# ANÁLISE FATORIAL EXPLORATÓRIA APLICADA A DADOS DE AVALIAÇÃO DO ESTADO NUTRICIONAL DE IDOSOS

**Elen Daniele Paulino**

Dissertação apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro - 2012

# ANÁLISE FATORIAL EXPLORATÓRIA APLICADA A DADOS DE AVALIAÇÃO DO ESTADO NUTRICIONAL DE IDOSOS

**Elen Daniele Paulino**

Orientador: Prof. Dr. **José Silvio Govone**

Co-orientadora: Prof$^{\underline{a}}$ Dr$^{\underline{a}}$ **Maria Rita Marques de Oliveria**

Dissertação apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" para a obtenção do título de Mestre em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro - 2012

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO DE AQUIS. E TRAT. DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ***SULAMITA SELMA CLEMENTE COLNAGO***

Paulino, Elen Daniele.
Análise fatorial exploratória aplicada a dados de avaliação do estado nutricional de idosos / Elen Daniele Paulino. – Botucatu : [s.n.], 2012

Dissertação (mestrado) – Universidade Estadual Paulista, Instituto de Biociências de Botucatu
Orientador: José Silvio Govone
Co-orientador: Maria Rita Marques de Oliveira
Capes: 10202056

1. Nutrição - Avaliação. 2. Idosos.

Palavras-chave: Análise de componentes principais; Análise fatorial; Avaliação Nutricional; Idosos.

# Dedicatória

Dedico esta dissertação à minha
avó Benedita Caroli Fregonesi e
aos meus pais José Carlos Paulino e
Aurélia Aparecida Fregonesi Paulino.

# Agradecimentos

Agradeço primeiro a Deus, por não me deixar perder a fé.

Agradeço à minha avó Benedita Caroli Fregonesi, o amor da minha vida, pelo seu imenso amor.

Agradeço aos meus pais Aurélia Aparecida Fregonesi Paulino e José Carlos Paulino pela ajuda, incentivo, amor e carinho.

Agradeço ao meu irmão Eder e à minha irmãzinha Noemi pelo amor, apoio e amizade.

Agradeço aos meus tios Araci, Luis Antônio, Amélia, Antônio e aos meus padrinhos Ana Lúcia e João pelo incentivo e ajuda.

Agradeço ao meu orientador professor Silvio pela ajuda no desenvolvimento do trabalho, pelas palavras de incentivo e por confiar no meu trabalho desde a graduação.

Agradeço à minha co-orientadora professora Maria Rita pela paciência em me ensinar a parte da nutrição e pelo apoio no desenvolvimento da dissertação.

Agradeço à professora Lilian pelos dados e pelo apoio no desenvolvimento do trabalho.

Agradeço aos meus amigos Léo, Davi, Marina, Gabriela Dorth, Fabio e a todos os outros que não citei, pela contribuição que cada um deu ao trabalho e pela amizade.

Agradeço aos professores e funcionários da pós-graduação em Biometria - UNESP - Botucatu.

Agradeço à CAPES pelo apoio financeiro para o desenvolvimento desta dissertação.

# Sumário

Página

**LISTA DE FIGURAS** **vii**

**LISTA DE TABELAS** **ix**

**RESUMO** **xii**

**SUMMARY** **xiv**

**1 INTRODUÇÃO** **1**

**2 REVISÃO DE LITERATURA** **5**

2.1 Análise de componentes principais . . . . . 5

2.1.1 Procedimento para a construção das componenetes principais . . . . . 7

2.2 Análise fatorial exploratória . . . . . 10

2.2.1 Procedimento para a obtenção dos fatores . . . . . 14

2.2.2 Análise de fatores de componentes principais . . . . . 16

**3 MATERIAL E MÉTODOS** **19**

3.1 Obtenção dos dados . . . . . 19

3.1.1 Antropometria . . . . . 20

3.1.2 Exames Bioquímicos . . . . . 22

3.2 Procedimento para uma análise de componentes principais . . . . . 22

3.3 Procedimento para uma análise de fatores de componentes principais . . 25

**4 RESULTADOS** **27**

4.1 Análise de componentes principais . . . . . 27

4.1.1 Análise de componentes principais realizada com 15 variáveis, para todos os idosos . . . . . 27

4.1.2 Análise de componentes principais realizada com 11 variáveis, para todos os idosos . . . . . 31

4.1.3 Análise de componentes principais realizada para os homens, com 11 variáveis . . . . . 34

4.1.4 Análise de componentes principais realizada para as mulheres, com 11 variáveis . . . . . 36

4.1.5 Análise de componentes principais realizada para 13 variáveis, para todos os idosos . . . . . 40

4.1.6 Análise de componentes principais realizada com 13 variáveis, para os idosos de 60-79 anos . . . . . 43

4.1.7 Análise de componentes principais realizada com 13 variáveis, para os idosos de 80-100 anos . . . . . 46

4.1.8 Análise de componentes principais realizada com variáveis antropométricas primárias e as bioquímicas, para todos os idosos . . . . . 48

4.2 Análise fatorial exploratória . . . . . 51

4.2.1 Análise fatorial realizada com 15 variáveis, para todos os idosos . . . . 51

4.2.2 Análise fatorial realizada com 11 variáveis, para todos os idosos . . . . 54

4.2.3 Análise fatorial realizada para os homens, com 11 variáveis . . . . . 55

4.2.4 Análise fatorial realizada para as mulheres, com 11 variáveis . . . . . 57

4.2.5 Análise fatorial realizada para 13 variáveis, para todos os idosos . . . . 58

4.2.6 Análise fatorial realizada com 13 variáveis, para os idosos de 60-79 anos 60

4.2.7 Análise fatorial realizada com 13 variáveis, para os idosos de 80-100 anos 62

4.2.8 Análise fatorial realizada com variáveis antropométricas primárias e as bioquímicas, para todos os idosos . . . . . 62

4.3 Software R usado nas análises . . . . . 64

**5 DISCUSSÃO** **65**

**6 CONCLUSÕES** **69**

**REFERÊNCIAS BIBLIOGRÁFICAS** **71**

# Lista de Figuras

Página


1 Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos. . . . . 29

2 Correlações agrupadas entre as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos. . . . . 30

3 Correlações entre as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos. 31

4 Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos. . . . . 32

5 Correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos. 33

6 Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos. . . . . 35

7 Correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino. . . . . 36

8 Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino. . . . . 38

9 Correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino. . . . . 39

10 Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino. . . . . 41

11 Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos. . . . . 42

12 Correlações entre as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos. 43

13 Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos. . . . . 45
14 Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 60-79 anos. . . . . 46
15 Correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 60-79 anos. 47
16 Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos de 60-79 anos. . . . . 49
17 Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 80-100 anos. . . . . 50
18 Correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 80-100 anos. 51
19 Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos de 80-100 anos. . . . . 53
20 Correlações agrupadas entre as variáveis antropométricas primárias e as bioquímicas, considerando os todos os idosos. . . . . 54
21 Correlações entre as variáveis antropométricas primárias e as bioquímicas, considerando os todos os idosos. . . . . 55
22 Scree-plot dos autovalores *versus* componentes principais para as variáveis antropométricas primárias e as bioquímicas, considerando todos os idosos. 57

# Lista de Tabelas

Página


1 Dados com variáveis $X_1$, ..., $X_p$ e observações em $n$ casos. . . . . 7

2 Correlações entre escores de testes para meninos de uma escola preparatória (Spearman, 1904). . . . . 11

3 Dados originais para variáveis $X_1$, ..., $X_p$, com $n$ idosos. . . . . 23

4 Dados padronizados para variáveis $X_1$, ..., $X_p$, com $n$ idosos. . . . . 24

5 Matriz de correlações entre as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos. . . . . 27

6 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 15 componentes, considerando todos os idosos. . . . . 28

7 Componentes principais (PC) para as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos. . . . . 30

8 Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos. . . . . 32

9 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as variáveis $X_1, X_3, ..., X_{12}$, para todos os idosos. . . . . 34

10 Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos. . . . . 35

11 Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino. . . . . 37

12 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino. . . . . 37
13 Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino. . . . . 38
14 Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino. . . . . 40
15 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino. . . . . 40
16 Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino. . . . . 41
17 Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para todos os idosos. 42
18 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 13 componentes, considerando todos os idosos . . . . . 44
19 Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos. . . . . 45
20 Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 60-79 anos. . . . . 46
21 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 13 componentes, considerando os idosos de 60-79 anos. . . . . 48
22 Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos de 60-79 anos. . . . . 49
23 Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 80-100 anos. . . . . 50

24 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 13 componentes, considerando os idosos de 80-100 anos. . . . . 52
25 Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos 80-100 anos. . . . . 53
26 Matriz de correlações entre as variáveis antropométricas primárias e as bioquímicas, para todos os idosos. . . . . 54
27 Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 10 componentes, considerando os todos os idosos. . . . . 56
28 Componentes principais (PC) para as variáveis antropométricas primárias e as bioquímicas, para todos os idosos. . . . . 56
29 Fatores para as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso. . . . . 58
30 Fatores para as variáveis $X_1, X_3, ..., X_{12}$, para todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso. . . . . 59
31 Fatores para as variáveis $X_1, X_2, ..., X_{15}$, para os homens, para cada variável, assinalado em negrito, o fator de maior peso. . . . . 59
32 Fatores para as variáveis $X_1, X_2, ..., X_{15}$, para as mulheres, para cada variável, assinalado em negrito, o fator de maior peso. . . . . 60
33 Fatores para as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso. . . . . 61
34 Fatores para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos 60-79 anos, para cada variável, assinalado em negrito, o fator de maior peso. . 61
35 Fatores para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos 80-100 anos, para cada variável, assinalado em negrito, o fator de maior peso. . 63
36 Fatores para as variáveis antropométricas primárias e as bioquímicas, considerando todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso. . . . . 64

# ANÁLISE FATORIAL EXPLORATÓRIA APLICADA A DADOS DE AVALIAÇÃO DO ESTADO NUTRICIONAL DE IDOSOS

Autora: ELEN DANIELE PAULINO

Orientador: Prof. Dr. JOSÉ SILVIO GOVONE

Co-orientadora: Profª Drª MARIA RITA MARQUES DE OLIVERIA

## RESUMO

Os idosos normalmente tem baixo consumo alimentar, devido o processo de envelhecimento, resultando em um grave problema, a desnutrição, que associada a idade elevada merece muita atenção. Para um diagnóstico nutricional claro, que possibilite uma intervenção nutricional adequada, deve ser realizada uma avaliação precisa do estado nutricional dos idosos. Essa avaliação é feita com os indicadores objetivos e subjetivos. Os indicadores utilizados para o presente trabalho foram os objetivos: antropométricos e bioquímicos. As variáveis antropométricas foram massa corporal total, estatura, circunferência do braço, circunferência muscular do braço, área de gordura do braço, as dobras cutâneas tricipital, bicipital, subescapular e supra-ilíaca, porcentagem de gordura corporal, circunferência da panturrilha e índice de massa corpórea; e as bioquímicas foram albumina, transferrina e proteína C reativa. Um grande desafio para a nutrição é agrupar e interpretar

as variáveis antropométricas e bioquímicas conjuntamente. Logo, o objetivo do presente trabalho foi realizar a análise de componentes principais nos dados de uma amostra de 113 idosos, para explicar com um menor número de variáveis a maior parte das informações no conjunto original de dados, seguida da análise fatorial exploratória nos mesmos dados, para obter fatores que agrupem as variáveis originais em subconjuntos de novas variáveis mutuamente não correlacionadas. As análises foram realizadas para todos os idosos, com todas as variáveis, e depois, com a retirada das variáveis que possuem baixas correlações com as demais, foram realizadas novas análises para todos os idosos, para os idosos divididos por faixa etária, e por sexo. Com a análise de componentes principais foi possível concluir que a variável circunferência do braço é a mais importante, ou seja, a que possui o maior peso. Na análise fatorial, para todos os idosos sem as variáveis estatura e proteína C reativa, os fatores foram divididos em "indicadores de gordura corporal", "indicadores de massa muscular" e "indicadores bioquímicos do estado nutricional".

Palavras-chave: análise de componentes principais, análise fatorial, avaliação nutricional, idosos.

**EXPLORATORY FACTOR ANALYSIS APPLIED TO NUTRITIONAL ASSESSMENT STATUS DATA OF ELDERLY**

Author: ELEN DANIELE PAULINO
Adviser: Prof. Dr. JOSÉ SILVIO GOVONE
Co-adviser: Profª Drª MARIA RITA MARQUES DE OLIVERIA

## SUMMARY

Elderly people usually have low nourishment consumption, due to the aging process, which results in a serious problem, malnutrition, and, when it is associated with age, it's certainly an issue that deserves special attention. In order to achieve a clear nutritional diagnosis, that is, one that allows an adequate nutritional intervention, it is necessary to develop a nutritional assessment of the elderly. Such assessment can be carried out using objective and subjective indicators. Some of the objectives indicators, anthropometric and biochemical indicators, were approached in this study. The anthropometric variables considered were total body mass; height; arm circumference; arm muscular circumference; arm fat area; triceps, biceps, subscapular and suprailiac skinfold thickness; body fat percentage; calf circumference and body mass index. The biochemical variables considered in this study were albumin, transferrin and C-reactive protein. Both grouping and

interpreting the biochemical and anthropometric variables taken together constitute a major challenge for nutrition. Thus, the objective of this study was performing a principal component analysis for a sample of 113 elders, in order to maximize the explanation of the original data set using fewer variables as possible and then perform an exploratory factor analysis, over the same data, to extract the factors that group up the original variables in a subset of new, mutually non-correlated variables. The analysis were carried out for all elders using all variables and then new analysis were held for all elders, divided into age and sex classes, using the remaining variables after the removal of the lowest correlated ones. Through the principal component analysis, it was possible to conclude that the arm circumference is the most important, that is, the variable which has the biggest weight. Through the factor analysis, for all elders without using height and C-reactive protein variables, the factors were separated in "body fat indicators", "muscular mass indicators" and "biochemical indicators of nutritional status".

Keywords: principal component analysis, factorial analysis, nutritional assessment, elderly.

# 1 INTRODUÇÃO

O envelhecimento populacional é um fenômeno global. Considerando a população mundial, estima-se que o número de pessoas com 60 anos ou mais irá crescer mais de 300% nos próximos 50 anos, indo de 606 milhões em 2000 para quase dois bilhões em 2050. O crescimento será ainda mais expressivo nos países menos desenvolvidos, onde esta população deverá aumentar mais do que 4 vezes, de 374 milhões em 2000 para 1,6 bilhões em 2050 (Scazufca et al., 2002).

Segundo Scazufca et al. (2002) o Brasil é um dos países em que o envelhecimento da população está ocorrendo mais rapidamente. Em 1950 o número de idosos era de aproximadamente 2 milhões, e correspondia a 4,1% da população total, já no ano 2000 esta população aumentou para 13 milhões, equivalente a 7,8% da população.

As estimativas indicam que, em 2025, o país terá mais do que o dobro do número de idosos, ou seja, 35 milhões (15,4% da população total) versus 16 milhões (8,8% da população total) em 2005 (Souza & Guariento, 2009). Nos próximos 50 anos estima-se que a população idosa será de 58 milhões (23,6% da população total) (Scazufca et al., 2002) e que a expectativa de vida ao nascer chegue a 85 anos de idade (Souza & Guariento, 2009).

Junto com o aumento do número de idosos, existe a elevada ocorrência de doenças crônicas nesta faixa etária (Sampaio, 2004). Assim, a maior expectativa de vida associada à frequência dessas doenças tem ocasionado um crescimento das taxas de idosos com incapacidades cognitivas e mecânicas (Karsch, 2003).

Com o processo de envelhecimento, as mudanças fisiológicas, metabólicas e a redução da capacidade funcional resultam na alteração das necessidades

nutricionais dos idosos (Souza & Guariento, 2009), sendo a desnutrição um problema preocupante, relacionado ao baixo consumo alimentar, característico do processo de envelhecimento (Galesi, 2010). Nos idosos mais velhos, a desnutrição é o distúrbio nutricional que se deve dar mais importância, por estar associada ao aumento da mortalidade e da vulnerabilidade a infecções (Souza & Guariento, 2009).

A avaliação do estado nutricional do idoso é considerada complexa em consequência da influência de vários fatores, os quais necessitam ser investigados para um diagnóstico nutricional preciso, que possibilite uma intervenção nutricional adequada (Sampaio, 2004).

Os indicadores que têm sido propostos para avaliar o estado nutricional são antropométricos, físicos, laboratoriais e dietéticos (Souza & Guariento, 2009).

Os indicadores antropométricos mais empregados em estudos com idosos são as medidas de massa corporal total, estatura, circunferência do braço e dobras cutâneas; os indicadores bioquímicos são as proteínas totais, albumina, transferrina, hematócrito, hemoglobina, contagem total de linfócitos e a proteína C reativa. Esta utilizada para excluir a possibilidade de uma inflamação interferir na interpretação dos valores da albumina (Galesi, 2010).

As vantagens da antropometria para a avaliação do estado nutricional do idoso são a simplicidade, os equipamentos de fácil aquisição e o baixo custo operacional para a realização das medidas (Cervi et al., 2005). No entanto, com o envelhecimento ocorrem alterações que podem comprometer a determinação de um diagnóstico antropométrico acurado, caso não sejam tomados cuidados específicos para neutralizar ou amenizar o efeito dessas alterações sobre a avaliação (Sampaio, 2004).

A avaliação nutricional do idoso apresenta limitações nos dados dos padrões de referência para as medidas antropométricas. Nos países desenvolvidos, os estudos para a construção desses padrões foram realizados a partir de amostras representativas, mas com limite superior de idade (Sampaio, 2004). Nos países em desenvolvimento, não há padrões para esse grupo etário (Sampaio, 2004), como é o

caso do Brasil, onde não existem valores antropométricos de referência para idosos (Rauen et al., 2008).

Para os indicadores bioquímicos há falta de referenciais metodológicos específicos para a avaliação do estado nutricional de idosos, e a maioria compõe-se apenas dos pontos de corte, sendo poucas as tabelas de distribuição percentual disponibilizadas (Galesi, 2010).

Um grande desafio para a ciência da nutrição é agrupar e interpretar de forma conjunta as variáveis antropométricas e bioquímicas, pois essas variáveis se complementam para um melhor diagnóstico do estado nutricional do idoso (Félix & Souza, 2009).

No trabalho de Galesi (2010), realizado com dados de 150 idosos de uma Instituição Asilar de Longa Permanência situada na cidade de Piracicaba - SP, foram discutidas algumas das principais questões relacionadas à nutrição no processo de envelhecimento. Este trabalho tinha como objetivo elaborar e validar um instrumento efetivo de vigilância nutricional para uso de cuidadores não formados, voltado para o idoso residente em instituições de longa permanência, a partir de indicadores objetivos e subjetivos.

Para os indicadores objetivos, Galesi (2010) utilizou para a classificação do estado nutricional a soma dos parâmetros nutricionais em percentual de adequação dividido pelo número de parâmetros avaliados e, em seguida classificou este percentual conforme proposto por Blackburn et al. (1977), mas somente utilizou-se as variáveis que foram encontradas referência em percentis na literatura, sendo uma das desvantagens dessa classificação.

Portanto, o objetivo do presente trabalho é realizar uma análise multivariada nos dados dos indicadores objetivos do trabalho de Galesi (2010) para tentar diagnosticar algumas relações importantes, de forma a agrupar as variáveis envolvidas e colaborar com a avaliação do estado nutricional de idosos por nutricionistas ou outros profissionais da área da saúde que realizem esta avaliação. Esta análise multivariada é composta das aplicações: da análise de componentes principais, para

explicar com um menor número de variáveis, que tenham alta correlação, a maior parte das informações no conjunto original de dados; e da análise fatorial, para obter fatores que agrupem as variáveis originais em subconjuntos de novas variáveis mutuamente não correlacionadas.

