

CENTRO DE ESTATÍSTICA APLICADA CEA USP
RELATÓRIO DE ANÁLISE ESTATÍSTICA CÓDIGO 08P31

TÍTULO: Análise lingüística das expressões em pacientes com insuficiência cardíaca.

PESQUISADORES: Alfredo José Mansur, Alexandre Kazuhiko Suzuki, Marcos Lopes e Maria José Baraldi.

INSTITUIÇÕES: Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da USP (InCor/USP) e Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH/USP).

FINALIDADE: Publicação.

RESPONSÁVEIS PELA ANÁLISE: Viviana Giampaoli, Elisete da Conceição Quintaneiro Aubin, Felipe Junqueira Ribeiro e Karina Gernhardt Nakamura.

REFERÊNCIA DESTE TRABALHO: Aubin, E.C.Q., Giampaoli, V., Nakamura, K.G. e Ribeiro, F.J. **Relatório de Análise Estatística sobre o projeto “Análise lingüística das expressões em pacientes com insuficiência cardíaca”**. São Paulo, IME-USP, 2008 (RAE – CEA – 08P31).

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS

Aubin, E.C.Q., Giampaoli, V., Amaral, M.H. e Beraldi, F. **Relatório de Análise Estatística sobre o projeto “Análise lingüística das expressões de sintomas em pacientes com insuficiência cardíaca”**. São Paulo, IME-USP, 2007 (RAE – CEA - 07P04).

Beh, E.J. (2004). Simple Correspondence Analysis: A Bibliographic Review. **International Statistical Review**, 72,2, 257-284.

Bussab, W.O. e Morettin, P.A. (2006). **Estatística Básica**. 5 ed. São Paulo:Saraiva.

Davies, M. K., Gibbs, C. R., Lip G. Y. H. (2006). **ABC da Insuficiência Cardíaca**. 1 ed. TECMEDD.

Falk K., Swedberg K., Gaston-Johansson F.,Ekman I. (2006). Fatigue and anemia in patients with chronic heart failure. **European Journal of Heart Failure**. 2006, 744–749.

Jurafsky, D. e Martin, J.H. (2000). **Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 1 ed. Prentice Hall.

Kaufman, L. & Rousseeuw, P. J. (1990), **Finding Groups in Data - An Introduction to Cluster Analysis**, Wiley, NY.

Rousseeuw, P. J. (1987), **Silhouettes: a graphical aid to the interpretation and validation of cluster analysis**, Journal of Computational and Applied Mathematics 20, 53–65.

Shannon, C.E. & Weaver, W. (1949), **The Mathematical Theory of Communications**, University of Illinois Press, Urbana.

Souza, E.F. (2007). **Comparação de escolha de agrupamentos: uma proposta utilizando a entropia**. Dissertação de Mestrado. São Paulo: IME – USP.

PROGRAMAS COMPUTACIONAIS UTILIZADOS

Access for Windows (versão XP);

Excel for Windows (versão XP);

Word for Windows (versão XP);

MINITAB 14;

SPAD 5.6;

SPSS 13;

R 2.7.2.

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Outros (06:990)

Análise de Associação e Dependência de Dados Quantitativos (06:010)

Associação e Dependência de Dados Qualitativos (06:020)

ÁREA DE APLICAÇÃO

Medicina Epidemiologia (14:040)

Lingüística (14:110)

ÍNDICE

Resumo.....	5
1. Introdução	6
2. Descrição do estudo	7
3. Descrição das variáveis.....	8
3.1. Variáveis sócio-demográficas.....	9
3.2. Variáveis clínicas.....	9
3.3. Variáveis lingüísticas.....	10
4. Análise descritiva	14
4.1. Análise dos dados sócio-demográficos.....	14
4.2. Análise dos dados clínicos.....	15
4.3. Análise dos dados lingüísticos categorizados.....	16
4.4. Análise de correspondência entre variáveis lingüísticas, clínicas e sócio-demográficas.....	17
5. Análise inferencial.....	19
5.1 Agrupamento dos pacientes.....	20
5.1.1 Agrupamento dos pacientes segundo a variável “Tematização média”.....	20
5.1.2 Agrupamento dos pacientes baseado nas variáveis clínicas	20
5.2 Definições	23
6. Resultados finais.....	25
7. Conclusões.....	26
APÊNDICE A	28
APÊNDICE B	43
APÊNDICE C	49

Resumo

No dia a dia o diagnóstico da insuficiência cardíaca (IC) é dado através da análise de exames clínicos. No momento da consulta, no entanto, o médico realiza algumas perguntas no intuito de saber como está a saúde do paciente e se a IC impõe algumas limitações no cotidiano de suas vidas.

O objetivo deste estudo, portanto, é avaliar se os relatos dos pacientes aos médicos podem ajudá-los a descobrir algo mais do que os exames podem indicar.

Com base nas respostas de um questionário com perguntas abertas sobre a saúde e o cotidiano dos pacientes, a equipe de lingüistas desenvolveu a **análise de conteúdo**, o que permitiu a construção de novas variáveis lingüísticas. Além disso, foram feitos diferentes exames clínicos e foram colhidas informações sócio-demográficas de cada paciente, completando o banco de dados.

Para analisar e identificar algumas associações entre as variáveis sócio-demográficas, clínicas e lingüísticas, na primeira parte do estudo foi desenvolvida uma análise descritiva, incluindo análises de correspondência. Na segunda parte foram utilizadas medidas oriundas da Teoria da Informação, que quantifica a informação que se obtém quando relacionamos o agrupamento lingüístico e o agrupamento clínico. O agrupamento lingüístico levou em conta a variável *tematização média*, baseada no relato de cada paciente, e o clínico foi construído utilizando técnicas estatísticas e considerando algumas das variáveis clínicas observadas. Desta análise, concluiu-se que há um ganho de informação para o médico quando, conhecidos os fatores clínicos, analisa-se o relato dos doentes quanto às suas dificuldades.

1. Introdução

A insuficiência cardíaca (IC), usualmente dada como a fase final de várias doenças cardíacas, é o termo médico utilizado quando ocorre falha em uma das duas metades do coração, incapacitando-o de bombear o sangue que recebe para o resto do corpo.

Com essa deficiência do sistema circulatório, os pacientes apresentam limitações em suas atividades diárias, como dificuldade para falar e dormir. As queixas mais relatadas aos médicos são a freqüente falta de ar e o inchaço dos membros devido à retenção de líquidos pelo corpo (edema).

Basicamente são duas as causas dessa doença. A que mais comumente provoca a IC é o conjunto de doenças que diminuem a força de contração do miocárdio (músculo cardíaco), como, por exemplo, a *cardiopatía chagástica*, a *hipertensiva* e a *isquêmica*. A outra causa seria referente às anomalias que levam o coração a uma sobrecarga de trabalho excessiva, devido a uma má formação cardíaca ou doenças que aumentam o metabolismo geral do organismo. Não há cura para a IC mas sim tratamentos que amenizam os seus sintomas (Davies, 2006).

O diagnóstico é dado através de indicadores clínicos, obtidos com o ecocardiograma, a radiografia do tórax, a medição da pressão arterial e a taxa de triglicérides, entre outros. A análise destes exames encontra-se bastante explorada na literatura, principalmente quando relacionadas com os relatos dos pacientes, porém, tais estudos são baseados em formulários com questões fechadas e alternativas definidas de resposta pré-categorizadas e limitadas.

Nesses tipos de formulários, não há possibilidade de o sujeito apresentar respostas diferentes das pré-existentes. Além disso, pressupõe-se uma habilidade do respondente em preencher o formulário, o que não ocorre em alguns casos. A dificuldade aumenta quando se tenta aplicar, em nosso país, modelos de formulário estrangeiro, que não necessariamente refletem as características da população brasileira.

Dessa forma, para a realização de um estudo completo dos sintomas relacionados à IC e à conseqüente qualidade de vida dos doentes em uma casuística brasileira, com base nos relatos manifestados, a utilização de entrevistas compostas por questões

abertas parece ser uma boa forma de obter-se uma grande quantidade (e qualidade) de informações.

Para a análise das informações contidas nos relatos dos pacientes fez-se necessário um processo “padronizado” que possibilitou a quantificação da informação coletada. Este processo do estudo no qual se interpreta e se extrai de cada discurso as informações mais relevantes para os fins da pesquisa é chamado de **análise de conteúdo**.

Para realizar a análise de conteúdo, um grupo de lingüistas classificou, de maneira individual, independente e com revisão de concordância entre eles no final, o relato dos pacientes em categorias pré-definidas, denominadas, neste trabalho, de **variáveis lingüísticas**. Nesta etapa, utilizou-se o auxílio do *software* SPAD (Système Portable pour l’Analyse de Données).

A motivação deste trabalho é a inexistência de estudos lingüísticos aplicados à área médica, somada à rica qualidade de informações que podem daí ser extraídas.

Finalmente, temos que o objetivo geral deste trabalho é entender como o relato dos pacientes associa-se com o seu verdadeiro estado clínico.

2. Descrição do estudo

Os 266 pacientes deste estudo foram entrevistados e clinicados no período de 2001 a 2006 no Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (InCor/USP). Todos foram classificados em uma das seguintes etiologias cardiovasculares: Doença de Chagas, hipertensão arterial, isquemia e inespecífica (etiologia não identificada).

Três tipos de informações foram coletadas: dados cadastrais e informações sócio-demográficas, dados clínicos e o relato dos pacientes. Os dados cadastrais/sócio-demográficos têm a finalidade de caracterizar a amostra de pacientes (detalhamento na Seção 3.1). Já os dados clínicos referem-se aos resultados de diversos exames aos quais os pacientes foram submetidos, tais como, diabetes, pressão arterial, colesterol, triglicérides, taxa de hemoglobina, hematócrito, uréia, entre outros (detalhamento na

Seção 3.2). Por fim, as variáveis lingüísticas tiveram origem no relato dos pacientes a cada uma das seguintes questões:

- 1) Como é que a sua doença o(a) atrapalha?
- 2) Como o Sr.(a) se desloca até o hospital?
- 3) O Sr.(a) consegue fazer suas atividades domésticas sem dificuldade?
- 4) Como é o seu sono noturno?
- 5) O Sr.(a) tem cansaço?
- 6) Há dificuldade para respirar?
- 7) Como é a sua alimentação?

As respostas dos pacientes foram gravadas e transcritas em arquivos de texto pela equipe médica. Esses arquivos foram enviados à equipe de lingüistas, que analisaram o conjunto dos relatos, sem acesso às informações clínicas e sociais dos pacientes, evitando possível viés nas análises dos textos.

Analisamos, neste projeto, apenas a questão número 6 (“Há dificuldade para respirar?”). A partir da análise das respostas a essa questão, construiu-se o banco de dados lingüístico. É válido destacar que a diversidade de respostas e a procura de achados interessantes ou ignorados nas mesmas tornou o processo extremamente lento pelo cuidado requerido.

Para realizar as análises estatísticas, criou-se um banco de dados geral unindo a base de dados lingüística com as bases de dados clínica e sócio-demográfica.

3. Descrição das variáveis

A base de dados é composta por três grupos de variáveis: sócio-demográficas, clínicas e lingüísticas. Os dados do primeiro e segundo grupo foram fornecidos pelos pesquisadores do InCor e os dados do terceiro grupo foram desenvolvidos pela equipe de lingüistas para a sexta questão aberta da entrevista.

3.1. Variáveis sócio-demográficas

- **Idade** (em anos, no momento em que o paciente entrou no estudo);
- **Faixa etária** (até 30 anos, de 30 a 40 anos, de 40 a 50 anos, de 50 a 60 anos e mais de 60 anos);
- **Sexo** (masculino, feminino);
- **Raça** (branca, amarela, negra, parda, indígena);
- **Situação conjugal** (casado ou com companheiro, solteiro ou separado, viúvo);
- **Escolaridade** (nenhuma, ensino fundamental incompleto, ensino fundamental completo, ensino médio incompleto, ensino médio completo, ensino superior completo ou incompleto);
- **Anos de escolaridade** (0, 1, 2,...);
- **Renda familiar** (até 3 salários mínimos, de 3 a 6 salários mínimos, mais de 6 salários mínimos).

3.2. Variáveis clínicas

- **Etiologia da IC** (Chagásica, Hipertensiva, Isquêmica e Inespecífica);
- **Diabetes melito** (sim, não);
- **Anemia** (sim, não);
- **Dependência da Insulina** (sim, não);
- **Peso** (kg);
- **Altura** (m);
- **Índice de Massa Corpórea** (kg/m²);

- **Faixa de IMC** (abaixo do peso, peso normal, sobrepeso, obesidade grau I, obesidade grau II e obesidade grau III);
- **Pressão arterial sistólica** (mmHg);
- **Pressão arterial diastólica** (mmHg);
- **Colesterol** (mg/dl);
- **Triglicérides** (mg/dl);
- **Glicemia** (mg/dl);
- **Diâmetro do ventrículo direito** (mm)
- **Óbito** (sim, não), decorrente da IC. Para obter essa variável, o paciente foi acompanhado durante três anos após sua inclusão no estudo. Óbitos após esse período não foram considerados;
- **Tempo de sobrevida.** É a diferença entre a data de óbito e a data de inclusão no estudo. Foi dividido nas seguintes categorias: 0|-6 meses, 6|-18 meses, 18|-36 meses e maior ou igual a 36 meses.

3.3. Variáveis lingüísticas

As variáveis lingüísticas podem ser classificadas em dois grandes grupos: variáveis categorizadas e variáveis textuais. As variáveis categorizadas são aquelas em que as respostas possíveis ou categorias já foram pré-estabelecidas. Por exemplo, para a variável “intensidade da dificuldade” há as categorias *sem dificuldade*, *incógnita*, *muita*, *mediana* e *pouca*. Contrariamente, nas variáveis textuais, as possíveis respostas não são pré-fixadas.

A seguir, apresentamos as variáveis lingüísticas definidas para a Questão 6:

- **Grau de dificuldade:** designa a situação do paciente quanto ao tipo ou grau da dificuldade respiratória no momento da entrevista. A dificuldade foi classificada com base nos seguintes níveis ou categorias:
 - sem dificuldade;
 - com dificuldade, sem mais detalhes;
 - com dificuldade, às vezes;
 - com dificuldade condicionada;
 - dificuldade no passado;
 - dificuldade às vezes no passado;
 - dificuldade condicionada no passado.
- **Condição da dificuldade:** é uma variável textual. Refere-se às causas manifestadas pelo paciente como motivo do surgimento do(s) sintoma(s). Por exemplo, “cansaço”, “deitar”, “suspirar”, entre outras.
- **Aceleração:** variável dicotômica (sim/não). Trata-se de uma intensificação de uma ação ou estado qualquer que leva ao aparecimento do(s) sintoma(s). Por exemplo, “andar rápido” recebe “sim” nesta variável.
- **Intensidade:** refere-se à intensidade da dificuldade. Foi considerada quando o paciente a menciona diretamente ou, em alguns poucos casos, quando há uma inferência segura sobre ela. As respostas foram classificadas nas seguintes categorias:
 - sem dificuldade;
 - incógnita ;
 - pouca ;
 - mediana;
 - muita.

