca, dor nas pernas, entre outros, ou qualquer outro sintoma que o profissional de saúde avaliador julgar como sintoma relacionado ao sistema circulatório.

- **Apresenta dispneia** – foi modificada para **Dispneia em repouso**, com as respostas ‘sim’ e ‘não’, contemplando apenas a situação de dispneia que representa maior gravidade e que é maior indicativo de fragilidade, que é a situação de presença de dificuldade para respirar ou “falta de ar” mesmo durante o repouso.

- **Uso de cigarros** – manteve-se conforme original. A justificativa de inclusão dessa variável consiste no fato de o hábito tabágico implicar em grave comprometimento das funções cardiovasculares e respiratórias, sobretudo em idosos (LIMA e FAUSTINO, 2013).

- **Paciente acamado** – manteve-se conforme original. Os diferentes motivos que originaram o confinamento do paciente idoso ao leito é forte indicativo de fragilidade, por ocasionar grande dependência de cuidador(es) na realização das atividades diárias, como por exemplo alimentação e higiene pessoal.

- **Limitação de força** – como para essa questão é possível a existência de mais de uma opção de resposta (por exemplo: um paciente pode apresentar limitação de força em mais de um membro), ela foi transformada em duas variáveis:

- **Limitação de força em membros superiores** - com as respostas 'sim' e 'não';

- **Limitação de força em membros inferiores** - com as respostas 'sim' e 'não'.

Essa divisão justifica-se pelo fato de não haver distinção de gravidade no fato de a limitação acontecer no membro esquerdo ou direito, e sim, no fato de a limitação ser nos membros inferiores ou superiores, uma vez que limitação em membros inferiores implica prejuízo da locomoção, o que indica maior dependência e gravidade quando comparada a limitação em membros superiores, que implica em limitação de outras atividades que não a locomoção (COLLINS *et al.*, 2004; SHEPHARD e BOUCHARD, 1994; *apud* JUNIOR e GUERRA, 2008).

- **Distúrbio de marcha** - manteve-se conforme original. A presença de distúrbio de marcha implica em maior risco para quedas, que é alto preditor de fragilidade em idosos (REZENDE *et al.*, 2012; PIOVESAN, 2011).

- **Classificação do risco nutricional** – foi transformada na variável '**Alto risco nutricional**', com as respostas 'sim' e 'não', uma vez que a gravidade do risco nutricional encontra-se na classificação igual a alto.

As implicações em relação à presença das variáveis consideradas nas análises apresentadas nessa monografia estão relatadas brevemente na Tabela 3.

Tabela 3 – Consequências da presença das variáveis analisadas em relação à saúde

Condição	Implicação em relação à saúde
Idade igual ou maior que 80 anos	aumento do estado de debilidade geral do organismo, devido ao processo natural de envelhecimento
Presença de úlcera de pressão	expõe o paciente a maior susceptibilidade de infecção
Cuidados paliativos	estado de fragilidade geral extrema, sem prognóstico de tratamento
Sintoma do sistema circulatório	risco aumentado de ocorrência de evento cardiovascular final (por exemplo: infarto e acidente vascular cerebral, entre outros)
Dispneia em repouso	indicativo de quadro grave de patologia do sistema respiratório
Uso de cigarros	repercute em comprometimento das funções cardiovasculares e respiratórias
Paciente acamado	implica em alto grau de dependência (exige presença 24h de cuidador) e risco aumentado de desenvolvimento de úlcera de pressão
Limitação de força em membros inferiores	implica em prejuízo da locomoção
Limitação de força em membros superiores	impõe restrições na realização de atividades da vida diária
Alto risco nutricional	implica em comprometimento do sistema imunológico, na situação de desnutrição; e em aumento do risco de presença de eventos cardiovasculares (por exemplo: infarto agudo do miocárdio e acidente vascular cerebral), na situação de obesidade
Distúrbio de marcha	implica em risco aumentado para quedas

A partir do exposto anteriormente, foram onze as variáveis binárias consideradas nas Análises de Agrupamentos realizadas para a definição dos níveis de complexidade em idosos fragilizados:

1. **Idade categorizada;**
2. **Úlcera de pressão;**
3. **Cuidados paliativos;**
4. **Sintoma sistema circulatório;**
5. **Dispneia em repouso;**
6. **Uso de cigarros;**
7. **Paciente acamado;**
8. **Limitação de força em membros superiores;**
9. **Limitação de força em membros inferiores;**
10. **Distúrbio de marcha;**
11. **Alto risco nutricional.**

O banco de dados da análise constituiu-se de 481 beneficiários idosos de uma única operadora privada de plano de saúde do estado do Rio Grande do Sul, incorporados pela medicina preventiva em programa de atenção a idosos fragilizados.

5 ANÁLISE ESTATÍSTICA DOS DADOS

5.1 Análise Descritiva

As variáveis selecionadas para Análise de Agrupamentos e distribuição de frequências das respectivas respostas, bem como o histograma dos valores de idade dos pacientes considerados estão expostos na Tabela 4 e Figura 1, para fins de análise exploratória. Todos os elementos da amostra considerada apresentaram respostas para todas as variáveis em análise.

Tabela 4 – Distribuição de frequências das variáveis consideradas para Análise de Agrupamentos

Variável	Resposta	Frequência absoluta	Frequência relativa (%)
Idade categorizada	60 a 79 anos	300	62
	80 anos ou mais	181	38
Úlcera de pressão	não	445	93
	sim	36	7
Cuidados paliativos	não	355	74
	sim	126	26
Sintoma sistema circulatório	não	381	79
	sim	100	21
Dispneia em repouso	não	445	93
	sim	36	7
Tabagismo	não	332	69
	sim	149	31
Paciente acamado	não	439	91
	sim	42	9
Limitação de força em membros inferiores	não	292	61
	sim	189	39
Limitação de força em membros superiores	não	189	39
	sim	292	61
Distúrbio marcha	não	358	74
	sim	123	26
Alto risco nutricional	não	284	59
	sim	197	41

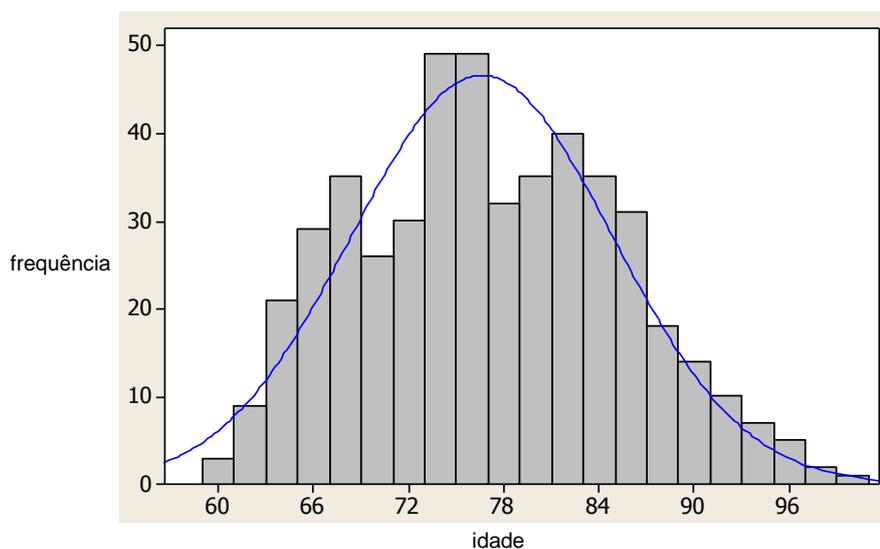


Figura 1 – Histograma da variável quantitativa “Idade” na amostra

Foi realizado teste de normalidade de Anderson-Darling para a variável “Idade”, que, sob hipótese nula de normalidade dos dados, resultou num p-valor $<0,005$, e conforme padrão observado no histograma assume distribuição não normal. As estatísticas descritivas para idade são as seguintes: mínimo de 60 anos, máximo de 99 anos, mediana de 76 anos, média de 76,61 anos, desvio-padrão de 8,24 anos, coeficiente de assimetria (*skewness*) 0,18 e de curtose (*kurtosis*) -0,59. Verifica-se ainda no histograma, a existência de 3 modas, uma em torno de 68 anos, outra por volta de 75 e anos e a terceira em torno de 82 anos. Sobre tal constatação, foi realizada breve análise da relação entre a variável quantitativa “Idade” em cada um dos grupos formados na Análise de Agrupamentos com o melhor método obtido, para $k=3$ (ver seção 5.4.2).

Em relação à distribuição de frequências das variáveis categóricas, verificou-se que, para 10 das 11 variáveis consideradas, a maioria dos pacientes considerados não apresentou a condição clínica relacionada à respectiva variável. Exceção foi feita para a variável “Limitação de força em membros superiores”, presente em 61% dos pacientes. Uma das possíveis causas para esse valor tão elevado pode ser devido ao fato dessa informação ser autorreferida, não havendo no momento da avaliação clínica realizada pelo profissional de saúde alguma mensuração de força dos membros superiores por meio de algum teste que utilize o peso corporal ou qualquer aparelho para esse fim. O impacto nas respostas dos respondentes da maneira autorreferida usada para coleta dessa informação pode sofrer influência dos seguintes aspectos: a forma como o profissional avaliador repassa a questão para o paciente; o entendimento do paciente a respeito da definição do que seja limitação de força em membros superiores; a situação psicológica do paciente (verifica-se frequentemente que pacientes poliqueixosos tendem a enxergar o próprio estado de saúde de uma maneira mais negativa do que realmente é). É possível ainda que quem responda vários itens da avaliação clínica do paciente seja um familiar próximo, mas que não acompanha o paciente por um período de tempo suficiente para poder definir a presença ou não de limitação de força em membros superiores.

Devido à alta prevalência encontrada para a variável “Limitação de força em membros superiores”, indicando provável erro na obtenção do dado, foi realizada também Análise de Agrupamentos desconsiderando-se essa variável,

somente com o método eleito como o melhor dentre os utilizados nessa monografia, análise que consta na seção de discussão dos resultados (ver seção 5.4.1).

5.2 Análise de Associação entre Variáveis

Foi realizado o teste estatístico qui-quadrado (SOARES e SIQUEIRA, 2002) para avaliar a possível existência de associações significativas entre algumas das variáveis selecionadas. A 5% de significância, foi encontrada associação estatisticamente significativa entre “Idade categorizada” e as seguintes variáveis: “Cuidados paliativos”, “Uso de cigarros”, “Paciente acamado”, “Limitação de força em membros inferiores”, “Limitação de força em membros superiores” e “Distúrbio de marcha” (ver resultados pormenorizados dos testes no Anexo II).

Tabela 5 – Teste estatístico qui-quadrado entre “Idade categorizada” e demais variáveis

Variável	Valor da estatística do teste	p-valor
Úlcera de pressão	3,936	0,047*
Cuidados paliativos	21,350	0,000*
Sintoma sistema circulatório	0,143	0,705
Dispneia em repouso	0,770	0,380
Uso de cigarros	7,076	0,008*
Paciente acamado	9,399	0,002*
Limitação de força em membros inferiores	28,864	0,000*
Limitação de força em membros superiores	28,864	0,000*
Distúrbio de marcha	28,430	0,000*
Risco nutricional	0,548	0,459

(*) significativo ao nível de significância de 5%

Da Tabela 5 depreende-se que nesta população a prevalência de cuidados paliativos, paciente acamado, limitação de força em membros inferiores, e

distúrbio de marcha é maior no grupo dos indivíduos com 80 anos ou mais, caracterizando o quadro de debilidade maior nos idosos mais velhos.