# 2 REVISÃO DE LITERATURA

Para a realização deste trabalho foram revisadas as técnicas de análise de componentes principais e de análise fatorial exploratória, com o objetivo de posterior aplicação.

## 2.1 Análise de componentes principais

A análise de componentes principais foi inicialmente descrita por Pearson (1901), que propôs um método para o cálculo com duas ou três variáveis apenas. Mais tarde Hotelling (1933) descreveu métodos práticos para muitas variáveis, mas os cálculos eram complexos porque tinham que ser feitos manualmente. Somente com os computadores eletrônicos é que a técnica de componentes principais passou a ser amplamente usada (Manly, 2005).

Segundo Lattin et al. (2011) a análise de componentes principais é um método para reexpressar dados multivariados, ela permite que os dados sejam reorganizados de modo que as primeiras poucas dimensões expliquem o maior número possível de informações disponíveis. Se houver a presença de redundância no conjunto de dados, pode ser possível explicar com um número pequeno de dimensões a maior parte das informações no conjunto original de dados. Essa redução de dimensões torna a visualização dos dados mais direta e a facilita uma análise subsequente.

Geometricamente, a análise de componentes principais consiste em reescrever o sistema de coordenadas da amostra em outro sistema de coordenadas com um eixo mais conveniente para a análise dos dados (Johnson & Wichern, 1986); essa análise permite a construção de gráficos bidimensionais contendo maior informação

estatística (Moita & Neto, 1998).

O objetivo principal é o de explicar a estrutura de variância e covariância de um vetor aleatório composto de $p$-variáveis, através de combinações lineares das variáveis originais; estas combinações são chamadas de componentes principais e são não correlacionadas entre si (Mingoti, 2005).

Tal análise toma $p$ variáveis $X_1$, $X_2$, ..., $X_p$ e obtém as combinações destas, que são as componentes principais $Z_1$, $Z_2$, ..., $Z_p$, cuja principal característica é que estas são obtidas em ordem decrescente de maior variância, ou seja, a componente principal $Z_1$ tem mais informação estatística que a componente principal $Z_2$, e esta tem mais informação estatística que a componente principal $Z_3$, e assim por diante (Anderson, 1958).

A solução de componentes principais possui a propriedade de que as componentes não são correlacionadas (Lattin et al., 2011). Essa falta de correlação das componentes principais significa que os índices estão medindo diferentes "dimensões" dos dados, com a propriedade que $Var(Z_1) \geq Var(Z_2) \geq ... \geq Var(Z_p)$, sendo que a variância de $Z_i$ é denotada por Var($Z_i$), com $i = 1, 2, ..., p$ (Manly, 2005).

No entanto, deseja-se reduzir o número de variáveis a serem avaliadas e interpretar as combinações lineares construídas, ou seja, a informação contida nas $p$-variáveis é substituída pela informação contida em $m$ componentes principais não correlacionadas, com $m < p$. Desta forma, o sistema de variabilidade das $p$-variáveis originais é aproximado pelo sistema de variabilidade das $m$ componentes principais (Mingoti, 2005).

Se as $m$ primeiras componentes principais explicam uma grande parte da variância total do vetor $\mathbf{X}$, pode-se restringir o foco de atenção apenas ao vetor aleatório ($Z_1$, $Z_2$, ..., $Z_m$)'. Desta forma um conjunto $m$-dimensional de variáveis aleatórias poderá ser examinado sem que, com isto, se perca muita informação sobre a estrutura de variâncias e covariâncias do vetor $\mathbf{X}$ (Mingoti, 2005).

Quando as variáveis originais são não correlacionadas, nem sempre reduz o número de variáveis originais a um pequeno número de variáveis transforma-

das. Os melhores resultados são obtidos quando as variáveis originais são altamente correlacionadas, positivamente ou negativamente, ou seja, existe uma quantidade de redundância nas variáveis originais (Manly, 2005).

### 2.1.1 Procedimento para a construção das componenetes principais

Para a análise de componentes principais, com $p$ variáveis para $n$ indivíduos, os dados são dispostos, inicialmente, na Tabela 1.

Tabela 1. Dados com variáveis $X_1$, ..., $X_p$ e observações em $n$ casos.

| Caso | $X_1$ | $X_2$ | $\cdots$ | $X_p$ |
|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{12}$ | $\cdots$ | $x_{1p}$ |
| 2 | $x_{21}$ | $x_{22}$ | $\cdots$ | $x_{2p}$ |
| . | . | . | $\cdots$ | . |
| . | . | . | $\cdots$ | . |
| . | . | . | $\cdots$ | . |
| n | $x_{n1}$ | $x_{n2}$ | $\cdots$ | $x_{np}$ |

Segundo Morrison (1967) e Manly (2005) o procedimento para obter-se as componentes principais é como segue :

Calcula-se o primeiro componente principal como uma combinação linear das variáveis $X_1$, $X_2$, ..., $X_p$, logo

$$Z_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p$$

e esse varia tanto quanto possível para os indivíduos sujeitos à condição

$$a_{11}^2 + a_{12}^2 + ... + a_{1p}^2 = 1.$$

Assim Var($Z_1$) é tão grande quanto possível, dada esta restrição sobre as constantes $a_{1j}$, com $j = 1, 2, ..., p$. A restrição é introduzida, senão a Var($Z_1$) pode ser aumentada fazendo crescer qualquer um dos valores $a_{1j}$.

O segundo componente principal é da seguinte forma

$$Z_2 = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p$$

e é escolhido de modo que Var($Z_2$) seja tão grande quanto possível, sujeito à restrição de que

$$a_{21}^2 + a_{22}^2 + ... + a_{2p}^2 = 1$$

e também à condição de que $Z_1$ e $Z_2$ tenham correlação zero entre si.

O terceiro componente principal é da forma que se segue

$$Z_3 = a_{31}X_1 + a_{32}X_2 + ... + a_{3p}X_p$$

e é tal que a Var($Z_3$) seja tão grande quanto possível, sujeito à restrição de que

$$a_{31}^2 + a_{32}^2 + ... + a_{3p}^2 = 1$$

e também que $Z_3$ seja não correlacionada com ambas $Z_1$ e $Z_2$.

De maneira análoga obtém-se $Z_4, Z_5, ..., Z_p$. Como são $p$-variáveis, pode-se obter no máximo $p$ componentes principais.

A análise de componentes principais envolve encontrar os autovalores da matriz de covariância amostral, a qual é simétrica e tem a forma:

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}$$

na qual o elemento $c_{ii}$ na diagonal é a variância de $X_i$, e o elemento $c_{ij}$, que não está na diagonal, é a covariância entre as variáveis $X_i$ e $X_j$.

Os autovalores encontrados na matriz $\mathbf{C}$ são as variâncias dos componentes principais. Existem $p$ autovalores, alguns dos quais podem ser zero, mas não podem ser negativos. Os autovalores são ordenados como $\lambda_1 \geq \lambda_2 \geq ...\lambda_p \geq 0$, em que $\lambda_i$ corresponde ao $i$-ésimo componente principal que é da seguinte forma (Manly, 2005):

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + ... + a_{ip}X_p$$

sujeito à restrição de que

$$a_{i1}^2 + a_{i2}^2 + ... + a_{ip}^2 = 1.$$

Temos que $Var(Z_i) = \lambda_i$, e as constantes $a_{i1}, a_{i2}, ..., a_{ip}$ são os elementos do autovetor correspondente aos autovalores $\lambda_1, \lambda_2, ..., \lambda_p$.

Uma propriedade importante é que a soma dos autovalores é igual à soma dos elementos da diagonal(traço) da matriz **C**. Isto é,

$$\lambda_1 + \lambda_2 + ... + \lambda_p = c_{11} + c_{22} + ... + c_{pp}.$$

Como $c_{ii}$ é a variância de $X_i$, e $\lambda_i$ é a variância de $Z_i$, então a soma das variâncias dos componentes principais é igual a soma das variâncias das variáveis originais. Portanto, os componentes principais contém toda a variação dos dados originais (Manly, 2005).

As componentes principais obtidas da matriz de covariância **C** são influenciadas pelas variáveis de maior variância, não sendo de muita utilidade nos casos em que existe uma discrepância muito acentuada entre essas variâncias. A discrepância pode ser muitas vezes causada pela diferença de unidades de medidas das variáveis. Este problema pode ser resolvido aplicando-se uma transformação nos dados originais, de modo a equilibrar os valores de variância (Mingoti, 2005). Neste caso é usual codificar ou transformar as variáveis de modo que cada variável tenha média zero e variância igual a um (Manly, 2005).

Esta padronização expressa cada observação como o número de desvios padrões da média como segue (Moita & Neto, 1998):

$$y_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j}$$

em que $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ e ${s_j}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$.

Logo a matriz de covariância **C** é transformada numa matriz de cor-

relação **D** da seguinte forma (Manly, 2005):

$$\mathbf{D} = \begin{bmatrix} 1 & d_{12} & \cdots & d_{1p} \\ d_{21} & 1 & \cdots & d_{2p} \\ \vdots & \vdots & & \vdots \\ d_{p1} & d_{p2} & \cdots & 1 \end{bmatrix}$$

na qual $d_{ij} = d_{ji}$ é a correlação entre $X_i$ e $X_j$. A análise de componentes principais envolve encontrar os autovalores da matriz de correlação. Neste caso, a soma dos termos da diagonal, e a soma dos autovalores, é igual a $p$, o número de variáveis X (Manly, 2005).

Após calcular as componentes $Z_1$ e $Z_2$ pode-se obter o gráfico de $Z_1$ *versus* $Z_2$, que fornece informações estatísticas para observação dos pontos no espaço bi-dimensional (Moita & Neto, 1998).

Para a utilização das componentes principais para uma outra análise, por exemplo uma análise fatorial, deseja-se saber quantas componentes principais devem ser suficientes para explicar grande parte da variância total das variáveis originais. Para Lattin et al. (2011) o pesquisador deve decidir quantos componentes principais reter para a análise subsequente, fazendo uma escolha entre a simplicidade (isto é, um pequeno número de dimensões) e completude (isto é, um grande número de dimensões que captam uma quantidade maior de informações disponíveis). Segundo Manly (2005) é usual somente os primeiros poucos componentes principais, desde que a soma de suas variâncias seja uma porcentagem alta (por exemplo, 80% ou mais) das somas das variâncias para todos os $p$ componentes. Para Moita & Neto (1998) embora a informação estatística presente nas $p$-variáveis originais seja a mesma dos $p$ componentes principais, é comum obter em apenas 2 ou 3 das primeiras componentes principais mais que 90% desta informação.

## 2.2 Análise fatorial exploratória

A análise fatorial surgiu com Spearman (1904), a partir dos esforços do psicólogo em obter uma melhor compreensão da 'inteligência'. Os testes de inte-

ligência contêm questões que dependem de habilidade verbal, habilidade matemática, memorização, etc. Tal técnica foi desenvolvida para analisar estes testes, e assim determinar se a 'inteligência' é medida por um único fator principal envolvendo todos os testes ou por diversos fatores, medindo atributos como 'habilidade matemática'.

Em seu trabalho, Spearman (1904) estudava as correlações entre escores destes testes, e notou que muitas correlações observadas poderiam estar contidas em um modelo simples. Em um dos casos ele obteve uma matriz de correlações (Tabela 2) para os escores dos testes em clássicos, francês, inglês, matemática, dicriminação de tom e música, para meninos de uma escola preparatória . Ele notou que quaisquer duas linhas desta matriz eram quase proporcionais, com exceção das diagonais. Para as linhas clássicos e inglês na Tabela 2, ele obteve as razões:

$$\frac{0,83}{0,67} \approx \frac{0,70}{0,64} \approx \frac{0,66}{0,54} \approx \frac{0,63}{0,51} \approx 1,2.$$

Tabela 2. Correlações entre escores de testes para meninos de uma escola preparatória (Spearman, 1904).

| | Clássicos | Francês | Inglês | Matemática | Discriminação de tom | Música |
|---|---|---|---|---|---|---|
| Clássicos | 1,00 | 0,83 | 0,78 | 0,70 | 0,66 | 0,63 |
| Francês | 0,83 | 1,00 | 0,67 | 0,67 | 0,65 | 0,57 |
| Inglês | 0,78 | 0,67 | 1,00 | 0,64 | 0,54 | 0,51 |
| Matemática | 0,70 | 0,67 | 0,64 | 1,00 | 0,45 | 0,51 |
| Discriminação de tom | 0,66 | 0,65 | 0,54 | 0,45 | 1,00 | 0,40 |
| Música | 0,63 | 0,57 | 0,51 | 0,51 | 0,40 | 1,00 |

Com essa observação, Spearman (1904) sugeriu que os seis escores fossem descritos pela equação

$$X_i = a_i F + e_i$$

em que $X_i$ é o $i$-ésimo escore, sendo que os dados para todos os meninos foram padronizados para ter uma média zero e um desvio-padrão um; o $a_i$ é uma constante; F é um valor "fator"; e $e_i$ é a parte de $X_i$ que é específica para o $i$-ésimo teste somente. Ele mostrou que a razão constante entre as linhas de uma matriz de correlações segue como uma consequência destas suposições, e com isso, este é um modelo plausível para os dados.

A variância de $X_i$ é dada por

$$\begin{aligned} Var(X_i) &= Var(a_iF + e_i) \\ &= Var(a_iF) + Var(e_i) \\ &= a_i^2 Var(F) + Var(e_i) \\ &= a_i^2 + Var(e_i) \end{aligned}$$

pois $a_i$ é uma constante, e assume-se que F e $e_i$ são independentes, e a variância de F é unitária. Como $Var(X_i) = 1$, segue

$$1 = a_i^2 + Var(e_i).$$

Portanto, o quadrado da constante $a_i$, a qual é chamada de carga de fator, é a proporção explicada da variância de $X_i$.

Então, Spearman (1904) formulou a teoria de dois fatores de testes mentais. Nesta teoria, cada resultado do teste é composto de duas partes, uma que é comum a todos os testes (inteligência geral), e outra que é específica para o teste. Isto resulta no modelo de análise de fatores geral:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + ... + a_{im}F_m + e_i$$

no qual $X_i$ é o $i$-ésimo escore do teste com média zero e variância unitária, $a_{i1}$, $a_{i2}$,..., $a_{im}$ são as cargas dos fatores para o $i$-ésimo teste, $F_1$, $F_2$,..., $F_m$ são m fatores comuns não correlacionados, cada um com média zero e variância unitária, e $e_i$ é um fator específico somente para o $i$-ésimo teste que não é correlacionado com qualquer dos fatores comuns e tem média zero.

A variância de $X_i$ para esse modelo é:

$$\begin{aligned} Var(X_i) &= 1 = a_{i1}^2 Var(F_1) + a_{i2}^2 Var(F_2) + ... + a_{im}^2 Var(F_m) + Var(e_i) \\ &= a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2 + Var(e_i) \end{aligned}$$

em que $a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2$ é chamado a comunalidade de $X_i$, é a parte da variância que está relacionada aos fatores comuns, e Var($e_i$) é chamada a especificidade de $X_i$, é a parte da variância que não é relacionada aos fatores comuns.

A correlação entre $X_i$ e $X_j$ é dada por

$$r_{ij} = a_{i1}a_{j1} + a_{i2}a_{j2} + ... + a_{im}a_{jm}.$$

Assim, dois escores de teste podem ser altamente correlacionados se eles têm altas cargas nos mesmos fatores. Como a comunalidade não pode exceder a um, vale que $-1 \leq a_{ij} \leq +1$.

A ideia básica da análise de fatores é que pode-se descrever um conjunto de $p$ variáveis $X_1$, $X_2$, ..., $X_p$ em termos de um número menor de índices ou fatores, e no processo obter uma melhor compreensão da relação destas variáveis (Manly, 2005), essa redução do número de dimensões facilita a visualização dos dados (Lattin et al., 2011).

A diferença da análise de fatores para a análise de componentes principais é que esta não é baseada em um modelo estatístico particular, enquanto que a análise de fatores é baseada em um modelo (Manly, 2005). No modelo fatorial comum existem os fatores específicos, esses fatores específicos são mutuamente não correlacionados e independentes dos fatores comuns subjacentes (Lattin et al., 2011).

A análise fatorial tem como objetivo principal descrever a variabilidade original do vetor aleatório $\mathbf{X}$, com um número menor m de variáveis aleatórias, chamadas de fatores comuns, e que estão relacionadas com o vetor original $\mathbf{X}$ através de um modelo linear. Neste modelo, uma parte da variabilidade de $\mathbf{X}$ é atribuída aos fatores comuns, e a outra parte é atribuída às variáveis que não foram incluídas no modelo, ou seja, ao erro aleatório. A análise fatorial encontra fatores que agrupam

as variáveis originais em subconjuntos de novas variáveis mutuamente não correlacionadas (Mingoti, 2005).

O modelo de fator comum assume pressupostos explícitos sobre como é medida cada variável no conjunto de dados. O modelo sustenta que a variância observada em cada medida pode ser atribuída a um número relativamente pequeno de fatores comuns e a um fator específico. Embora possa haver, mais de um fator específico para cada variável, é impossível distinguir estatisticamente entre um e mais de um fator específico. Assume-se que a variância no fator específico reflete a quantidade de variância no erro, assim quanto menor o erro da variância, mais confiável é a medida (Lattin et al., 2011). Esta técnica é utilizada para a obtenção de índices que podem preservar de forma relevante a totalidade das informações obtidas por meio das variáveis originais (Possoli, 1984).

### 2.2.1 Procedimento para a obtenção dos fatores

Para uma análise de fatores, com $p$ variáveis para $n$ indivíduos, os dados são dispostos da mesma forma que para a análise de componentes principais (Manly, 2005).

Uma abordagem começa com uma análise de componentes principais e os $m$ componentes principais são então tomados como sendo os $m$ fatores provisórios. Os fatores encontrados são não correlacionados entre si, e são também não correlacionados com os fatores específicos (Manly, 2005).

Se $F_1$, $F_2$, ..., $F_m$ são os fatores provisórios, então suas combinações lineares são da seguinte forma

$$\begin{aligned}
F_1^* &= d_{11}F_1 + d_{12}F_2 + ... + d_{1m}F_m \\
F_2^* &= d_{21}F_1 + d_{22}F_2 + ... + d_{2m}F_m \\
&\vdots \\
F_m^* &= d_{m1}F_1 + d_{m2}F_2 + ... + d_{mm}F_m
\end{aligned}$$

eles são construídos não sendo correlacionados e explicam os dados tão bem quanto os

fatores provisórios. Logo, há uma infinidade de soluções alternativas para o modelo de análise de fatores (Manly, 2005).

Isto leva a um próximo passo na análise, em que os fatores provisórios são transformados para encontrar novos fatores que sejam mais fáceis de se interpretar, que é chamado de rotação de fator. Girar ou transformar neste contexto significa escolher os valores $d_{ij}$. Por último calcula-se os escores dos fatores, que são os valores dos fatores rotacionados $F_1^*$, $F_2^*$, ..., $F_m^*$ para cada um dos $n$ indivíduos para os quais os dados estão disponíveis (Manly, 2005).

A rotação de fatores pode ser ortogonal ou oblíqua; na rotação ortogonal os novos fatores são não correlacionados, como os fatores provisórios; já na rotação oblíqua, os novos fatores são correlacionados (Manly, 2005).

A abordagem para encontrar-se uma matriz $\mathbf{T}$ de rotação baseia-se nos princípios da estrutura simples desenvolvida por Thurstone (1947). A idéia geral era encontrar agrupamentos de variáveis na medida do possível, com cada um definindo somente um fator (Lattin et al., 2011).