- **Queixas relacionadas:** é uma variável textual. Refere-se a outros fatores, que não a dificuldade para respirar, e que se manifestam nas respostas. Por exemplo, “dificuldade em dormir”, “dor no peito”, entre outras.
- **Fatores atenuantes:** é uma variável textual. Refere-se a qualquer fator que o paciente associe com o alívio do(s) sintoma(s). Por exemplo, “descansar”, “falar devagar”, entre outras.
- **Desaceleração:** é uma variável dicotômica (sim/não), e inversamente à “aceleração”, a desaceleração representa uma atenuação do processo relacionado ao surgimento do sintoma. Por exemplo, se “andar rápido” é marcado pela aceleração, “parar de andar” é uma desaceleração, recebendo “sim” nesta variável.
- **Tematização média:** é o “assunto” (tema) apresentado - aquilo de que o paciente falou. Foram identificados pelos lingüistas oito temas na questão:
 - **Corpo (ou partes do corpo):** citação de alguma parte do corpo, como por exemplo, em “*parece que meu **coração** está grande e me dá um bolo no **estômago***”. Abreviado por “C”.
 - **Atividades físicas:** qualquer atividade que incapacite o paciente de levar uma vida normal. Por exemplo, na frase “*às vezes, até conversando eu tenho **canseira***”. Abreviado por “A”.
 - **Fatores psicológicos:** o psicológico do(s) paciente(s) pode estar bastante abalado no geral. Frequentemente é relatada a palavra “morte”. Por exemplo, em “*sentia uma coisa horrível, uma vontade de morrer*”. Abreviado por “P”.
 - **Posição de dormir:** o paciente relata dificuldades em achar uma posição para dormir. Abreviado por “O”.

- **Medicamentos:** quando o paciente cita o uso de algum medicamento. Por exemplo, *“tomo os remédios direitinho e evito excesso de líquido”*. Abreviado por “M”.
 - **Fatores externos:** agentes externos atuam contra o paciente, gerando ou aumentando sua dificuldade. Por exemplo, em *“principalmente se alguém estiver fumando do meu lado”*. Abreviado por “E”.
 - **Alimentação:** o paciente relata incômodos e dificuldade de respirar em situações e/ou períodos relacionados à alimentação. Por exemplo, no trecho *“quando estou para me alimentar, sinto mais aquela fraqueza”*. Abreviado por “L”.
 - **Doenças:** o paciente relata alguma outra doença, exceto a IC. Por exemplo, em *“porque eu tenho diabetes também”*. Abreviado por “D”.
- **Figuras dos sintomas:** são segmentos de textos que expressam imagens sensoriais e/ou concretas ligadas aos temas. Por exemplo, *“eu tenho que puxar muito o ar”*, *“eu sentia assim apertar o peito”*, entre outras.
 - **Aspecto evolutivo:** trata da avaliação que o paciente pode relatar acerca de sua doença , isto é, se houve melhora ou piora, é definida nas seguintes categorias:
 - Não mencionado;
 - Paciente diz que se sente melhor (Positivo);
 - Paciente não diz que se sente melhor (Negativo).
 - **Figuras do tratamento:** é uma variável textual relacionada especificamente à terapia. Por exemplo, *“os médicos”*, *“os comprimidos”*, *“meu tratamento é muito bom”*, entre outras.

- **Observações:** não é uma variável, pois não passa por análises. Existe para o caso de ser necessário fazer anotações.

4. Análise descritiva

Realizou-se a análise descritiva (ver, por exemplo, Bussab e Morettin, 2006) dos três grupos de variáveis (clínicas, sócio-demográficas e lingüísticas). Os gráficos e tabelas estão apresentados no Apêndice A.

No Apêndice B, encontram-se os gráficos da análise de correspondência (Beh, 2004) entre algumas variáveis lingüísticas e as variáveis clínicas e sócio-demográficas. Com isso, buscou-se identificar possíveis associações entre o relato dos pacientes e o quadro clínico, objetivo do trabalho.

4.1. Análise dos dados sócio-demográficos

Da Tabela A.1 e Gráfico A.1, observamos a predominância do sexo masculino na amostra, 62% (166) dos pacientes. Na Tabela A.2 e Gráfico A.2 temos a distribuição dos pacientes segundo a raça, em que aproximadamente 62% (164) e 31% (83) dos pacientes se auto-declararam da raça branca e negra/parda, respectivamente.

Observamos a distribuição dos pacientes segundo a renda familiar na Tabela A.3 e Gráfico A.3. Dado que são pacientes atendidos na rede pública de saúde, é esperado que a renda familiar desses pacientes seja predominantemente baixa. Cerca de 44% (117) dos pacientes têm renda familiar de até 3 salários mínimos, 29% (78) com renda entre 3 e 6 salários mínimos e 23% (62) com mais de 6 salários mínimos.

Quanto aos anos de escolaridade (Tabela A.4 e Gráfico A.4), observa-se que 35,3% (94) dos pacientes possuem até 4 anos de estudo (inclusive), sendo a média e mediana de 5,8 e 5 anos, respectivamente, o que representa uma escolaridade baixa.

Na Tabela A.5 e Gráfico A.5, verifica-se que mais de 80% dos pacientes têm idade superior a 40 anos, sendo a idade mediana acima de 50 anos e inferior a 55 anos. A faixa

etária com maior frequência de pacientes é de 45 e 49 anos (16,2%). A idade média é de 51 anos, o desvio padrão é de 12,5 anos e coeficiente de variação é 24,5%.

4.2. Análise dos dados clínicos

Na Tabela A.6 e Gráfico A.6 temos a distribuição dos 266 pacientes segundo a etiologia da IC. A maioria dos pacientes possui etiologia hipertensiva (33,5%) ou inespecífica (31,6%). Juntas, representam aproximadamente 65% dos pacientes.

Vemos na Tabela A.7 e Gráfico A.7 que, dentre as mulheres, há uma incidência de 71% com etiologia hipertensiva ou inespecífica, sendo maior do que entre os homens (61,4%). Já a proporção de homens com etiologia isquêmica é de 26,5%, uma incidência maior quando comparada à das mulheres (17%).

Em relação ao óbito por IC, através da Tabela A.8 e Gráfico A.8, percebe-se que 28,2% dos pacientes faleceram durante o estudo. Pela Tabela A.9 e Gráfico A.9, parece haver associação entre óbito e etiologia, pois as proporções de óbitos em pacientes chagásticos (46,9%) e isquêmicos (44,3%) foram superiores às proporções de óbitos em outras etiologias.

Muitos pacientes da amostra foram classificados como fora do peso normal, segundo a variável “faixa de IMC”: 3,4% estavam abaixo e 55,8% acima do peso considerado ideal. Isso pode ser visto na Tabela A.10 e Gráfico A.10. Segundo os cardiologistas, a IC pode causar retenção de líquidos (edematização), o que eleva o peso, e, conseqüentemente, o índice de massa corpórea.

Analisando o tempo de sobrevivência (Tabela A.11 e Gráfico A.11), percebe-se que, dos pacientes que morreram durante o estudo, 68% vieram a falecer entre 6 e 36 meses após o início do estudo. Na Tabela A.12 e Gráfico A.12, é possível notar que os pacientes com as etiologias chagástica e inespecífica tiveram menor tempo de sobrevivência, pois cerca de 30% do total de pacientes de cada uma dessas etiologias morreram em menos de 6 meses. Por outro lado, cerca de 90% dos pacientes de etiologia hipertensiva tiveram tempo de sobrevivência maior ou igual a 6 meses.

Na Tabela A.13 e Gráfico A.13 nota-se que 76% (202) dos pacientes são diabéticos. Na Tabela A.14 e Gráfico A.14 percebe-se que 43,6% (116) dos pacientes têm pressão diastólica e sistólica acima do considerado normal, 130mmHg e 85mmHg, respectivamente.

Em Falk *et al.* (2006) a anemia é citada como um fator para a ocorrência da dificuldade respiratória. Na Tabela A.15 e Gráfico A.15 temos a distribuição dos pacientes segundo o diagnóstico (ou não) de anemia, em que nota-se que 82,3% dos pacientes apresentam anemia. Na Tabela A.16 e Gráfico A.16 observa-se que a proporção de pacientes que não relataram nenhuma dificuldade para respirar é maior naqueles que não apresentaram anemia (26,9%) quando comparados àqueles com anemia (13,3%). Já a proporção de pacientes que relataram dificuldade condicionada para respirar é maior naqueles que apresentaram anemia (37,8%) quando comparados àqueles sem anemia (16,4%).

4.3. Análise dos dados lingüísticos categorizados

A distribuição dos pacientes segundo a tematização média é apresentada na Tabela A.17 e Gráfico A.17. Observa-se que, das 241 tematizações médias relatadas pelos pacientes, 30,3% (73) foram relacionadas à alguma parte do corpo e 24,9% (60) foram citações de dificuldade ao realizar atividade física. As tematizações médias *medicamentos*, *doenças* e *alimentação* apresentaram as menores freqüências, 7,5% (18), 2,5% (6) e 1,7% (4), respectivamente. Estas tematizações são consideradas pelos médicos como as mais “surpreendentes”, pois esperam que tragam bastante informação por não serem esperadas no relato de um paciente.

Na Tabela A.18 e Gráfico A.18, é apresentada a distribuição dos pacientes segundo o grau de dificuldade para respirar, em que cerca de 25% dos pacientes relataram não terem dificuldades para respirar e 35% apresentaram dificuldade condicionada (ao realizar exercícios físicos, por exemplo). Os demais pacientes relataram dificuldades somente no passado (11,3%) ou outros tipos de dificuldade (28,7%). Já na Tabela A.19 temos a distribuição dos pacientes segundo grau da intensidade e tematização média.

Observa-se que pacientes com dificuldade condicionada ou com dificuldade de respirar, sem mais detalhes, citaram mais partes do corpo (48,1% e 50%, respectivamente) enquanto que pacientes com “dificuldade às vezes” ou com “dificuldade condicionada no passado” citaram mais a tematização “atividades físicas” (30,9% e 36,4%, respectivamente). Pacientes com dificuldades no passado relataram mais a tematização média “medicamentos” (36,4%).

Através da Tabela A.20 e Gráfico A.20, é apresentada a distribuição dos pacientes segundo a etiologia por tematização média. Nota-se que, dentre os pacientes que apresentaram tematização média “E” (fatores externos), a maior proporção ocorre para pacientes com etiologia isquêmica (44,4%). Destaca-se também que, dos seis pacientes com tematização “L” (alimentação), cinco deles têm etiologia hipertensiva. Observando a distribuição da etiologia, dentro da tematização “O” (posição de dormir), 11 (47,8%) pacientes possuem etiologia hipertensiva.

Da Tabela A.21, nota-se que os pacientes com as etiologias chagástica, inespecífica e isquêmica citaram mais a tematização “corpo”, enquanto que os pacientes com etiologia hipertensiva citaram mais “atividades físicas”.

4.4. Análise de correspondência entre variáveis lingüísticas, clínicas e sócio-demográficas

Com o propósito de analisar não somente se existem relações entre as categorias das variáveis lingüísticas, clínicas e sócio-demográficas, mas também como se dão, optou-se por realizar uma análise descritiva baseada na análise de correspondência. Todos os gráficos encontram-se no Apêndice B.

Ressaltamos que, neste trabalho, considerou-se a análise da variável “tematização média”, pois já foram realizados estudos prévios com o mesmo corpus de texto (Aubin *et al*, 2007). Esse tipo de tematização leva em consideração somente o assunto global do relato do paciente, fazendo com que muita informação seja perdida.

No Gráfico B.1, percebe-se que a tematização média “posição para dormir” está bem próxima e associada com as faixas etárias “entre 50 e 60 anos” e “mais de 60 anos”. Uma

possível justificativa seria porque pessoas mais velhas normalmente sentem mais dificuldade para dormir e esse quadro acaba piorando com a freqüente falta de ar noturna. Pode-se notar também que a tematização “atividade física” está próxima à faixa etária “de 40 a 50 anos”. Isso pode ser explicado pelo fato de que nessa idade as pessoas, ainda bastante ativas, começam a sentir os primeiros efeitos da idade.

Do gráfico que relaciona tematização média com a etiologia da IC (Gráfico B.2), pode-se notar que “posição de dormir” está próxima das etiologias isquêmica e inespecífica e que a etiologia chagástica está próxima das tematizações “atividade física”, “corpo” e “alimentação”. A etiologia hipertensiva não se encontra próxima a nenhuma tematização, possivelmente porque essa etiologia leva o paciente a não ter sintomas muito graves.

Através do Gráfico B.3, percebe-se que a etiologia isquêmica está muito associada ao óbito (o que já era esperado, pois dentre os pacientes que faleceram, a maior porcentagem era isquêmico) e que as etiologias hipertensiva e inespecífica estão próximas à tematização “posição de dormir” e ao “não óbito”.

Do Gráfico B.4, pode-se concluir que a categoria “6 a 18 meses” de tempo de sobrevida está próximo à tematização média “atividade física” e que o tempo de sobrevida “mais de 36 meses” é a categoria que fica mais distante de todas as tematizações. Normalmente, os pacientes com um maior tempo de sobrevida estão em um nível mais inicial da doença, fazendo com que os sintomas estejam menos evidentes e, portanto, não haja tantas queixas.