Outras associações importantes foram verificadas entre “Uso de cigarros” e “Dispneia em repouso”, “Paciente acamado” e “Limitação de força em membros inferiores”, “Distúrbio de marcha” e “Limitação de força em membros inferiores”, “Paciente acamado” e “Úlcera de pressão”, conforme Tabela 6, e serão discutidas na sequência (ver Anexo II).

Tabela 6 – Teste estatístico qui-quadrado entre pares de variáveis escolhidas

Variáveis	Valor da estatística do teste	p-valor
“Uso de cigarros” e “Dispneia em repouso”	4,803	0,028*
“Paciente acamado” e “Limitação de força em membros inferiores”	19,922	0,000*
“Distúrbio de marcha” e “Limitação de força em membros inferiores”	55,042	0,000*
“Paciente acamado” e “Úlcera de pressão”	0,008	0,930

(*) significativo ao nível de significância de 5%

Um resultado interessante encontrado no banco de dados analisado foi que as variáveis “Paciente acamado” e “Úlcera de pressão” não mostraram associação significativa, contrariando a literatura de saúde que indica maior prevalência e susceptibilidade de úlcera de pressão em pacientes restritos ao leito. Tal fato pode ser resultado do pequeno número de pacientes acamados dentro da população (8,73%), e sinaliza que, para esta amostra, as ações educativas de prevenção à úlcera por pressão, deve ser dirigida à toda amostra e não apenas aos pacientes acamados.

Observou-se associação existente entre “Paciente acamado” e “Limitação de força em membros inferiores”, conforme esperado, uma vez que tanto a limitação de força em membros inferiores leva à situação de acamado, quanto a própria situação de acamado leva à limitação de força em membros inferiores.

Uma relação análoga existe entre as variáveis “Distúrbio de marcha” e “Limitação de força em membros inferiores”, uma vez que tanto a limitação de

força em membros inferiores leva ao distúrbio de marcha, quanto o distúrbio de marcha leva à limitação de força em membros inferiores.

5.3 Análise de Agrupamentos

Nesta monografia, foram realizadas análises estatísticas envolvendo métodos de Análise de *Cluster* Hierárquicos e não Hierárquicos. Como já existe na metodologia proposta pela empresa “A” a definição de estratificação em 3 níveis de cuidados para os idosos fragilizados, com as devidas ações de cuidado dirigidas a cada um desses estratos, nessa monografia foi utilizado inicialmente o valor de $k=3$ como número de grupos. Além disso foi utilizado também o valor $k=4$, para verificar a influência da inserção de mais um grupo ao modelo de análise.

Na Tabela 7 é exibido o vetor de modas das variáveis de análise para o conjunto dos 481 elementos amostrais considerados, e o respectivo número de dissimilaridades, para efeito de comparação com as Análises de Agrupamentos que foram realizadas. Como número de dissimilaridades, entenda-se o número de elementos com resposta diferente da moda da variável respectiva; e como dissimilaridade total, o somatório de todas as dissimilaridades para cada variável.

Tabela 7 – Vetor de modas e número de dissimilaridades para o conjunto dos 481 pacientes selecionados para análise

Variáveis	Moda	nº de dissimilaridades(*)
Idade	0	181
Úlcera de pressão	0	36
Cuidados paliativos	0	126
Sintoma sistema circulatório	0	100
Dispneia em repouso	0	36
Uso de cigarros	0	149
Paciente acamado	0	42
Limitação de força em membros inferiores	0	189
Limitação de força em membros superiores	1	189
Distúrbio de marcha	0	123
Alto risco nutricional	0	197
DISSIMILARIDADE TOTAL	-	1.368

(*) dissimilaridade em relação à moda da variável

A maior parte das informações técnicas e detalhes relacionados aos métodos de agrupamentos utilizados que serão abordados a seguir foram alocados no Anexo III, com o intuito de conferir maior fluidez na leitura desta monografia.

5.3.1 Análise de Agrupamentos Hierárquicos

Na Análise de Agrupamentos através de métodos Hierárquicos foram utilizados inicialmente os métodos de ligação simples, ligação completa, ligação das médias das distâncias e ligação de Ward (EVERITT 2011). Os detalhes matemáticos de cada um desses métodos constam nos itens 2.1 a 2.4 do Anexo I. A dissimilaridade média encontrada para cada um dos métodos de ligação para o caso de $k=3$ está exibido na Tabela 8, considerando-se as 11 variáveis descritas nas Tabelas 3 e 7.

Tabela 8 – Dissimilaridade média por método de agrupamento Hierárquico - $k=3$

Método de ligação	Dissimilaridade média(*)
Ligação simples	449,00
Ligação completa	307,33
Ligação das médias	323,33
Ligação de Ward	306,67

(*) Dissimilaridade média: somatório de todas as dissimilaridades por *cluster* dividido pelo número de *clusters* da partição. O número de dissimilaridades em cada *cluster* é calculado em relação ao vetor de modas do *cluster*.

Considerando-se que quanto menor o valor da dissimilaridade melhor é a qualidade da partição resultante, o método no qual obteve-se melhor resultado e o que foi aqui considerado como representante do modelo Hierárquico, foi o da ligação de Ward, embora o valor esteja próximo do obtido no método de ligação média. Todos os métodos foram implementados considerando-se a distância Euclidiana ao quadrado como medida de comparação entre elementos e *clusters*, em função do fato de haver uma relação entre o coeficiente de dissimilaridade e a distância Euclidiana ao quadrado no caso de dados binários (ver item 1 do Anexo I).

O vetor de modas e o número de dissimilaridades total por *cluster* encontrados para $k=3$ e $k=4$ na Análise de Agrupamentos pelo método Hierárquico da ligação de Ward constam nas Tabelas A e B do Anexo III.

5.3.2 Análise de Agrupamentos não Hierárquico – k-Médias

Para este método, foram utilizadas sementes de inicialização aleatórias apenas uma vez e sementes obtidas pela partição resultante do método Hierárquico de ligação de Ward, ou seja, os k vetores de modas produzidos pelo método de Ward, conforme resultado obtido na seção 5.3.1. O vetor de modas e os valores de dissimilaridade por *cluster* encontrados para cada um dos valores de $k=3$ e $k=4$, na Análise de Agrupamentos não Hierárquicos utilizando-se o método das k -médias constam nas Tabelas C a F do Anexo III.

5.3.3 Análise de Agrupamentos não Hierárquico – k-Modas

Para este método, também foram utilizadas sementes de inicialização aleatórias geradas apenas uma única vez pelo próprio *software* estatístico utilizado³ e sementes obtidas pelas partições resultantes do método Hierárquico de ligação de Ward e pelo método Hierárquico k -Médias, conforme resultados obtidos nas seções 5.3.1. e 5.3.2. O vetor de modas e os valores de dissimilaridade média e por *cluster* encontrados para $k=3$ e $k=4$ na Análise de Agrupamentos não Hierárquicos produzido pelo método das k -Modas constam na Tabela 10 e nas Tabelas G a M do Anexo III.

5.4 Discussão dos Resultados

Nesta seção será realizada a comparação dos valores de dissimilaridade de todas as Análises de Agrupamentos com $k=3$ e $k=4$, apresentados na seção 5.3, conforme Tabela 9.

Para $k=3$, o método que apresentou menor valor de dissimilaridade isoladamente foi o k -Modas com sementes de inicialização provenientes do

³ Para a análise do k -Modas foi utilizado o *software* estatístico “R”, pacote “klaR” (2013).

método k-Médias (sementes aleatórias), que foi o método com o menor percentual de dissimilaridade em relação à dissimilaridade total sem partição. O segundo melhor método para k=3 foi o método não Hierárquico k-Médias, com sementes aleatórias.

Ao ser adicionado mais um grupo na análise (k=4), o método que apresentou menor valor de dissimilaridade foi o k-Médias com sementes de inicialização aleatórias, seguido do método k-Modas com sementes de inicialização provenientes do método k-Médias (sementes da ligação de Ward).

Como esperado, a inclusão de mais um grupo na análise melhorou a qualidade do agrupamento. Porém a pequena melhoria observada no número de dissimilaridades (5%) não justifica a escolha por k=4, uma vez que já existe aplicabilidade para o número de grupos igual a 3, na rotina de trabalho da empresa “A”.

Tabela 9 – Comparação da dissimilaridade nas diversas Análises de Agrupamentos realizadas, para k=3 e k=4

Método	Sementes	Dissimila ridade total (k=3)	% dissimilaridade total em relação à dissimilaridade total sem partição	Dissimila ridade total (k=4)	% dissimilaridade total em relação à dissimilaridade total sem partição
Hierárquico - ligação Ward	-	920	49,38	850	45,63
k-Médias	aleatórias	867	46,54	778	41,76
k-Médias	ligação de Ward	899	48,26	809	43,42
k-Modas	aleatórias	1042	55,04	839	45,03
k-Modas*	ligação de Ward	905	48,58	796	42,73
k-Modas	k-Médias – sem. aleatórias	613	32,90	864	46,38
k-Modas*	k-Médias – sem. ligação de Ward	905	48,58	796	42,73

Dissimilaridade total sem partição: 1.368

* Os resultados encontrados nos métodos k-Modas com sementes de inicialização provenientes do método da ligação de Ward, e k-Modas com sementes de inicialização provenientes do método k-Médias (sementes da ligação de Ward) foram exatamente iguais, tanto para k=3 quanto para k=4, em virtude do vetor de modas dos métodos da ligação de Ward e do método k-Médias com sementes de inicialização provenientes do método de ligação de Ward serem iguais, para ambos os valores de k considerados.

Foi realizada análise descritiva dos grupos formados a partir da melhor partição encontrada para k=3, conforme Figura 2, e para k=4, conforme Figura 3.

Partindo-se do pressuposto de que a presença das características exibidas no eixo horizontal das Figuras 2 e 3 implica em condição de maior prejuízo à manutenção da saúde, deverá ser considerado como maior complexidade o grupo no qual há um maior número de características com maior percentual de prevalência e assim sucessivamente.

De acordo com a Figura 2, observa-se em geral que o **grupo 2 (n=135)** é constituído por maior percentual de pacientes clinicamente mais complexos (comparativamente aos demais grupos). Nesse grupo encontra-se a maioria de pacientes com idade maior ou igual a 80 anos, em cuidados paliativos, com limitação de força em membros inferiores e em alto risco nutricional, além de ser o grupo também com maior percentual de pacientes acamados e com distúrbio de marcha.