Com a rotação deseja-se que as cargas de fator para os novos fatores sejam ou próximas de zero ou muito diferentes de zero. Isto significa que para um $a_{ij}$ próximo de zero, então $X_i$ não é fortemente relacionada com o fator $F_j$; e um $a_{ij}$ com um grande valor positivo ou negativo, significa que $X_i$ é determinado em grande parte por $F_j$. Os fatores são identificados mais facilmente quando cada escore de teste é fortemente relacionado com alguns fatores, mas não relacionado com outros (Manly, 2005).

Um método de rotação de fatores ortogonal que é muito usado é chamado de rotação varimax, ele é baseado na suposição de que a interpretabilidade do $j$-ésimo fator pode ser medida pela variância dos quadrados de suas cargas de fator, isto é, a variância de $a_{1j}^2$, $a_{2j}^2$, ..., $a_{mj}^2$. Se esta variância for grande, então os valores $a_{ij}$ tendem a ser ou próximos de zero ou próximos da unidade (Manly, 2005).

A rotação varimax maximiza a soma destas variâncias para todos os fatores. Kaiser (1958) foi o primeiro que sugeriu esta rotação. Mais tarde, ele

modificou-a normalizando as cargas de fator antes de maximizar as variâncias de seus quadrados, porque isto da melhores resultados (Kaiser, 1958). Logo, a rotação varimax pode ser aplicada com ou sem a normalização de Kaiser. Inúmeros outros métodos de rotação ortogonal têm sido propostos (Manly, 2005).

### 2.2.2 Análise de fatores de componentes principais

O procedimento que se segue para a realização da análise de fatores com $p$ variáveis, logo tem-se $p$ componentes principais, é da seguinte forma Manly (2005):

$$\begin{aligned} Z_1 &= b_{11}X_1 + b_{12}X_2 + ... + b_{1p}X_p \\ Z_2 &= b_{21}X_1 + b_{22}X_2 + ... + b_{2p}X_p \\ &\vdots \\ Z_p &= b_{p1}X_1 + b_{p2}X_2 + ... + b_{pp}X_p \end{aligned}$$

em que os valores $b_{ij}$ são obtidos dos autovetores da matriz de correlações. Como esta transformação dos valores de X para valores de Z é ortogonal, o relacionamento inverso é dado por

$$\begin{aligned} X_1 &= b_{11}Z_1 + b_{21}Z_2 + ... + b_{p1}Z_p \\ X_2 &= b_{12}Z_1 + b_{22}Z_2 + ... + b_{p2}Z_p \\ &\vdots \\ X_p &= b_{1p}Z_1 + b_{2p}Z_2 + ... + b_{pp}Z_p \end{aligned}$$

O próximo passo para conduzir a análise fatorial é a estimação de $m$. Isto consiste em encontrar os autovalores da matriz de correlação e ordená-los em ordem decrescente. Observa-se quais autovalores são os mais importantes em termos de grandeza numérica, utilizando os seguintes critérios (Mingoti, 2005):

Critério 1: a proporção da variância total relacionada com cada autovalor $\lambda_i$ é dado por $\lambda_i/p$, com $i = 1, 2..., p$ (Johnson & Wichern, 1986). O valor

de $m$ será igual ao número de autovalores que representam maiores proporções da variância total;

Critério 2: o valor de $m$ será igual ao número de autovalores $\lambda_i$ maiores ou iguais a 1. Este critério foi proposto por Kaiser (1958);

Critério 3: este critério procura no gráfico "scree-plot" (Cattell, 1966) um "ponto de salto", este representa um decréscimo de importância em relação à variância total. Então, o valor de $m$ é igual ao número de autovalores anteriores ao "ponto de salto".

Os critérios levam em consideração a grandeza numérica dos autovalores. Uma escolha adequada do valor de $m$ deve levar em consideração a interpretabilidade dos fatores e a descrição da estrutura do vetor aleatório $\mathbf{Z}$ com um número pequeno de fatores (Mingoti, 2005).

Logo escolhido o valor de $m$, obtem-se $m$ ($m < p$) componentes principais, e assim as últimas equações tornam-se (Manly, 2005):

$$\begin{aligned}
X_1 &= b_{11}Z_1 + b_{21}Z_2 + ... + b_{m1}Z_m + e_1 \\
X_2 &= b_{12}Z_1 + b_{22}Z_2 + ... + b_{m2}Z_m + e_2 \\
&\vdots \\
X_p &= b_{1p}Z_1 + b_{2p}Z_2 + ... + b_{mp}Z_m + e_p
\end{aligned}$$

em que $e_i$ é uma combinação linear dos componentes principais $Z_{m+1}$ a $Z_p$.

Os componentes principais $Z_1$, $Z_2$, ..., $Z_m$ são escalonados para terem variâncias unitárias, como requerido pelos fatores. Isto é feito dividindo $Z_i$ pelo seu desvio-padrão ($\sqrt{\lambda_i}$), o qual é a raiz quadrada do correspondente autovalor na matriz de correlações. As equações ficam da seguinte forma (Manly, 2005):

$$\begin{aligned}
X_1 &= \sqrt{\lambda_1}b_{11}F_1 + \sqrt{\lambda_2}b_{21}F_2 + ... + \sqrt{\lambda_m}b_{m1}F_m + e_1 \\
X_2 &= \sqrt{\lambda_1}b_{12}F_1 + \sqrt{\lambda_2}b_{22}F_2 + ... + \sqrt{\lambda_m}b_{m2}F_m + e_2 \\
&\vdots \\
X_p &= \sqrt{\lambda_1}b_{1p}F_1 + \sqrt{\lambda_2}b_{2p}F_2 + ... + \sqrt{\lambda_m}b_{mp}F_m + e_p
\end{aligned}$$

em que $F_i = Z_i/\sqrt{\lambda_i}$. Obtem-se o modelo de fatores não rotacionado

$$\begin{aligned} X_1 &= a_{11}F_1 + a_{12}F_2 + ... + a_{1m}F_m + e_1 \\ X_2 &= a_{21}F_1 + a_{22}F_2 + ... + a_{2m}F_m + e_2 \\ &\vdots \\ X_p &= a_{p1}F_1 + a_{p2}F_2 + ... + a_{pm}F_m + e_p \end{aligned}$$

na qual $a_{ij} = \sqrt{\lambda_j}b_{ji}$. Os valores do $i$-ésimo fator não rotacionado são os valores do $i$-ésimo componente principal após eles terem sido escalonados.

Após uma rotação varimax ou outro tipo de rotação, tem-se uma nova solução da seguinte forma

$$\begin{aligned} X_1 &= g_{11}F_1^* + g_{12}F_2^* + ... + g_{1m}F_m^* + e_1 \\ X_2 &= g_{21}F_1^* + g_{22}F_2^* + ... + g_{2m}F_m^* + e_2 \\ &\vdots \\ X_p &= g_{p1}F_1^* + g_{p2}F_2^* + ... + g_{pm}F_m^* + e_p \end{aligned}$$

em que $F_i^*$ é o novo $i$-ésimo fator.

Os valores dos fatores rotacionados são dados pela equação matricial

$$\mathbf{F}^* = \mathbf{XG(G'G)}^{-1}$$

em que $\mathbf{F}^*$ é uma matriz $n$ x $m$ contendo os valores para os $m$ fatores rotacionados em suas colunas, com uma linha para cada uma das $n$ linhas originais de dados; $\mathbf{X}$ é a matriz $n$ x $p$ dos dados originais, após codificar as variáveis $X_1, X_2, ..., X_p$ para terem média zero e variância um; e $\mathbf{G}$ é a matriz $p$ x $m$ das cargas de fatores rotacionados (Manly, 2005).

A análise de fatores não é tão objetiva como muitos dos métodos estatísticos, pois não possui unicidade de soluções. Logo, alguns autores como Seber (2004), Hills (1977) e Chatfield & Collins (1980) criticam sua aplicação. Já Rencher (2002) discute o porque muitas vezes a análise fatorial não funciona. Ele observa que existem dados que a análise fatorial não deveria ser usada, mas que em outros o método é muito útil.

# 3 MATERIAL E MÉTODOS

Para a aplicação da metodologia descrita foram utilizados os dados obtidos do trabalho de Galesi (2010), referentes a avaliação do estado nutricional de idosos. Nestes dados foi realizada uma análise de componentes principais, e posteriormente uma análise fatorial exploratória.

## 3.1 Obtenção dos dados

No trabalho de Galesi (2010) foram analisados dados de idosos de ambos os sexos, com idade igual ou superior a 60 anos, de uma Instituição Asilar de Longa Permanência situada na cidade de Piracicaba - SP. Esta instituição foi escolhida por apresentar condições estruturais e o número de idosos necessários para o desenvolvimento da pesquisa.

A amostra calculada foi de 150 idosos. O delineamento amostral utilizado foi o probabilístico estratificado e o tamanho amostral foi calculado segundo Cochran (1977), considerando a população de 238 internos. Para o cálculo do tamanho amostral, Galesi (2010) utilizou uma pesquisa prévia realizada na instituição no ano de 2005, na qual foi observada a prevalência de desnutrição de 20% entre os idosos, utilizando a Mini-Avaliação Nutricional (MAN) como método de avaliação (Guigoz et al., 1994).

O estudo de Galesi (2010) teve um número maior de participantes do sexo feminino (52%), o que condiz com a população brasileira acima de 60 anos, que é composta por 55,1% de mulheres (Brasil, 2003). A média de idade da amostra foi de 77 anos; para os homens, a média foi de 73 anos e para as mulheres foi de 80 anos,

o que mostra que entre os idosos residentes na instituição o número de mulheres com idade mais avançada é maior (Galesi, 2010).

A avaliação do estado nutricional dos idosos foi constituída pelas variáveis objetivas antropométricas e bioquímicas. As antropométricas são massa corporal total ($X_1$), estatura ($X_2$), a circunferência do braço (CB - $X_3$), circunferência muscular do braço (CMB - $X_4$), área de gordura do braço (AGB - $X_5$), as dobras cutâneas tricipital (DCT - $X_6$), bicipital (DCB - $X_7$), subescapular (DCSE - $X_8$) e supra-ilíaca (DCSI - $X_9$), porcentagem de gordura corporal ($X_{10}$), circunferência da panturrilha (CP - $X_{11}$) e índice de massa corpórea (IMC - $X_{12}$); e as bioquímicas são albumina ($X_{13}$), transferrina ($X_{14}$) e proteína C reativa ($X_{15}$) (Galesi, 2010).

A amostra inicial era composta de 150 idosos, mas para a análise de componentes principais e análise fatorial exploratótria apenas as observações completas são utilizadas (Mingoti, 2005), com isso a amostra passou a ter 113 idosos. Destes, 60 são mulheres e 53 são homens, o que continua com um número maior de participantes do sexo feminino.

A redução da amostra ocorreu porque alguns idosos se recusaram a participar da retirada de sangue no dia da coleta, ou já haviam falecido, ou no caso de idosos acamados sem possibilidade de mover-se, houve o comprometimento da mensuração da DCSE (Galesi, 2010).

### 3.1.1 Antropometria

Os dados antropométricos foram obtidos por (Galesi, 2010) da seguinte forma:

A massa corporal total foi obtida com a utilização de uma balança eletrônica portátil tipo plataforma com capacidade para 150kg e precisão de 100g. Os acamados ou utilizadores de cadeiras de rodas foram pesados em cima de suas cadeiras de rodas, em uma balança eletrônica digital com capacidade de 300kg adaptada a uma rampa, e depois o peso da cadeira foi descontado. Já para os idosos

amputados, seguiu o que sugere Kamimura et al. (2002): acrescentar o peso da extremidade amputada ao peso obtido na balança (Galesi, 2010).

Para a medida da estatura, Galesi (2010) utilizou um estadiômetro móvel; já os idosos impossibilitados de permanecer em pé, a medida da estatura foi tomada com o idoso deitado em posição supina com uma fita métrica medindo da extremidade da cabeça à base do pé; e nos casos em que a realização dessa medida não foi possível, utilizou o cálculo estimado da estatura de acordo com Chumlea et al. (1985); obtendo a medida o comprimento da perna até a altura do joelho. A medida da estatura é obtida das seguintes fórmulas (Acuña & Cruz, 2004):

$$H = 64,19 - (0,04.I) + (0,02.AJ)$$

$$M = 84,88 - (0,24.I) + (1,83.AJ)$$

em que I = idade, H = estatura homem (cm), M = estatura mulher (cm) e AJ = altura do joelho (cm).

O Índice de Massa Corporal foi calculado dividindo a massa corporal total (em kg) pelo quadrado da estatura em metros (Massa corporal total/$Estatura^2$ - $kg/m^2$ ) (Sampaio, 2004).

Para a realização das medidas das circunferências do braço (CB) e da panturrilha (CP) em cm, utilizou-se a fita métrica inelástica e flexível com precisão de 1 mm; e para as dobras cutâneas tricipital (DCT), bicipital (DCB), subescapular (DCSE) e supra-ilíaca (DCSI) em mm, foi utilizado o adipômetro (Galesi, 2010). A circunferência da panturrilha (CP) tem sido considerada um bom indicador de desnutrição (Rauen et al., 2008).

As dobras cutâneas tricipital (DCT), bicipital (DCB), subescapular (DCSE) e supra-ilíaca (DCSI) foram mensuradas para a estimativa da porcentagem de gordura corporal (Galesi, 2010). A CB e a DCT foram utilizadas para a obtenção da circunferência muscular do braço (CMB) em cm e da área de gordura do braço (AGB) em $cm^2$. A equação para estimar a CMB é da seguinte forma (Sampaio,

2004):

$$CMB = CB - (\pi.DCT)$$

### 3.1.2 Exames Bioquímicos

Os indicadores bioquímicos para a classificação do estado nutricional foram obtidos de exames de sangue, com exceção dos idosos que se recusaram a participar da retirada do sangue justamente no dia da coleta, ou que já haviam falecido (Galesi, 2010).

Os idosos fizeram um jejum mínimo de 3 horas para a realização da coleta de amostras de sangue, feita por pessoal técnico habilitado. Estas amostras foram transportadas para o laboratório de análises, em que foram analisadas utilizando metodologia respaldada com controle de qualidade interno e externo pelos seguintes métodos: colorimétrico enzimático para análise de transferrina e das proteínas totais e frações e imunoaglutinação direta em látex para proteína C reativa (Galesi, 2010).

## 3.2 Procedimento para uma análise de componentes principais

Para a análise de componentes principais nos dados de Galesi (2010), com $p$ variáveis para $n$ indivíduos, os dados amostrais seguem o esquema da Tabela 3, em que cada $x_{ij}$ corresponde a um valor de medida da variável $X_j$ realizado em um idoso $i$, com $i = 1, 2, ..., n$ e $j = 1, 2..., p$.

Existe uma discrepância entre as variâncias das variáveis originais, pois as unidades de medida não são as mesmas para todas as variáveis. Assim, cada componente passa a ser dominada por uma variável, o que torna as componentes sem muita utilidade prática. Logo, é necessário que seja feita alguma transformação nos dados, de modo a equilibrar melhor as variâncias (Mingoti, 2005).

Então o primeiro passo para análise de componentes principais é padronizar os dados com variância um e média zero (Manly, 2005). A padronização

Tabela 3. Dados originais para variáveis $X_1$, ..., $X_p$, com $n$ idosos.

| Idosos | $X_1$ | $X_2$ | $\cdots$ | $X_p$ |
|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{12}$ | $\cdots$ | $x_{1p}$ |
| 2 | $x_{21}$ | $x_{22}$ | $\cdots$ | $x_{2p}$ |
| . | . | . | $\cdots$ | . |
| . | . | . | $\cdots$ | . |
| . | . | . | $\cdots$ | . |
| n | $x_{n1}$ | $x_{n2}$ | $\cdots$ | $x_{np}$ |

para cada observação é feita da seguinte forma (Moita & Neto, 1998):

$$y_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j}$$

em que $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ e ${s_j}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$.

Logo, com a padronização dos dados dos $n$ indivíduos, as variáveis $X_1$, $X_2$, ...., $X_p$ obtém-se novos dados, que estão dispostos na Tabela 4, em que cada $y_{ij}$ corresponde a um valor de medida da variável $X_j$ realizado em um idoso $i$.

A análise de componentes principais é feita sobre a matriz de correlação **D**, esta é da seguinte forma (Manly, 2005):

$$\mathbf{D} = \begin{bmatrix} 1 & d_{12} & \cdots & d_{1p} \\ d_{21} & 1 & \cdots & d_{2p} \\ \vdots & \vdots & & \vdots \\ d_{p1} & d_{p2} & \cdots & 1 \end{bmatrix}$$

em que $d_{ij} = d_{ji}$ é a correlação entre $X_i$ e $X_j$. Logo, a soma dos termos da diagonal, e a soma dos autovalores, é igual a $p$, que é o número de variáveis X (Manly, 2005). Da matriz de correlação encontram-se os autovalores $\lambda_1$, $\lambda_2$, ..., $\lambda_p$, e seus correspondentes autovetores $a_1$, $a_2$, ..., $a_p$.

Tabela 4. Dados padronizados para variáveis $X_1$, ..., $X_p$, com $n$ idosos.

| Idosos | $X_1$ | $X_2$ | $\cdots$ | $X_p$ |
|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{12}$ | $\cdots$ | $y_{1p}$ |
| 2 | $y_{21}$ | $y_{22}$ | $\cdots$ | $y_{2p}$ |
| . | . | . | $\cdots$ | . |
| . | . | . | $\cdots$ | . |
| . | . | . | $\cdots$ | . |
| n | $y_{n1}$ | $y_{n2}$ | $\cdots$ | $y_{np}$ |

Portanto, obtém-se $p = 15$ componentes principais, que são da seguinte forma:

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p \\ Z_2 &= a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p \\ &\vdots \\ Z_p &= a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p \end{aligned}$$

em que os valores $a_{ij}$ são dados pelos autovetores da matriz de correlação $\mathbf{D}$.

Para a utilização das componentes principais para a análise fatorial, deseja-se saber quantas componentes principais devem ser suficientes para serem usadas nesta análise. Utiliza-se o critério de proporção da variância total relacionada com cada autovalor $\lambda_i$ ($\lambda_i/p$, com $i = 1, 2...p$) (Johnson & Wichern, 1986), o valor de $m$ será igual ao número de autovalores que representam maiores proporções da variância total; outro critério usado é que o valor de $m$ será igual ao número de autovalores $\lambda_i$ maiores ou iguais a 1, este critério foi proposto por Kaiser (1958); e mais um critério é o de procurar no gráfico "scree-plot" (Cattell, 1966) um "ponto de salto", então o valor de $m$ é igual ao número de autovalores anteriores ao "ponto de salto".

## 3.3 Procedimento para uma análise de fatores de componentes principais

Com as $p$ componentes principais, dadas da seguinte forma:

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p \\ Z_2 &= a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p \\ &\vdots \\ Z_p &= a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p \end{aligned}$$

em que os valores $a_{ij}$ são dados pelos autovetores da matriz de correlação $\mathbf{D}$. Pode-se obter o relacionamento inverso, pois esta transformação dos valores X para valores Z é ortogonal, que é como segue:

$$\begin{aligned} X_1 &= b_{11}Z_1 + b_{21}Z_2 + ... + b_{p1}Z_p \\ X_2 &= b_{12}Z_1 + b_{22}Z_2 + ... + b_{p2}Z_p \\ &\vdots \\ X_p &= b_{1p}Z_1 + b_{2p}Z_2 + ... + b_{pp}Z_p \end{aligned}$$

O próximo passo para conduzir a análise fatorial é a estimação de $m$. Observa-se quais autovalores são os mais importantes, utilizando os seguintes critérios (Mingoti, 2005):

Critério 1: a proporção da variância total relacionada com cada autovalor $\lambda_i$ é dado por $\lambda_i/p$, com i = 1, 2 ...p (Johnson & Wichern, 1986). O valor de $m$ será igual ao número de autovalores que representam maiores proporções da variância total;

Critério 2: o valor de $m$ será igual ao número de autovalores $\lambda_i$ maiores ou iguais a 1, este critério foi proposto por Kaiser (1958);

Critério 3: este critério procura no gráfico "scree-plot" (Cattell, 1966) um "ponto de salto", este representa um decréscimo de importância em relação à variância total. Então, o valor de $m$ é igual ao número de autovalores anteriores ao "ponto de salto".