Já ao relacionar o tempo de sobrevida com as condições da dificuldade (Gráfico B.5), concluímos que a categoria “de 18 a 36 meses” está próxima às condições “nariz tampado”, “estômago vazio”, “dormir”, “dia frio”, “alergia no nariz” e “calor”, possivelmente porque pessoas com maior tempo de sobrevida supostamente estão em um estágio mais inicial da doença. Nessa fase, os pacientes queixam-se de coisas mais simples e mais subjetivas. Já a categoria “de 6 a 18 meses” de sobrevida está associada com as condições “subir escada”, “esforçar”, “descer rampa”, “ambiente abafado” e “andar”. Provavelmente isto ocorre, pois pessoas com um tempo de vida um pouco menor (ou seja, em um estágio da doença um pouco mais avançado) já começam a sentir dificuldades realizando esforços físicos. Por fim, percebe-se que a categoria “de 0 a 6 meses” está próxima às condições “suspirar” e “inchaço”, indicando que os pacientes em

um estágio da doença muito avançado sentem dificuldade até em suspirar e também se queixam de inchaço nos membros. Com exceção do Gráfico B.4, destacamos que seria necessário, segundo as tabelas de inércia apresentadas no Apêndice B, a análise dos gráficos de pelo menos a combinação de seis eixos para ter uma inércia acumulada próxima à 70%, com a qual a interpretação dos resultados seria confusa e pouco informativa. Assim, essa abordagem por meio da análise de correspondência é meramente ilustrativa.

5. Análise inferencial

Conforme mencionado anteriormente, o objetivo do trabalho é investigar possíveis associações do relato do paciente com o seu quadro clínico. A solução proposta foi utilizar alguns conceitos da Teoria da Informação, um campo de estudos surgido originalmente na área de engenharia das comunicações há pouco mais de 50 anos e que tem servido para a criação de novos métodos de comparação e avaliação de agrupamentos. Para maiores informações, consultar Shannon & Weaver (1949).

Para a utilização destes conceitos, considerou-se a variável *tematização média* para o “agrupamento lingüístico” (*L*), sendo que cada combinação determinou um grupo diferente. No total, somaram 34 grupos (Tabela C.1). Foi feito também um outro agrupamento, denominado “agrupamento clínico” (*C*), considerando somente algumas variáveis clínicas. Assim, a partir destas duas segmentações, algumas medidas foram calculadas de maneira a avaliar a presença ou não de associação entre o quadro clínico e o relato dos pacientes.

5.1 Agrupamento dos pacientes

5.1.1 Agrupamento dos pacientes segundo a variável lingüística “Tematização média”

O agrupamento dos pacientes considerando a variável lingüística “Tematização média”, detalhada na Seção 3.3, foi realizado de forma que fosse preservada a quantidade de tematizações ditas por cada paciente, ou seja, cada combinação de tematização determinou um grupo diferente, totalizando 34 grupos descritos na Tabela C.1, sendo “a: atividades físicas”, “c: corpo (ou partes do corpo)”, “p: fatores psicológicos”, “o: posição de dormir”, “e: fatores externos”, “m: medicamentos”, “d: doenças” e “l: alimentação”.

5.1.2 Agrupamento dos pacientes baseado nas variáveis clínicas

A análise de agrupamentos é uma técnica estatística de análise multivariada, utilizada para encontrar ou revelar a existência de grupos de indivíduos em uma amostra, de tal modo que sejam similares entre si e diferentes daqueles pertencentes aos demais grupos, de acordo com as variáveis consideradas.

Assim, um segundo agrupamento foi constituído utilizando apenas algumas variáveis clínicas consideradas mais relevantes, sendo realizado de forma distinta e independente do agrupamento lingüístico.

- **Escolha das variáveis utilizadas**

Para o agrupamento dos pacientes, dentre todas as variáveis clínicas (ver Seção 3.2) foram consideradas as seguintes variáveis:

- Pressão arterial sistêmica;
- Pressão arterial diastólica;
- HDL colesterol;
- LDL colesterol;

- Colesterolemia;
- Triglicérides.

A escolha por tais variáveis foi devido a estas serem as que tiveram maior poder de discriminação, ou seja, utilizando-as foi possível identificar grupos bem definidos de pacientes, semelhantes entre si e distintos entre outros grupos.

- **Técnica utilizada**

Conforme detalha Souza (2007), nos últimos 40 anos houve um abundante desenvolvimento de técnicas para realizar a análise de agrupamentos, a fim de acomodar a diversidade de estruturas que se pode definir com a expressão “tipos de grupos”. Dentre essa abundância de técnicas analíticas para construir agrupamentos, utilizamos neste trabalho o método de partição *PAM* (Kaufman & Rousseeuw, 1990).

O método *PAM* (do inglês *Partitioning Around Medoids*), comparado ao tradicional método de agrupamento *k-means*, tem as seguintes características:

- a) Também aceita uma matriz de dissimilaridade;
- b) É mais robusto, pois minimiza a soma das dissimilaridades, ao invés de uma soma de distâncias euclidianas;
- c) Proporciona o gráfico de silhueta;
- d) Aceita a entrada de valores faltantes (comumente chamados de *missings*).

Devido às diferentes origens e ordem de grandeza, o agrupamento foi realizado com as variáveis padronizadas, utilizando o desvio médio absoluto, definido como

$$z_i = \frac{1}{n} \sum_{i=1}^n |x_i - med(X)|,$$

em que $med(X)$ é a mediana da variável X .

Uma diferença do método *PAM* para os métodos de agrupamento hierárquicos tradicionais é que o número de grupos deve ser informado antes da execução do algoritmo, ou seja, devemos defini-lo *a priori*, utilizando algum critério, como por exemplo o método das silhuetas, descrito a seguir.

- **Determinação do tamanho dos grupos**

Para a determinação do tamanho dos grupos, foi utilizado o método das silhuetas. Esse método, proposto por Rousseeuw (1987), auxilia o pesquisador a escolher o número mais adequado de grupos e, ao mesmo tempo, permitir que se construa uma representação gráfica do agrupamento encontrado.

Para a escolha do número de grupos, Kaufman & Rousseeuw (1990) sugerem construir todos os respectivos agrupamentos com $k = 2, 3, \dots, n-1$ grupos e escolher aquele com o maior *coeficiente de silhueta média (CSM)*, desde que esse valor seja maior que 0,25. O *CSM* é um índice de qualidade para todo o agrupamento, que varia no intervalo $[-1;1]$, além de ser adimensional. Quanto mais próximo de 1 é o valor do *CSM*, melhor definido está o grupo. Valores próximos de -1 e próximos de zero indicam, respectivamente, indivíduos mal classificados e indivíduos em ponto intermediário entre os grupos. Maiores detalhes e interpretações foram omitidos e podem ser encontrados em Souza (2007).

Através do Gráfico C.1 observamos que o maior valor do coeficiente de silhueta média foi conseguido com $k = 3$ grupos ($CSM = 0,31$). Os pacientes foram alocados (através da técnica *PAM*) em três grupos de tamanhos 126, 94 e 32.

- **Caracterização dos grupos**

No Gráfico C.2 apresentamos o *boxplot* das variáveis pressão arterial sistêmica e diastólica, segundo o grupo de pacientes. Temos 22,5% (27), 44,3% (39) e 40,0% (12) pacientes nos grupos G1, G2 e G3, respectivamente, acima do limite de normalidade (130mmHg) da pressão arterial sistêmica e 25,8% (31), 37,5% (33) e 26,7% (8) pacientes em cada grupo, respectivamente, acima do limite de normalidade (85mmHg) da pressão arterial diastólica. Nas Tabelas C.2.1 e C.2.2 temos as medidas resumo para cada grupo. Há indícios de que os grupos 2 e 3 são formados por pacientes com um quadro clínico mais desfavorável, pois são os que possuem uma maior porcentagem deles com pressão fora do limite de normalidade.

No Gráfico C.3 temos o *boxplot* das variáveis HDL Colesterol e LDL Colesterol segundo o grupo de pacientes. Vemos que os grupos 1 e 2 apresentam valores de HDL Colesterol superiores aos do grupo 3. O grupo 2 é o que apresenta os maiores valores

de LDL Colesterol. São considerados níveis normais de colesterol acima de 60mg/dl (HDL) e abaixo de 130mg/dl (LDL). Novamente, os grupos 2 e 3 são os que possuem maior quantidade de pacientes fora dos níveis recomendados. Nas Tabelas C.3.1 e C.3.2 temos as medidas resumo para cada grupo.

No Gráfico C.4 apresentamos o *boxplot* da variável taxa de triglicérides, segundo o grupo de pacientes. Vemos que o grupo 1 é composto por aqueles pacientes com as menores (e mais adequadas) taxas de triglicérides e, também, o mais homogêneo. O grupo 3 é composto por pacientes com as maiores taxas de triglicérides dentre os três grupos, e é também o que apresenta maior variabilidade dos valores. Temos 57,1% (72), 48,9% (46) e 100,0% (32) pacientes em cada grupo, respectivamente, fora dos limites de normalidade (entre 80 e 150mg/dl) na taxa de triglicérides. Na Tabela C.4 temos as medidas resumo para cada grupo.

No Gráfico C.5 temos o *boxplot* da variável taxa de colesterolemia, segundo o grupo de pacientes. Dado que a taxa de colesterolemia é considerada normal até 200mg/dl, temos 2,4% (3), 94,7% (89) e 56,3% (18) pacientes em cada grupo, respectivamente, acima dos limites de normalidade desta variável. Na Tabela C.5 temos as medidas resumo para cada grupo. Parece que os grupos 2 e 3 possuem os pacientes com pior quadro clínico, com taxas de colesterolemia mais afastadas do considerado ideal.

De maneira geral, as tematizações “c”, “a” e “p” são as mais freqüentes nos três grupos, nessa ordem. Na Tabela C.6, considerando aqueles pacientes que realizaram alguma tematização, temos a distribuição das tematizações médias segundo cada um dos grupos. Claramente a tematização mais freqüente é a “c: corpo (ou partes do corpo)” para os grupos 1 e 2. Para o grupo 3, as tematizações “a: atividades físicas” e “c: corpo (ou partes do corpo)” são praticamente igualmente freqüentes. Os grupos 2 e 3, aparentemente formados por pacientes com um quadro clínico mais desfavorável, possuem as tematizações “o” e “e” com maiores freqüências que o grupo 1.

5.2 Definições

Para comparar os agrupamentos lingüístico (L) e clínico (C), utilizou-se algumas medidas da Teoria da Informação, descritas a seguir.

(i) Entropia

É uma medida que avalia a incerteza (ou “surpresa”) da informação. Por isso, quanto maior a probabilidade de ocorrência de uma informação, menor sua informatividade. A entropia é calculada por:

$$H(A) = -\sum_{k=1}^K P(k) \log [P(k)],$$

em que A é o agrupamento em estudo, K é o número de grupos do agrupamento A e $P(k)$ é a probabilidade de um indivíduo pertencer ao k -ésimo grupo de A , sendo $A = L$ ou $A = C$.

(ii) Informação mútua entre C e L

É uma medida que representa quanta informação contém um agrupamento a respeito do outro, e é determinada por:

$$I(C, L) = \sum_{k=1}^K \sum_{k'=1}^{K'} P(k, k') \log \left[\frac{P(k, k')}{P(k) \cdot P(k')} \right],$$

em que C é o agrupamento clínico, L é o agrupamento lingüístico, K é o número de grupos de C , K' é o número de grupos de L , $P(k, k')$ é a probabilidade de um indivíduo pertencer, simultaneamente, aos dois agrupamentos, $P(k)$ é probabilidade de um indivíduo pertencer ao grupo k do agrupamento C e $P(k')$ é probabilidade de um indivíduo pertencer ao grupo k' do agrupamento L .

(iii) Entropia condicional

Representa a quantidade média de informação que ainda resta para comunicar C dado que L já é conhecido, sendo obtida por:

$$H(C | L) = - \sum_{k=1}^K \sum_{k'=1}^{K'} P(k, k') \log [P(k | k')],$$

em que C , L , K , K' e $P(k, k')$ são como antes e $P(k | k')$ é probabilidade de um indivíduo pertencer ao agrupamento k de C , dado que pertence ao agrupamento k' de L .

(iv) Variação da informação

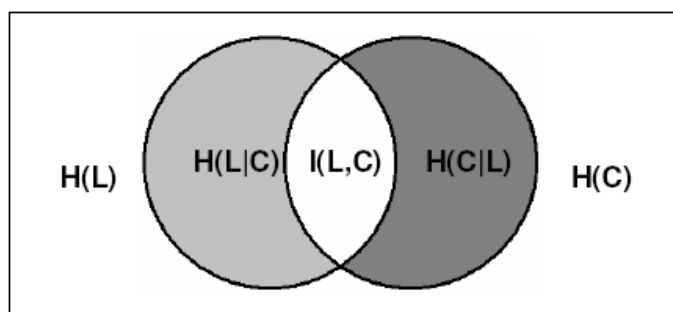
Utilizando esses conceitos, define-se um índice que mede a variação total da informação induzida pelos agrupamentos C e L , denominado *variação da informação*, sendo calculado por:

$$\begin{aligned} VI(C, L) &= H(C) + H(L) - 2 \times I(C, L) \\ &= H(C | L) + H(L | C) \end{aligned}$$

A partir da variação da informação, pode-se também medir quanta informação sobre C será perdida e quanta informação sobre L será ganha em uma mudança do primeiro para o segundo agrupamento, através de suas respectivas entropias condicionais.

A Figura 5.1.1 ilustra cada uma destas medidas.

Figura 5.1.1: Ilustração gráfica dos conceitos.



6. Resultados

finais

Dadas as duas classificações, construídas de forma independente, distintas e obtidas através das variáveis lingüísticas e clínicas, aplicou-se as medidas de “Entropia

Condicional”, “Informação Mútua” e “Variação da Informação” apresentadas do Tópico 5.2. Os valores dessas medidas a esse conjunto de dados foram:

$$H(C|L) = 0,39$$

$$H(L|C) = 0,15$$

$$I(L,C) = I(C,L) = 0,02$$

$$VI(L,C) = VI(C,L) = 0,54$$

Tais resultados indicam que a quantidade de informação, ao acrescentarmos o relato dos pacientes ao relatório clínico do paciente, é maior do que o contrário (acrescentarmos as medidas clínicas tendo o relato dos pacientes), já que $H(L|C) < H(C|L)$. Há também pouca informação contida simultaneamente nos dois agrupamentos, já que a informação mútua foi baixa (0,02). Por fim, tem-se um ganho considerável de informação ao utilizarmos os dois agrupamentos considerados nesse estudo, já que a variação da informação apresentou-se razoavelmente distante de zero (0,54).

7. Conclusões

Os pacientes com insuficiência cardíaca, pertencentes à amostra desse projeto, possuem idade média de 51 anos. A maioria deles (58,8%) encontra-se acima do peso considerado “ideal”.

Dentre as quatro etiologias investigadas, as mais freqüentes foram a hipertensiva e inespecífica, com proporções de 33,5% e 31,6% respectivamente. Há indícios de que as etiologias da IC estejam relacionadas com o óbito dos pacientes. Os dados sugerem que pacientes das etiologias chagástica e isquêmica têm uma maior incidência do óbito (Tabela A.9). Por outro lado, as etiologias hipertensiva e inespecífica parecem relacionar-se com a tematização “posição de dormir” e ao “não óbito” do paciente (Gráfico B.3).