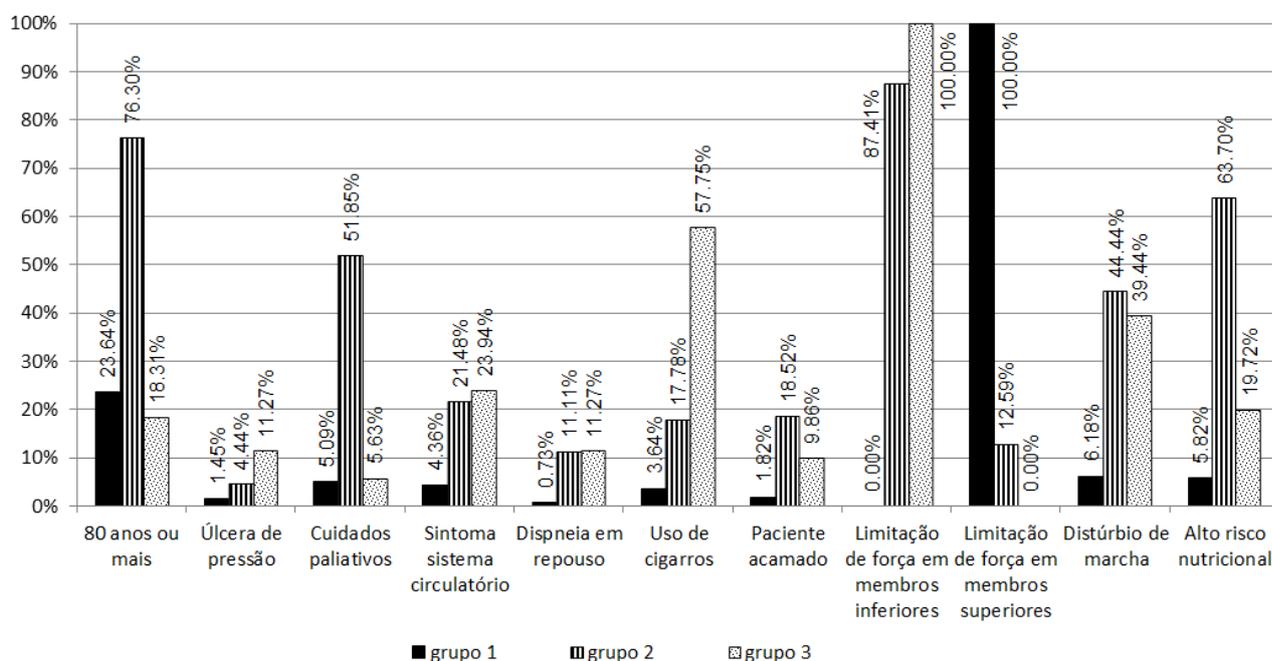


Figura 2 – Comparação das características dos indivíduos constituintes de cada um dos grupos formados pelo método k-Modas com sementes de inicialização provenientes da partição gerada pelo método k-Médias (sementes aleatórias) – escolhida como a melhor partição, para k=3

O **grupo 3 (n=71)**, para $k=3$, apresenta maior percentual de pacientes mais complexos comparativamente ao grupo 1, e menor percentual que o grupo 2, em relação à maioria das características clínicas avaliadas. O grupo 3 é o grupo que concentrou a maioria de tabagistas, o maior percentual de pacientes com dispneia em repouso (percentual próximo ao grupo 2), o maior percentual também de portadores de úlcera de pressão e sintomas do sistema circulatório.

Já o **grupo 1 (n=275)**, para $k=3$, que foi o grupo mais volumoso obtido a partir da Análise de Agrupamentos, abrange os pacientes menos complexos, pois apresenta o menor percentual de pacientes com as características consideradas, de acordo com a Figura 2, exceto limitação de força em membros superiores, que conforme já discutido na seção 4, é a variável que implica em menor comprometimento da saúde em geral, e especialmente em relação à limitação de força em membros inferiores.

Com o acima exposto, de acordo com a Figura 2, poder-se-ia direcionar o conjunto de pacientes constituintes do grupo 2 para o nível de maior complexidade (alta complexidade), e respectivo plano de cuidados, de acordo com a Tabela 1; o grupo 3 corresponderia aos pacientes de complexidade intermediária; e o grupo 1 seria aquele dos pacientes de baixa complexidade.

Quando se insere mais um grupo ($k=4$), de acordo com análise da Figura 3, o grupo no qual existe a maior prevalência das características em questão, que deve ser considerado o grupo com os pacientes de alta complexidade, é o **grupo 4 (n=106)**. Quando comparado aos demais grupos, esse foi o grupo com maior percentual de pacientes com sintoma do sistema circulatório, dispneia em repouso, tabagistas e acamados, além de todos os pacientes incluídos nesse grupo apresentarem limitação de força em membros inferiores.

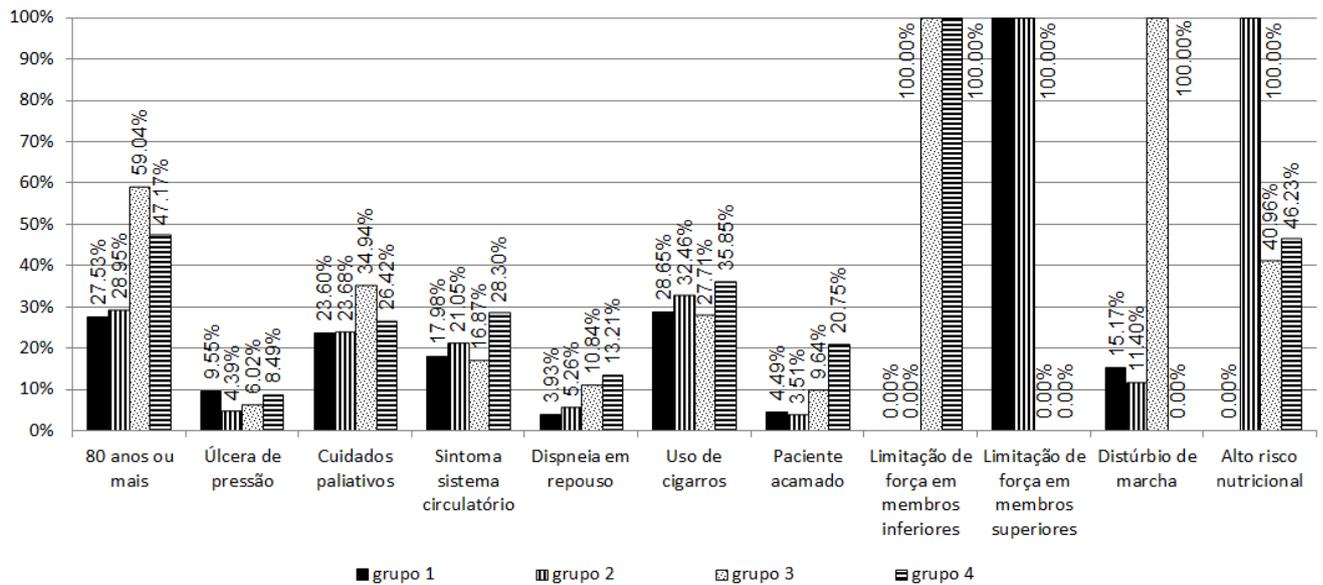


Figura 3 – Comparação das características dos indivíduos constituintes de cada um dos grupos formados pelo método k-Médias (sementes aleatórias) – escolhida como a melhor partição, para k=4

Seguindo uma ordenação decrescente de complexidade, para k=4, na sequência, o **grupo 3 (n=83)** seria o de segunda maior complexidade, constituído pelo maior percentual de pacientes acima de 80 anos, em cuidados paliativos, com limitação de força em membros inferiores e distúrbio de marcha, comparativamente aos demais grupos.

O próximo grupo na ordem decrescente de complexidade seria o **grupo 2 (n=114)**, que apresenta percentual de presença das características consideradas nos pacientes pertencentes a esse grupo ligeiramente maior que o **grupo 1 (n=178)**, exceto as características de paciente acamado e distúrbio de marcha. Embora os percentuais de quase todas as condições analisadas seja ligeiramente superior no grupo 2 quando comparado ao grupo 1, a diferença desses percentuais nesses dois grupos é muito pequena (exceto para alto risco nutricional, que está presente apenas no grupo 2), ou seja, esses grupos são parecidos entre si, o que pode implicar na junção desses grupos em um mesmo nível de complexidade, sem maiores prejuízos.

Após análise dos grupos para $k=4$, pode ser definido que o grupo 4 seja aquele com os pacientes mais graves (alta complexidade), o grupo 3 seja composto pelos pacientes de complexidade intermediária (média complexidade) e os grupos 2 e 1 sejam unidos, para formar o grupo dos pacientes menos complexos (baixa complexidade).

Considerando-se as duas análises discutidas nessa seção, foi verificado que, mesmo para $k=4$, em ambas as análises chegou-se em 3 grupos, consoante à metodologia já preconizada pela empresa “**A**”, que estabelece 3 planos de cuidados distintos, para 3 níveis de complexidade. Nesse contexto, a melhor opção encontrada foi a partição oferecida pelo método k-Modas com sementes de inicialização provenientes do método k-Médias (sementes aleatórias), que foi o método com o menor percentual de dissimilaridade em relação à dissimilaridade total sem partição, não apenas considerando-se $k=3$, mas incluindo as análises realizadas também para $k=4$, conforme Tabela 9.

Na Tabela 10 apresentam-se as modas e dissimilaridades dos *clusters* obtidos pela aplicação do método k-Modas com sementes de inicialização provenientes do método de k-Médias (sementes aleatórias).

Tabela 10⁵ - Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento não Hierárquico k-modas com sementes do vetor de modas do k-médias (sementes aleatórias)

<i>Cluster</i>	1				2				3			
Total de elementos	275				135				71			
Variáveis	0	1	Moda	n ^o dissimilaridades	0	1	Moda	n ^o dissimilaridades	0	1	Moda	n ^o dissimilaridades
Idade categorizada	210	65	0	65	32	103	1	32	58	13	0	13
Úlcera de pressão	271	4	0	4	129	6	0	6	63	8	0	8
Cuidados paliativos	261	14	0	14	65	70	1	65	67	4	0	4
Sintoma sistema circulatório	263	12	0	12	106	29	0	29	54	17	0	17
Dispneia em repouso	273	2	0	2	120	15	0	15	63	8	0	8
Uso de cigarros	265	10	0	10	111	24	0	24	30	41	1	30
Paciente acamado	270	5	0	5	110	25	0	25	64	7	0	7
Limitação de força em membros inferiores	275	0	0	0	17	118	1	17	0	71	1	0
Limitação de força em membros superiores	0	275	1	0	118	17	0	17	71	0	0	0
Distúrbio de marcha	258	17	0	17	75	60	0	60	43	28	0	28
Alto risco nutricional	259	16	0	16	49	86	1	49	57	14	0	14
DISSIMILARIDADE	145				339				129			

(*) Dissimilaridade total = 145 + 339 + 129 = 613

(**) Dissimilaridade média = (145 + 339 + 129) / 3 = 204,33

⁵ As respostas das variáveis constantes nas tabelas 10 e 11 receberam os valores '0' e '1', sendo que '0' significa 'não' e '1' significa 'sim', exceto para a variável 'idade', para a qual '0' significa idade menor que 80 anos, e '1' significa idade maior ou igual a 80 anos. Como número de dissimilaridades entenda-se a dissimilaridade em relação à moda da variável.