Uma escolha adequada do valor de $m$ deve levar em consideração a interpretabilidade dos fatores e os critérios 1, 2 e 3 (Mingoti, 2005).

Escolhido o valor de $m$, obtem-se o modelo de fatores não rotacionado (Manly, 2005):

$$\begin{aligned}
X_1 &= a_{11}F_1 + a_{12}F_2 + ... + a_{1m}F_m + e_1 \\
X_2 &= a_{21}F_1 + a_{22}F_2 + ... + a_{2m}F_m + e_2 \\
&\vdots \\
X_p &= a_{p1}F_1 + a_{p2}F_2 + ... + a_{pm}F_m + e_p
\end{aligned}$$

em que $a_{ij} = \sqrt{\lambda_j} b_{ji}$.

Após uma rotação varimax ou outro tipo de rotação, uma nova solução tem a forma (Manly, 2005):

$$\begin{aligned}
X_1 &= g_{11}F_1^* + g_{12}F_2^* + ... + g_{1m}F_m^* + e_1 \\
X_2 &= g_{21}F_1^* + g_{22}F_2^* + ... + g_{2m}F_m^* + e_2 \\
&\vdots \\
X_p &= g_{p1}F_1^* + g_{p2}F_2^* + ... + g_{pm}F_m^* + e_p
\end{aligned}$$

em que $F_i^*$ representa o novo $i$-ésimo fator.

# 4 RESULTADOS

## 4.1 Análise de componentes principais

### 4.1.1 Análise de componentes principais realizada com 15 variáveis, para todos os idosos

Aplicou-se a análise de componentes principais aos dados de 113 idosos para as variáveis massa corporal total ($X_1$), estatura ($X_2$), CB ($X_3$), CMB ($X_4$), AGB ($X_5$), DCT ($X_6$), DCB ($X_7$), DCSE ($X_8$), DCSI ($X_9$), porcentagem de gordura corporal ($X_{10}$), CP ($X_{11}$), IMC ($X_{12}$), albumina ($X_{13}$), transferrina ($X_{14}$) e proteína C reativa ($X_{15}$). Os dados foram padronizados porque as variáveis possuem diferentes unidades de medida. Logo, a análise de componentes principais foi feita sobre a matriz de correlações, dada pela Tabela 5.

Tabela 5: Matriz de correlações entre as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos.

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | $X_{11}$ | $X_{12}$ | $X_{13}$ | $X_{14}$ | $X_{15}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 1.00 | 0.41 | 0.76 | 0.77 | 0.57 | 0.41 | 0.38 | 0.57 | 0.53 | 0.38 | 0.72 | 0.80 | -0.04 | -0.16 | -0.20 |
| $X_2$ | 0.41 | 1.00 | -0.05 | 0.14 | -0.22 | -0.27 | -0.14 | -0.19 | -0.24 | -0.40 | 0.14 | -0.21 | -0.08 | -0.20 | -0.18 |
| $X_3$ | 0.76 | -0.05 | 1.00 | 0.87 | 0.89 | 0.74 | 0.60 | 0.75 | 0.71 | 0.70 | 0.67 | 0.85 | 0.03 | -0.00 | -0.03 |
| $X_4$ | 0.77 | 0.14 | 0.87 | 1.00 | 0.56 | 0.31 | 0.34 | 0.57 | 0.51 | 0.35 | 0.61 | 0.74 | 0.00 | -0.00 | -0.11 |
| $X_5$ | 0.57 | -0.22 | 0.89 | 0.56 | 1.00 | 0.95 | 0.70 | 0.73 | 0.72 | 0.85 | 0.54 | 0.76 | 0.06 | 0.02 | 0.06 |
| $X_6$ | 0.41 | -0.27 | 0.74 | 0.31 | 0.95 | 1.00 | 0.70 | 0.66 | 0.66 | 0.86 | 0.46 | 0.62 | 0.06 | -0.00 | 0.09 |
| $X_7$ | 0.38 | -0.14 | 0.60 | 0.34 | 0.70 | 0.70 | 1.00 | 0.55 | 0.58 | 0.70 | 0.42 | 0.48 | 0.00 | 0.01 | 0.14 |
| $X_8$ | 0.57 | -0.19 | 0.75 | 0.57 | 0.73 | 0.66 | 0.55 | 1.00 | 0.72 | 0.77 | 0.54 | 0.74 | 0.06 | -0.01 | 0.03 |
| $X_9$ | 0.53 | -0.24 | 0.71 | 0.51 | 0.72 | 0.66 | 0.58 | 0.72 | 1.00 | 0.83 | 0.43 | 0.72 | 0.01 | -0.01 | 0.10 |
| $X_{10}$ | 0.38 | -0.40 | 0.70 | 0.35 | 0.85 | 0.86 | 0.70 | 0.77 | 0.83 | 1.00 | 0.44 | 0.66 | 0.04 | 0.05 | 0.17 |
| $X_{11}$ | 0.72 | 0.14 | 0.67 | 0.61 | 0.54 | 0.46 | 0.42 | 0.54 | 0.43 | 0.44 | 1.00 | 0.66 | 0.03 | -0.15 | -0.26 |
| $X_{12}$ | 0.80 | -0.21 | 0.85 | 0.74 | 0.76 | 0.62 | 0.48 | 0.74 | 0.72 | 0.66 | 0.66 | 1.00 | 0.02 | -0.02 | -0.09 |
| $X_{13}$ | -0.04 | -0.08 | 0.03 | 0.00 | 0.06 | 0.06 | 0.00 | 0.06 | 0.01 | 0.04 | 0.03 | 0.02 | 1.00 | 0.03 | -0.05 |
| $X_{14}$ | -0.16 | -0.20 | -0.00 | -0.00 | 0.02 | -0.00 | 0.01 | -0.01 | -0.01 | 0.05 | -0.15 | -0.02 | 0.03 | 1.00 | -0.08 |
| $X_{15}$ | -0.20 | -0.18 | -0.03 | -0.11 | 0.06 | 0.09 | 0.14 | 0.03 | 0.10 | 0.17 | -0.26 | -0.09 | -0.05 | -0.08 | 1.00 |

Da matriz de correlações (Tabela 5) obtém-se os autovalores e a suas porcentagens de variância total explicada, esses valores são dados na Tabela 6. Com os autovalores e seus correspondentes autovetores, obtém-se as 15 componentes principais.

Nota-se também, que na matriz de correlações (Tabela 5) as variáveis $X_2, X_{13}, X_{14}$ e $X_{15}$ possuem baixas correlações com as demais variáveis, fato que influencia nos resultados, pois a análise de componentes principais produz melhores resultados quando as variáveis são altamente correlacionadas.

Tabela 6. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 15 componentes, considerando todos os idosos.

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 7,4631 | 49,75% | 49,75% |
| Componente 2 | 2,1495 | 14,33% | 64,08% |
| Componente 3 | 1,1577 | 7,72% | 71,80% |
| Componente 4 | 0,9859 | 6,57% | 78,37% |
| Componente 5 | 0,7853 | 5,24% | 83,61% |
| Componente 6 | 0,7336 | 4,89% | 88,50% |
| Componente 7 | 0,4325 | 2,88% | 91,38% |
| Componente 8 | 0,3853 | 2,57% | 93,95% |
| Componente 9 | 0,3374 | 2,25% | 96,20% |
| Componente 10 | 0,2729 | 1,82% | 98,02% |
| Componente 11 | 0,2151 | 1,43% | 99,46% |
| Componente 12 | 0,0675 | 0,45% | 99,91% |
| Componente 13 | 0,0079 | 0,05% | 99,96% |
| Componente 14 | 0,0062 | 0,04% | 100,00% |
| Componente 15 | 0 | 0,00% | 100,00% |

Da Tabela 6, tem-se que os 3 primeiros autovalores são $\lambda_1 = 7,4631$, $\lambda_2 = 2,1495$ e $\lambda_3 = 1,1577$, ou seja, os 3 primeiros autovalores são maiores que 1, e eles explicam 49,75%, 14,33% e 7,72% da variância total, respectivamente. Os 3 autovalores juntos explicam 71,80% da variância total.

A escolha do número de componentes principais suficientes para explicar grande parte da variância total das variáveis originais, segue o critério dos autovalores que são maiores que 1. No gráfico scree-plot (Figura 1) pode-se observar que a partir da terceira componente principal, os autovalores decaem mais lenta-

mente, o que também contribui para a decisão do número de componentes. As 3 componentes principais são dadas na Tabela 7.

A primeira componente é dada por:

$$Z_1 = 0,2725X_1 - 0,054X_2 + 0,346X_3 + 0,2654X_4 + 0,3385X_5 + 0,3039X_6 + 0,2609X_7 + 0,3099X_8 + 0,3028X_9 + 0,3103X_{10} + 0,2585X_{11} + 0,3265X_{12} + 0,0144X_{13} - 0,0093X_{14} - 0,0006X_{15}$$

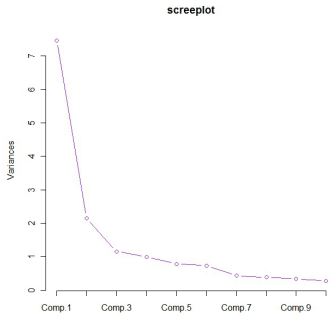


Figura 1 - Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos.

A primeira componente principal contrasta as variáveis massa corporal total, CB, CMB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura, CP, IMC e albumina, com as variáveis estatura, transferrina e proteína C reativa. Assim, um idoso que possui valores altos para as variáveis massa corporal total, CB, CMB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura, CP, IMC e albumina, e valores baixos para as variáveis estatura, transferrina e proteína C reativa, tende a ter um escore alto. E caso inverso, tende a ter um escore baixo.

As variáveis $X_2, X_{13}, X_{14}$ e $X_{15}$ apresentam um peso muito pequeno na componente, o que era de se esperar pelo fato de possuírem baixa correlações com as outras variáveis. As variáveis CB e AGB são as que possuem o maior peso nesta componente, sendo as de maior importânca.

Tabela 7. Componentes principais (PC) para as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos.

| Variável | PC1 | PC2 | PC3 |
|---|---|---|---|
| Massa corporal | 0,2725 | -0,4063 | -0,0623 |
| Estatura | -0,054 | -0,5261 | -0,2397 |
| CB | **0,346** | -0,0976 | 0,0357 |
| CMB | 0,2654 | -0,3075 | 0,0896 |
| AGB | **0,3385** | 0,1253 | -0,0034 |
| DCT | 0,3039 | 0,2287 | -0,0528 |
| DCB | 0,2609 | 0,1719 | -0,1491 |
| DCSE | 0,3099 | 0,0547 | 0,0223 |
| DCSI | 0,3028 | 0,1309 | -0,0536 |
| % G | 0,3103 | 0,2916 | -0,0398 |
| CP | 0,2585 | -0,2931 | 0,0462 |
| IMC | 0,3265 | -0,0901 | 0,1017 |
| Albumina | 0,0144 | 0,0655 | 0,4272 |
| Transferrina | -0,0093 | 0,1827 | 0,6412 |
| Proteína C Reativa | -0,0006 | 0,1827 | -0,5416 |

As componentes principais 2 e 3 explicam uma porcentagem de variância muito baixa (14,33% e 7,72%, respectivamente), logo não apresentam resultados muito bons.

Como as variáveis $X_2, X_{13}, X_{14}$ e $X_{15}$ possuem baixas correlações (Tabela 5, Figura 2 e Figura 3) com as demais, foram realizadas outras análises de componentes principais excluindo essas variáveis.

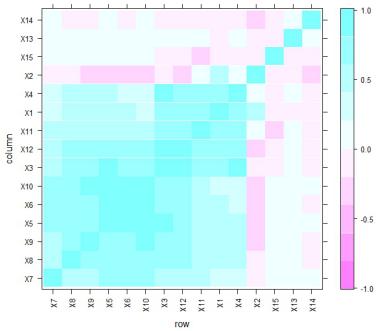


Figura 2 - Correlações agrupadas entre as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos.

Realizou-se a análise de componentes principais com as todas as

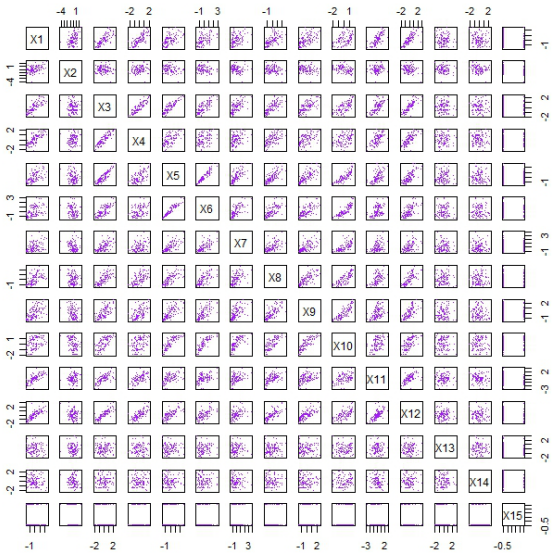


Figura 3 - Correlações entre as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos.

variáveis separados por sexo. Obteve-se conclusões semelhantes à situação quando considerados todos os idosos.

### 4.1.2 Análise de componentes principais realizada com 11 variáveis, para todos os idosos

Com a retirada das variáveis $X_2, X_{13}, X_{14}$ e $X_{15}$ pode-se observar nas Figuras 4 e 5 que as variáveis $X_1, X_3, ..., X_{12}$ possuem altas correlações entre si. Na Figura 4 não observa-se valores de baixas correlações, ou seja, não tem valores da cor rosa; e na Figura 5 pode-se observar uma pequena dispersão dos pontos ao redor de uma reta inclinada, ou seja, as correlações são altas.

Realizou-se uma nova análise de componentes principais nos dados dos 113 idosos, agora somente com as 11 variáveis $(X_1, X_3, ..., X_{12})$.

Baseada na matriz de correlações dada pela Tabela 8, obtém-se os 2 primeiros autovalores, que são $\lambda_1 = 7,4424$ e $\lambda_2 = 1,4699$, e eles explicam 67,66% e 13,36% da variância, respectivamente; e os 2 autovalores juntos explicam 81,02% da

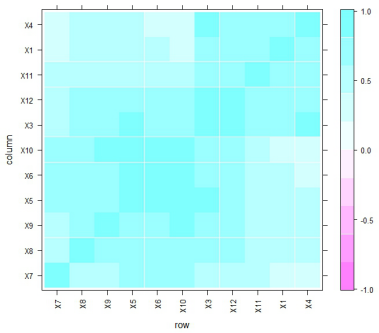


Figura 4 - Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos.

Tabela 8: Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos.

| | $X_1$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | $X_{11}$ | $X_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 1.00 | 0.76 | 0.77 | 0.57 | 0.41 | 0.38 | 0.57 | 0.53 | 0.38 | 0.72 | 0.80 |
| $X_3$ | 0.76 | 1.00 | 0.87 | 0.89 | 0.74 | 0.60 | 0.75 | 0.71 | 0.70 | 0.67 | 0.85 |
| $X_4$ | 0.77 | 0.87 | 1.00 | 0.56 | 0.31 | 0.34 | 0.57 | 0.51 | 0.35 | 0.61 | 0.74 |
| $X_5$ | 0.57 | 0.89 | 0.56 | 1.00 | 0.95 | 0.70 | 0.73 | 0.72 | 0.85 | 0.54 | 0.76 |
| $X_6$ | 0.41 | 0.74 | 0.31 | 0.95 | 1.00 | 0.70 | 0.66 | 0.66 | 0.86 | 0.46 | 0.62 |
| $X_7$ | 0.38 | 0.60 | 0.34 | 0.70 | 0.70 | 1.00 | 0.55 | 0.58 | 0.70 | 0.42 | 0.48 |
| $X_8$ | 0.57 | 0.75 | 0.57 | 0.73 | 0.66 | 0.55 | 1.00 | 0.72 | 0.77 | 0.54 | 0.74 |
| $X_9$ | 0.53 | 0.71 | 0.51 | 0.72 | 0.66 | 0.58 | 0.72 | 1.00 | 0.83 | 0.43 | 0.72 |
| $X_{10}$ | 0.38 | 0.70 | 0.35 | 0.85 | 0.86 | 0.70 | 0.77 | 0.83 | 1.00 | 0.44 | 0.66 |
| $X_{11}$ | 0.72 | 0.67 | 0.61 | 0.54 | 0.46 | 0.42 | 0.54 | 0.43 | 0.44 | 1.00 | 0.66 |
| $X_{12}$ | 0.80 | 0.85 | 0.74 | 0.76 | 0.62 | 0.48 | 0.74 | 0.72 | 0.66 | 0.66 | 1.00 |

variância total (Tabela 9).

Na Figura 6, pode-se visualizar no gráfico scree-plot que a partir da segunda componente principal, os autovalores decaem mais lentamente, o que também contribui para a decisão do número de componentes. As 2 componentes principais são dadas na Tabela 10.

A primeira componente é dada por:

$$Z_1 = 0,277X_1 + 0,3473X_3 + 0,268X_4 + 0,3381X_5 + 0,3027X_6 + 0,2608X_7 + 0,3097X_8 + 0,3021X_9 + 0,3083X_{10} + 0,2608X_{11} + 0,3264X_{12}$$

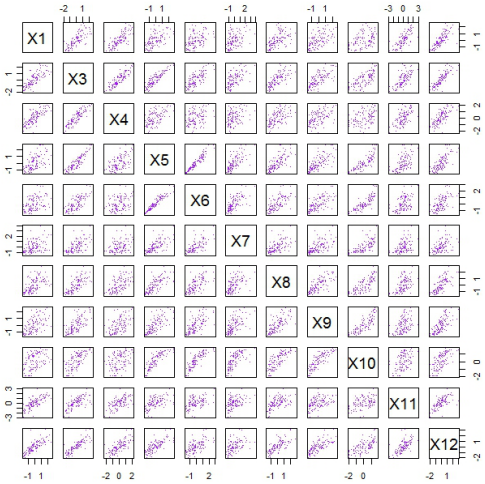


Figura 5 - Correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos.

A primeira componente é um índice global das variáveis, em que todas as variáveis são importantes, de acordo com os valores numéricos dos respectivos coeficientes na combinação linear; a CB e a AGB são as variáveis mais importantes, pois elas possuem os maiores pesos. Essa componente representa os pesos das variáveis $X_1, X_3, ..., X_{12}$, esses pesos podem ser utilizados na classificação do estado nutricional do idoso.

A segunda componente contrasta massa corporal total, CB, CMB, CP e IMC, com AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura. Assim, um idoso que possui valores altos para as variáveis massa corporal total, CB, CMB, CP e IMC, e valores baixos para as variáveis AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura, tende a ter um escore baixo. E caso inverso, tende a ter um escore alto.

Na sequência, aplicou-se a análise de componentes principais nas 11 variáveis considerando os dois sexos separadamente.

Tabela 9. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as variáveis $X_1, X_3, ..., X_{12}$, para todos os idosos.