Somente 24,9% dos pacientes relataram não ter nenhum tipo de dificuldade para respirar, enquanto que 75,1% apresentaram essa dificuldade com alguma intensidade

(Gráfico A.18). Os pacientes com mais de 50 anos relataram dificuldade em achar uma posição para dormir (Gráfico B.1).

Ao analisarmos os relatos dos pacientes, utilizando as tematizações médias, observou-se que as tematizações “corpo” e “atividades físicas” foram as mais relatadas. As tematizações “corpo”, “atividades físicas” e “alimentação” parecem ter associação com a etiologia chagástica (Gráfico B.2).

O tempo de sobrevivência do paciente, aparentemente, mostrou-se relacionado com a tematização média, pois, à medida que o tempo de sobrevivência diminuiu, condições mais simples acabam ocasionando dificuldades para respirar.

Na segunda parte do trabalho, optou-se por utilizar algumas medidas da Teoria da Informação para avaliar a associação entre o relato do paciente (através da variável tematização média) e o seu quadro clínico (através do agrupamento de algumas variáveis clínicas). Através das entropias condicionais, concluiu-se que uma maior informação é obtida quando, conhecidas as variáveis lingüísticas, acrescentam-se as variáveis clínicas, do que o inverso, como é esperado.

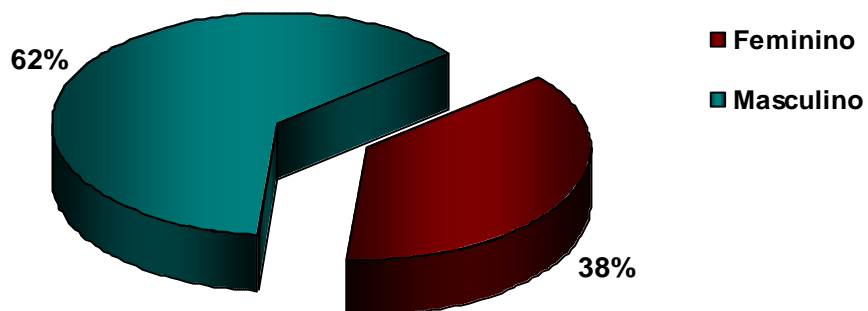
Temos que a informação mútua, intersecção das medidas $H(C)$ e $H(L)$, foi de 0,02, ou seja, é pequena a quantidade de informação contida simultaneamente nos dois agrupamentos. Por fim, obtemos de maneira quantitativa com estas técnicas, o ganho de informação que existe quando o médico ouve o relato do paciente, o qual será útil na hora de realizar comparações tanto em relação às outras perguntas quanto ao outro tipo de classificação lingüística.

APÊNDICE A

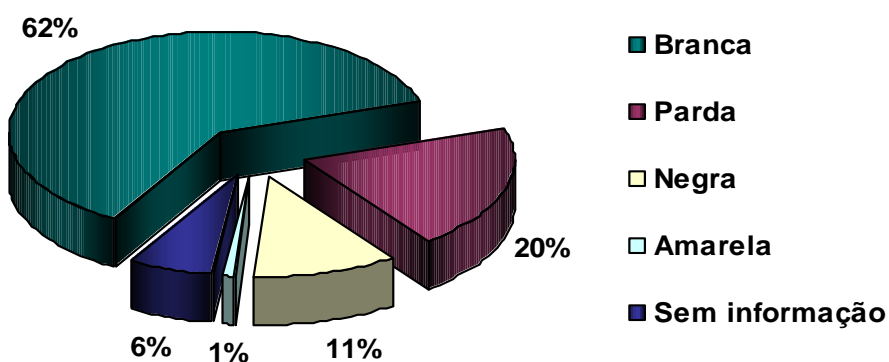
Tabelas e Gráficos da Análise Descritiva

Tabela A.1 – Distribuição dos pacientes segundo o gênero.

Gênero	Frequência	%
Homem	166	62,4
Mulher	100	37,6
TOTAL	266	100,0

Gráfico A.1 – Distribuição dos pacientes segundo o gênero.**Tabela A.2** – Distribuição dos pacientes segundo a raça.

Raça	Frequência	%
Branca	164	61,7
Parda	53	19,9
Negra	30	11,3
Amarela	2	0,8
Sem informação	17	6,4
TOTAL	266	100,0

Gráfico A.2 – Distribuição dos pacientes segundo a raça.**Tabela A.3** – Distribuição dos pacientes segundo a renda familiar.

Faixa salarial	Frequência	%
Até 3 salários mínimos	117	44,0
Entre 3 e 6 salários mínimos	78	29,3
Mais de 6 salários mínimos	62	23,3
Sem informação	9	3,4

TOTAL	266	100,0
--------------	------------	--------------

Gráfico A.3 – Distribuição dos pacientes segundo a renda familiar.



Tabela A.4 – Distribuição dos pacientes segundo anos de escolaridade.

Faixa de anos de estudo	Freqüência	%
0 a 4 anos	94	35,3
5 a 8 anos	68	25,6
9 a 12 anos	32	12,0
13 anos ou mais	4	1,5
Sem resposta	68	25,6
TOTAL	266	100,0

Gráfico A.4 – Histograma de freqüência dos pacientes segundo anos de escolaridade.

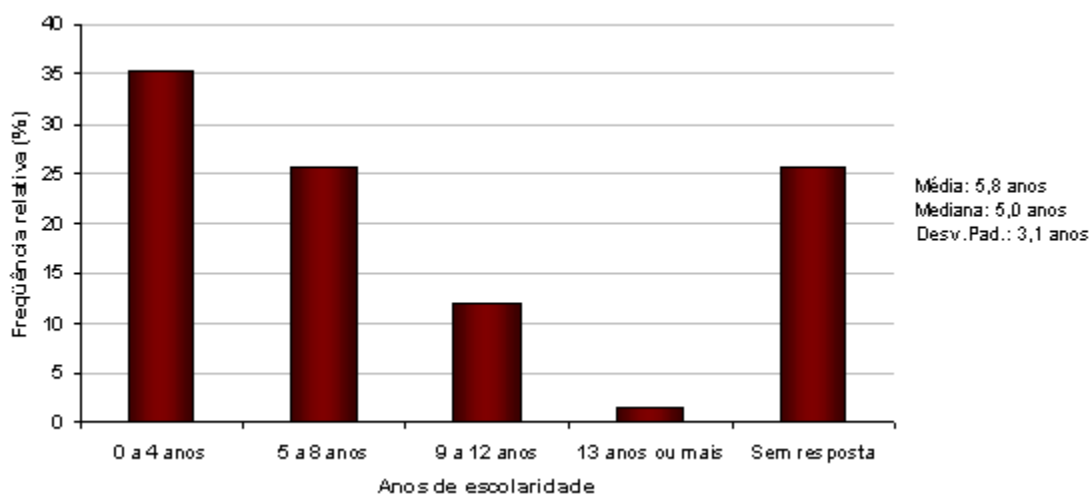


Tabela A.5 – Distribuição dos pacientes segundo a idade.

Idade (em anos)	Frequência	%	% Acumulado
15 20 anos	4	1,5	1,5
20 25 anos	2	0,8	2,3
25 30 anos	8	3,0	5,3
30 35 anos	8	3,0	8,3
35 40 anos	24	9,0	17,3
40 45 anos	29	10,9	28,2
45 50 anos	43	16,2	44,4
50 55 anos	40	15,0	59,4
55 60 anos	39	14,7	74,1
60 65 anos	31	11,7	85,7
65 70 anos	21	7,9	93,6
70 75 anos	10	3,8	97,4
75 80 anos	5	1,9	99,2
80 85 anos	2	0,8	100,0
TOTAL	266	100,0	

Gráfico A.5 – Histograma de frequência dos pacientes segundo a idade.

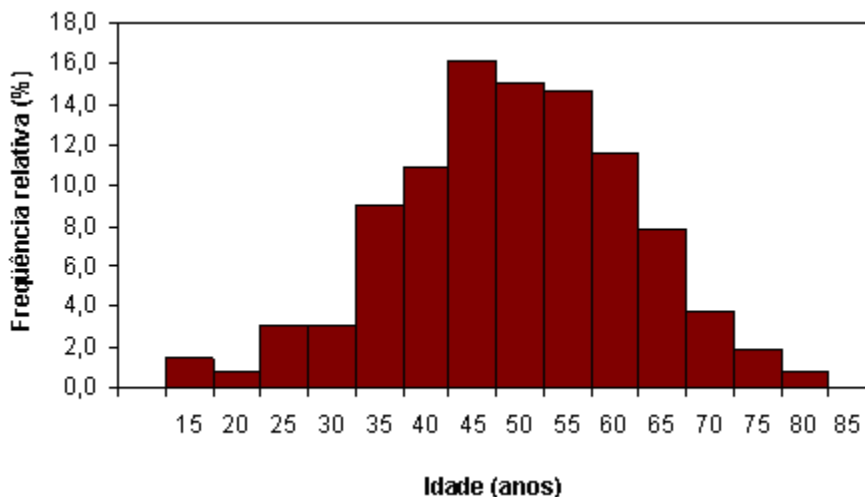


Tabela A.6 – Distribuição dos pacientes segundo a etiologia da IC.

Etiologia da IC	Frequência	%
Chagástica	32	12,0
Hipertensiva	89	33,5
Isquêmica	61	22,9
Inespecífica	84	31,6
TOTAL	266	100,0

Gráfico A.6 – Distribuição dos pacientes segundo a etiologia da IC.

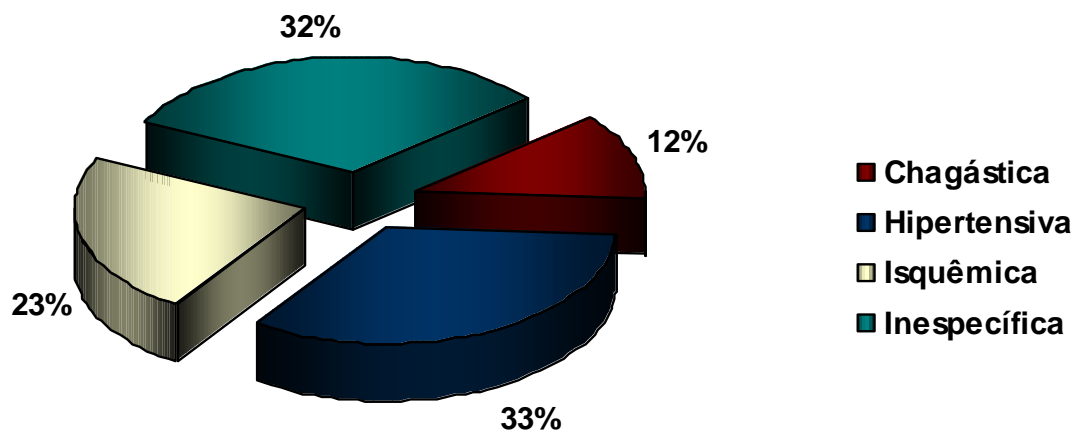


Tabela A.7 – Distribuição dos pacientes segundo a etiologia da IC e sexo.

Etiologia da IC	Masculino		Feminino		Total	
	Frequência	%	Frequência	%	Frequência	%
Chagástica	20	12,0	12	12,0	32	12,0
Isquêmica	44	26,5	17	17,0	61	22,9
Inespecífica	47	28,3	37	37,0	84	31,6
Hipertensiva	55	33,1	34	34,0	89	33,5
TOTAL	166	100,0	100	100,0	266	100,0

Gráfico A.7 – Distribuição dos pacientes segundo a etiologia da IC e sexo.

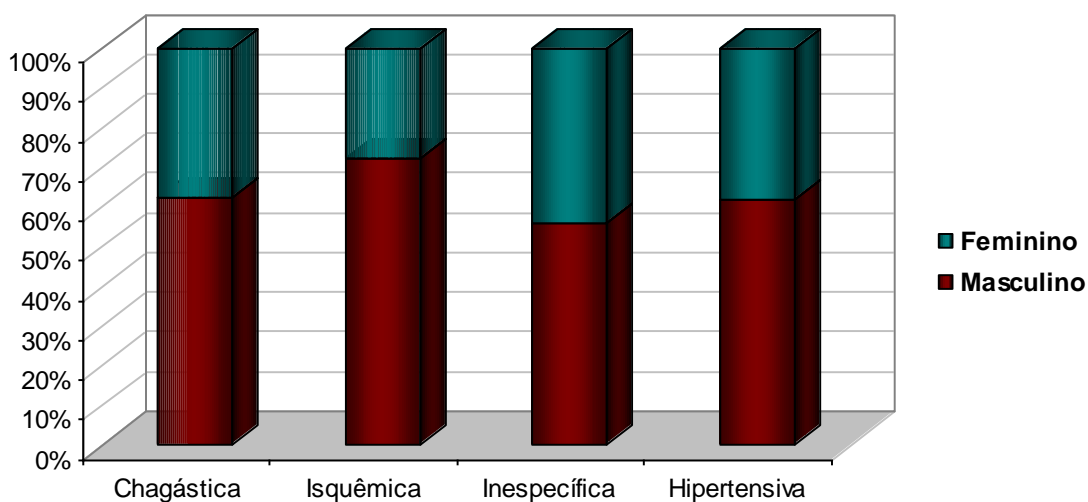


Tabela A.8 – Distribuição dos pacientes segundo o óbito, até três anos após 1ª consulta.

Situação	Frequência	%
Óbito	75	28,2
Não óbito	191	71,8
TOTAL	266	100,0

Gráfico A.8 – Distribuição dos pacientes segundo o óbito, até três anos após a primeira consulta.

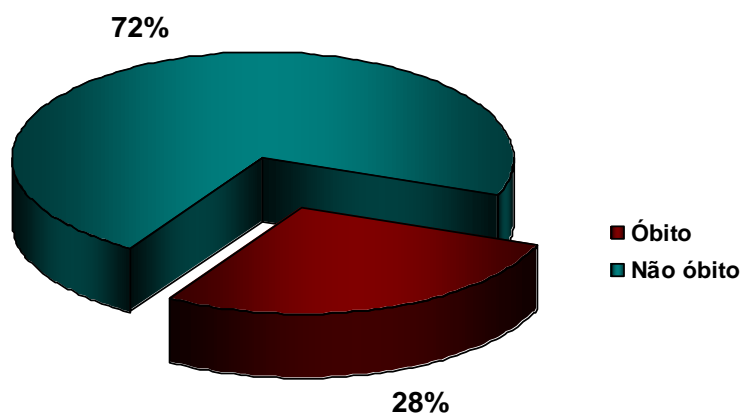


Tabela A.9 – Distribuição dos pacientes segundo a etiologia da IC e óbito.