5.4.1 Análise de Agrupamentos desconsiderando-se a variável “Limitação de força em membros superiores”

Em função dos erros de medida que podem estar presentes nas respostas da variável “Limitação de força em membros superiores” (devido à alta prevalência encontrada na amostra considerada para análise), possivelmente devido ao fato de que o método de medição para essa variável é auto-referido, como mencionado na seção 5.1, realizou-se a Análise de *Cluster* sem considerar essa variável.

Foi realizada a Análise de Agrupamentos por meio do método k-Modas com sementes de inicialização provenientes do método k-Médias (sementes aleatórias), para $k=3$, que foi o método eleito como o melhor dentre todas as análises realizadas considerando as 11 variáveis selecionadas, conforme seção 5.4.

A dissimilaridade total encontrada para o método eleito como o melhor, com $k=3$, sem utilizar a variável limitação de força em membros superiores foi maior (834) que aquela encontrada na análise incluindo a variável limitação de força em membros superiores (613), o que indica que a análise considerando a inclusão da referida variável implicou em grupos mais homogêneos, sugerindo melhor partição dos elementos.

Foi realizada a análise das características dos grupos obtidos por meio da análise aqui considerada, sem a variável “Limitação de força em membros superiores”, conforme Figura 4.

Nesta partição, o grupo com maior percentual de presença das características representadas pelas variáveis consideradas foi o **grupo 3 (n=118)**, no qual encontra-se o maior percentual de pacientes com 80 anos ou mais, com úlcera de pressão, em cuidados paliativos, com dispneia em repouso, com limitação de força em membros inferiores e com distúrbio de marcha, comparativamente aos demais grupos. Esse seria o grupo dos pacientes mais complexos.

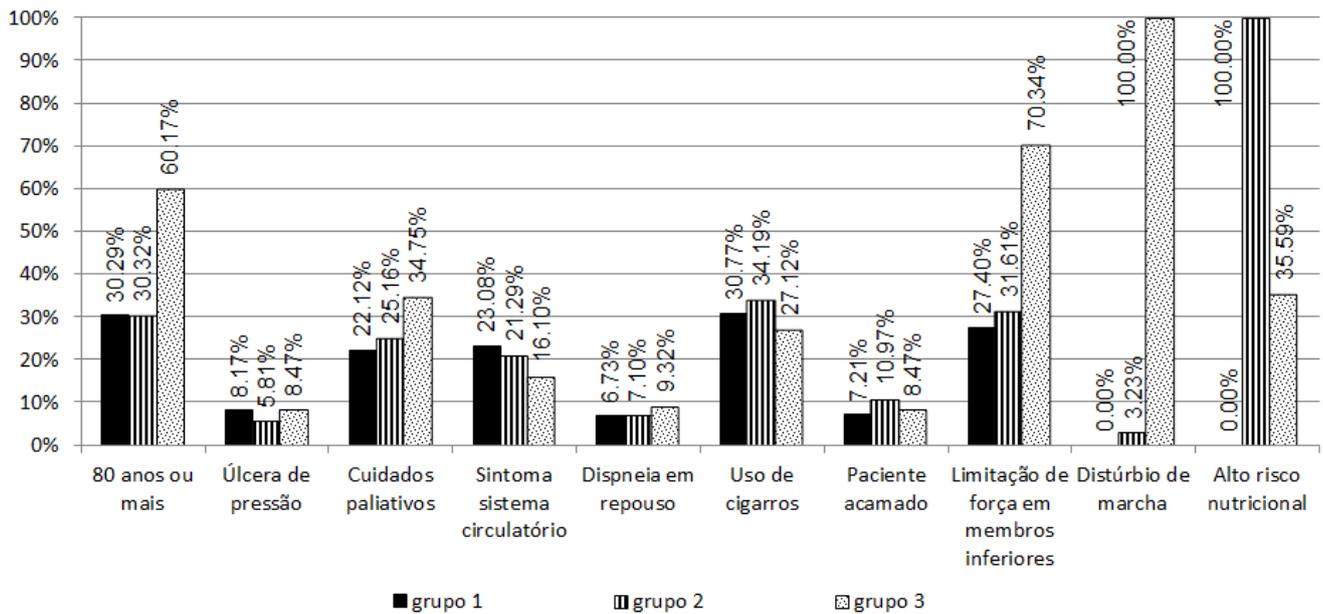


Figura 4 - Comparação das características dos indivíduos constituintes de cada um dos grupos formados pelo método k-Modas com sementes de inicialização provenientes da partição gerada pelo método k-Médias (sementes aleatórias) – com k=3, desconsiderando-se a variável “Limitação de força em membros superiores”

O grupo 2 (n=155) e o grupo 1 (n=208) apresentam percentuais bastante próximos de presença das características representadas pelas variáveis consideradas na análise. Porém é possível definir que o grupo 2 apresenta um maior percentual de pacientes mais complexos que o grupo 1, pois o grupo 2 possui maior percentual de pacientes com a presença de todas as variáveis consideradas, exceto úlcera de pressão, quando comparado ao grupo 1, sendo que para a variável alto risco nutricional a diferença de percentual entre os grupos 2 e 1 é grande, pois no grupo 2, todos os pacientes apresentam alto risco nutricional, enquanto que no grupo 1 não há nenhum paciente com essa condição.

Na Tabela 11 apresentam-se as modas e dissimilaridades dos *clusters* obtidos, desconsiderando-se a variável limitação dos membros superiores, pela aplicação do método k-Modas com sementes de inicialização provenientes do método de k-Médias (sementes aleatórias).

Tabela 11 - Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento não Hierárquico k-modas com sementes do vetor de modas do k-médias (sementes aleatórias) - desconsiderando-se a variável “Limitação de força em membros superiores” na formação dos grupos.

<i>Cluster</i>	1				2				3			
Total de elementos	208				155				118			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	145	63	0	63	108	47	0	47	47	71	1	47
Úlcera de pressão	191	17	0	17	146	9	0	9	108	10	0	10
Cuidados paliativos	162	46	0	46	116	39	0	39	77	41	0	41
Sintoma sistema circulatório	160	48	0	48	122	33	0	33	99	19	0	19
Dispneia em repouso	194	14	0	14	144	11	0	11	107	11	0	11
Uso de cigarros	144	64	0	64	102	53	0	53	86	32	0	32
Paciente acamado	193	15	0	15	138	17	0	17	108	10	0	10
Limitação de força em membros inferiores	151	57	0	57	106	49	0	49	35	83	1	35
Distúrbio de marcha	208	0	0	0	150	5	0	5	0	118	1	0
Alto risco nutricional	208	0	0	0	0	155	1	0	76	42	0	42
DISSIMILARIDADE	324				263				247			

(*) Dissimilaridade total = 324 + 263 + 247 = 834

(**) Dissimilaridade média = (324 + 263 + 247) / 3 = 278,00

Apesar da partição desconsiderando a variável “Limitação de força em membros superiores” ter fornecido grupos menos homogêneos (maior dissimilaridade) que a partição incluindo essa variável, ainda assim ela deve ser considerada como opção para definição de grau de complexidade dos pacientes, devido ao provável problema constatado nos dados obtidos dessa variável.

5.4.2 Análise da variável quantitativa “Idade” no melhor agrupamento obtido com k=3

Conforme observado na Figura 1 (histograma dos valores de idade) e comentado na seção 5.1, a distribuição dos valores de idade apresentou aspecto trimodal. Em virtude de tal fato, foi realizada a análise comparativa da variável idade em cada um dos grupos formados a partir do método eleito como melhor (k-modas com sementes de inicialização do k-médias, este por sua vez obtido por meio de sementes aleatórias), para k=3, incluindo e também excluindo a variável “Limitação de força em membros superiores”, conforme Figuras 5 e 6.

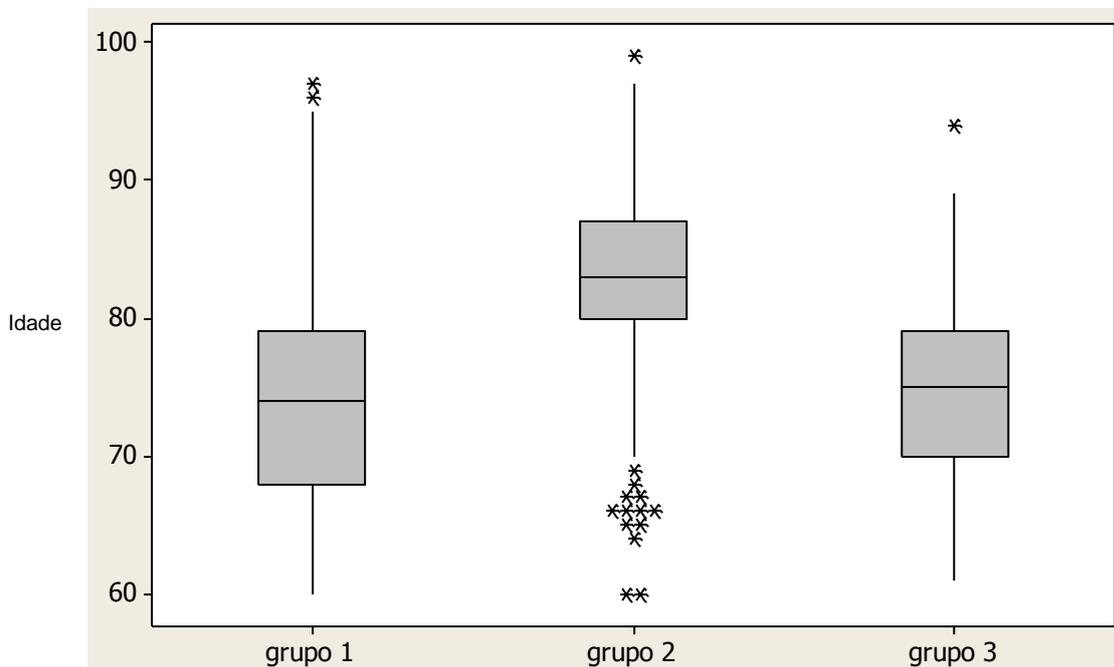


Figura 5 – Box Plot da variável quantitativa “Idade” em cada um dos 3 grupos formados pelo método k-Modas com sementes de inicialização provenientes da partição gerada pelo método k-Médias (sementes aleatórias), incluindo a variável “Limitação de força em membros superiores”