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 7,4424 | 67,66% | 67,66% |
| Componente 2 | 1,4699 | 13,36% | 81,02% |
| Componente 3 | 0,5557 | 5,05% | 86,07% |
| Componente 4 | 0,4131 | 3,76% | 89,83% |
| Componente 5 | 0,3942 | 3,58% | 93,41% |
| Componente 6 | 0,2839 | 2,58% | 95,99% |
| Componente 7 | 0,2264 | 2,06% | 98,05% |
| Componente 8 | 0,1362 | 1,24% | 99,29% |
| Componente 9 | 0,0702 | 0,64% | 99,93% |
| Componente 10 | 0,0081 | 0,07% | 100,00% |
| Componente 11 | 0 | 0,00% | 100,00% |

### 4.1.3 Análise de componentes principais realizada para os homens, com 11 variáveis

Na Tabela 11 e Figura 7, pode-se observar que as variáveis $X_1, X_3, ..., X_{12}$ são altamente correlacionadas entre si, considerando-se somente dados de idosos do sexo masculino.

Realizou-se a análise de componentes principais nos dados dos 53 idosos do sexo masculino para as 11 variáveis $(X_1, X_3, ..., X_{12})$. Esta análise foi feita sobre a matriz de correlações dada pela Tabela 11.

Pode-se observar na Tabela 12 que os 2 primeiros autovalores são $\lambda_1 = 7,7733$ e $\lambda_2 = 1,4561$, e eles explicam 70,67% e 13,24% da variância total, respectivamente; e os 2 autovalores juntos explicam 83,90% da variância total.

O gráfico scree-plot (Figura 8) mostra que a partir da segunda componente principal, os autovalores decaem mais lentamente, logo, tem-se 2 componentes principais, que são dadas na Tabela 13.

A primeira componente é da seguinte forma:

$$Z_1 = 0,3287X_1 + 0,3381X_3 + 0,2729X_4 + 0,3196X_5 + 0,2668X_6 +$$

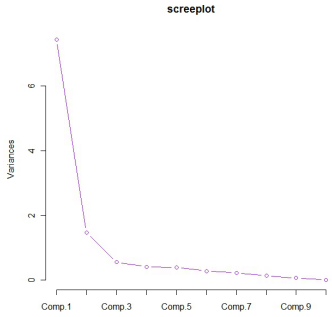


Figura 6 - Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos.

Tabela 10. Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{12}$, considerando todos os idosos.

| Variável | PC1 | PC2 |
|---|---|---|
| Massa corporal | 0,277 | -0,4378 |
| CB | **0,3473** | -0,1363 |
| CMB | 0,268 | -0,4642 |
| AGB | **0,3381** | 0,1948 |
| DCT | 0,3027 | 0,3664 |
| DCB | 0,2608 | 0,3083 |
| DCSE | 0,3097 | 0,0595 |
| DCSI | 0,3021 | 0,1564 |
| % G | 0,3083 | 0,3754 |
| CP | 0,2608 | -0,319 |
| IMC | 0,3264 | -0,191 |

$$0,2261X_7 + 0,3047X_8 + 0,2946X_9 + 0,3398X_{10} + 0,2876X_{11} + 0,3176X_{12}$$

A primeira componente é um índice global das variáveis, em que todas as variáveis são importantes, de acordo com os valores numéricos dos respectivos coeficientes na combinação linear; sendo que os maiores valores referem-se a porcentagem de gordura e CB, essas são as mais importantes. Essa componente representa os pesos de $X_1, X_3, ..., X_{12}$, esses pesos podem ser utilizados na classificação do estado nutricional dos idosos do sexo masculino.

A segunda componente contrasta massa corporal total, CB, CMB,

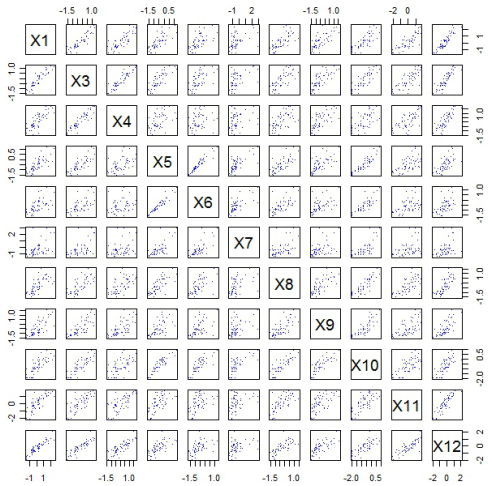


Figura 7 - Correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino.

DCSI, CP e IMC, com AGB, DCT, DCB, DCSE e porcentagem de gordura. Assim, um idoso do sexo masculino que possui valores altos para as variáveis AGB, DCT, DCB, DCSE e porcentagem de gordura, e valores baixos para as variáveis massa corporal total, CB, CMB, DCSI, CP e IMC, tende a ter um escore alto. E caso contrário, tende a ter um escore baixo.

### 4.1.4 Análise de componentes principais realizada para as mulheres, com 11 variáveis

Na Tabela 14 e na Figura 9, pode-se observar que as variáveis $X_1, X_3, ..., X_{12}$ possuem altos valores de correlações para as mulheres.

Realizou-se a análise de componentes principais nos dados das 60 idosas para as 11 variáveis $(X_1, X_3, ..., X_{12})$.

Da matriz de correlações (Tabela 14) obtém-se os autovalores e suas porcentagens de variância, que são dados na Tabela 15, tem-se que o primeiro autovalor é $\lambda_1 = 8,0391$, e ele explica 73,08% da variância total.

Observa-se no gráfico scree-plot (Figura 10) que a partir da segunda

Tabela 11: Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino.

| | $X_1$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | $X_{11}$ | $X_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 1.00 | 0.89 | 0.78 | 0.75 | 0.58 | 0.45 | 0.72 | 0.72 | 0.78 | 0.79 | 0.91 |
| $X_3$ | 0.89 | 1.00 | 0.90 | 0.81 | 0.61 | 0.51 | 0.74 | 0.75 | 0.83 | 0.74 | 0.86 |
| $X_4$ | 0.78 | 0.90 | 1.00 | 0.48 | 0.21 | 0.26 | 0.58 | 0.65 | 0.58 | 0.67 | 0.80 |
| $X_5$ | 0.75 | 0.81 | 0.48 | 1.00 | 0.95 | 0.70 | 0.74 | 0.65 | 0.91 | 0.58 | 0.66 |
| $X_6$ | 0.58 | 0.61 | 0.21 | 0.95 | 1.00 | 0.69 | 0.62 | 0.50 | 0.82 | 0.44 | 0.48 |
| $X_7$ | 0.45 | 0.51 | 0.26 | 0.70 | 0.69 | 1.00 | 0.44 | 0.45 | 0.71 | 0.36 | 0.37 |
| $X_8$ | 0.72 | 0.74 | 0.58 | 0.74 | 0.62 | 0.44 | 1.00 | 0.70 | 0.85 | 0.66 | 0.72 |
| $X_9$ | 0.72 | 0.75 | 0.65 | 0.65 | 0.50 | 0.45 | 0.70 | 1.00 | 0.82 | 0.59 | 0.72 |
| $X_{10}$ | 0.78 | 0.83 | 0.58 | 0.91 | 0.82 | 0.71 | 0.85 | 0.82 | 1.00 | 0.69 | 0.75 |
| $X_{11}$ | 0.79 | 0.74 | 0.67 | 0.58 | 0.44 | 0.36 | 0.66 | 0.59 | 0.69 | 1.00 | 0.82 |
| $X_{12}$ | 0.91 | 0.86 | 0.80 | 0.66 | 0.48 | 0.37 | 0.72 | 0.72 | 0.75 | 0.82 | 1.00 |

Tabela 12. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino.

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 7,7733 | 70,67% | 70,67% |
| Componente 2 | 1,4561 | 13,24% | 83,90% |
| Componente 3 | 0,4702 | 4,27% | 88,18% |
| Componente 4 | 0,4435 | 4,03% | 92,21% |
| Componente 5 | 0,3425 | 3,11% | 95,32% |
| Componente 6 | 0,2486 | 2,26% | 97,58% |
| Componente 7 | 0,1491 | 1,36% | 98,94% |
| Componente 8 | 0,0791 | 0,72% | 99,66% |
| Componente 9 | 0,0309 | 0,28% | 99,94% |
| Componente 10 | 0,0067 | 0,06% | 100,00% |
| Componente 11 | 0 | 0,00% | 100,00% |

componente principal, os autovalores decaem mais lentamente o que também contribui para a decisão do número de componentes.

Como o $\lambda_1$ é maior que um, obtém-se 1 componente principal, mas a proporção da variância acumulada é 73,08%, logo considera-se 2 componentes principais, como mostra na Tabela 16, e que tem proporção da variância acumulada 80,02%.

A primeira componente é dada por:

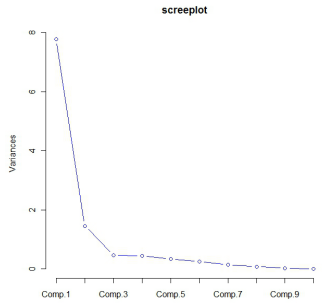


Figura 8 - Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino.

Tabela 13. Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo masculino.

| Variável | PC1 | PC2 |
|---|---|---|
| Massa corporal | 0,3287 | -0,177 |
| CB | **0,3381** | -0,1464 |
| CMB | 0,2729 | -0,4484 |
| AGB | 0,3196 | 0,3212 |
| DCT | 0,2668 | 0,5006 |
| DCB | 0,2261 | 0,466 |
| DCSE | 0,3047 | 0,0096 |
| DCSI | 0,2946 | -0,0769 |
| % G | **0,3398** | 0,1897 |
| CP | 0,2876 | -0,2385 |
| IMC | 0,3176 | -0,2762 |

$$Z_1 = 0,309X_1 + 0,3373X_3 + 0,2971X_4 + 0,3329X_5 + 0,3035X_6 + 0,2576X_7 + 0,2885X_8 + 0,2794X_9 + 0,3272X_{10} + 0,2571X_{11} + 0,3142X_{12}$$

A primeira componente é um índice global das variáveis, em que todas as variáveis são importantes, de acordo com os valores numéricos dos respectivos coeficientes na combinação linear; sendo que os maiores valores referem-se CB e AGB, e essas são as mais importantes. Para as mulheres as variáveis mais importantes coincidem com as variáveis na análise para todos os idosos com as 11 variáveis. Essa componente representa os pesos das variáveis $X_1, X_3, ..., X_{12}$, e esses pesos podem

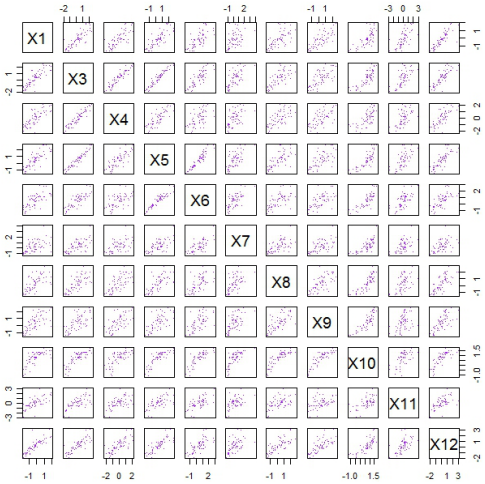


Figura 9 - Correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino.

ser utilizados na classificação do estado nutricional das idosas.

A segunda componente contrasta massa corporal total, CB, CMB, AGB, CP e IMC, com DCT, DCB, DCSE, DCSI e porcentagem de gordura. Assim, uma idosa que possui valores altos para as variáveis massa corporal total, CB, CMB, AGB, CP e IMC, e valores baixos para as variáveis DCT, DCB, DCSE, DCSI e porcentagem de gordura, tende a ter um escore baixo. E no caso inverso, tende a ter um escore alto.

Na próxima análise de componentes principais realizada nos dados de todos os idosos foram retiradas as variáveis estatura ($X_2$) e proteína C reativa ($X_{15}$). A variável estatura foi retirada porque possui baixa correlação com o IMC (Cervi et al., 2005). Já a variável proteína C reativa também possui baixos valores de correlações com as demais variáveis, e na avaliação nutricional dos idosos essa variável foi utilizada somente para excluir a possibilidade de uma inflamação interferir na interpretação dos valores da albumina (Galesi, 2010).

Tabela 14: Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino.

| | $X_1$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | $X_{11}$ | $X_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 1.00 | 0.83 | 0.77 | 0.81 | 0.69 | 0.47 | 0.66 | 0.68 | 0.74 | 0.69 | 0.85 |
| $X_3$ | 0.83 | 1.00 | 0.93 | 0.96 | 0.83 | 0.64 | 0.73 | 0.68 | 0.82 | 0.68 | 0.84 |
| $X_4$ | 0.77 | 0.93 | 1.00 | 0.79 | 0.56 | 0.50 | 0.67 | 0.61 | 0.68 | 0.59 | 0.77 |
| $X_5$ | 0.81 | 0.96 | 0.79 | 1.00 | 0.94 | 0.68 | 0.68 | 0.65 | 0.83 | 0.66 | 0.82 |
| $X_6$ | 0.69 | 0.83 | 0.56 | 0.94 | 1.00 | 0.69 | 0.61 | 0.58 | 0.81 | 0.64 | 0.70 |
| $X_7$ | 0.47 | 0.64 | 0.50 | 0.68 | 0.69 | 1.00 | 0.56 | 0.59 | 0.75 | 0.52 | 0.52 |
| $X_8$ | 0.66 | 0.73 | 0.67 | 0.68 | 0.61 | 0.56 | 1.00 | 0.67 | 0.84 | 0.52 | 0.73 |
| $X_9$ | 0.68 | 0.68 | 0.61 | 0.65 | 0.58 | 0.59 | 0.67 | 1.00 | 0.85 | 0.41 | 0.71 |
| $X_{10}$ | 0.74 | 0.82 | 0.68 | 0.83 | 0.81 | 0.75 | 0.84 | 0.85 | 1.00 | 0.60 | 0.79 |
| $X_{11}$ | 0.69 | 0.68 | 0.59 | 0.66 | 0.64 | 0.52 | 0.52 | 0.41 | 0.60 | 1.00 | 0.59 |
| $X_{12}$ | 0.85 | 0.84 | 0.77 | 0.82 | 0.70 | 0.52 | 0.73 | 0.71 | 0.79 | 0.59 | 1.00 |

Tabela 15. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino.

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 8,0391 | 73,08% | 73,08% |
| Componente 2 | 0,763 | 6,94% | 80,02% |
| Componente 3 | 0,7004 | 6,37% | 86,39% |
| Componente 4 | 0,4296 | 3,91% | 90,29% |
| Componente 5 | 0,3709 | 3,37% | 93,66% |
| Componente 6 | 0,331 | 3,01% | 96,67% |
| Componente 7 | 0,1946 | 1,77% | 98,44% |
| Componente 8 | 0,1199 | 1,09% | 99,53% |
| Componente 9 | 0,0477 | 0,43% | 99,97% |
| Componente 10 | 0,0038 | 0,03% | 100,00% |
| Componente 11 | 0 | 0,00% | 100,00% |

### 4.1.5 Análise de componentes principais realizada para 13 variáveis, para todos os idosos

Com a retirada das variáveis $X_2$ e $X_{15}$ pode-se observar nas Figuras 11 e 12, e na Tabela 17 que as variáveis $X_1, X_3, ..., X_{12}$ possuem altos valores de correlações entre si, e as variáveis albumina $X_{13}$ e transferrina $X_{14}$ não possuem valores tão altos de correlações.

Realizou-se a análise de componentes principais nos dados de todos os idosos, para as 13 variáveis $(X_1, X_3, ..., X_{14})$.

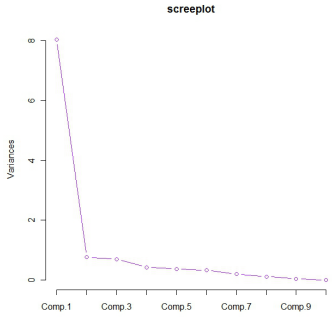


Figura 10 - Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino.

Tabela 16. Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{12}$, considerando os idosos do sexo feminino.

| Variável | PC1 | PC2 |
|---|---|---|
| Massa corporal | 0,309 | -0,3316 |
| CB | **0,3373** | -0,1897 |
| CMB | 0,2971 | -0,362 |
| AGB | **0,3329** | -0,0728 |
| DCT | 0,3035 | 0,127 |
| DCB | 0,2576 | 0,5602 |
| DCSE | 0,2885 | 0,1684 |
| DCSI | 0,2794 | 0,3472 |
| % G | 0,3272 | 0,3252 |
| CP | 0,2571 | -0,3221 |
| IMC | 0,3142 | -0,1714 |

Observa-se na Tabela 18, que os 3 primeiros autovalores são $\lambda_1 = 7,4445$, $\lambda_2 = 1,5209$ e $\lambda_3 = 1,0125$, e eles explicam 57,27%, 11,70% e 7,79% da variância total, respectivamente; e os 3 autovalores juntos explicam 76,75% da variância total.

Na Figura 13, pode-se visualizar no gráfico scree-plot que a partir da terceira componente principal, os autovalores decaem mais lentamente, o que também contribui para a decisão do número de componentes. As 3 componentes principais são dadas na Tabela 19.

A primeira componente é dada por:

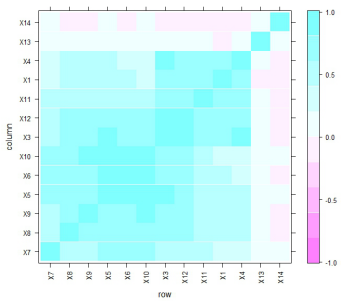


Figura 11 - Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos.

Tabela 17: Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para todos os idosos.

| | X1 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 1.00 | 0.76 | 0.77 | 0.57 | 0.41 | 0.38 | 0.57 | 0.53 | 0.38 | 0.72 | 0.80 | -0.04 | -0.16 |
| X3 | 0.76 | 1.00 | 0.87 | 0.89 | 0.74 | 0.60 | 0.75 | 0.71 | 0.70 | 0.67 | 0.85 | 0.03 | -0.00 |
| X4 | 0.77 | 0.87 | 1.00 | 0.56 | 0.31 | 0.34 | 0.57 | 0.51 | 0.35 | 0.61 | 0.74 | 0.00 | -0.00 |
| X5 | 0.57 | 0.89 | 0.56 | 1.00 | 0.95 | 0.70 | 0.73 | 0.72 | 0.85 | 0.54 | 0.76 | 0.06 | 0.02 |
| X6 | 0.41 | 0.74 | 0.31 | 0.95 | 1.00 | 0.70 | 0.66 | 0.66 | 0.86 | 0.46 | 0.62 | 0.06 | -0.00 |
| X7 | 0.38 | 0.60 | 0.34 | 0.70 | 0.70 | 1.00 | 0.55 | 0.58 | 0.70 | 0.42 | 0.48 | 0.00 | 0.01 |
| X8 | 0.57 | 0.75 | 0.57 | 0.73 | 0.66 | 0.55 | 1.00 | 0.72 | 0.77 | 0.54 | 0.74 | 0.06 | -0.01 |
| X9 | 0.53 | 0.71 | 0.51 | 0.72 | 0.66 | 0.58 | 0.72 | 1.00 | 0.83 | 0.43 | 0.72 | 0.01 | -0.01 |
| X10 | 0.38 | 0.70 | 0.35 | 0.85 | 0.86 | 0.70 | 0.77 | 0.83 | 1.00 | 0.44 | 0.66 | 0.04 | 0.05 |
| X11 | 0.72 | 0.67 | 0.61 | 0.54 | 0.46 | 0.42 | 0.54 | 0.43 | 0.44 | 1.00 | 0.66 | 0.03 | -0.15 |
| X12 | 0.80 | 0.85 | 0.74 | 0.76 | 0.62 | 0.48 | 0.74 | 0.72 | 0.66 | 0.66 | 1.00 | 0.02 | -0.02 |
| X13 | -0.04 | 0.03 | 0.00 | 0.06 | 0.06 | 0.00 | 0.06 | 0.01 | 0.04 | 0.03 | 0.02 | 1.00 | 0.03 |
| X14 | -0.16 | -0.00 | -0.00 | 0.02 | -0.00 | 0.01 | -0.01 | -0.01 | 0.05 | -0.15 | -0.02 | 0.03 | 1.00 |

$$Z_1 = 0,2771X_1 + 0,3472X_3 + 0,268X_4 + 0,3381X_5 + 0,3027X_6 + 0,2607X_7 + 0,3097X_8 + 0,302X_9 + 0,3081X_{10} + 0,261X_{11} + 0,3264X_{12} + 0,0137X_{13} - 0,0112X_{14}$$

Na primeira componente principal a variável transferrina contrasta com as demais, essa variável possui baixa correlação com as demais. E a albumina, que também possui baixa correlação, tem um peso baixo na componente. Já a variável mais importante, a que possui o maior peso é a CB, seguida da AGB que possui o

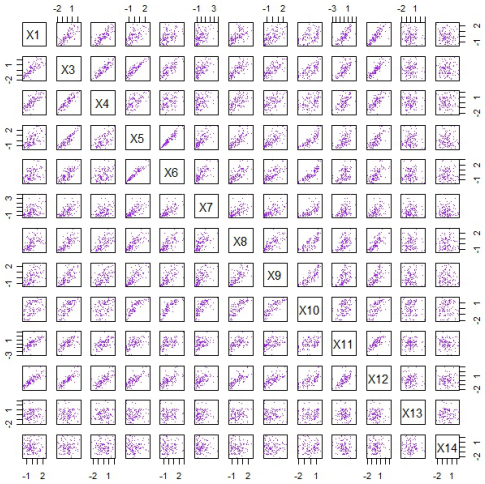


Figura 12 - Correlações entre as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos.

segundo maior valor.