Etiologia da IC	Óbito		Não óbito		Total	
	Frequência	%	Frequência	%	Frequência	%
Chagástica	15	46,9	17	53,1	32	100,0
Isquêmica	27	44,3	34	55,7	61	100,0
Inespecífica	21	25,0	63	75,0	84	100,0
Hipertensiva	12	13,5	77	86,5	89	100,0
TOTAL	75	28,2	191	71,8	266	100,0

Gráfico A.9 – Distribuição dos pacientes segundo a etiologia da IC e óbito.

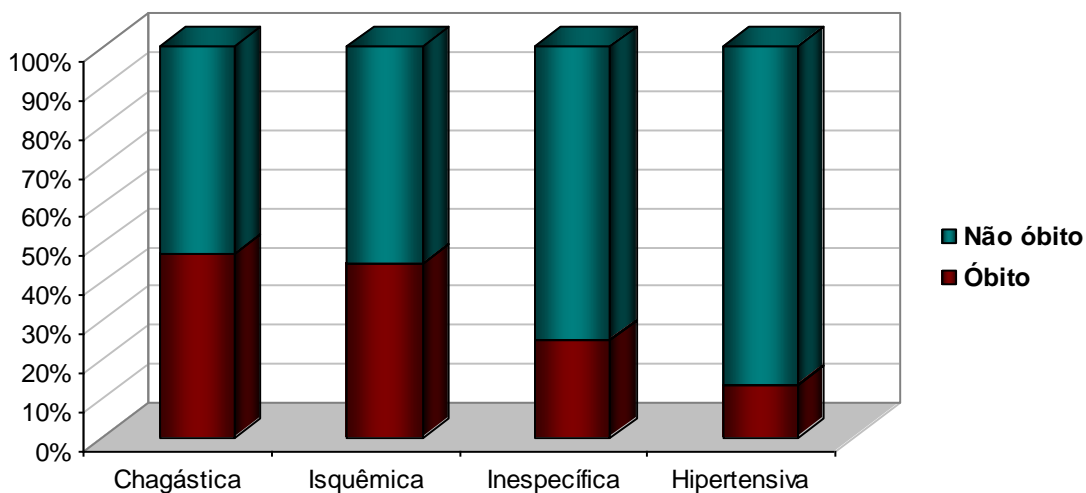
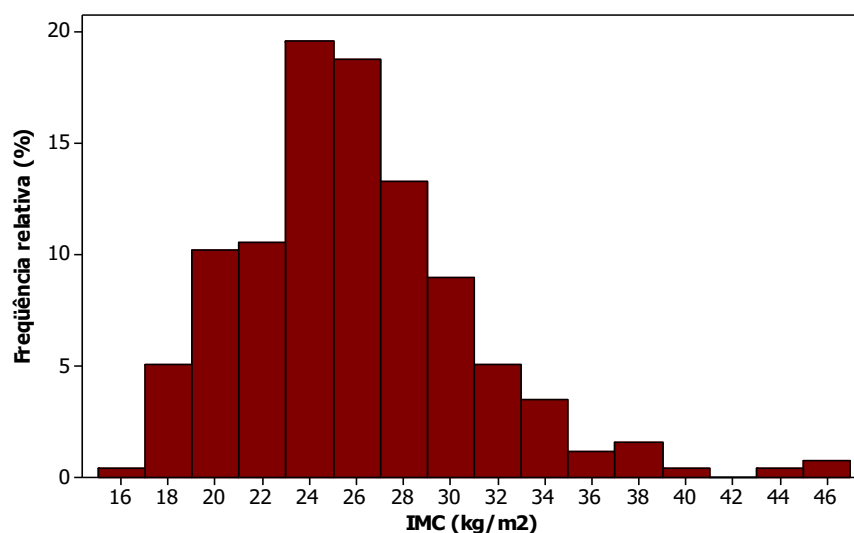


Tabela A.10 – Distribuição dos pacientes segundo a faixa do IMC.

Faixa IMC	IMC	Frequência	%
Abaixo do peso	< 18,5 kg/m ²	9	3,4
Peso normal	≥ 18,5 e < 24,5 kg/m ²	108	40,8
Sobrepeso	≥ 24,5 e < 30,0 kg/m ²	89	33,6
Obesidade Grau I	≥ 30,0 e < 35,0 kg/m ²	38	14,0
Obesidade Grau II	≥ 35,0 e < 40,0 kg/m ²	8	3,0
Obesidade Grau III	≥ 40,0 kg/m ²	3	1,0
Missing		11	4,2
TOTAL		266	100,0

Gráfico A.10 – Histograma de freqüência dos pacientes segundo o IMC.**Tabela A.11** – Distribuição dos pacientes segundo o tempo de sobrevida.

Sobrevida	Frequência	%
0 a 6 meses	17	22,4
6 a 18 meses	24	31,6
18 a 36 meses	28	36,8
Mais de 36 meses	7	9,2
TOTAL	76	100

Gráfico A.11 – Distribuição dos pacientes segundo o tempo de sobrevida.

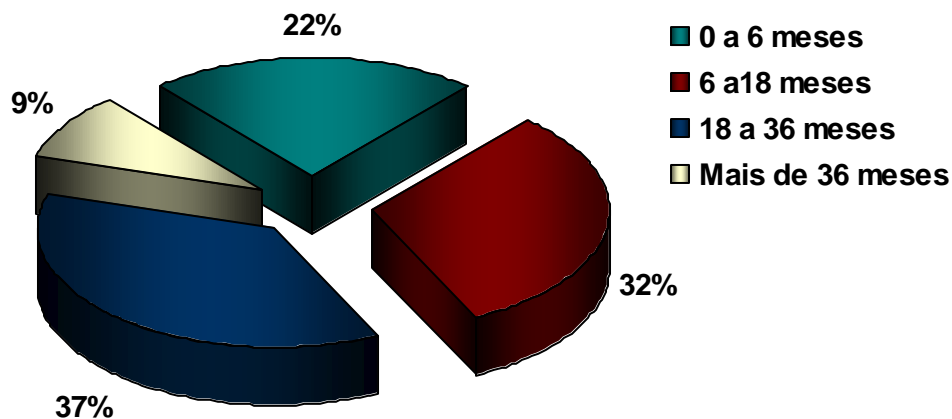


Tabela A.12 – Distribuição dos pacientes segundo a etiologia e tempo de sobrevida.

Sobrevida	Etiologia				Total
	Chagásica	Hipertensiva	Inespecífica	Isquêmica	
0 a 6 meses	5	1	7	4	17
6 a 18 meses	4	6	6	8	24
18 a 36 meses	7	3	7	11	28
Mais de 36 meses	0	2	1	4	7
TOTAL	16	12	21	27	76

Gráfico A.12 – Distribuição dos pacientes segundo a etiologia e tempo de sobrevida.

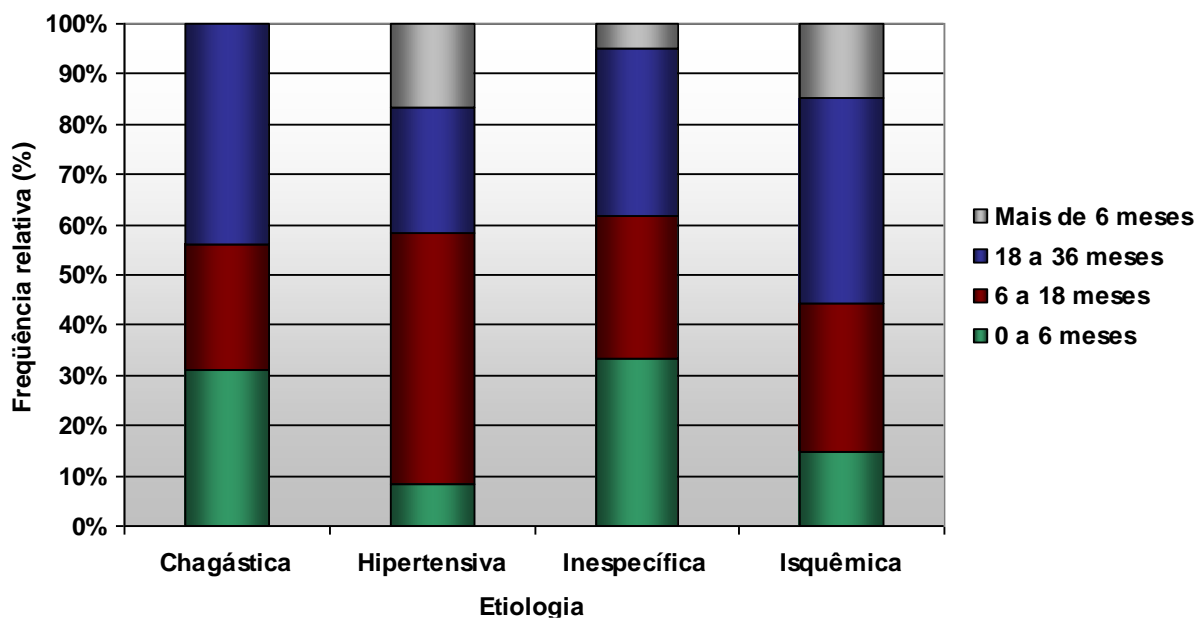
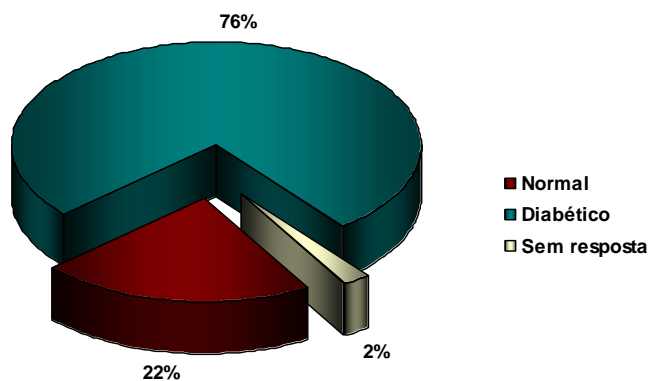


Tabela A.13 – Distribuição dos pacientes segundo diabetes.

Situação	Frequência	%
Normal	58	21,8
Diabético	202	75,9
Sem resposta	6	2,3
TOTAL	266	100,0

Gráfico A.13 – Distribuição dos pacientes segundo diabetes.**Tabela A.14** – Distribuição dos pacientes segundo o nível da pressão arterial.

Nível da pressão arterial	Frequência	%
Ideal	78	29,3
Normal	58	21,8
Hipertensão	116	43,6
Missing	14	5,3
TOTAL	266	100,0

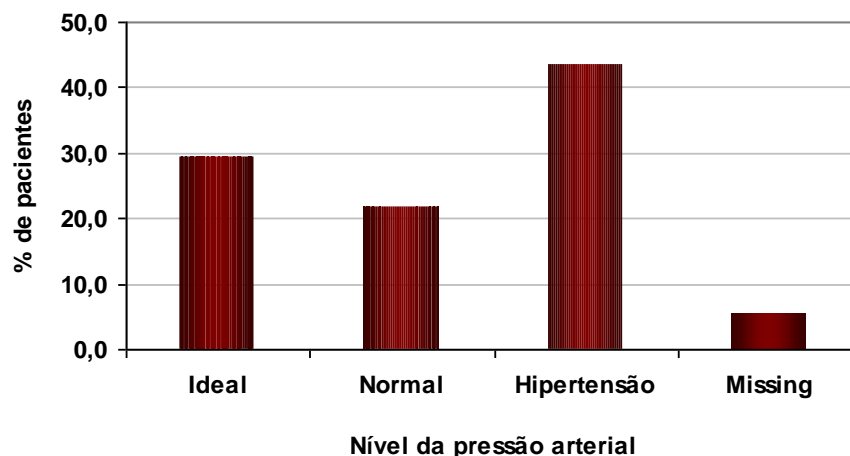
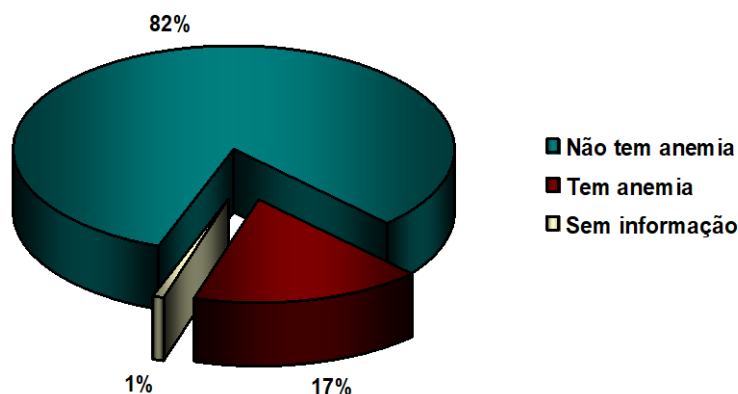
Gráfico A.14 – Distribuição dos pacientes segundo o nível da pressão arterial.

Tabela A.15 – Distribuição dos pacientes segundo diagnóstico de anemia.

Anemia	Frequência	%
Não tem anemia	219	82,3
Tem anemia	45	16,9
Sem informação	2	0,8
TOTAL	266	100,0

Gráfico A.15 – Distribuição dos pacientes segundo diagnóstico de anemia.**Tabela A.16** – Distribuição dos pacientes segundo diagnóstico de anemia e grau de dificuldade da IC.

Grau de dificuldade	Não tem anemia		Tem anemia		Total ⁽¹⁾	
	Frequência	%	Frequência	%	Frequência	%
Sem dificuldade	59	26,9	6	13,3	66	24,9
Com dificuldade condicionada	36	16,4	17	37,8	53	19,9
Com dificuldade, sem mais detalhes	23	10,5	1	2,2	24	9,0
Com dificuldade, às vezes	75	34,2	17	37,8	93	35,0
Dificuldade no passado	17	7,8	3	6,7	20	7,5
Dificuldade condicionada no passado	9	4,1	1	2,2	10	3,8
TOTAL	219	100,0	45	100,0	266	100,0

Nota: (1) Inclusive os pacientes sem informação sobre anemia.

Gráfico A.16 – Distribuição dos pacientes segundo diagnóstico de anemia e grau de dificuldade da IC.

