

Mery Natali Silva Abreu

**USO DE MODELOS DE REGRESSÃO LOGÍSTICA ORDINAL EM  
EPIDEMIOLOGIA: UM EXEMPLO USANDO A QUALIDADE DE VIDA**

Universidade Federal de Minas Gerais  
Programa de Pós-Graduação em Saúde Pública  
Belo Horizonte – MG  
2007

Mery Natali Silva Abreu

**USO DE MODELOS DE REGRESSÃO LOGÍSTICA ORDINAL EM  
EPIDEMIOLOGIA: UM EXEMPLO USANDO A QUALIDADE DE VIDA**

Dissertação apresentada ao Programa de Pós-graduação em Saúde Pública da Universidade Federal de Minas Gerais, como requisito parcial para Obtenção do título de Mestre em Saúde Pública (área de concentração em Epidemiologia)

Orientadora: Armindia Lucia Siqueira

Co-orientadora: Waleska Teixeira Caiaffa

Belo Horizonte – MG  
2007

A162u Abreu, Mery Natali Silva  
Uso de modelos de regressão logística ordinal em epidemiologia:  
um exemplo usando a qualidade de vida/Mery Natali Silva Abreu.  
Belo Horizonte, 2007.  
179f.  
Dissertação.(mestrado) - Universidade Federal de Minas Gerais.  
Faculdade de Medicina.  
Área de concentração: Epidemiologia  
Orientadora: Arminda Lucia Siqueira  
Co-orientadora: Waleska Teixeira Caiaffa  
1.Epidemiologia 2.Modelos estatísticos 3.Análise de regressão  
4.Qualidade de vida 5.Indicadores de qualidade de vida 6.Saúde  
pública/estatística & dados numéricos I.Título

NLM: WA 105  
CDU: 614.2-036.22

**UNIVERSIDADE FEDERAL DE MINAS GERAIS****Reitor**

Prof. Ronaldo Tadêu Pena

**Vice-Reitora**

Prof<sup>a</sup>. Heloisa Maria Murgel Starling

**Pró-Reitor de Pós-Graduação**

Prof. Jaime Arturo Ramirez

**Pró-Reitor de Pesquisa**

Prof. Carlos Alberto Pereira Tavares

**FACULDADE DE MEDICINA****Diretor**

Prof. Francisco José Penna

**Chefe do Departamento de Medicina Preventiva e Social**

Prof<sup>a</sup>. Maria da Conceição Juste Werneck Cortes

**PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE PÚBLICA****Coordenador**

Prof. Mark Drew Crosland Guimarães

**Sub-Coordenadora**

Prof<sup>a</sup>. Sandhi Maria Barreto

**Colegiado**

Prof<sup>a</sup>. Ada Ávila Assunção

Prof<sup>a</sup>. Elizabeth Barboza França

Prof. Fernando Augusto Proietti

Prof. Francisco de Assis Acurcio

Prof<sup>a</sup>. Maria Fernanda F. L. Costa

Prof<sup>a</sup>. Mariângela Leal Cherchiglia

Prof<sup>a</sup>. Soraya Almeida Belisário

Prof<sup>a</sup>. Sandhi Maria Barreto

Prof. Tarcísio Márcio Magalhães Pinheiro

Prof<sup>a</sup>. Waleska Teixeira Caiaffa

Cristiane Amorim Andrade (Representante Discente Titular - Doutorado)

Aline Dayrell Ferreira (Representante Discente Suplente - Doutorado)

## **Declaração de defesa**

**À minha mãe, sempre presente nos momentos felizes e difíceis dando seu apoio e incentivo, que foram essenciais para que eu chegasse até aqui.**

## **Agradecimentos**

Às professoras Waleska e Arminda pela orientação, disponibilidade e apoio, fundamentais para o desenvolvimento do trabalho.

À Clareci pelo constante apoio e incentivo, além de disponibilizar banco de dados utilizado no trabalho referente à qualidade de vida na esquizofrenia.

À Bruna Oliveira por disponibilizar o banco de dados sobre qualidade de vida de pacientes portadores de marca passo.

Aos professores do Programa de Pós-Graduação em Saúde Pública pelas contribuições para minha formação.

Aos funcionários do Centro de Pós-Graduação e do Programa de Pós-Graduação em Saúde Pública pelo respeito e atenção.

Ao Grupo de Pesquisa em Epidemiologia pela estrutura de trabalho.

Aos colegas do GPW (“Grupo de Estudos dos Orientandos da Waleska”) pelas contribuições para a dissertação.

Aos meus pais, irmão, familiares e amigos, que estavam sempre presentes nas horas difíceis, me apoiando, incentivando com carinho e amizade.

A Deus por mais essa conquista.