As componentes principais 2 e 3 explicam uma porcentagem de variância muito baixa, 11,70% e 7,79%, respectivamente, logo os resultados não são muito bons.

Nas duas análises de componentes principais que seguem o grupo de idosos foi dividido, por faixa etária, em dois grupos: um de 60-79 anos e o outro de 80-100 anos. Nestas análises também foram retiradas as variáveis estatura ($X_2$) e proteína C reativa ($X_{15}$).

### 4.1.6 Análise de componentes principais realizada com 13 variáveis, para os idosos de 60-79 anos

Para os idosos de 60-79 anos, pode-se observar nas Figuras 14 e 15, e na Tabela 20, que com a retirada das variáveis $X_2$ e $X_{15}$ as variáveis $X_1, X_3, ..., X_{12}$ possuem altos valores de correlações, e as variáveis albumina $X_{13}$ e transferrina $X_{14}$ possuem baixos valores de correlações.

Realizou-se uma análise de componentes principais nos dados dos ido-

Tabela 18. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 13 componentes, considerando todos os idosos

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 7,4445 | 57,27% | 57,27% |
| Componente 2 | 1,5209 | 11,70% | 68,96% |
| Componente 3 | 1,0125 | 7,79% | 76,75% |
| Componente 4 | 0,9824 | 7,56% | 84,31% |
| Componente 5 | 0,5439 | 4,18% | 88,49% |
| Componente 6 | 0,3951 | 3,04% | 91,53% |
| Componente 7 | 0,3899 | 3,00% | 94,53% |
| Componente 8 | 0,2775 | 2,13% | 96,67% |
| Componente 9 | 0,2243 | 1,73% | 98,39% |
| Componente 10 | 0,1318 | 1,01% | 99,41% |
| Componente 11 | 0,0693 | 0,53% | 99,94% |
| Componente 12 | 0,0079 | 0,06% | 100,00% |
| Componente 13 | 0 | 0,00% | 100,00% |

sos de 60-79 anos, com as 13 variáveis $(X_1, X_3, ..., X_{14})$.

Pode-se observar na Tabela 21, que os 2 primeiros autovalores são $\lambda_1 = 7,4724$ e $\lambda_2 = 1,8088$, e eles explicam 57,48% e 13,91% da variância total, respectivamente; e os 2 autovalores juntos explicam 71,39% da variância total.

Na Figura 16, pode-se visualizar no gráfico scree-plot que a partir da segunda componente principal, os autovalores decaem mais lentamente, o que também contribui para a decisão do número de componentes. As 2 componentes principais são dadas na Tabela 22.

A primeira componente é dada por:

$$Z_1 = 0,263X_1 + 0,3473X_3 + 0,251X_4 + 0,3392X_5 + 0,3069X_6 + 0,263X_7 + 0,3178X_8 + 0,3033X_9 + 0,3132X_{10} + 0,2684X_{11} + 0,323X_{12} + 0,0004X_{13} - 0,0338X_{14}$$

A primeira componente contrasta a variável transferrina com as demais. Sendo a variável mais importante a CB, a que possui o maior peso, seguida da AGB que possui o segundo maior valor. As variáveis albumina e transferrina,

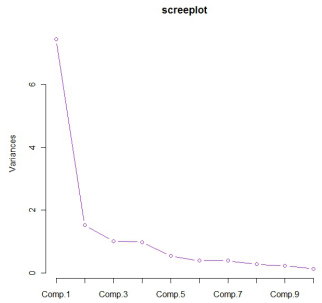


Figura 13 - Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos.

Tabela 19. Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos.

| Variável | PC1 | PC2 | PC3 |
|---|---|---|---|
| Massa corporal | 0,2771 | -0,4358 | 0,0212 |
| CB | **0,3472** | -0,1171 | 0,0877 |
| CMB | 0,268 | -0,4188 | 0,211 |
| AGB | **0,3381** | 0,1918 | -0,0198 |
| DCT | 0,3027 | 0,3418 | -0,1172 |
| DCB | 0,2607 | 0,2844 | -0,156 |
| DCSE | 0,3097 | 0,062 | 0,0227 |
| DCSI | 0,302 | 0,1461 | -0,063 |
| % G | 0,3081 | 0,3601 | -0,0959 |
| CP | 0,261 | -0,3242 | 0,0143 |
| IMC | 0,3264 | -0,1743 | 0,0744 |
| Albumina | 0,0137 | 0,1205 | 0,7401 |
| Transferrina | -0,0112 | 0,2856 | 0,5845 |

que possuem baixos valores de correlações, são as variáveis que possuem os menores pesos. Para os idosos de 60-79 anos os resultados da primeira componente principal coincidem com os da análise com 13 variáveis realizada para todos os idosos.

A segunda componente contrasta massa corporal total, CB, CMB, CP e IMC, com AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura, albumina e transferrina. Assim, um idoso que possui valores altos para as variáveis massa corporal total, CB, CMB, CP e IMC, e valores baixos para as variáveis AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura albumina e transferrina, tende a ter

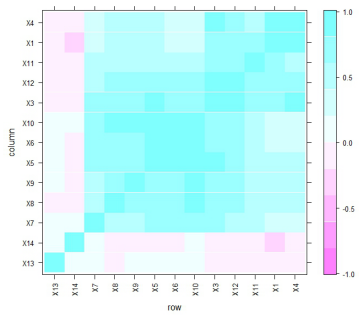


Figura 14 - Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 60-79 anos.

Tabela 20: Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 60-79 anos.

| | X1 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 1.00 | 0.77 | 0.82 | 0.52 | 0.35 | 0.32 | 0.49 | 0.50 | 0.35 | 0.74 | 0.74 | -0.14 | -0.26 |
| X3 | 0.77 | 1.00 | 0.84 | 0.87 | 0.74 | 0.61 | 0.76 | 0.71 | 0.70 | 0.71 | 0.87 | -0.00 | -0.11 |
| X4 | 0.82 | 0.84 | 1.00 | 0.47 | 0.26 | 0.27 | 0.50 | 0.49 | 0.30 | 0.60 | 0.74 | -0.10 | -0.14 |
| X5 | 0.52 | 0.87 | 0.47 | 1.00 | 0.97 | 0.75 | 0.80 | 0.71 | 0.86 | 0.60 | 0.75 | 0.08 | -0.04 |
| X6 | 0.35 | 0.74 | 0.26 | 0.97 | 1.00 | 0.74 | 0.74 | 0.65 | 0.87 | 0.53 | 0.63 | 0.12 | -0.02 |
| X7 | 0.32 | 0.61 | 0.27 | 0.75 | 0.74 | 1.00 | 0.58 | 0.60 | 0.73 | 0.41 | 0.47 | 0.00 | 0.03 |
| X8 | 0.49 | 0.76 | 0.50 | 0.80 | 0.74 | 0.58 | 1.00 | 0.78 | 0.85 | 0.52 | 0.75 | -0.01 | -0.00 |
| X9 | 0.50 | 0.71 | 0.49 | 0.71 | 0.65 | 0.60 | 0.78 | 1.00 | 0.84 | 0.46 | 0.71 | 0.00 | -0.06 |
| X10 | 0.35 | 0.70 | 0.30 | 0.86 | 0.87 | 0.73 | 0.85 | 0.84 | 1.00 | 0.47 | 0.66 | 0.08 | 0.03 |
| X11 | 0.74 | 0.71 | 0.60 | 0.60 | 0.53 | 0.41 | 0.52 | 0.46 | 0.47 | 1.00 | 0.65 | -0.05 | -0.16 |
| X12 | 0.74 | 0.87 | 0.74 | 0.75 | 0.63 | 0.47 | 0.75 | 0.71 | 0.66 | 0.65 | 1.00 | -0.02 | -0.06 |
| X13 | -0.14 | -0.00 | -0.10 | 0.08 | 0.12 | 0.00 | -0.01 | 0.00 | 0.08 | -0.05 | -0.02 | 1.00 | 0.05 |
| X14 | -0.26 | -0.11 | -0.14 | -0.04 | -0.02 | 0.03 | -0.00 | -0.06 | 0.03 | -0.16 | -0.06 | 0.05 | 1.00 |

um escore baixo. E no caso inverso, tende a ter um escore alto.

### 4.1.7 Análise de componentes principais realizada com 13 variáveis, para os idosos de 80-100 anos

Para os idosos de 80-100 anos, pode-se observar nas Figuras 17 e 18, e na Tabela 23, que com a retirada das variáveis $X_2$ e $X_{15}$ as variáveis $X_1, X_3, ..., X_{12}$ possuem altos valores de correlações, e as variáveis albumina $X_{13}$ e transferrina $X_{14}$ não possuem valores tão altos de correlações.

Realizou-se uma análise de componentes principais nos dados idosos

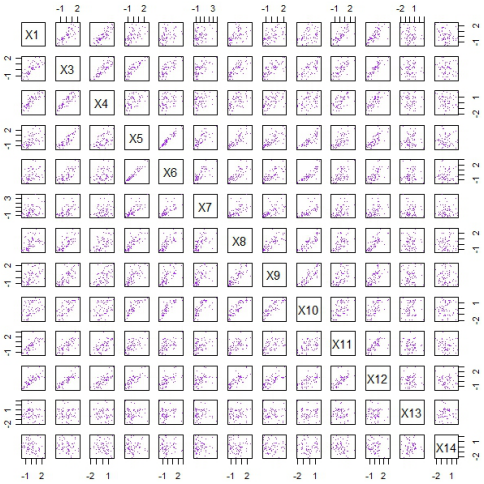


Figura 15 - Correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 60-79 anos.

de 80-100 anos, com as 13 variáveis $(X_1, X_3, ..., X_{14})$.

Observa-se na Tabela 24 que os 3 primeiros autovalores são $\lambda_1 = 7,6414$, $\lambda_2 = 1,1084$ e $\lambda_3 = 1,0965$, e eles explicam 58,78%, 8,53% e 8,43% da variância total, respectivamente; e os 3 autovalores juntos explicam 75,74% da variância total.

Na Figura 19, pode-se visualizar no gráfico scree-plot que a partir da terceira componente principal, os autovalores decaem mais lentamente, o que também contribui para a decisão do número de componentes. As 3 componentes principais são dadas na Tabela 25.

A primeira componente é dada por:

$$Z_1 = 0,2976X_1 + 0,3436X_3 + 0,2872X_4 + 0,3388X_5 + 0,3053X_6 + 0,2561X_7 + 0,2927X_8 + 0,2939X_9 + 0,3057X_{10} + 0,2535X_{11} + 0,3265X_{12} + 0,0195X_{13} + 0,0228X_{14}$$

A primeira componente é um índice global das variáveis, em que todas

Tabela 21. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 13 componentes, considerando os idosos de 60-79 anos.

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 7,4724 | 57,48% | 57,48% |
| Componente 2 | 1,8088 | 13,91% | 71,39% |
| Componente 3 | 0,969 | 7,45% | 78,85% |
| Componente 4 | 0,9165 | 7,05% | 85,90% |
| Componente 5 | 0,5825 | 4,48% | 90,38% |
| Componente 6 | 0,387 | 2,98% | 93,36% |
| Componente 7 | 0,3287 | 2,53% | 95,88% |
| Componente 8 | 0,1859 | 1,43% | 97,31% |
| Componente 9 | 0,1598 | 1,23% | 98,54% |
| Componente 10 | 0,1255 | 0,97% | 99,51% |
| Componente 11 | 0,0571 | 0,44% | 99,95% |
| Componente 12 | 0,0068 | 0,05% | 100,00% |
| Componente 13 | 0 | 0,00% | 100,00% |

as variáveis são importantes, de acordo com os valores numéricos dos respectivos coeficientes na combinação linear; sendo que os maiores valores referem-se a CB e a AGB, e essas são as mais importantes. As variáveis albumina e transferrina, que possuem baixos valores de correlações, são as variáveis que possuem os menores pesos. Para os idosos de 80-100 anos os resultados da primeira componente principal coincidem com os da análises com 13 variáveis realizadas para todos os idosos e para os idosos 60-79 anos.

Como as componentes principais 2 e 3 explicam uma porcentagem de variância muito baixa (8,53% e 8,43%, respectivamente), os resultados apresentados por elas não são muito bons.

### 4.1.8 Análise de componentes principais realizada com variáveis antropométricas primárias e as bioquímicas, para todos os idosos

Nesta análise foram utilizadas as variáveis antrométricas primárias (variáveis que foram obtidas de forma direta sem fórmula): massa corporal total, estatura, CB, DCT, DCB, DCSE, DCSI e CP, e as bioquímicas: albumina e transferrina.

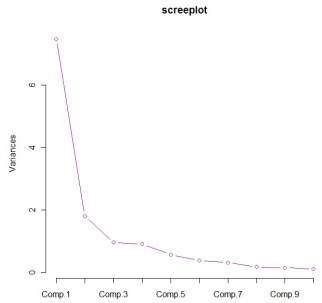


Figura 16 - Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos de 60-79 anos.

Tabela 22. Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos de 60-79 anos.

| Variável | PC1 | PC2 |
|---|---|---|
| Massa corporal | 0,263 | -0,4469 |
| CB | **0,3473** | -0,1321 |
| CMB | 0,251 | -0,4413 |
| AGB | **0,3392** | 0,1801 |
| DCT | 0,3069 | 0,3089 |
| DCB | 0,263 | 0,277 |
| DCSE | 0,3178 | 0,1227 |
| DCSI | 0,3033 | 0,1097 |
| % G | 0,3132 | 0,3154 |
| CP | 0,2684 | -0,2501 |
| IMC | 0,323 | -0,1477 |
| Albumina | 0,0004 | 0,2693 |
| Transferrina | -0,0338 | 0,3163 |

Pode-se observar nas Figuras 20 e 21, e na Tabela 26, que as variáveis estatura $X_2$, albumina $X_{13}$ e transferrina $X_{14}$ possuem baixos valores de correlações, e demais variáveis possuem altos valores de correlações.

Realizou-se a análise de componentes principais nos dados de todos os idosos, para as variáveis antropométricas primárias e as bioquímicas.

Observa-se na Tabela 27 que os 2 primeiros autovalores são $\lambda_1 = 4,6219$ e $\lambda_2 = 1,6552$, e eles explicam 46,22% e 16,55% da variância total, respectivamente; e os 2 autovalores juntos explicam 62,77% da variância total.

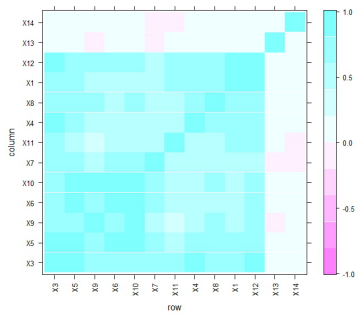


Figura 17 - Correlações agrupadas entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 80-100 anos.

Tabela 23: Matriz de correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 80-100 anos.

| | X1 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 1.00 | 0.76 | 0.70 | 0.69 | 0.58 | 0.50 | 0.67 | 0.57 | 0.49 | 0.69 | 0.88 | 0.04 | 0.02 |
| X3 | 0.76 | 1.00 | 0.91 | 0.93 | 0.78 | 0.62 | 0.74 | 0.71 | 0.72 | 0.64 | 0.83 | 0.06 | 0.13 |
| X4 | 0.70 | 0.91 | 1.00 | 0.70 | 0.45 | 0.47 | 0.65 | 0.55 | 0.48 | 0.60 | 0.74 | 0.07 | 0.18 |
| X5 | 0.69 | 0.93 | 0.70 | 1.00 | 0.94 | 0.64 | 0.69 | 0.75 | 0.82 | 0.54 | 0.79 | 0.06 | 0.09 |
| X6 | 0.58 | 0.78 | 0.45 | 0.94 | 1.00 | 0.63 | 0.61 | 0.70 | 0.84 | 0.47 | 0.67 | 0.02 | 0.00 |
| X7 | 0.50 | 0.62 | 0.47 | 0.64 | 0.63 | 1.00 | 0.50 | 0.57 | 0.65 | 0.47 | 0.52 | -0.00 | -0.01 |
| X8 | 0.67 | 0.74 | 0.65 | 0.69 | 0.61 | 0.50 | 1.00 | 0.61 | 0.68 | 0.55 | 0.71 | 0.11 | 0.00 |
| X9 | 0.57 | 0.71 | 0.55 | 0.75 | 0.70 | 0.57 | 0.61 | 1.00 | 0.84 | 0.38 | 0.72 | -0.01 | 0.06 |
| X10 | 0.49 | 0.72 | 0.48 | 0.82 | 0.84 | 0.65 | 0.68 | 0.84 | 1.00 | 0.49 | 0.69 | 0.03 | 0.06 |
| X11 | 0.69 | 0.64 | 0.60 | 0.54 | 0.47 | 0.47 | 0.55 | 0.38 | 0.49 | 1.00 | 0.67 | 0.03 | -0.12 |
| X12 | 0.88 | 0.83 | 0.74 | 0.79 | 0.67 | 0.52 | 0.71 | 0.72 | 0.69 | 0.67 | 1.00 | 0.01 | 0.04 |
| X13 | 0.04 | 0.06 | 0.07 | 0.06 | 0.02 | -0.00 | 0.11 | -0.01 | 0.03 | 0.03 | 0.01 | 1.00 | 0.02 |
| X14 | 0.02 | 0.13 | 0.18 | 0.09 | 0.00 | -0.01 | 0.00 | 0.06 | 0.06 | -0.12 | 0.04 | 0.02 | 1.00 |

Na Figura 22, pode-se visualizar no gráfico scree-plot que a partir da segunda componente principal, os autovalores decaem mais lentamente, o que também contribui para a decisão do número de componentes. As 2 componentes principais são dadas na Tabela 28.