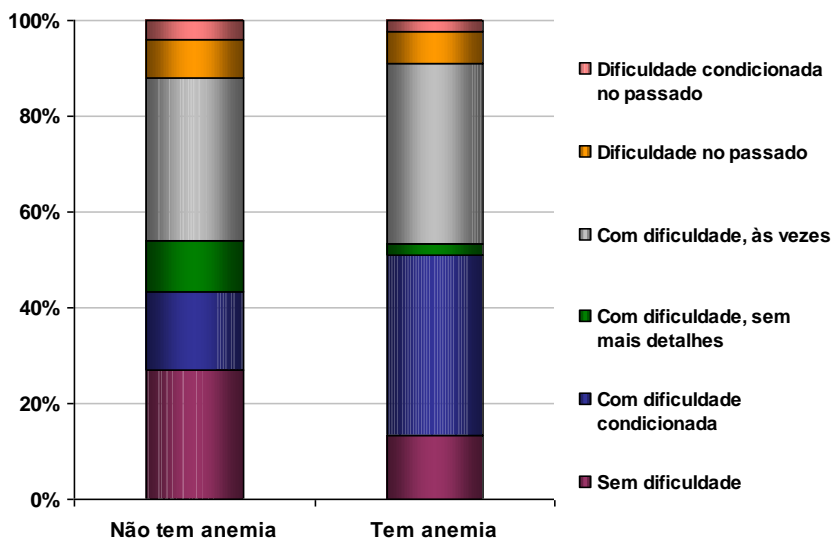
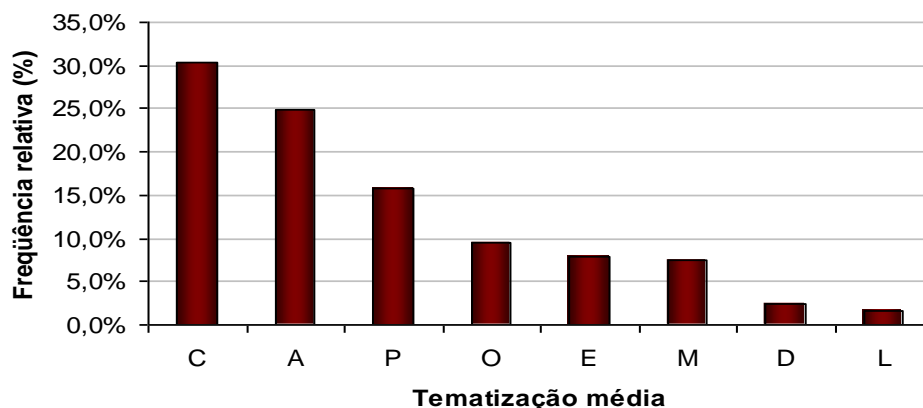


Tabela A.17 – Número e porcentagem de pacientes segundo a tematização média.

Tematização média	Frequência	%
C - Corpo (ou partes do corpo)	73	30,3
A - Atividades físicas	60	24,9
P - Fatores psicológicos	38	15,8
O - Posição (de dormir)	23	9,5
E - Fatores externos	19	7,9
M - Medicamentos	18	7,5
D - Doenças	6	2,5
L - Alimentação	4	1,7
TOTAL	241	100,0
Sem tematização média	144	

Gráfico A.17 – Porcentagem de pacientes segundo a tematização média.**Tabela A.18** – Distribuição dos pacientes segundo o grau de dificuldade para respirar.

Dificuldade	Frequência	%
Sem dificuldade	66	24,9
Com dificuldade condicionada	93	35,1
Com dificuldade, sem mais detalhes	53	20,0
Com dificuldade, às vezes	24	8,7
Dificuldade no passado	20	7,5
Dificuldade condicionada no passado	10	3,8
TOTAL	266	100,0

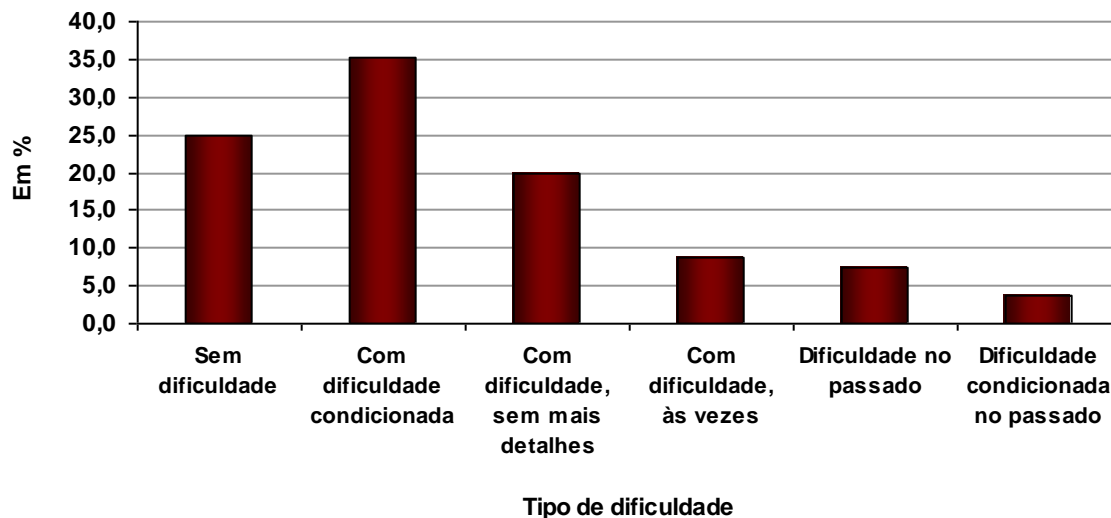
Gráfico A.18 – Distribuição dos pacientes segundo o tipo de dificuldade.

Tabela A.19 – Distribuição dos pacientes segundo a tematização média e o grau de dificuldade.

Grau de dificuldade	C		A		P		O	
	Frequência	%	Frequência	%	Frequência	%	Frequência	%
Sem dificuldade	1	50,0	1	50,0	0	0,0	0	0,0
Com dificuldade condicionada	25	48,1	7	13,5	10	19,2	3	5,8
Com dificuldade, sem mais detalhes	8	50,0	0	0,0	3	18,8	0	0,0
Com dificuldade, às vezes	35	23,5	46	30,9	21	14,1	17	11,4
Dificuldade no passado	2	18,2	2	18,2	3	27,3	0	0,0
Dificuldade condicionada no passado	2	18,2	4	36,4	1	9,1	3	27,3
TOTAL	73	30,3	60	24,9	38	15,8	23	9,5

Grau de dificuldade	M		E		L		D		TOTAL	
	Frequência	%	Frequência	%	Frequência	%	Frequência	%	Frequência	%
Sem dificuldade	0	0,0	0	0,0	0	0,0	0	0,0	2	100,0
Com dificuldade condicionada	3	5,8	1	1,9	2	3,8	1	1,9	52	100,0
Com dificuldade, sem mais detalhes	3	18,8	1	6,3	0	0,0	1	6,3	16	100,0
Com dificuldade, às vezes	9	6,0	15	10,1	4	2,7	2	1,3	149	100,0
Dificuldade no passado	4	36,4	0	0,0	0	0,0	0	0,0	11	100,0
Dificuldade condicionada no passado	0	0,0	1	9,1	0	0,0	0	0,0	11	100,0
TOTAL	19	7,9	18	7,5	6	2,5	4	1,7	241	100,0

Tabela A.20 – Distribuição dos pacientes segundo a etiologia por tematização média.

Etiologia	C		A		P		O	
	Frequência	%	Frequência	%	Frequência	%	Frequência	%
Chagástica	15	20,5	8	13,3	6	15,8	0	0,0
Hipertensiva	22	30,1	24	40,0	13	34,2	11	47,8
Inespecífica	22	30,1	18	30,0	13	34,2	6	26,1
Isquêmica	14	19,2	10	16,7	6	15,8	6	26,1
TOTAL	73	100,0	60	100,0	38	100,0	23	100,0

Etiologia	M		E		L		D	
	Freqüência	%	Freqüência	%	Freqüência	%	Freqüência	%
Chagástica	1	5,3	3	16,7	1	16,7	2	50,0
Hipertensiva	6	31,6	3	16,7	5	83,3	1	25,0
Inespecífica	6	31,6	4	22,2	0	0,0	1	25,0
Isquêmica	6	31,6	8	44,4	0	0,0	0	0,0
TOTAL	19	100,0	18	100,0	6	100,0	4	100,0

Gráfico A.20 – Distribuição dos pacientes segundo a etiologia por tematização média.

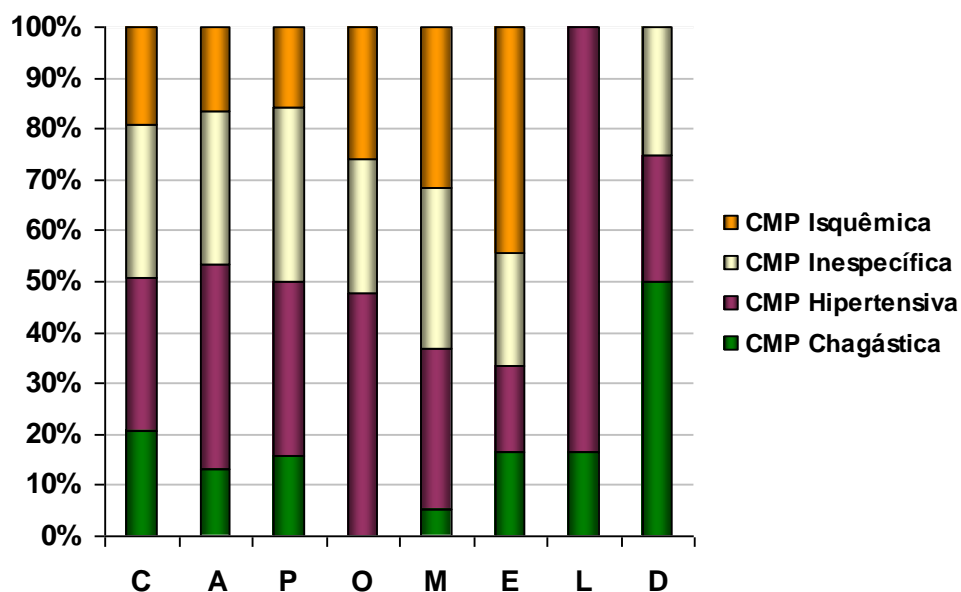


Tabela A.21 – Distribuição dos pacientes segundo a tematização média por etiologia.

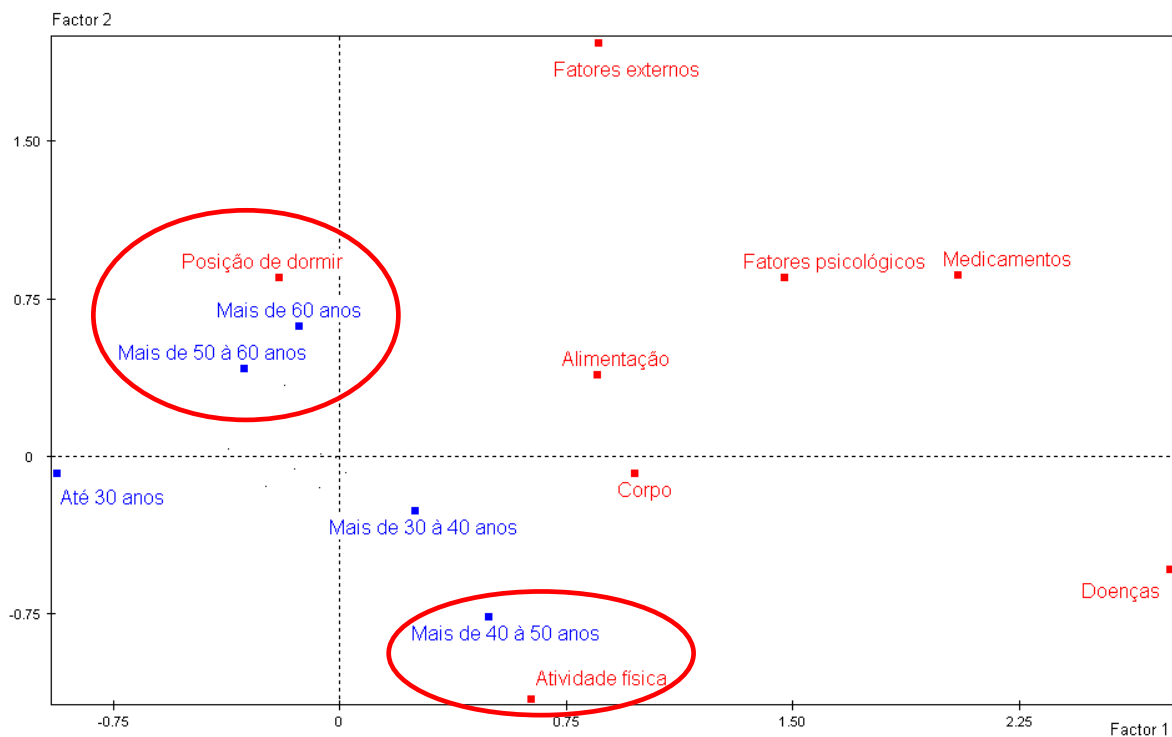
Etiologia	C		A		P		O	
	Freqüência	%	Freqüência	%	Freqüência	%	Freqüência	%
Chagástica	15	41,7	8	22,2	6	16,7	0	0,0
Hipertensiva	22	25,9	24	28,2	13	15,3	11	12,9
Inespecífica	22	31,4	18	25,7	13	18,6	6	8,6
Isquêmica	14	28,0	10	20,0	6	12,0	6	12,0
TOTAL	73	30,3	60	24,9	38	15,8	23	9,5

Etiologia	M		E		L		D		TOTAL	
	Freqüência	%	Freqüência	%	Freqüência	%	Freqüência	%	Freqüência	%
Chagástica	1	2,8	3	8,3	1	2,8	2	5,6	36	100,0
Hipertensiva	6	7,1	3	3,5	5	5,9	1	1,2	85	100,0
Inespecífica	6	8,6	4	5,7	0	0,0	1	1,4	70	100,0
Isquêmica	6	12,0	8	16,0	0	0,0	0	0,0	50	100,0
TOTAL	19	7,9	18	7,5	6	2,5	4	1,7	241	100,0