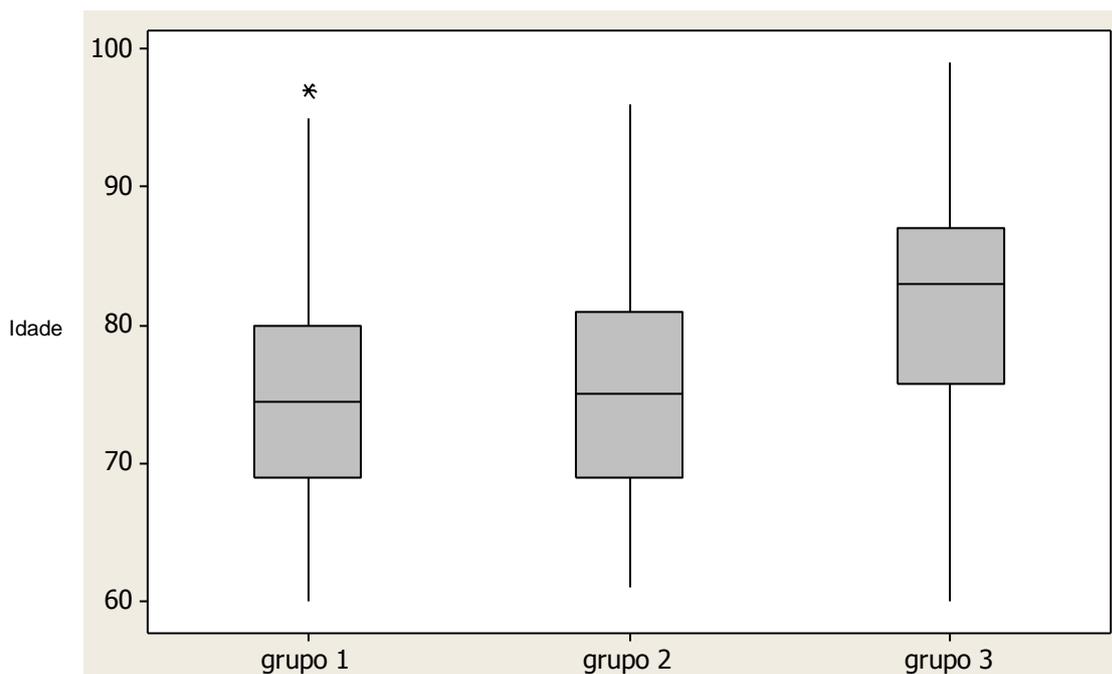


Figura 6 – Box Plot da variável quantitativa “Idade” em cada um dos 3 grupos formados pelo método k-Modas com sementes de inicialização provenientes da partição gerada pelo método k-Médias (sementes aleatórias), excluindo a variável “Limitação de força em membros superiores”

Observa-se que nos grupos formados, conforme Tabela 12, tanto na situação de exclusão da variável “Limitação de força em membros superiores”, quanto na situação de inclusão dessa variável, as modas não refletem as modas encontradas no histograma da Figura 1, considerando a totalidade dos 481 elementos amostrais. Verifica-se que em ambas situações, dois grupos apresentam modas próximas (grupos 1 e 3, análise incluindo “Limitação de força em membros superiores”; e grupos 1 e 2, na análise excluindo a mesma variável).

Tabela 12 – Modas de cada um dos 3 grupos formados pelo método k-Modas com sementes de inicialização provenientes da partição gerada pelo método k-Médias (sementes aleatórias), incluindo e excluindo a variável “Limitação de força em membros superiores”

	Moda (incluindo variável “Limitação de força em membros superiores”)	Moda (excluindo variável “Limitação de força em membros superiores”)
Grupo 1	74	74
Grupo 2	83	74; 75
Grupo 3	76	85

6 CONSIDERAÇÕES FINAIS

A partir das análises realizadas, constatou-se a superioridade da qualidade dos métodos não Hierárquicos (k-Médias e k-Modas), em relação aos métodos Hierárquicos, para o banco de dados analisado, devido à característica dos dados (dados binários).

Uma vantagem da utilização de métodos estatísticos de agrupamentos em relação à definição de critérios fixos para definição dos graus de complexidade de pacientes consiste no fato de que a definição dos graus de complexidade pela Análise de Agrupamentos é realizada comparativamente entre todos os pacientes participantes do programa, de maneira que permite que sempre existam representantes em cada um dos grupos definidos, favorecendo a utilização racional dos recursos de saúde disponíveis.

Outro estudo pode ser feito analisando-se dentre as variáveis consideradas quais implicam em maior prejuízo à saúde do paciente, e com isso aplicando pesos diferenciados às variáveis durante realização de Análise de Agrupamentos. Foi observado em várias análises (inclusive na análise apresentada na Figura 2) que as variáveis “Limitação de força em membros superiores” e “Limitação de força em membros inferiores” foram definidoras de *clusters* (em um grupo todos os indivíduos apresentavam limitação de força em membros inferiores e em outro grupo nenhum apresentava essa limitação). Porém, conforme discutido nas seções 4 e 5.1 respectivamente, a variável limitação de força em membros superiores, apesar de oferecer prejuízo à qualidade de vida do indivíduo, não é grande definidora de alta gravidade do quadro clínico do paciente; e possivelmente houve algum problema na obtenção dos dados obtidos em relação a essa variável, devido à alta prevalência verificada para a presença de limitação de força em membros superiores em toda a amostra.

Analisando-se a partição obtida por meio do método k-Modas com sementes de inicialização provenientes do método k-Médias (sementes aleatórias) com $k=3$, de acordo com o vetor de modas de cada um dos grupos,

observou-se que as variáveis úlcera de pressão, cuidados paliativos e alto risco nutricional apresentam moda 1 apenas no grupo 2 (Tabela 10), ou seja, essas variáveis tiveram bastante influência na formação dos grupos, e de fato, definem um conjunto de características que traduzem em substancial agravamento do quadro clínico de um paciente idoso fragilizado. A condição de paciente acamado está intrinsecamente relacionada à presença de úlcera de pressão, cuidados paliativos e alto risco nutricional, e mesmo a moda dessa variável sendo 0 em todos os grupos formados, foi no grupo 2 que se verificou o maior percentual de pacientes acamados, 18,5%, contra 9,9% no grupo 3 e 1,8% no grupo 1. A opção de incrementar a análise com outras variáveis, além daquelas aqui utilizadas, seria válida também para melhorar a distinção dos grupos.

Deve ser observado que as características encontradas em cada um dos grupos, bem como as características da amostra sem partição, sofre influência da localização espacial dos pacientes, ou seja, reflete possível presença de regionalismos, em virtude da grande extensão territorial brasileira. Nesta monografia foram analisados os dados dos pacientes do estado do Rio Grande do Sul. Outras Análises de Agrupamentos realizadas com pacientes de outras localidades geográficas podem apresentar perfis distintos daquele encontrado nesta monografia. Assim, o modelo aqui apresentado não pode ser generalizado para todas as localidades em que a empresa “A” atua. Para cada localidade, um modelo de Análise de *Cluster* em particular deverá ser ajustado.

Deve ser recomendado à operadora cujos dados foram aqui analisados, que obtenha as respostas das condições clínicas avaliadas, em especial a condição de limitação de força em membros superiores, de maneira mais criteriosa, para que possam ser evitadas inconsistências externas, uma vez que foi encontrado um percentual elevado de portadores de limitação em membros superiores na amostra considerada, resultado divergente do esperado.

Após novas avaliações de novos pacientes e reavaliações de pacientes já acompanhados no programa de atenção ao idoso fragilizado, um maior volume de dados amostrais será obtido ao longo do tempo. Para esse novo montante de

dados é importante realizar o ajuste via Análise de *Cluster* para verificação da composição dos grupos, se eles estão mantendo a mesma conformação e se a partição em 3 grupos ainda permanecerá como melhor opção.

Como o *software* desenvolvido e comercializado pela empresa “**A**” já utiliza rotina de *scripts* integrados para realização de análise estatística descritiva de banco de dados de saúde no *software* estatístico R, a possibilidade de implementação com sucesso de *script* para realização de Análise de Agrupamentos como definidora de graus de complexidade de pacientes é grande.

A realização do trabalho apresentado nessa monografia foi importante, pois ampliou conhecimentos sobre os métodos de agrupamentos e permite a disseminação e aplicação desses conhecimentos na empresa “**A**”, também para outras variáveis e demais instrumentos de avaliação clínica de pacientes utilizados, em outros bancos de dados do cliente selecionado para este trabalho em particular e também em banco de dados de todos os demais clientes.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAUJO, D. Polarização epidemiológica no Brasil. *Epidemiologia e Serviços de Saúde*, v. 21, n. 4, p.6, 2012.

BUSSAB, W. O.; MIAZAKI, E. S.; ANDRADE, D. F. *Introdução à Análise de Agrupamentos*. São Paulo: IME – USP, 1990.

CAMARANO, A. A. *Envelhecimento da população brasileira: uma contribuição demográfica*. Rio de Janeiro: Instituto de Pesquisa Econômica e Aplicada, 2002.

CHAIMOWICZ, F. A saúde dos idosos brasileiros às vésperas do século XXI: problemas, projeções e alternativas. *Revista de Saúde Pública*. v. 31, n. 2, p. 184-200, 1997.

COLLINS, K. ROONEY, B. L. SMALLEY, K. J. HAVENS, S. Functional fitness, disease and independence in community-Dwelling older adults in Western Wisconsin, *Wisconsin Medical Journal*. v. 103, p. 42-3, 2004.

DUARTE, Y. A. O.; DIOGO, M. J. D. *Atendimento Domiciliar: um enfoque gerontológico*. São Paulo: Atheneu, 2000.

EVERITT, B. S. *Cluster Analysis*. New York: John Wiley & Sons, Inc., 2011.

FRIED, L. P.; TANGEN, C. M.; WALTSON, J.; NEWMAN, A.B.; HIRSCH, C.; GOTTDIENER, J.; SEEMAN, T. S.; TRACY, R.; KOP, W. J.; BURKE, G.; MCBURNIE, M. A. Frailty in older adults: Evidence for a phenotype. *Journal of Gerontology*, v. 56, n. 3, p. 146-156, 2001.

HAIR, J. F.; ANDERSON, R.E.; TATHAM, R. L.; BLACK, W. C. *Análise multivariada de dados*. Porto Alegre: Bookman, 2005.

HUANG, Z.; NG, M. K. A fuzzy k-modes algorithm for clustering categorical data. *IEEE Transactions on Fuzzy Systems*, v. 7, p. 446-452, 1999.

JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. New Jersey: Prentice Hall, 2007.

LIMA, P. V. S. F.; FAUSTINO, A. M. Aspectos Gerais do Tabagismo entre Idosos no Brasil: Revisão de Literatura. *Revista Eletrônica Gestão & Saúde*, v. 4, n. 1, p.1860-1877, 2013.

MATOS, R. A. *Comparação de Metodologias de Análise de Agrupamentos na Presença de Variáveis Categóricas e Contínuas*. 2007. Dissertação (Mestrado) – Departamento de Estatística, Universidade Federal de Minas Gerais.

MEDEIROS, A. B. F.; LOPES, C. H. A. F.; JORGE, M. S. B. Análise da prevenção e tratamento das úlceras por pressão propostos por enfermeiros. *Revista da Escola de Enfermagem da USP*. v. 43, n. 1, p. 223-228, 2009.

MINGOTI, S. A. *Análise de Dados Através de Métodos de Estatística Multivariada: Uma abordagem Aplicada*. Belo Horizonte: Editora UFMG, 2005.

MORAIS, E. P.; RODRIGUES, R. A. P.; GERHARDT, T. E. Os idosos mais velhos no meio rural: realidade de vida e saúde de uma população do interior gaúcho. *Texto Contexto e Enfermagem*, v. 17, n. 2, p. 374-383, 2008.

PIOVESAN, A. C. Fatores que Predisõem a Quedas em Idosos Residentes na Região Oeste de Santa Maria, RS. *Revista Brasileira de Geriatria e Gerontologia*, v. 14, n. 1, p. 75-78, 2011.