A primeira componente é dada por:

$$Z_1 = 0,3526X_1 - 0,0394X_2 + 0,4305X_3 + 0,3832X_6 + 0,3453X_7 + 0,396X_8 + 0,3828X_9 + 0,3438X_{11} + 0,0178X_{13} - 0,0289X_{14}$$

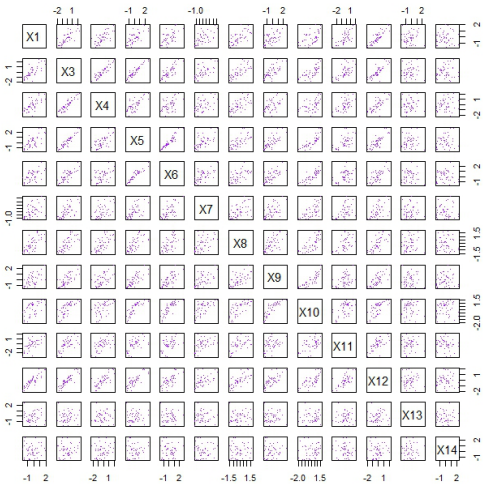


Figura 18 - Correlações entre as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 80-100 anos.

Na primeira componente principal as variáveis estatura e transferrina contrastam com as demais. E a albumina, que também possui baixa correlação, tem um peso baixo na componente. A variável mais importante, a que possui o maior peso é CB, seguida da DCSE que possui o segundo maior valor.

Como a segunda componente principal explica uma porcentagem de variância muito baixa (16,55%), ela não apresenta resultado muito bom.

## 4.2 Análise fatorial exploratória

### 4.2.1 Análise fatorial realizada com 15 variáveis, para todos os idosos

O primeiro passo para realizar uma análise fatorial exploratória nas variáveis $X_1, X_2, ..., X_{15}$, para todos os idosos, é a escolha do número $m$ de fatores. Esta escolha foi baseada nos critérios: o autovalor ser maior que 1, o número de autovalores que representam uma alta proporção de variância total e o número das componente principais do gráfico "scree-plot", quando os autovalores decaem mais lentamente.

Tabela 24. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 13 componentes, considerando os idosos de 80-100 anos.

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 7,6414 | 58,78% | 58,78% |
| Componente 2 | 1,1084 | 8,53% | 67,31% |
| Componente 3 | 1,0965 | 8,43% | 75,74% |
| Componente 4 | 0,9831 | 7,56% | 83,30% |
| Componente 5 | 0,5363 | 4,13% | 87,43% |
| Componente 6 | 0,411 | 3,16% | 90,59% |
| Componente 7 | 0,4045 | 3,11% | 93,70% |
| Componente 8 | 0,3536 | 2,72% | 96,42% |
| Componente 9 | 0,3073 | 2,36% | 98,79% |
| Componente 10 | 0,1056 | 0,81% | 99,60% |
| Componente 11 | 0,0461 | 0,36% | 99,95% |
| Componente 12 | 0,0061 | 0,05% | 100,00% |
| Componente 13 | 0 | 0,00% | 100,00% |

Como visto na Tabela 6 tem-se que os 3 primeiros autovalores são maiores que 1 e os 3 autovalores juntos explicam 71,80% da variância total; e no gráfico scree-plot (Figura 1) observa-se que a partir da terceira componente principal os autovalores decaem mais lentamente. Então, tem-se $m$ é igual a 3 fatores.

Na Tabela 29, obtém-se os 3 fatores rotacionados para as variáveis $X_1, X_2, ..., X_{15}$. Estes fatores agrupam as variáveis da seguinte forma:

Fator 1: CB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura e IMC.

Fator 2: massa corporal total, estatura, CMB, CP e proteína C reativa.

Fator 3: albumina e transferrina.

O Fator 1 tem cargas positivas altas para as variáveis CB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura e IMC; ele será chamado de "indicadores de gordura corporal".

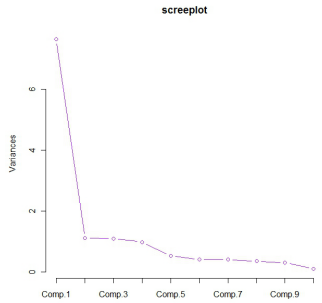


Figura 19 - Scree-plot dos autovalores *versus* componentes principais para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos de 80-100 anos.

Tabela 25. Componentes principais (PC) para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos 80-100 anos.

| Variável | PC1 | PC2 | PC3 |
|---|---|---|---|
| Massa corporal | 0,2976 | -0,1345 | -0,3097 |
| CB | **0,3436** | -0,1377 | -0,0103 |
| CMB | 0,2872 | -0,3537 | -0,1179 |
| AGB | **0,3388** | 0,0327 | 0,1581 |
| DCT | 0,3053 | 0,2356 | 0,2449 |
| DCB | 0,2561 | 0,2264 | 0,1274 |
| DCSE | 0,2927 | -0,0448 | -0,1104 |
| DCSI | 0,2939 | 0,1376 | 0,2697 |
| % G | 0,3057 | 0,2147 | 0,3055 |
| CP | 0,2535 | 0,0033 | -0,4614 |
| IMC | 0,3265 | -0,0666 | -0,1444 |
| Albumina | 0,0195 | -0,4382 | -0,177 |
| Transferrina | 0,0228 | -0,683 | 0,5726 |

O Fator 2 tem cargas positivas altas para as variáveis massa corporal total, estatura, CMB e CP, e carga negativa alta para proteína C reativa; ele será chamado de "indicadores de massa muscular e bioquímico".

O Fator 3 tem cargas positivas altas para as variáveis albumina e transferrina; ele será chamado de "indicadores bioquímicos do estado nutricional".

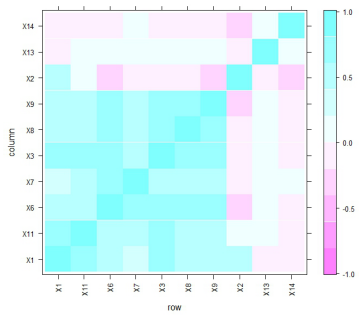


Figura 20 - Correlações agrupadas entre as variáveis antropométricas primárias e as bioquímicas, considerando os todos os idosos.

Tabela 26: Matriz de correlações entre as variáveis antropométricas primárias e as bioquímicas, para todos os idosos.

| | X1 | X2 | X3 | X6 | X7 | X8 | X9 | X11 | X13 | X14 |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 1.00 | 0.41 | 0.76 | 0.41 | 0.38 | 0.57 | 0.53 | 0.72 | -0.04 | -0.16 |
| X2 | 0.41 | 1.00 | -0.05 | -0.27 | -0.14 | -0.19 | -0.24 | 0.14 | -0.08 | -0.20 |
| X3 | 0.76 | -0.05 | 1.00 | 0.74 | 0.60 | 0.75 | 0.71 | 0.67 | 0.03 | -0.00 |
| X6 | 0.41 | -0.27 | 0.74 | 1.00 | 0.70 | 0.66 | 0.66 | 0.46 | 0.06 | -0.00 |
| X7 | 0.38 | -0.14 | 0.60 | 0.70 | 1.00 | 0.55 | 0.58 | 0.42 | 0.00 | 0.01 |
| X8 | 0.57 | -0.19 | 0.75 | 0.66 | 0.55 | 1.00 | 0.72 | 0.54 | 0.06 | -0.01 |
| X9 | 0.53 | -0.24 | 0.71 | 0.66 | 0.58 | 0.72 | 1.00 | 0.43 | 0.01 | -0.01 |
| X11 | 0.72 | 0.14 | 0.67 | 0.46 | 0.42 | 0.54 | 0.43 | 1.00 | 0.03 | -0.15 |
| X13 | -0.04 | -0.08 | 0.03 | 0.06 | 0.00 | 0.06 | 0.01 | 0.03 | 1.00 | 0.03 |
| X14 | -0.16 | -0.20 | -0.00 | -0.00 | 0.01 | -0.01 | -0.01 | -0.15 | 0.03 | 1.00 |

### 4.2.2 Análise fatorial realizada com 11 variáveis, para todos os idosos

Observa-se na Tabela 9 que os 2 primeiros autovalores são maiores que 1 e os 2 autovalores juntos explicam 81,02% da variância total; e no gráfico scree-plot (Figura 6) observa-se a partir da segunda componente principal os autovalores decaem mais lentamente . Então, tem-se que $m$ igual 2 fatores.

Na Tabela 30, obtém-se os 2 fatores rotacionados para as variáveis $X_1, X_3, ..., X_{12}$, de todos os idosos. Esses fatores agrupam as variáveis da seguinte forma:

Fator 1: AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura.

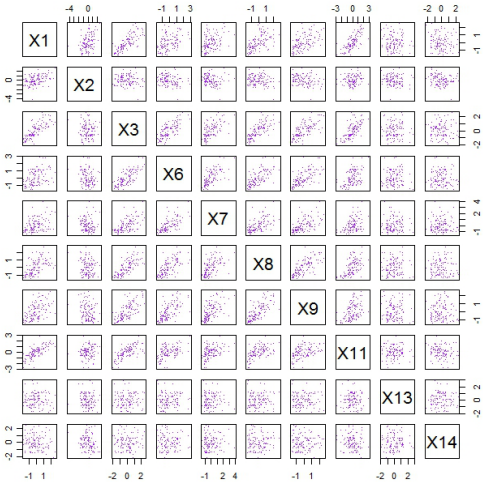


Figura 21 - Correlações entre as variáveis antropométricas primárias e as bioquímicas, considerando os todos os idosos.

Fator 2: Massa corporal total, CB, CMB, CP e IMC.

O Fator 1 tem cargas positivas altas para as variáveis AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura; ele será chamado de "indicadores de gordura corporal".

O Fator 2 tem cargas positivas altas para as variáveis massa corporal total, CB, CMB, CP e IMC; ele será chamado de "indicadores de massa muscular".

### 4.2.3 Análise fatorial realizada para os homens, com 11 variáveis

Como visto na Tabela 12 tem-se que os 2 primeiros autovalores são maiores que 1 e os 2 autovalores juntos explicam 83,90% da variância total; e no gráfico scree-plot (Figura 8) observa-se que os autovalores decaem mais lentamente na segunda componente principal. Então, tem-se $m$ igual 2 fatores.

Na Tabela 31, obtém-se os 2 fatores rotacionados para as variáveis $X_1, X_3, ..., X_{12}$, dos homens. Esses fatores agrupam as variáveis da seguinte forma:

Tabela 27. Autovalores, porcentagem de variância explicada e porcentagem de variância explicada acumulada para as 10 componentes, considerando os todos os idosos.

| | Autovalores | % de variância | % de variância acumulada |
|---|---|---|---|
| Componente 1 | 4,6219 | 46,22% | 46,22% |
| Componente 2 | 1,6552 | 16,55% | 62,77% |
| Componente 3 | 0,9928 | 9,93% | 72,70% |
| Componente 4 | 0,891 | 8,91% | 81,61% |
| Componente 5 | 0,5594 | 5,59% | 87,20% |
| Componente 6 | 0,4498 | 4,50% | 91,70% |
| Componente 7 | 0,2961 | 2,96% | 94,66% |
| Componente 8 | 0,2667 | 2,67% | 97,33% |
| Componente 9 | 0,1768 | 1,77% | 99,10% |
| Componente 10 | 0,0901 | 0,90% | 100,00% |

Tabela 28. Componentes principais (PC) para as variáveis antropométricas primárias e as bioquímicas, para todos os idosos.

| Variável | PC1 | PC2 |
|---|---|---|
| Massa corporal | 0,3526 | -0,4241 |
| Estatura | -0,0394 | -0,6775 |
| CB | **0,4305** | -0,0361 |
| DCT | 0,3832 | 0,2239 |
| DCB | 0,3453 | 0,1602 |
| DCSE | **0,396** | 0,1017 |
| DCSI | 0,3828 | 0,1617 |
| CP | 0,3438 | -0,281 |
| Albumina | 0,0178 | 0,1511 |
| Transferrina | -0,0289 | 0,3818 |

Fator 1: Massa corporal total, CB, CMB, DCSE, DCSI, CP e IMC.

Fator 2: AGB, DCT, DCB e porcentagem de gordura.

O Fator 1 tem cargas positivas altas para as variáveis massa corporal total, CB, CMB, DCSE, DCSI, CP e IMC; ele será chamado de "indicadores de gordura corporal e massa muscular".

O Fator 2 tem cargas positivas altas para as variáveis AGB, DCT, DCB

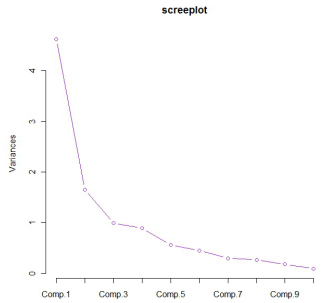


Figura 22 - Scree-plot dos autovalores *versus* componentes principais para as variáveis antropométricas primárias e as bioquímicas, considerando todos os idosos.

e porcentagem de gordura; ele será chamado de "indicadores de gordura corporal".

### 4.2.4 Análise fatorial realizada para as mulheres, com 11 variáveis

Como visto na Tabela 15 tem-se que o primeiro autovalor é maior que 1, mas os 2 primeiros autovalores juntos explicam 80,02% da variância total; e no gráfico scree-plot (Figura 10) visualiza que os autovalores decaem mais lentamente na segunda componente principal. Então, tem-se que $m$ igual 2 fatores.

Os 2 fatores rotacionados, são dados na Tabela 32 para as variáveis $X_1, X_3, ..., X_{12}$, das mulheres. Esses fatores agrupam as variáveis da seguinte forma:

Fator 1: Massa corporal total, CB, CMB, AGB, CP e IMC.

Fator 2: DCT, DCB, DCSE, DCSI e porcentagem de gordura.

O Fator 1 tem cargas positivas altas para as variáveis massa corporal total, CB, CMB, AGB, CP, IMC; ele será chamado de "indicadores de massa muscular e gordura corporal".

O Fator 2 tem cargas positivas altas para as variáveis DCT, DCB,

Tabela 29. Fatores para as variáveis $X_1, X_2, ..., X_{15}$, considerando todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 | Fator3 |
|---|---|---|---|
| Massa corporal | 0.39 | **0.85** | -0.18 |
| Estatura | -0.45 | **0.55** | -0.42 |
| CB | **0.76** | 0.58 | 0.02 |
| CMB | 0.42 | **0.75** | 0.01 |
| AGB | **0.90** | 0.28 | 0.05 |
| DCT | **0.89** | 0.09 | 0.03 |
| DCB | **0.76** | 0.09 | -0.09 |
| DCSE | **0.78** | 0.33 | 0.05 |
| DCSI | **0.82** | 0.21 | -0.01 |
| % G | **0.95** | 0.02 | 0.06 |
| CP | 0.42 | **0.71** | -0.03 |
| IMC | **0.71** | 0.55 | 0.09 |
| Albumina | 0.03 | 0.02 | **0.47** |
| Transferrina | 0.02 | -0.12 | **0.73** |
| Proteína. C Reativa | 0.30 | **-0.54** | -0.46 |

DCSE, DCSI e porcentagem de gordura; ele será chamado de "indicadores de gordura corporal".

### 4.2.5 Análise fatorial realizada para 13 variáveis, para todos os idosos

Como visto na Tabela 18 tem-se que os 3 primeiros autovalores são maiores que 1 e os 3 autovalores juntos explicam 76,75% da variância total; e no gráfico scree-plot (Figura 13) observa-se que os autovalores decaem mais lentamente na terceira componente principal. Então, tem-se $m$ igual 3 fatores.

Na Tabela 33, obtém-se os 3 fatores rotacionados para as variáveis $X_1, X_3, ..., X_{14}$, de todos os idosos. Esses fatores agrupam as variáveis da seguinte forma:

Fator 1: AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura.

Fator 2: massa corporal total, CB, CMB, CP e IMC.

Fator 3: albumina e transferrina.

Tabela 30. Fatores para as variáveis $X_1, X_3, ..., X_{12}$, para todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 |
|---|---|---|
| Massa corporal | 0.22 | **0.90** |
| CB | 0.60 | **0.75** |
| CMB | 0.18 | **0.91** |
| AGB | **0.85** | 0.43 |
| DCT | **0.91** | 0.21 |
| DCB | **0.78** | 0.19 |
| DCSE | **0.68** | 0.50 |
| DCSI | **0.74** | 0.40 |
| % G | **0.93** | 0.21 |
| CP | 0.28 | **0.76** |
| IMC | 0.52 | **0.76** |

Tabela 31. Fatores para as variáveis $X_1, X_2, ..., X_{15}$, para os homens, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 |
|---|---|---|
| Massa corporal | **0.85** | 0.40 |
| CB | **0.85** | 0.45 |
| CMB | **0.93** | 0.05 |
| AGB | 0.46 | **0.86** |
| DCT | 0.21 | **0.94** |
| DCB | 0.15 | **0.83** |
| DCSE | **0.66** | 0.54 |
| DCSI | **0.70** | 0.44 |
| % G | 0.60 | **0.77** |
| CP | **0.81** | 0.27 |
| IMC | **0.90** | 0.29 |

O Fator 1 tem cargas positivas altas para as variáveis AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura; ele será chamado de "indicadores de gordura corporal".

O Fator 2 tem cargas positivas altas para as variáveis massa corporal total, CB, CMB, CP e IMC; ele será chamado de "indicadores de massa muscular".

O Fator 3 tem cargas positivas altas para as variáveis albumina e transferrina; ele será chamado de "indicadores bioquímicos do estado nutricional".

Tabela 32. Fatores para as variáveis $X_1, X_2, ..., X_{15}$, para as mulheres, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 |
|---|---|---|
| Massa corporal | **0.85** | 0.37 |
| CB | **0.82** | 0.51 |
| CMB | **0.84** | 0.32 |
| AGB | **0.75** | 0.58 |
| DCT | 0.57 | **0.65** |
| DCB | 0.22 | **0.85** |
| DCSE | 0.51 | **0.65** |
| DCSI | 0.39 | **0.75** |
| % G | 0.50 | **0.83** |
| CP | **0.73** | 0.27 |
| IMC | **0.77** | 0.48 |

### 4.2.6 Análise fatorial realizada com 13 variáveis, para os idosos de 60-79 anos

Observa-se na Tabela 21 que os 2 primeiros autovalores são maiores que 1 e os 2 autovalores juntos explicam 71,39% da variância total; e no gráfico scree-plot (Figura 16) observa-se a partir da segunda componente principal os autovalores decaem mais lentamente . Então, tem-se que $m$ igual 2 fatores.

Na Tabela 34, obtém-se os 2 fatores rotacionados para as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 60-79 anos. Esses fatores agrupam as variáveis da seguinte forma:

Fator 1: CB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura e IMC.

Fator 2: massa corporal total, CMB, CP, albumina e transferrina.

O Fator 1 tem cargas positivas altas para as variáveis CB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura e IMC; ele será chamado de "indicadores de gordura corporal".

Tabela 33. Fatores para as variáveis $X_1, X_3, ..., X_{14}$, considerando todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 | Fator3 |
|---|---|---|---|
| Massa corporal | 0.23 | **0.89** | -0.13 |
| CB | **0.61** | 0.74 | 0.05 |
| CMB | 0.19 | **0.90** | 0.05 |
| AGB | **0.85** | 0.43 | 0.06 |
| DCT | **0.91** | 0.21 | 0.02 |
| DCB | **0.79** | 0.18 | -0.04 |
| DCSE | **0.68** | 0.50 | 0.05 |
| DCSI | **0.75** | 0.39 | 0.00 |
| % G | **0.93** | 0.21 | 0.05 |
| CP | 0.29 | **0.76** | -0.10 |
| IMC | 0.53 | **0.75** | 0.01 |
| Albumina | -0.03 | 0.08 | **0.76** |
| Transferrina | 0.08 | -0.14 | **0.67** |

O Fator 2 tem cargas positivas altas para as variáveis massa corporal total, CMB e CP, e cargas negativas altas para as variáveis albumina e transferrina; ele será chamado de "indicadores de massa muscular e bioquímicos".