APÊNDICE B

Tabelas e Gráficos da Análise de Correspondência

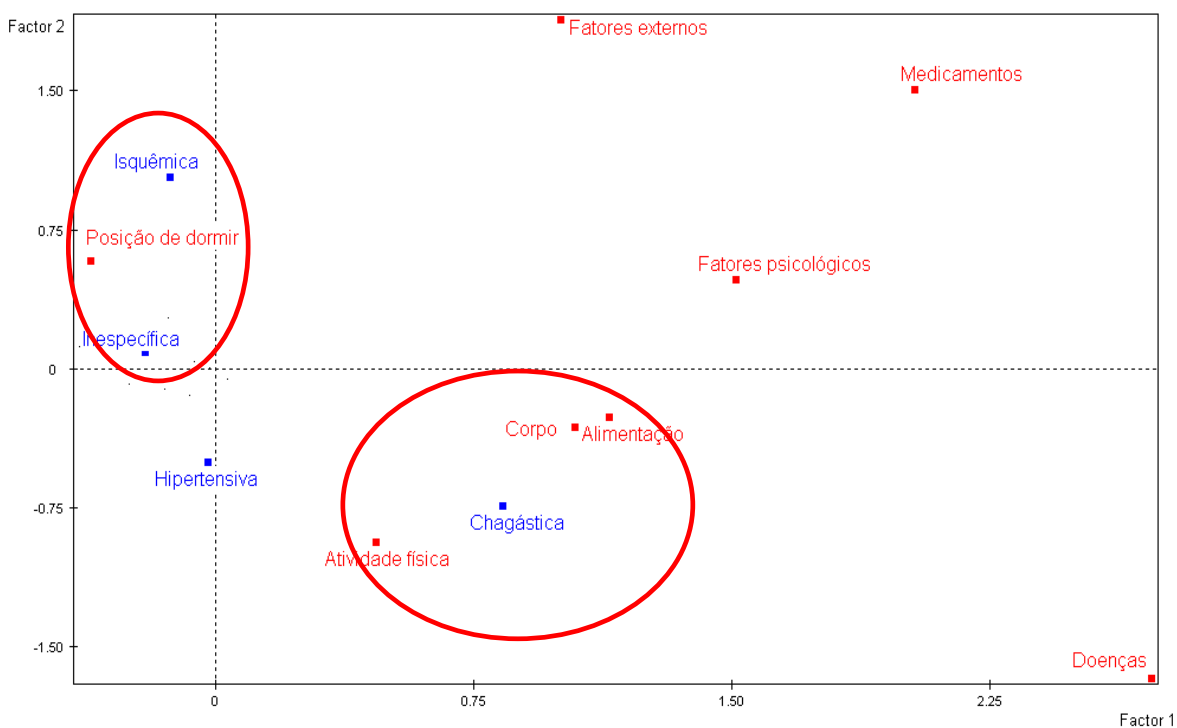
Gráfico B.1 – Análise de correspondência entre as variáveis “tematização média” e “faixa etária”.



Inércia Total: 1,3750.

Eixo	Autovalor	%	% Acumulado
1	0,1941	14,12	14,12
2	0,1585	11,53	25,64
3	0,1468	10,67	36,32
4	0,1321	9,61	45,93
5	0,1248	9,07	55,00
6	0,1195	8,69	63,69
7	0,1140	8,29	71,98
8	0,1119	8,14	80,12
9	0,0993	7,22	87,34
10	0,0885	6,44	93,78
11	0,0855	6,22	100,00

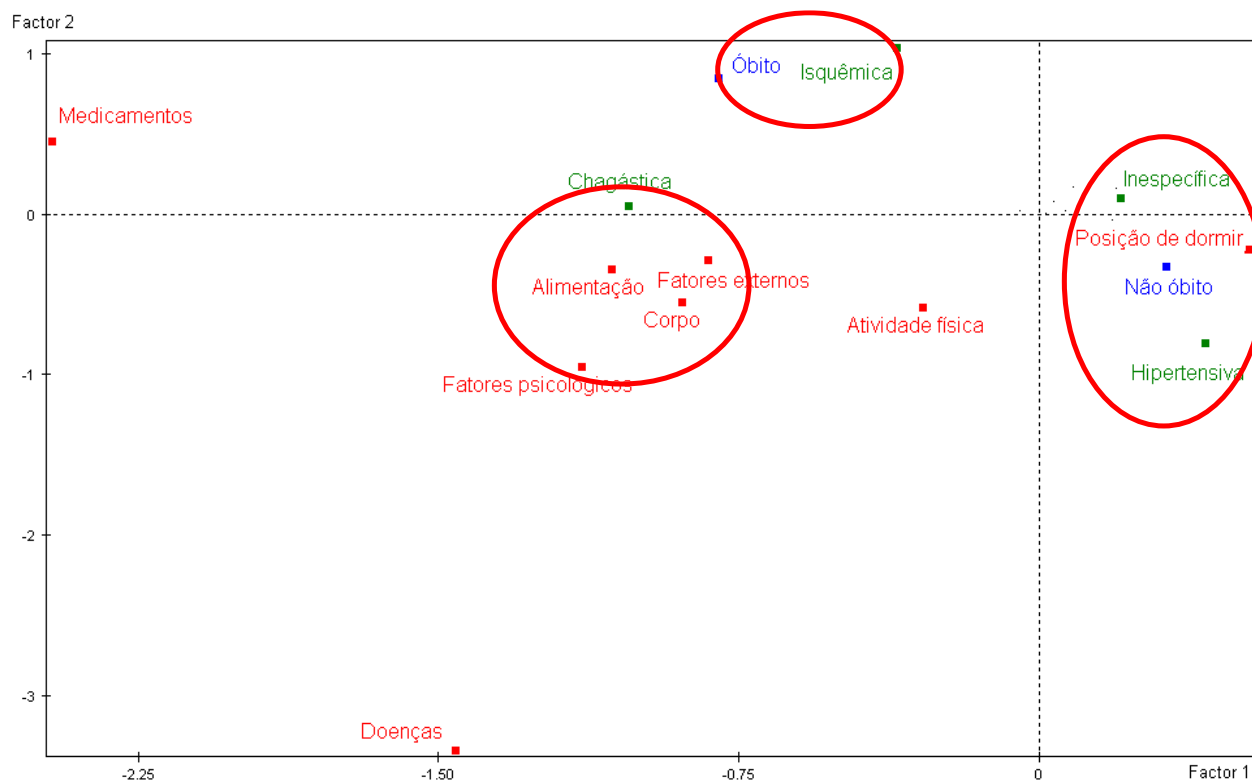
Gráfico B.2 – Análise de correspondência entre as variáveis “tematização média” e “etiologia”.



Inércia Total: 1,2500.

Eixo	Autovalor	%	% Acumulado
1	0,1911	15,29	15,29
2	0,1585	12,68	27,97
3	0,1473	11,78	39,75
4	0,1314	10,51	50,26
5	0,1292	10,33	60,59
6	0,1189	9,51	70,11
7	0,1073	8,59	78,69
8	0,0953	7,62	86,32
9	0,0878	7,02	93,34
10	0,0833	6,66	100,00

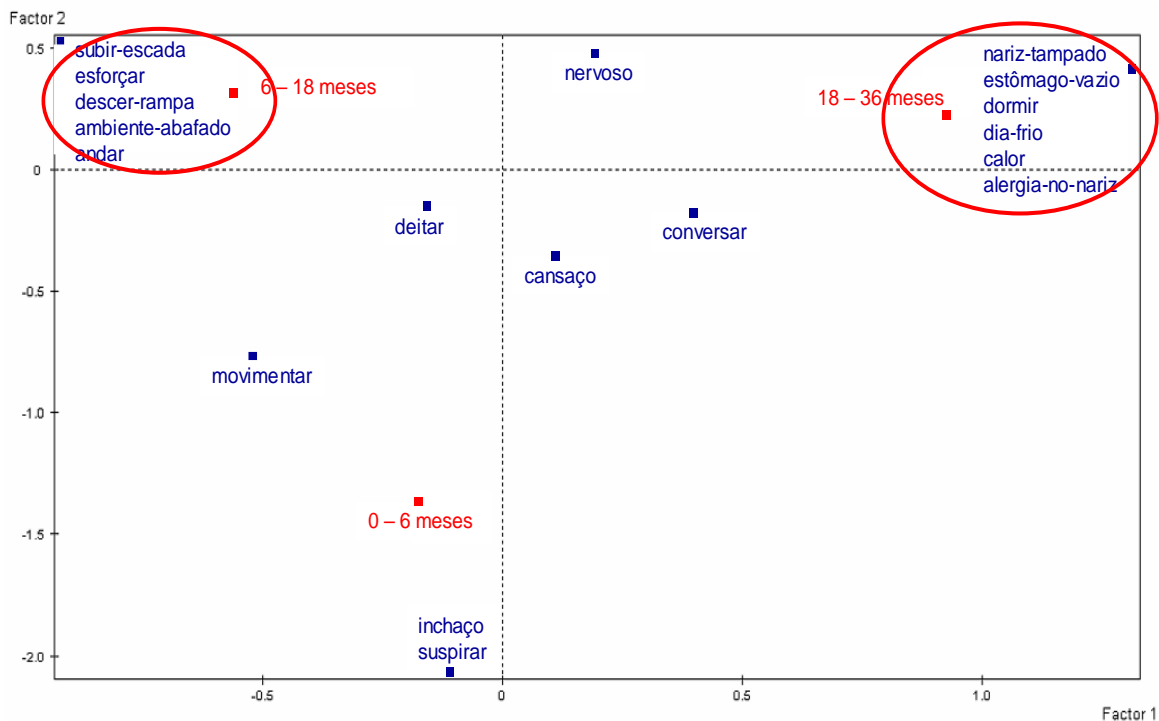
Gráfico B.3 – Análise de correspondência entre as variáveis “tematização média”, “etiologia” e “óbito”.



Inércia Total: 1,2222.

Eixo	Autovalor	%	% Acumulado
1	0,1771	14,49	14,49
2	0,1561	12,77	27,26
3	0,1381	11,30	38,56
4	0,1199	9,81	48,37
5	0,1153	9,43	57,80
6	0,1059	8,66	66,46
7	0,0954	7,81	74,27
8	0,0860	7,04	81,30
9	0,0841	6,88	88,19
10	0,0765	6,26	94,45
11	0,0679	5,55	100,00

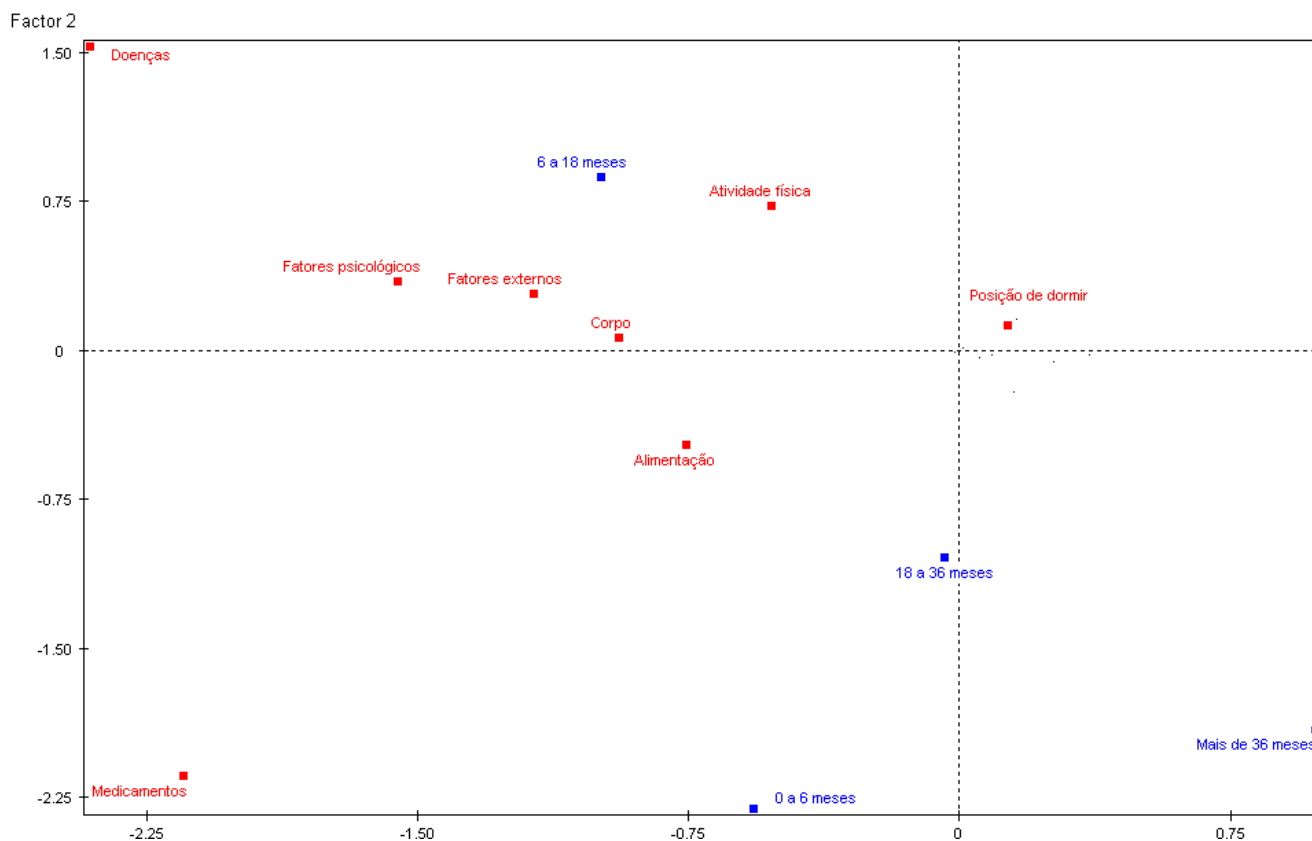
Gráfico B.4 – Análise de correspondência entre as variáveis “tempo de sobrevivida” e “grau de dificuldade”.



Inércia Total: 1,1785.

Eixo	Autovalor	%	% Acumulado
1	0,6186	52,49	52,49
2	0,5599	47,51	100,00

Gráfico B.5 – Análise de correspondência entre as variáveis “tempo de sobrevivência” e “tematização média”.



Inércia Total: 1,0000.

Eixo	Autovalor	%	% Acumulado
1	0,2136	21,36	21,36
2	0,1654	16,54	37,90
3	0,1482	14,82	52,72
4	0,1342	13,42	66,13
5	0,1282	12,82	78,96
6	0,1086	10,86	89,32
7	0,1018	10,18	100,00

APÊNDICE C

Tabelas e Gráficos da Análise Inferencial

Tabela C.1 – Agrupamento baseado na variável “Tematização média”.