REZENDE, A. A. B.; LIMA e SILVA, I.; CARDOSO, F. B.; BERESFORD, H. Medo do idoso em sofrer quedas recorrentes: a marcha como fator determinante da independência funcional. *Acta Fisiátrica*, v. 17, n.3, p. 117-121, 2010.

SHEPHARD, R. J.; BOUCHARD, C. Principal components of fitness: relationship to physical activity and lifestyle. *Canadian Journal of Applied Physiology*, v. 19, p. 200-214, 1994.

SOARES, J. F. SIQUEIRA, A. L. *Introdução à Estatística Médica*. Belo Horizonte: Coopmed, 2008.

SOKAL, R. R.; SNEATH, P. H. A. *Principles of numerical taxonomy*. London: Freeman, 1963.

TEIXEIRA, I. N. D. O. Percepções de profissionais de saúde sobre duas definições de fragilidade no idoso. *Ciência & Saúde Coletiva*, v. 13, n. 4, p. 1181-1188, 2008.

VIRTUOSO JÚNIOR, J. S.; GUERRA, R. O. Fatores associados às limitações funcionais em idosos de baixa renda. *Revista da Associação Médica Brasileira*, v. 54, n. 5, p. 430-435, 2008.

WORLD HEALTH ORGANIZATION. *The World Health report 2002: Reducing risks, promoting healthy life*. Geneva: WHO. 2002

ANEXO I – Fórmulas Matemáticas da Medida de Distância e dos Métodos de Agrupamentos Utilizados

1. Relação entre distância Euclidiana ao quadrado e coeficiente de dissimilaridade para dados binários

Seja p o número de variáveis do conjunto de dados analisados, n o número de elementos desse conjunto e X_{ij} o dado referente à variável i e ao elemento j . A distância Euclidiana ao quadrado média entre dois elementos amostrais E_l e E_k é dada por:

$$d(E_l, E_k) = \left[\sum_{i=1}^p \frac{1}{p} (X_{i,l} - X_{i,k})^2 \right]$$

em que $l, k = 1, 2, \dots, n; l \neq k$

Neste caso, quanto menor o valor da distância, maior será a similaridade dos elementos comparados. Portanto, a distância Euclidiana é um índice de dissimilaridade.

No caso em que as variáveis são binárias a distância Euclidiana ao quadrado média entre os elementos E_l e E_k é expressa como:

$$d(E_l, E_k) = \frac{r}{p}$$

em que r é o número de variáveis observadas para as quais os elementos E_l e E_k diferem, ou seja, o número de discordâncias entre esses elementos. O mínimo valor de r é 0, enquanto que o máximo é p . Assim a distância Euclidiana ao quadrado média é equivalente ao coeficiente de dissimilaridade simples (SOKAL e SNEATH, 1963).

Para efeito de agrupamento, basta calcular a distância Euclidiana ao quadrado para comparar os elementos amostrais, no caso de dados binários, já que a partição produzida será igual àquela produzida ao se utilizar a distância Euclidiana média ou o coeficiente de dissimilaridade de Sokal e Sneath (1963).

2. Descrição dos métodos de agrupamentos utilizados na monografia

2.1. Método de Ligação Simples (*Single Linkage*)

No método de Ligação Simples, a cada estágio do processo de agrupamento os *clusters* são comparados através da medida de distância definida como:

$$d(C_i, C_j) = \min \{ d(X_q, X_z), q \neq z, q \in \{1, 2, \dots, n_i\} \text{ e } z \in \{1, 2, \dots, n_j\} \}$$

sendo n_i, n_j , os tamanhos amostrais dos conglomerados C_i e C_j , respectivamente, $i \neq j$.

2.2. Método de Ligação Completa (*Complete Linkage*)

No método de Ligação Completa, a cada estágio do processo de agrupamento os *clusters* são comparados através da medida de distância definida como:

$$d(C_i, C_j) = \max \{ d(X_q, X_z), q \neq z, q \in \{1, 2, \dots, n_i\} \text{ e } z \in \{1, 2, \dots, n_j\} \}$$

sendo n_i, n_j , os tamanhos amostrais dos conglomerados C_i e C_j , respectivamente, $i \neq j$.

2.3. Método de Ligação das Médias (*Average Linkage*)

No método de Ligação das Médias, a cada estágio do processo de agrupamento os *clusters* são comparados através da medida de distância definida como:

$$d(C_i, C_j) = \frac{1}{n_i n_j} \sum_{q \in C_i} \sum_{z \in C_j} d(X_q, X_z)$$

sendo n_i, n_j , os tamanhos amostrais dos conglomerados C_i e C_j , respectivamente, $i \neq j$.

2.4. Método de ligação de Ward

O método da ligação de Ward utiliza a seguinte função de distância para comparação dos conglomerados C_i e C_j , nos passos do algoritmo de agrupamento:

$$d(C_i, C_j) = \left[\frac{n_i n_j}{n_i + n_j} \right] (\bar{X}_i - \bar{X}_j)' (\bar{X}_i - \bar{X}_j)$$

sendo \bar{X}_i e \bar{X}_j os vetores de médias dos conglomerados C_i e C_j , respectivamente, $i \neq j$.

Nota: Nos métodos de ligação apresentados em 2.1, 2.2, 2.3 e 2.4, em cada passo do algoritmo agrupa-se os dois *clusters* que produzem a menor distância.

2.5. Método das k-Médias

O método das k-Médias busca a partição dos dados em k *clusters* que minimiza a função objetivo dada por:

$$J = \sum_{l=1}^k \sum_{i=1}^n w_{i,l} d(X_i, V_l)$$

sujeito a $w_{i,l} \in \{0,1\}$, $i \in \{1,2, \dots, n\}$ e $l \in \{1,2, \dots, k\}$,

sendo V_l a semente (ou centróide) do conglomerado l , $l = 1, 2, \dots, k$, $d(\cdot)$ a distância Euclidiana ao quadrado. A cada passo do algoritmo os centróides são atualizados de acordo com os elementos que foram alocados em cada *cluster* no passo anterior, ou seja, calcula-se o vetor de médias amostral de cada grupo formado e esses vetores passam a ser as novas sementes. Os coeficientes w_{il} são os coeficientes de pertinência do elemento amostral i ao conglomerado l . No método k-Médias cada elemento somente pode ser alocado em um dos k grupos que serão formados, ou seja na solução final w_{il} deverá ser igual a 1 para apenas um dos conglomerados $l = 1, 2, \dots, k$.

2.6. Método k-Modas

O método k-Modas busca a partição dos dados em k *clusters* que minimiza a seguinte função de proximidade:

$$\begin{aligned}
 P(W, Q) &= \sum_{l=1}^k \sum_{i=1}^n w_{i,l} d(X_i, Q_l) \\
 &= \sum_{l=1}^k \sum_{i=1}^n \sum_{j=1}^m w_{i,l} \delta(x_{i,j}, q_{l,j}) \\
 &\text{sujeito a } w_{i,l} \in \{0,1\}, i \in \{1, 2, \dots, n\} \text{ e } l \in \{1, 2, \dots, k\} \\
 &\sum_{l=1}^k w_{i,l} = 1, i \in \{1, 2, \dots, n\}
 \end{aligned}$$

em que o vetor de modas da amostra, considerando l grupos e m variáveis, é $Q = \{Q_1, Q_2, \dots, Q_l\}$, sendo $Q_l = [q_{l,1}, q_{l,2}, \dots, q_{l,m}]$ o vetor de modas do conglomerado l , $w_{i,l}$ é o elemento da matriz partição W que indica se a i -ésima observação pertence ($w_{i,l} = 1$) ou não ($w_{i,l} = 0$) ao grupo l , k é o número de grupos e n é o número de observações amostrais. A medida de dissimilaridade $\delta(\cdot)$ é definida como:

$$\delta(x_{i,j}, q_{l,j}) = \begin{cases} 0 & \text{se } x_{i,j} = q_{l,j} \\ 1 & \text{se } x_{i,j} \neq q_{l,j} \end{cases}$$

O processo do k-Modas inicia-se com a definição do valor do número de grupos k a serem obtidos e seleção das k sementes iniciais. Na sequência, cada elemento deve ser alocado a um dos k grupos, de acordo com a medida de proximidade adotada. Após cada realocação, as modas são atualizadas e deve ser realizada novamente a comparação dos elementos às modas atualizadas. Caso existam elementos que estejam mais próximos a outras modas que não àquelas do próprio grupo, esses elementos devem ser realocados e as modas dos grupos formados devem ser novamente atualizadas. Esses passos devem ser repetidos até não ser possível fazer nenhuma outra alocação. Assim como o método k-Médias, no k-Modas cada elemento amostral é alocado em apenas um dos grupos formados.

ANEXO II – Tabelas Contendo os Resultados dos Testes estatísticos Qui-quadrado Apresentados na Seção 5.2

		Úlcera de pressão		
		não	sim	total
Idade categorizada	60 a 79 anos	272	28	300
		90,67%	9,33%	100,00%
		61,12%	77,78%	62,37%
	80 anos ou mais	173	8	181
		95,58%	4,42%	100,00%
		38,88%	22,22%	37,63%
Total	445	36	481	
	92,52%	7,48%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 3,936
p-valor = 0,047

		Cuidados paliativos		
		não	sim	Total
Idade categorizada	60 a 79 anos	243	57	300
		81,00%	19,00%	100,00%
		68,45%	45,24%	62,37%
	80 anos ou mais	112	69	181
		61,88%	38,12%	100,00%
		31,55%	54,76%	37,63%
Total	355	126	481	
	73,80%	26,20%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 21,350
p-valor = 0,000

		Sintoma sistema circulatório		
		não	sim	total
Idade categorizada	60 a 79 anos	236	64	300
		78,67%	21,33%	100,00%
		61,94%	64,00%	62,37%
	80 anos ou mais	145	36	181
		80,11%	19,89%	100,00%
		38,06%	36,00%	37,63%
Total	381	100	481	
	79,21%	20,79%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 0,143
p-valor = 0,705

		Dispneia em repouso		
		não	sim	Total
Idade categorizada	60 a 79 anos	280	20	300
		93,33%	6,67%	100,00%
		62,92%	55,56%	62,37%
	80 anos ou mais	165	16	181
		91,16%	8,84%	100,00%
		37,08%	44,44%	37,63%
Total	445	36	481	
	92,52%	7,48%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 0,770
p-valor = 0,380

		Uso cigarros		
		não	sim	total
Idade categorizada	60 a 79 anos	194	106	300
		64,67%	35,33%	100,00%
		58,43%	71,14%	62,37%
	80 anos ou mais	138	43	181
		76,24%	23,76%	100,00%
		41,57%	28,86%	37,63%
Total	332	149	481	
	69,02%	30,98%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 7,076
p-valor = 0,008

		Paciente acamado		
		não	sim	Total
Idade categorizada	60 a 79 anos	283	17	300
		94,33%	5,67%	100,00%
		64,46%	40,48%	62,37%
	80 anos ou mais	156	25	181
		86,19%	13,81%	100,00%
		35,54%	59,52%	37,63%
Total	439	42	481	
	91,27%	8,73%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 9,399
p-valor = 0,002