Tabela 34. Fatores para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos 60-79 anos, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 |
|---|---|---|
| Massa corporal | 0.43 | **0.84** |
| CB | **0.80** | 0.54 |
| CMB | 0.40 | **0.82** |
| AGB | **0.95** | 0.14 |
| DCT | **0.93** | -0.05 |
| DCB | **0.81** | -0.06 |
| DCSE | **0.86** | 0.19 |
| DCSI | **0.82** | 0.19 |
| % G | **0.95** | -0.05 |
| CP | 0.54 | **0.60** |
| IMC | **0.73** | 0.53 |
| Albumina | 0.14 | **-0.33** |
| Transferrina | 0.08 | **-0.43** |

### 4.2.7 Análise fatorial realizada com 13 variáveis, para os idosos de 80-100 anos

Observa-se na Tabela 24 que os 3 primeiros autovalores são maiores que 1 e os 3 autovalores juntos explicam 75,74% da variância total; e no gráfico scree-plot (Figura 19) observa-se a partir da terceira componente principal os autovalores decaem mais lentamente . Então, tem-se que $m$ igual 3 fatores.

Na Tabela 35, obtém-se os 3 fatores rotacionados para as variáveis $X_1, X_3, ..., X_{14}$, para os idosos de 80-100 anos. Esses fatores agrupam as variáveis da seguinte forma:

Fator 1: CB, AGB, DCT, DCB, DCSI e porcentagem de gordura.

Fator 2: massa corporal total, CMB, DCSE, CP, IMC e albumina.

Fator 3: transferrina.

O Fator 1 tem cargas positivas altas para as variáveis CB, AGB, DCT, DCB, DCSI e porcentagem de gordura; ele será chamado de "indicadores de gordura corporal".

O Fator 2 tem cargas positivas altas para as variáveis massa corporal total, CMB, DCSE, CP, IMC e albumina; ele será chamado de "indicadores de massa muscular e bioquímico".

O Fator 3 tem cargas positivas altas para as variáveis transferrina; ele será chamado de "indicador bioquímico".

### 4.2.8 Análise fatorial realizada com variáveis antropométricas primárias e as bioquímicas, para todos os idosos

Como visto na Tabela 27 tem-se que os 2 primeiros autovalores são maiores que 1 e os 2 autovalores juntos explicam 62,77% da variância total; e no

Tabela 35. Fatores para as variáveis $X_1, X_3, ..., X_{14}$, considerando os idosos 80-100 anos, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 | Fator3 |
|---|---|---|---|
| Massa corporal | 0.43 | **0.78** | -0.05 |
| CB | **0.69** | 0.66 | 0.13 |
| CMB | 0.40 | **0.77** | 0.22 |
| AGB | **0.83** | 0.45 | 0.08 |
| DCT | **0.88** | 0.24 | -0.04 |
| DCB | **0.71** | 0.25 | -0.11 |
| DCSE | 0.56 | **0.60** | -0.01 |
| DCSI | **0.83** | 0.25 | 0.06 |
| % G | **0.91** | 0.21 | 0.01 |
| CP | 0.31 | **0.75** | -0.26 |
| IMC | 0.61 | **0.69** | -0.01 |
| Albumina | -0.22 | **0.36** | 0.28 |
| Transferrina | 0.09 | -0.04 | **0.93** |

gráfico scree-plot (Figura 22) observa-se que os autovalores decaem mais lentamente na segunda componente principal. Então, tem-se $m$ igual 2 fatores.

Na Tabela 36, obtém-se os 2 fatores rotacionados para as variáveis antropométricas primárias e as bioquímicas, para todos os idosos. Esses fatores agrupam as variáveis da seguinte forma:

Fator 1: massa corporal total, CB , DCT, DCB, DCSE, DCSI e CP.

Fator 2: estatura, albumina e transferrina.

O Fator 1 tem cargas positivas altas para as variáveis massa corporal total, CB , DCT, DCB, DCSE, DCSI e CP; ele será chamado de "indicadores de gordura corporal".

O Fator 2 tem cargas positivas altas para as variáveis estatura, e carga negativa alta para albumina e transferrina; ele será chamado de "indicadores bioquímicos e estatura".

Tabela 36. Fatores para as variáveis antropométricas primárias e as bioquímicas, considerando todos os idosos, para cada variável, assinalado em negrito, o fator de maior peso.

| Variável | Fator1 | Fator2 |
|---|---|---|
| Massa corporal | **0.73** | 0.58 |
| Estatura | -0.12 | **0.87** |
| CB | **0.92** | 0.09 |
| DCT | **0.84** | -0.25 |
| DCB | **0.75** | -0.17 |
| DCSE | **0.86** | -0.09 |
| DCSI | **0.83** | -0.17 |
| CP | **0.72** | 0.39 |
| Albumina | 0.05 | **-0.19** |
| Transferrina | -0.04 | **-0.49** |

## 4.3 Software R usado nas análises

As análises de componentes principais e as análises fatorial foram realizadas pelo software R (Husson et al., 2010).

# 5 DISCUSSÃO

Na análise de componentes principais para as 15 variáveis e todos os idosos, algumas dessas variáveis possuem baixas correlações com as demais, fato que influencia nos resultados. Então essas variáveis foram retiradas e as análises foram realizadas novamente, para 11 variáveis e todos os idosos. Mas como todas as bioquímicas foram retiradas, realizou-se uma nova análise para 13 variáveis e todos os idosos, em que foram retiradas as variáveis proteína C reativa e a estatura.

Na análise com 15 variáveis, as variáveis estatura, transferrina e proteína C reativa, contrastam com as demais variáveis; e as variáveis que possuem baixas correlações com as demais, são as que tem os menores pesos na primeira componente principal (Tabela 7). Tem-se resultado semelhante na análise com 13 variáveis, na primeira componente principal a variável transferrina contrasta com as demais; e as variáveis transferrina e a albumina são as que apresentam os menores pesos (Tabela 19). Quando a análise é realizada para as 11 variáveis, na primeira componente principal obtem-se um índice global, em que as variáveis são ponderadas de acordo com a sua "importância"; já a segunda componente principal contrasta massa corporal total, CB, CMB, CP e IMC, que são indicadores de massa muscular, com AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura, que são indicadores de massa de gordura (Tabela 10). Nas 3 análises as variáveis que possuem os maiores são CB e AGB, ou seja, são as mais "importantes".

Na análise fatorial para as 15 variáveis, obtem-se 3 fatores, no fator 1 (CB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura e IMC) além dos indicadores de gordura, tem as variáveis CB e IMC; e a variável proteína C reativa aparece no fator 2 (massa corporal total, estatura, CMB, CP e proteína C reativa), e

não no fator 3 (albumina e transferrina), que tem os indicadores bioquímicos (Tabela 29). Mas na análise fatorial para 13 variáveis, tem-se no fator 1 (AGB, DCT, DCB, DCSE, DCSI e porcentagem de gordura) apenas os indicadores de gordura, no fator 2 (massa corporal total, CB, CMB, CP e IMC) os indicadores de massa muscular e no fator 3 (albumina e transferrina) todos os bioquímcos (Tabela 33). Na análise fatorial para 11 variáveis, como não tem as variáveis bioquímicas, não tem mais o indicador bioquímico, e os fatores 1 e 2 são iguais aos da análise com 13 variáveis (Tabela 30).

Retiradas as variáveis estatura, albumina, transferrina e proteína C reativa, além da análise para todos os idosos também foram realizadas análises por sexo. Para as mulheres a primeira componente principal também obtem um índice global, sendo as variáveis CB e AGB as que possuem o maior peso, o que coincide com a análise realizada para todos os idosos; mas na segunda componente principal as variáveis massa corporal total, CB, CMB, AGB, CP e IMC, contrastam com DCT, DCB, DCSE, DCSI e porcentagem de gordura, nessa componente o que difere da análise para todos os idosos é a AGB que passa a contrastar com os indicadores de gordura, deixando o outro grupo de variáveis com indicadores mistos, tanto de massa muscular quanto de gordura (Tabela 16). Já para os homens a primeira componente principal também apresenta um índice global, mas as variáveis mais "importantes" são porcentagem de gordura e CB; e na segunda componente principal a massa corporal total, CB, CMB, DCSI, CP e IMC, contrasta com AGB, DCT, DCB, DCSE e porcentagem de gordura, diferenciando na variável DCSI, neste caso o segundo grupo de variáveis apresenta indicadores de gordura e o primeiro grupo indicadores mistos, ou seja, indicadores de massa muscular e indicador de gordura (Tabela 13).

Na análise fatorial para homens, o fator 1 (Massa corporal total, CB, CMB, DCSE, DCSI, CP e IMC) passa a ter as variáveis DCSE e DCSI que para todos os idosos pertencia ao fator 2, deixando o fator 1 com indicadores de massa muscular e gordura; e o fator 2 (AGB, DCT, DCB e porcentagem de gordura) continua tendo apenas indicadores de gordura (Tabela 31). No caso da análise fatorial para as

mulheres, o fator 1 (Massa corporal total, CB, CMB, AGB, CP e IMC) passa a ter a variável AGB, fato que diferencia do fator 1 para todos os idosos, e também fica com indicadores de massa muscular e de gordura; e o fator 2 (DCT, DCB, DCSE, DCSI e porcentagem de gordura) também continua com somente indicadores de gordura (Tabela 32).

Para as 13 variáveis foram realizadas análises para todos os idosos, e para os idosos divididos por faixa etária, em dois grupos, um com idosos de 60-79 anos e o outro com idosos de 80-100 anos. Na análise de componentes principais para os idosos de 60-79 anos, a primeira componente principal é semelhante a primeira componente principal para todos os idosos; e nesta análise a segunda componente principal contrasta massa corporal total, CB, CMB, CP e IMC, que são indicadores de massa mucular, com AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura, albumina e transferrina, que possui os indicadores de gordura e o bioquímicos (Tabela 22). Para os idosos de 80-100 anos, na primeira componente principal obtem-se um índice global (Tabela 25). Nas 3 análises, as variáveis albumina e transferrina são as variáveis que possuem os menores pesos, e as variáveis CB e AGB as que possuem o maiores pesos.

Para os idosos de 60-79 anos obtem-se 2 fatores, o fator 1 (CB, AGB, DCT, DCB, DCSE, DCSI, porcentagem de gordura e IMC) tem além dos indicadores de gordura as variáveis CB e IMC; já o fator 2 (massa corporal total, CMB, CP, albumina e transferrina) tem a mais as variáveis albumina e transferrina (Tabela 34). Para os idosos de 80-100 anos, tem-se 3 fatores, o mesmo número de fator da análise para todos os idosos, o fator 1 (CB, AGB, DCT, DCB, DCSI e porcentagem de gordura) diferencia pela presença da variável CB e ausência da variável DCSE, já o fator 2 (massa corporal total, CMB, DCSE, CP, IMC e albumina) além da presença da DCSE, tem a variável albumina, deixando o fator com indicadores de massa muscular, de gordura e bioquímicos, e o fator 3 (transferrina) possui apenas um indicador bioquímico (Tabela 35).

Na análise para as variáveis antropométricas primárias, que são apenas

as variáveis obtidas de forma direta sem fórmulas, e as bioquímicas, para todos os idosos, obtem-se na primeira componente principal que as variáveis estatura e transferrina contrastam com as demais (Tabela 28). As variáveis que possuem baixas correlações continuam tendo os menores pesos, e a variável que possui o maior peso é a CB, seguida da DCSE com o segundo maior valor. Já na análise fatorial, o fator 1 (massa corporal total, CB , DCT, DCB, DCSE, DCSI e CP) tem as variáveis que possuem alta correlações entre si e o fator 2 (estatura, albumina e transferrina) as que possuem baixas correlações (Tabela 36).

Para as análises de componentes principais a CB é a variável que possui o maior peso, ou seja, é a variável mais "importante", isso acontece porque combinação com a medida da dobra cutânea do tríceps (DCT) permite, através da aplicação de fórmulas, calcular a circunferência muscular do braço (CMB) e a área de gordura do braço (AGB) que são correlacionadas com a massa muscular total, sendo utilizadas para diagnosticar alterações da massa muscular corporal total e, assim, o estado nutricional protéico (Acuña & Cruz, 2004).

Com a aplicação da análises de componentes principais, obtem-se peso para todas as variáveis, fato importante para a realização da avaliação nutricional. Já a aplicação da análise fatorial possibilitou ver como que os fatores se agrupam para cada caso.

# 6 CONCLUSÕES

Conclui-se, que com a análise de componentes principais, obtém-se pesos para todas as variáveis, de acordo com a "importância" das mesmas na avaliação do estado nutricional, fato que contribui para o trabalho de Galesi (2010), já que nesse trabalho utilizou-se apenas as variáveis que possuem referência em percentis na literatura, atribuindo o mesmo peso para todas as variáveis.

Nas análises de componentes principais realizadas, foi possível concluir que a variável circunferência do braço obteve o maior peso, ou seja, é a variável mais "importante".

Com os pesos das componentes principais pode-se obter um escore para cada idoso, e com o valor de cada escore é possível ordena-los. Mas não é possível fazer uma classificação do estado nutricional com esses escores, pois não temos critérios para classificar os idosos em eutróficos, levemente desnutridos, moderadamente desnutridos e gravemente desnutridos.

Como sugestão, num próximo estudo, seria uma avaliação do estado nutricional, com as mesmas variáveis antropométricas e bioquímicas, para uma população de idosos homogênea, com base em uma amostra representativa e que contenha somente idosos considerados bem nutridos. Com a aplicação da análise de componentes principais nestes dados, seria possível calcular um escore para cada um destes idosos, e utilizando o conjunto de escores, construir um intervalo que seria considerado o "padrão-áureo" para idosos bem nutridos. À seguir, para qualquer amostra representativa de idosos, como a de Galesi (2010), aplicar a análise de componentes principais e encontrar o escore de cada um destes idosos, verificando a posição do mesmo em relação ao intervalo "padrão-áureo" construído. O estado

nutricional de cada idoso seria classificado de acordo com a posição do escore em relação ao intervalo.

A análise fatorial possibilitou agrupar as variáveis em fatores, mostrando como que os indicadores de gordura, de massa muscular e os bioquímicos se agrupam.

Na análise fatorial, para todos os idosos sem as variáveis estatura e proteína C reativa, apresenta os fatores divididos em "indicadores de gordura corporal", "indicadores de massa muscular" e "indicadores bioquímicos do estado nutricional".

# REFERÊNCIAS BIBLIOGRÁFICAS


ACUÑA, K.; CRUZ, T. Avaliação do Estado Nutricional de Adultos e Idosos e Situação Nutricional da População Brasileira. **Arquivos Brasileiros de Endocrinologia e Metabologia**, v.48, n.3, p.345–361, 2004.

ANDERSON, T. W. **An Introduction to Multivariate Statistical Analysis**. New York: John Wiley and Sons, 1958. 374p.

BLACKBURN, G. L.; BISTRION, B. R.; MOINI, B. S.; SCHULAMN, H. T.; SMITH, M. F. Nutritional and metabolic assessment of the hospitalized patient. **Journal of Parenteral and Enteral Nutrition**, v.1, p.11–22, 1977.

BRASIL. **Censo demográfico 2000: características gerais da população Brasil 2000**. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2003.

CATTELL, R. B. The scree test for the number of factors. **Multivariate Behavioral Research**, v.1, p.245–276, 1966.

CERVI, A.; FRANCESCHINI, S. C. C.; PRIORE, S. E. Análise crítica do uso do índice de massa corporal para idosos. **Revista de Nutrição**, v.18, n.6, p.765–775, 2005.

CHATFIELD, C.; COLLINS, A. J. **Introduction to multivariate analysis**. London: Chapman and Hall, 1980.

CHUMLEA, W. C.; ROCHE, A. F.; STEINBAUGH, M. L. Estimating stature from knee height for persons 60 to 90 years of age. **Journal of the American Geriatric Society**, v.33, n.2, p.116–120, 1985.

COCHRAN, W. G. The estimation of sample size. In: COCHRAN, W. G. (Ed.). **Sampling techniques**. New York: John Willey, 1977. p.72–90.

FÉLIX, L. N.; SOUZA, E. M. T. Avaliação nutricional de idosos em uma instituição por diferentes instrumentos. **Revista de Nutrição**, v.22, n.4, p.571–580, 2009.

GALESI, L. F. Vigilância Nutricional em Idosos: Proposta para Cuidadores sem Formação específica. Araraquara, 2010. 105p. Tese (Doutorado) - Faculdade de Ciências Farmacêuticas, UNESP.

GUIGOZ, Y.; VELLAS, B. J.; GARRY, P. J. Mini Nutritional Assessment: A Practical Assessment Tool for Grading the Nutritional State of Elderly Patients. **Facts and Research in Gerontology**, v.4, n.2, p.15–59, 1994.

HILLS, M. Applied Statistics. **Book Review**, v.26, p.339–340, 1977.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v.24, n.7, p.498–520, 1933.

HUSSON, F.; LÊ, S.; PAGÈS, J. **Exploratory multivariate analysis by example using R**. CRC Press, 2010. 240p.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice Hall, 1986. 594p.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v.23, n.3, p.187–200, 1958.

KAMIMURA, M. A.; BAXMANN, A.; SAMPAIO, L. R.; CUPPARI, L. Avaliação Nutricional. In: CUPPARI, L. (Ed.). **Nutrição Clínica no Adulto. Guias de Medicina Ambulatorial e Hospitalar – UNIFESP/ Escola Paulista de Medicina**. Barueri: Manole, 2002. p.71–109.

KARSCH, U. M. Idosos dependentes: famílias e cuidadores. **Cadernos de Saúde Pública**, v.19, n.3, p.861–866, 2003.

LATTIN, J. M.; CARROLL, J. D.; GREEN, P. E. **Análise de dados multivariados**. São Paulo: CENGAGE Learning, 2011. 455p.

MANLY, B. J. F. **Multivariate statistical methods: a pimer**. Boca Raton: Chapman and Hall, 2005. 214p.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005. 297p.

MOITA, G. C.; NETO, J. M. M. Uma introdução à análise exploratória de dados multivariados. **Química Nova**, v.21, n.4, p.467–469, 1998.

MORRISON, D. F. **Multivariate statistical methods**. EUA: McGraw-Hill, 1967. 338p.

PEARSON, K. On lines and planes of closest fit to a system of points in space. **Philosophical Magazine**, v.2, n.6, p.559–572, 1901.

POSSOLI, S. Técnicas de análise multivariada para avaliação das condições de saúde dos municípios do Rio Grande do Sul, Brasil. **Revista Saúde Pública**, v.18, p.288–300, 1984.

RAUEN, M. S.; MOREIRA, E. A. M.; CALVO, M. C. M.; LOBO, A. S. Avaliação do estado nutricional de idosos institucionalizados. **Revista de Nutrição**, v.21, n.3, p.303–310, 2008.

RENCHER, A. C. **Methods of multivariate analysis**. New York: John Wiley, 2002. 708p.

SAMPAIO, L. R. Avaliação Nutricional e Envelhecimento. **Revista de Nutrição**, v.17, n.4, p.507–514, 2004.

SCAZUFCA, M.; CERQUEIRA, A. T. A. R.; MENEZES, P. R. Investigações epidemiológicas sobre demência nos países em desenvolvimento. **Revista de Saúde Pública**, v.36, n.6, p.773–778, 2002.

SEBER, G. A. F. **Multivariate observations**. New York: John Wiley, 2004. 686p.

SOUZA, V. M. C.; GUARIENTO, M. E. Avaliação do idoso desnutrido. **Revista Brasileira de Clínica Médica**, v.7, p.46–49, 2009.

SPEARMAN, C. "General intelligence," objectively determined and measured. **American Journal of Psychology**, v.15, p.201–293, 1904.

THURSTONE, L. L. **Multiple factor analysis**. Chicago: University of Chicago Press, 1947. 535p.