Tematização média	Frequência	%
Sem tematização	116	43,6
Atividades físicas [a]	26	9,8
Corpo (ou partes do corpo) [c]	23	8,6
Posição de dormir [o]	13	4,9
Fatores externos [e]	9	3,4
Fatores psicológicos [p]	6	2,3
Medicamentos [m]	5	1,9
a,c	17	6,4
c,p	8	3,0
a,p	6	2,3
c,l	3	1,1
c,o	3	1,1
e,p	3	1,1
a,o	2	0,8
c,d	2	0,8
c,m	2	0,8
c,a,p	2	0,8
a,c,m	2	0,8
...
TOTAL DE PACIENTES	266	100,0

Gráfico C.1 – Gráfico de silhuetas para k = 2, 3, 4 e 5 grupos.

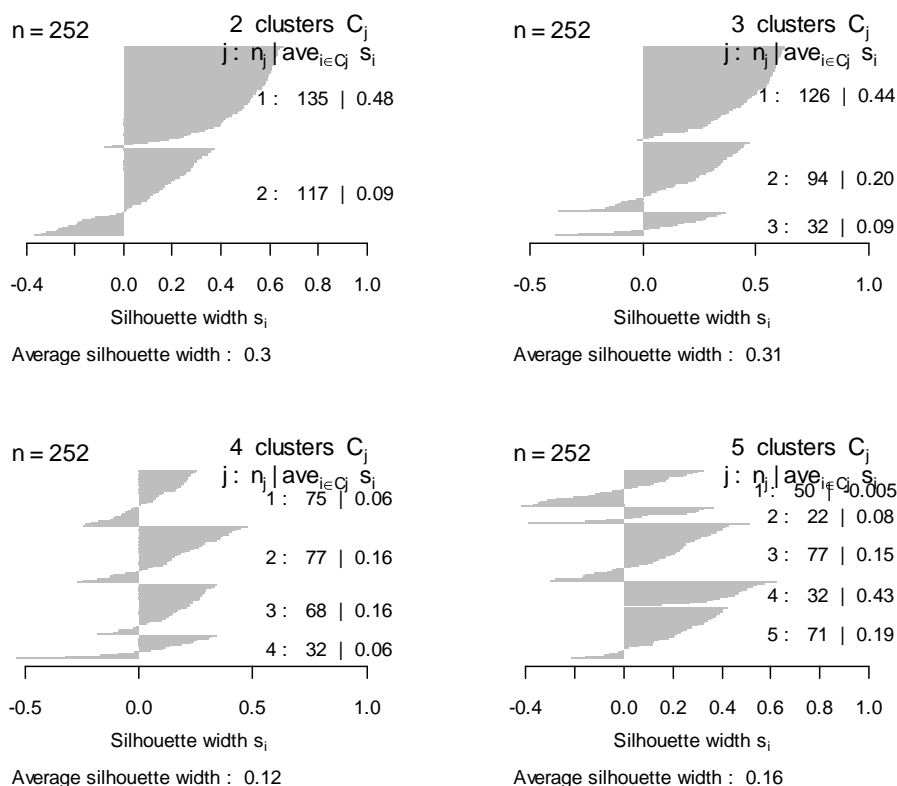
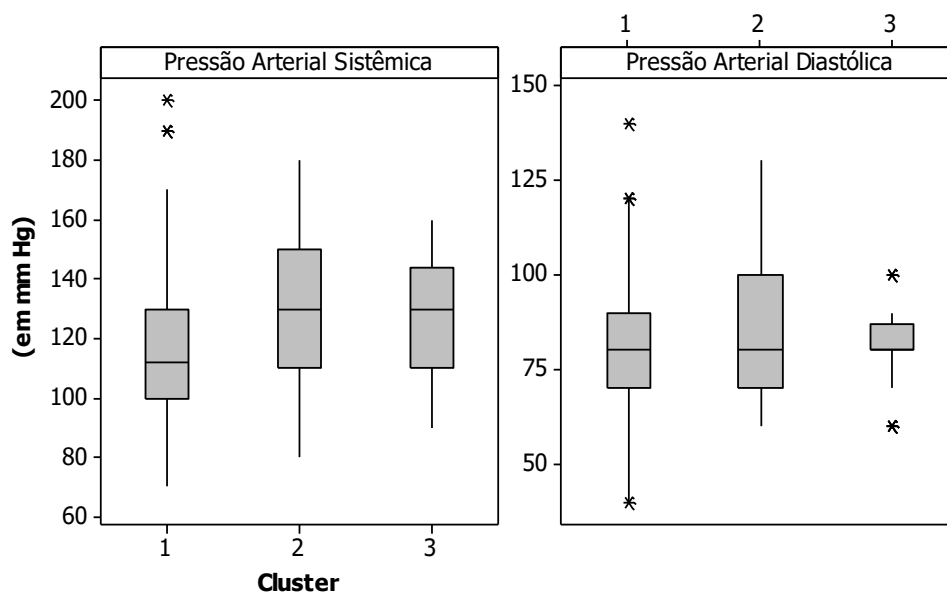


Gráfico C.2 – Boxplot da Pressão Arterial Sistêmica e Diastólica vs Cluster.**Tabela C.2.1 – Medidas descritivas da pressão sistêmica vs grupos.**

Medida	Grupo 1	Grupo 2	Grupo 3
N	126	94	32
Média	119,1	129,9	128,8
Desvio padrão	24,5	24,6	18,6
Coefficiente de variação	20,5	18,9	14,4
Mínimo	70,0	80,0	90,0
Q1	100,0	110,0	110,0
Mediana	112,0	130,0	130,0
Q3	130,0	150,0	144,0
Máximo	200,0	180,0	160,0

Tabela C.2.2 – Medidas descritivas da pressão diastólica vs grupos.

Medida	Grupo 1	Grupo 2	Grupo 3
N	126	94	32
Média	79,2	84,3	80,3
Desvio padrão	15,6	16,5	9,9
Coefficiente de variação	19,7	19,5	12,3
Mínimo	40,0	60,0	60,0
Q1	70,0	70,0	80,0
Mediana	80,0	80,0	80,0
Q3	90,0	100,0	87,0
Máximo	140,0	130,0	100,0

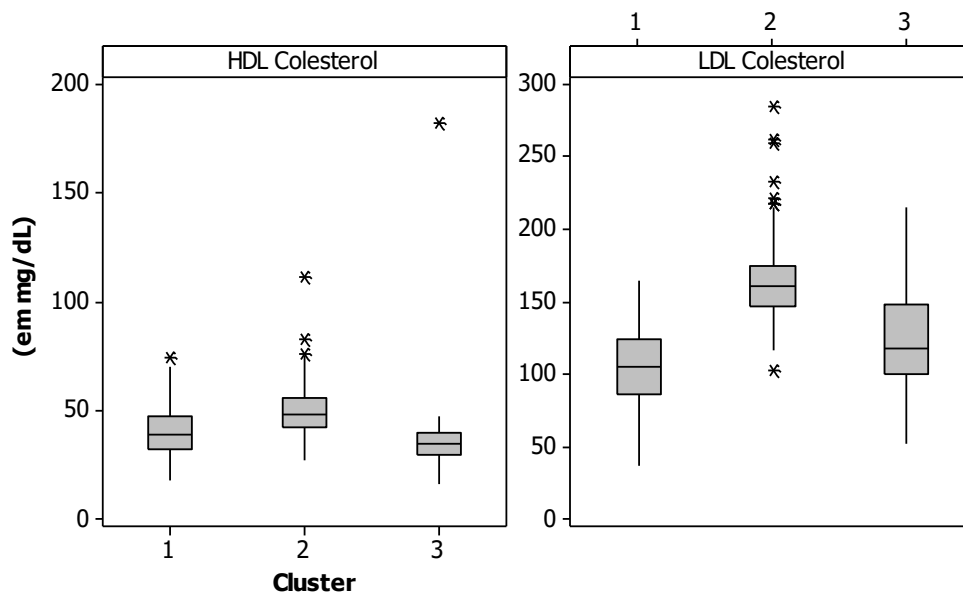
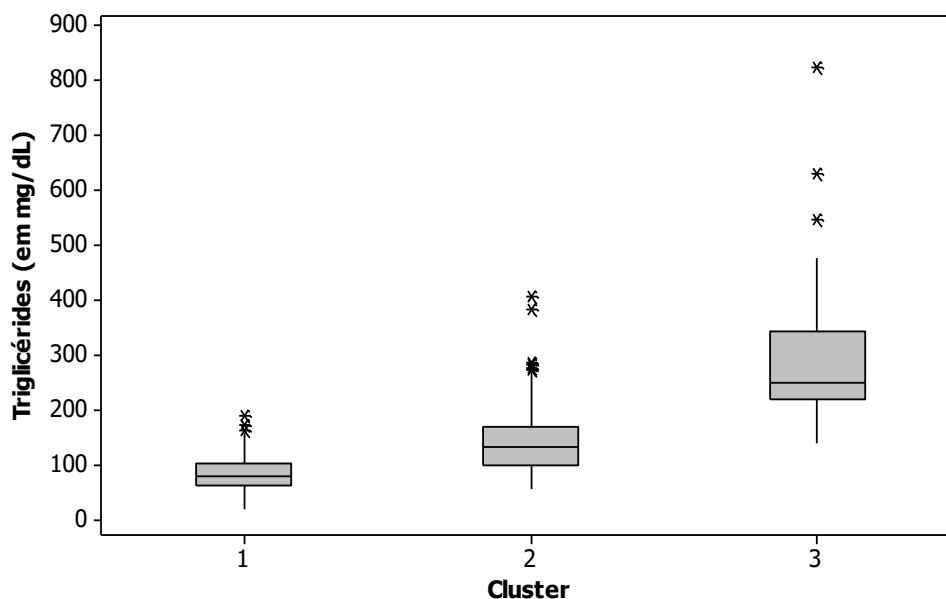
Gráfico C.3 – Boxplot do HDL Colesterol e LDL Colesterol vs grupos.

Tabela C.3.1 – Medidas descritivas do HDL Colesterol vs grupos.

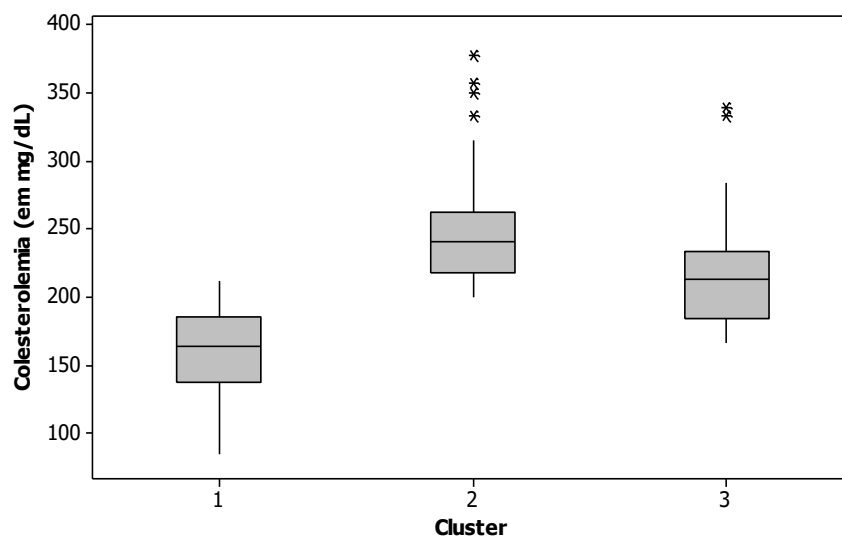
Medida	Grupo 1	Grupo 2	Grupo 3
N	126	94	32
Média	40,1	49,7	39,0
Desvio padrão	11,1	12,8	27,0
Coeficiente de variação	27,8	25,6	69,3
Mínimo	18,0	27,0	16,0
Q1	32,0	42,0	29,3
Mediana	39,0	48,0	35,0
Q3	47,3	55,5	39,8
Máximo	74,0	111,0	182,0

Tabela C.3.2 – Medidas descritivas do LDL Colesterol vs grupos.

Medida	Grupo 1	Grupo 2	Grupo 3
N	126	94	32
Média	103,0	165,3	124,8
Desvio padrão	26,2	31,2	31,8
Coeficiente de variação	25,5	18,9	25,5
Mínimo	37,0	102,0	52,0
Q1	85,8	146,5	100,0
Mediana	104,5	160,5	117,0
Q3	124,0	174,5	148,0
Máximo	165,0	284,0	215,0

Gráfico C.4 – Boxplot da taxa de Triglicérides vs grupos.**Tabela C.4 – Medidas descritivas da taxa de Triglicérides vs grupos.**

Medida	Grupo 1	Grupo 2	Grupo 3
N	126	94	32
Média	85,2	146,0	300,5
Desvio padrão	32,5	32,5	144,1
Coeficiente de variação	38,1	22,2	48,0
Mínimo	20,0	57,0	139,0
Q1	61,0	99,0	218,5
Mediana	78,5	131,5	250,0
Q3	101,0	168,3	343,5
Máximo	190,0	405,0	821,0

Gráfico C.5 – Boxplot da taxa de Colesterolemia vs grupos.**Tabela C.5** – Medidas descritivas da taxa de Colesterolemia vs grupos.

Medida	Grupo 1	Grupo 2	Grupo 3
N	126	94	32
Média	160,2	244,9	219,0
Desvio padrão	29,6	35,0	44,3
Coefficiente de variação	18,5	14,3	20,2
Mínimo	84,0	200,0	166,0
Q1	136,8	218,0	184,5
Mediana	164,0	240,0	213,0
Q3	185,0	261,5	233,5
Máximo	212,0	377,0	339,0

Tabela C.6 – Tematizações médias vs grupos.

Tematização média	Grupo 1 (126 pacientes)		Grupo 2 (94 pacientes)		Grupo 3 (32 pacientes)	
	Frequência	%	Frequência	%	Frequência	%
Atividades físicas [a]	26	35,1	24	42,9	7	53,8
Corpo (ou partes do corpo) [c]	34	45,9	25	44,6	8	61,5
Doenças [d]	2	2,7	1	1,8	2	15,4
Fatores externos [e]	8	10,8	7	12,5	3	23,1
Alimentação [l]	3	4,1	0	0,0	1	7,7
Medicamentos [m]	8	10,8	7	12,5	1	7,7
Posição de dormir [o]	13	17,6	7	12,5	2	15,4
Fatores psicológicos [p]	17	23,0	13	23,2	4	30,8
Total de pacientes com tematização	74		56		13	