		Limitação de força em MMII ⁶		
		não	sim	total
Idade categorizada	60 a 79 anos	210	90	300
		70,00%	30,00%	100,00%
		71,92%	47,62%	62,37%
	80 anos ou mais	82	99	181
		45,30%	54,70%	100,00%
		28,08%	52,38%	37,63%
Total	292	189	481	
	60,71%	39,29%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 28,864

p-valor = 0,000

		Limitação de força em MMSS ⁷		
		não	sim	Total
Idade categorizada	60 a 79 anos	90	210	300
		30,00%	70,00%	100,00%
		47,62%	71,92%	62,37%
	80 anos ou mais	99	82	181
		54,70%	45,30%	100,00%
		52,38%	28,08%	37,63%
Total	189	292	481	
	39,29%	60,71%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 28,864

p-valor = 0,000

		Distúrbio marcha		
		não	sim	total
Idade categorizada	60 a 79 anos	248	52	300
		82,67%	17,33%	100,00%
		69,27%	42,28%	62,37%
	80 anos ou mais	110	71	181
		60,77%	39,23%	100,00%
		30,73%	57,72%	37,63%
Total	358	123	481	
	74,43%	25,57%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 28,430

p-valor = 0,000

		Alto risco nutricional		
		não	sim	total
Idade categorizada	60 a 79 anos	181	119	300
		60,33%	39,67%	100,00%
		63,73%	60,41%	62,37%
	80 anos ou mais	103	78	181
		56,91%	43,09%	100,00%
		36,27%	39,59%	37,63%
Total	284	197	481	
	59,04%	40,96%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 0,548

p-valor = 0,459

		Dispneia em repouso		
		não	sim	total
uso cigarros	não	313	19	332
		94,28%	5,72%	100,00%
		70,34%	52,78%	69,02%
	sim	132	17	149
		88,59%	11,41%	100,00%
		29,66%	47,22%	30,98%
Total	445	36	481	
	92,52%	7,48%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 4,803

p-valor = 0,028

		Limitação de força em MMII		
		não	sim	total
paciente acamado	não	280	159	439
		63,78%	36,22%	100,00%
		95,89%	84,13%	91,27%
	sim	12	30	42
		28,57%	71,43%	100,00%
		4,11%	15,87%	8,73%
Total	292	189	481	
	60,71%	39,29%	100,00%	
	100,00%	100,00%	100,00%	

qui-quadrado = 19,922

p-valor = 0,000

⁶ MMII = membros inferiores

⁷ MMSS = membros superiores

		Limitação de força em MMII		
		não	sim	total
distúrbio marcha	não	252	106	358
		70,39%	29,61%	100,00%
		86,30%	56,08%	74,43%
	sim	40	83	123
		32,52%	67,48%	100,00%
		13,70%	43,92%	25,57%
	Total	292	189	481
		60,71%	39,29%	100,00%
		100,00%	100,00%	100,00%

qui-quadrado = 55,042

p-valor = 0,000

		Úlcera de pressão		
		não	sim	total
paciente acamado	não	406	33	439
		92,48%	7,52%	100,00%
		91,24%	91,67%	91,27%
	sim	39	3	42
		92,86%	7,14%	100,00%
		8,76%	8,33%	8,73%
	Total	445	36	481
		92,52%	7,48%	100,00%
		100,00%	100,00%	100,00%

qui-quadrado = 0,008

p-valor = 0,930

ANEXO III – Distribuição de Frequências, Vetor de Modas e Dissimilaridade por *Cluster* para Cada Valor de k e Método de Agrupamento Utilizados

As respostas das variáveis constantes em todas as tabelas deste Anexo receberam os valores '0' e '1', sendo que '0' significa 'não' e '1' significa 'sim', exceto para a variável 'idade', para a qual '0' significa idade menor que 80 anos, e '1' significa idade maior ou igual a 80 anos. Como número de dissimilaridades entenda-se a dissimilaridade em relação à moda da variável.

Tabela A – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento Hierárquico utilizando ligação de Ward

<i>Cluster</i>	1				2				3			
Total de elementos	189				231				61			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	90	99	1	90	210	21	0	21	0	61	1	0
Úlcera de pressão	175	14	0	14	209	22	0	22	61	0	0	0
Cuidados paliativos	132	57	0	57	183	48	0	48	40	21	0	21
Sintoma sistema circulatório	145	44	0	44	180	51	0	51	56	5	0	5
Dispneia em repouso	166	23	0	23	220	11	0	11	59	2	0	2
Uso de cigarros	128	61	0	61	155	76	0	76	49	12	0	12
Paciente acamado	159	30	0	30	223	8	0	8	57	4	0	4
Limitação de força em membros inferiores	0	189	1	0	231	0	0	0	61	0	0	0
Limitação de força em membros superiores	189	0	0	0	0	231	1	0	0	61	1	0
Distúrbio de marcha	106	83	0	83	212	19	0	19	40	21	0	21
Alto risco nutricional	106	83	0	83	145	86	0	86	33	28	0	28
DISSIMILARIDADE	485				342				93			

(*) Dissimilaridade total = 485 + 342 + 93 = 920

(**) Dissimilaridade média = (485 + 342 + 93) / 3 = 306,67

Tabela B – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=4, no método de agrupamento Hierárquico utilizando ligação de Ward

<i>Cluster</i>	1				2				3				4			
Total de elementos	76				113				231				61			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	73	3	0	3	17	96	1	17	210	21	0	21	0	61	1	0
Úlcera de pressão	64	12	0	12	111	2	0	2	209	22	0	22	61	0	0	0
Cuidados paliativos	68	8	0	8	64	49	0	49	183	48	0	48	40	21	0	21
Sintoma sistema circulatório	57	19	0	19	88	25	0	25	180	51	0	51	56	5	0	5
Dispneia em repouso	73	3	0	3	93	20	0	20	220	11	0	11	59	2	0	2
Uso de cigarros	50	26	0	26	78	35	0	35	155	76	0	76	49	12	0	12
Paciente acamado	66	10	0	10	93	20	0	20	223	8	0	8	57	4	0	4
Limitação de força em membros inferiores	0	76	1	0	0	113	1	0	231	0	0	0	61	0	0	0
Limitação de força em membros superiores	76	0	0	0	113	0	0	0	0	231	1	0	0	61	1	0
Distúrbio de marcha	49	27	0	27	57	56	0	56	212	19	0	19	40	21	0	21
Alto risco nutricional	42	34	0	34	64	49	0	49	145	86	0	86	33	28	0	28
DISSIMILARIDADE	142				273				342				93			

(*) Dissimilaridade total = 142 + 273 + 342 + 93 = 850

(**) Dissimilaridade média = (142 + 273 + 342 + 93) / 4 = 212,50

Tabela C – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento não Hierárquico das k-Médias (sementes aleatórias)

<i>Cluster</i>	1				2				3			
Total de elementos	178				114				189			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	129	49	0	49	81	33	0	33	90	99	0	90
Úlcera de pressão	161	17	0	17	109	5	0	5	175	14	0	14
Cuidados paliativos	136	42	0	42	87	27	0	27	132	57	0	57
Sintoma sistema circulatório	146	32	0	32	90	24	0	24	145	44	0	44
Dispneia em repouso	171	7	0	7	108	6	0	6	166	23	0	23
Uso de cigarros	127	51	0	51	77	37	0	37	128	61	0	61
Paciente acamado	170	8	0	8	110	4	0	4	159	30	0	30
Limitação de força em membros inferiores	178	0	0	0	114	0	0	0	0	189	1	0
Limitação de força em membros superiores	0	178	1	0	0	114	1	0	189	0	0	0
Distúrbio de marcha	151	27	0	27	101	13	0	13	106	83	0	83
Alto risco nutricional	178	0	0	0	0	114	1	0	106	83	0	83
DISSIMILARIDADE	233				149				485			

(*) Dissimilaridade total = 233 + 149 + 485 = 867

(**) Dissimilaridade média = (233 + 149 + 485) / 3 = 289,00

Tabela D – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento não Hierárquico das k-Médias (sementes da ligação de Ward)

<i>Cluster</i>	1				2				3			
Total de elementos	189				210				82			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	90	99	1	90	210	0	0	0	0	82	1	0
Úlcera de pressão	175	14	0	14	192	18	0	18	78	4	0	4
Cuidados paliativos	132	57	0	57	172	38	0	38	51	31	0	31
Sintoma sistema circulatório	145	44	0	44	168	42	0	42	68	14	0	14
Dispneia em repouso	166	23	0	23	199	11	0	11	80	2	0	2
Uso de cigarros	128	61	0	61	136	74	0	74	68	14	0	14
Paciente acamado	159	30	0	30	205	5	0	5	75	7	0	7
Limitação de força em membros inferiores	0	189	1	0	210	0	0	0	82	0	0	0
Limitação de força em membros superiores	189	0	0	0	0	210	1	0	0	82	1	0
Distúrbio de marcha	106	83	0	83	192	18	0	18	60	22	0	22
Alto risco nutricional	106	83	0	83	129	81	0	81	49	33	0	33
DISSIMILARIDADE	485				287				127			

(*) Dissimilaridade total = 485 + 287 + 127 = 899

(**) Dissimilaridade média = (485 + 287 + 127) / 3 = 299,67

Tabela E – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=4, no método de agrupamento não Hierárquico das k-Médias (sementes aleatórias)

Cluster	1				2				3				4			
Total de elementos	178				114				83				106			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	129	49	0	49	81	33	0	33	34	49	1	34	56	50	0	50
Úlcera de pressão	161	17	0	17	109	5	0	5	78	5	0	5	97	9	0	9
Cuidados paliativos	136	42	0	42	87	27	0	27	54	29	0	29	78	28	0	28
Sintoma sistema circulatório	146	32	0	32	90	24	0	24	69	14	0	14	76	30	0	30
Dispneia em repouso	171	7	0	7	108	6	0	6	74	9	0	9	92	14	0	14
Uso de cigarros	127	51	0	51	77	37	0	37	60	23	0	23	68	38	0	38
Paciente acamado	170	8	0	8	110	4	0	4	75	8	0	8	84	22	0	22
Limitação de força em membros inferiores	178	0	0	0	114	0	0	0	0	83	1	0	0	106	1	0
Limitação de força em membros superiores	0	178	1	0	0	114	1	0	83	0	0	0	106	0	0	0
Distúrbio de marcha	151	27	0	27	101	13	0	13	0	83	1	0	106	0	0	0
Alto risco nutricional	178	0	0	0	0	114	1	0	49	34	0	34	57	49	0	49
DISSIMILARIDADE	233				149				156				240			

(*) Dissimilaridade total = 233 + 149 + 156 + 240 = 778

(**) Dissimilaridade média = (233 + 149 + 156 + 240) / 4 = 194,50

Tabela F – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=4, no método de agrupamento não Hierárquico das k-Médias (sementes da ligação de Ward)

Cluster	1				2				3				4			
Total de elementos	90				99				210				82			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	90	0	0	0	0	99	1	0	210	0	0	0	0	82	1	0
Úlcera de pressão	80	10	0	10	95	4	0	4	192	18	0	18	78	4	0	4
Cuidados paliativos	71	19	0	19	61	38	0	38	172	38	0	38	51	31	0	31
Sintoma sistema circulatório	68	22	0	22	77	22	0	22	168	42	0	42	68	14	0	14
Dispneia em repouso	81	9	0	9	85	14	0	14	199	11	0	11	80	2	0	2
Uso de cigarros	58	32	0	32	70	29	0	29	136	74	0	74	68	14	0	14
Paciente acamado	78	12	0	12	81	18	0	18	205	5	0	5	75	7	0	7
Limitação de força em membros inferiores	0	90	1	0	0	99	1	0	210	0	0	0	82	0	0	0
Limitação de força em membros superiores	90	0	0	0	99	0	0	0	0	210	1	0	0	82	1	0
Distúrbio de marcha	56	34	0	34	50	49	0	49	192	18	0	18	60	22	0	22
Alto risco nutricional	52	38	0	38	54	45	0	45	129	81	0	81	49	33	0	33
DISSIMILARIDADE	176				219				287				127			

(*) Dissimilaridade total = 176 + 219 + 287 + 127 = 809

(**) Dissimilaridade média = (176 + 219 + 287 + 127) / 4 = 202,25

Tabela G – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento não Hierárquico k-modas (sementes aleatórias)

Cluster	1				2				3			
Total de elementos	263				71				147			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	174	89	0	89	2	69	1	2	124	23	0	23
Úlcera de pressão	240	23	0	23	69	2	0	2	136	11	0	11
Cuidados paliativos	219	44	0	44	18	53	1	18	118	29	0	29
Sintoma sistema circulatório	207	56	0	56	58	13	0	13	116	31	0	31
Dispneia em repouso	245	18	0	18	62	9	0	9	138	9	0	9
Uso de cigarros	214	49	0	49	57	14	0	14	61	86	1	61
Paciente acamado	239	24	0	24	57	14	0	14	143	4	0	4
Limitação de força em membros inferiores	123	140	1	123	22	49	1	22	147	0	0	0
Limitação de força em membros superiores	140	123	0	123	49	22	0	22	0	147	1	0
Distúrbio de marcha	201	62	0	62	22	49	1	22	135	12	0	12
Alto risco nutricional	217	46	0	46	16	55	1	16	51	96	1	51
DISSIMILARIDADE	657				154				231			

(*) Dissimilaridade total = 657 + 154 + 231 = 1042

(**) Dissimilaridade média = (657 + 154 + 231) / 3 = 347,33

Tabela H – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento não Hierárquico k-modas com sementes do vetor de modas da ligação de Ward

Cluster	1				2				3			
Total de elementos	151				38				292			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	52	99	1	52	38	0	0	0	210	82	0	82
Úlcera de pressão	142	9	0	9	33	5	0	5	270	22	0	22
Cuidados paliativos	101	50	0	50	31	7	0	7	223	69	0	69
Sintoma sistema circulatório	116	35	0	35	29	9	0	9	236	56	0	56
Dispneia em repouso	132	19	0	19	34	4	0	4	279	13	0	13
Uso de cigarros	108	43	0	43	20	18	0	18	204	88	0	88
Paciente acamado	126	25	0	25	33	5	0	5	280	12	0	12
Limitação de força em membros inferiores	0	151	1	0	0	38	1	0	292	0	0	0
Limitação de força em membros superiores	151	0	0	0	38	0	0	0	0	292	1	0
Distúrbio de marcha	79	72	0	72	27	11	0	11	252	40	0	40
Alto risco nutricional	106	45	0	45	0	38	1	0	178	114	0	114
DISSIMILARIDADE	350				59				496			

(*) Dissimilaridade total = 350 + 59 + 496 = 905

(**) Dissimilaridade média = (350 + 59 + 496) / 3 = 301,67

Tabela I – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=3, no método de agrupamento não Hierárquico k-modas com sementes do vetor de modas do k-médias (sementes da ligação de Ward)

Cluster	1				2				3			
Total de elementos	151				38				292			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	52	99	1	52	38	0	0	0	210	82	0	82
Úlcera de pressão	142	9	0	9	33	5	0	5	270	22	0	22
Cuidados paliativos	101	50	0	50	31	7	0	7	223	69	0	69
Sintoma sistema circulatório	116	35	0	35	29	9	0	9	236	56	0	56
Dispneia em repouso	132	19	0	19	34	4	0	4	279	13	0	13
Uso de cigarros	108	43	0	43	20	18	0	18	204	88	0	88
Paciente acamado	126	25	0	25	33	5	0	5	280	12	0	12
Limitação de força em membros inferiores	0	151	1	0	0	38	1	0	292	0	0	0
Limitação de força em membros superiores	151	0	0	0	38	0	0	0	0	292	1	0
Distúrbio de marcha	79	72	0	72	27	11	0	11	252	40	0	40
Alto risco nutricional	106	45	0	45	0	38	1	0	178	114	0	114
DISSIMILARIDADE	350				59				496			

(*) Dissimilaridade total = 350 + 59 + 496 = 905

(**) Dissimilaridade média = (350 + 59 + 496) / 3 = 301,67

Tabela J – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=4, no método de agrupamento não Hierárquico k-modas com sementes aleatórias

Cluster	1				2				3				4			
Total de elementos	276				95				55				55			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	207	69	0	69	32	63	1	32	49	6	0	6	12	43	1	12
Úlcera de pressão	254	22	0	22	92	3	0	3	50	5	0	5	49	6	0	6
Cuidados paliativos	223	53	0	53	40	55	1	40	42	13	0	13	50	5	0	5
Sintoma sistema circulatório	222	54	0	54	72	23	0	23	44	11	0	11	43	12	0	12
Dispneia em repouso	265	11	0	11	85	10	0	10	49	6	0	6	46	9	0	9
Uso de cigarros	225	51	0	51	86	9	0	9	0	55	1	0	21	34	1	21
Paciente acamado	265	11	0	11	77	18	0	18	50	5	0	5	47	8	0	8
Limitação de força em membros inferiores	248	28	0	28	7	88	1	7	37	18	0	18	0	55	1	0
Limitação de força em membros superiores	28	248	1	28	88	7	0	7	18	37	1	18	55	0	0	0
Distúrbio de marcha	250	26	0	26	19	76	1	19	43	12	0	12	46	9	0	9
Alto risco nutricional	189	87	0	87	55	40	0	40	0	55	1	0	40	15	0	15
DISSIMILARIDADE	440				208				94				97			

(*) Dissimilaridade total = 440 + 208 + 94 + 97 = 839

(**) Dissimilaridade média = (440 + 208 + 94 + 97) / 4 = 209,75

Tabela K – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=4, no método de agrupamento não Hierárquico k-modas com sementes do vetor de modas da ligação de Ward

<i>Cluster</i>	1				2				3				4			
Total de elementos	140				71				235				35			
Variável	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	88	52	0	52	2	69	1	2	179	56	0	56	31	4	0	4
Úlcera de pressão	128	12	0	12	69	2	0	2	215	20	0	20	33	2	0	2
Cuidados paliativos	117	23	0	23	18	53	1	18	192	43	0	43	28	7	0	7
Sintoma sistema circulatório	105	35	0	35	58	13	0	13	189	46	0	46	29	6	0	6
Dispneia em repouso	125	15	0	15	62	9	0	9	225	10	0	10	33	2	0	2
Uso de cigarros	91	49	0	49	57	14	0	14	184	51	0	51	0	35	1	0
Paciente acamado	121	19	0	19	57	14	0	14	229	6	0	6	32	3	0	3
Limitação de força em membros inferiores	0	140	1	0	22	49	1	22	235	0	0	0	35	0	0	0
Limitação de força em membros superiores	140	0	0	0	49	22	0	22	0	235	1	0	0	35	1	0
Distúrbio de marcha	94	46	0	46	22	49	1	22	212	23	0	23	30	5	0	5
Alto risco nutricional	94	46	0	46	16	55	1	16	174	61	0	61	0	35	1	0
DISSIMILARIDADE	297				154				316				29			

(*) Dissimilaridade total = 297 + 154 + 316 + 29 = 796

(**) Dissimilaridade média = (242 + 381 + 114 + 69) / 4 = 199

Tabela L – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=4, no método de agrupamento não Hierárquico k-modas com sementes do vetor de modas do k-médias, a partir de sementes aleatórias

<i>Cluster</i>	1				2				3				4			
Total de elementos	311				39				63				68			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	239	72	0	72	11	28	1	11	20	43	1	20	30	38	1	30
Úlcera de pressão	284	27	0	27	39	0	0	0	62	1	0	1	60	8	0	8
Cuidados paliativos	244	67	0	67	8	31	1	8	48	15	0	15	55	13	0	13
Sintoma sistema circulatório	245	66	0	66	31	8	0	8	49	14	0	14	56	12	0	12
Dispneia em repouso	294	17	0	17	35	4	0	4	55	8	0	8	61	7	0	7
Uso de cigarros	215	96	0	96	27	12	0	12	43	20	0	20	47	21	0	21
Paciente acamado	298	13	0	13	12	27	1	12	62	1	0	1	67	1	0	1
Limitação de força em membros inferiores	282	29	0	29	2	37	1	2	0	63	1	0	8	60	1	8
Limitação de força em membros superiores	29	282	1	29	37	2	0	2	63	0	0	0	60	8	0	8
Distúrbio de marcha	279	32	0	32	33	6	0	6	18	45	1	18	28	40	1	28
Alto risco nutricional	207	104	0	104	14	25	1	14	63	0	0	0	0	68	1	0
DISSIMILARIDADE	552				79				97				136			

(*) Dissimilaridade total = 552 + 79 + 97 + 136 = 864

(**) Dissimilaridade média = (552 + 79 + 97 + 136) / 4 = 216,00

Tabela M – Distribuição de frequências, vetor de modas e dissimilaridade por *cluster* para k=4, no método de agrupamento não Hierárquico k-modas com sementes do vetor de modas do k-médias, a partir de sementes da ligação de Ward

Cluster	1				2				3				4			
Total de elementos	140				71				235				35			
Variáveis	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades	0	1	Moda	nº dissimilaridades
Idade categorizada	88	52	0	52	2	69	1	2	179	56	0	56	31	4	0	4
Úlcera de pressão	128	12	0	12	69	2	0	2	215	20	0	20	33	2	0	2
Cuidados paliativos	117	23	0	23	18	53	1	18	192	43	0	43	28	7	0	7
Sintoma sistema circulatório	105	35	0	35	58	13	0	13	189	46	0	46	29	6	0	6
Dispneia em repouso	125	15	0	15	62	9	0	9	225	10	0	10	33	2	0	2
Uso de cigarros	91	49	0	49	57	14	0	14	184	51	0	51	0	35	1	0
Paciente acamado	121	19	0	19	57	14	0	14	229	6	0	6	32	3	0	3
Limitação de força em membros inferiores	0	140	1	0	22	49	1	22	235	0	0	0	35	0	0	0
Limitação de força em membros superiores	140	0	0	0	49	22	0	22	0	235	1	0	0	35	1	0
Distúrbio de marcha	94	46	0	46	22	49	1	22	212	23	0	23	30	5	0	5
Alto risco nutricional	94	46	0	46	16	55	1	16	174	61	0	61	0	35	1	0
DISSIMILARIDADE	297				154				316				29			

(*) Dissimilaridade total = 297 + 154 + 316 + 29 = 796

(**) Dissimilaridade média = (297 + 154 + 316 + 29) / 4 = 199,00