**“O correr da vida embrulha tudo, a vida é assim:  
esquenta e esfria, aperta e daí afrouxa,  
sossega e depois desinquieta.  
O que ela quer da gente é coragem.”**

**Guimarães Rosa**

## Resumo

O tema qualidade de vida (QV) tem ganhado ênfase nos últimos anos e vem sendo utilizado na linguagem cotidiana e no contexto da pesquisa científica. Uma das formas de mensurar a QV é por meio de instrumentos desenvolvidos por especialistas da área e os resultados são mensurados por meio de escalas ordinais. Esse tipo de variável consiste em uma série de categorias com uma determinada ordenação. Para a análise dos dados ordinais, são necessários métodos estatísticos específicos, como os modelos de regressão ordinal, ainda pouco divulgados e utilizados. Tais modelos possibilitam estimar o risco de ocorrência de um evento adverso, por exemplo, uma baixa qualidade de vida. Entretanto, os modelos ordinais requerem alguns cuidados na análise dos dados, são cercados de pressupostos, e nem todos os tipos estão implementados nos pacotes estatísticos mais utilizados na saúde pública. O objetivo do trabalho é comparar e apresentar recomendações quanto ao uso e implementação computacional dos modelos de regressão ordinal, em particular dos modelos de chances proporcionais, de chances proporcionais parciais, de razão-contínua e estereótipo. Os resultados mostraram que os modelos de regressão ordinal investigados são apropriados para analisar os dados de QV, sendo que cada tipo de modelo investigado apresentou melhor adequação dependendo da forma como a variável resposta é apresentada. Além disso, uma implementação computacional adequada e o domínio dos comandos para execução dos modelos ordinais é essencial para que esses modelos sejam comparados. Gráficos de diagnóstico foram ferramentas importantes para a verificação do ajuste dos modelos. Em se tratando de estudos cuja variável resposta é ordinal, recomenda-se que sejam evitados procedimentos tais como dicotomizar e desconsiderar a ordenação, que podem resultar em perda da informação proporcionada pelos dados. Por fim, a escolha do melhor modelo depende do caráter da variável ordinal, adequação do modelo às suposições, qualidade do ajuste e reduzido número de parâmetros estimados.

Palavras-chave: qualidade de vida, dados ordinais, saúde pública, regressão logística ordinal.

## **Abstract**

In the last years, the emphasis on quality of life has been increasing and this concept has been used in the daily language and the context of the scientific research. The quality of life is often measured by means of instruments developed for specialists of the area and the results are measured by ordinal scales. This type of ordinal variable consists of a series of categories with one determined order. For the analysis of the ordinal data, specific statistical methods are necessary, such as the ordinal regression models, still little divulged and used. Such models make possible to estimate the risk of occurrence of an adverse event, for instance, the low quality of life. However, the ordinal models require some attention in the analysis of the data, they are surrounded of assumptions, and nor all the types are implemented in statistical packages more used in the public health. The objective of the work is compare and present recommendations to using and implementing of the models of ordinal regression, in particular the proportional odds model, the partial proportional odds model, the continuation-ratio model and the stereotype model. The results had shown that the ordinal regression models are appropriate to analyze the quality of life data, and each type of investigated model presented better adequacy depending on the form as the outcome is presented. Moreover, an appropriate computational implementation and the domain of the commands for execution of the ordinal models are essential so that these models are compared. Residual plots had been important tools for goodness of fit of the models. When the response variable is ordinal, one should avoid procedures such as dichotomizing and ignoring the order, that can result in loss of the proportionate information for the data. Finally, the choice of the best model depends on the character of the ordinal variable, adequacy of the assumptions model, goodness-of-fit and reduced number of estimated parameters.

**Key words:** quality of life; ordinal data; public health; ordinal logistic regression.

## SUMÁRIO

1. CONSIDERAÇÕES INICIAIS.....	12
2. OBJETIVOS.....	15
3. ARTIGO 1: Análise de estudos sobre qualidade de vida: aplicação dos modelos de regressão logística ordinal .....	16
INTRODUÇÃO.....	21
PLANEJAMENTO DE ESTUDOS SOBRE QUALIDADE DE VIDA .....	23
ANÁLISE DE DADOS DE QUALIDADE DE VIDA.....	24
Análise descritiva.....	24
Análise de tabelas de contingência.....	25
Análise multivariada: regressão logística ordinal.....	25
Modelo de Chances Proporcionais.....	26
Modelo de Chances Proporcionais Parciais.....	27
Modelo de Razão Contínua.....	28
Modelo Estereótipo.....	29
Implementação Computacional.....	30
EXEMPLOS DE APLICAÇÃO.....	31
Exemplo 1.....	32
Exemplo 2.....	34
Cálculo do Poder .....	35
DISCUSSÃO.....	36
REFERÊNCIAS BIBLIOGRÁFICAS.....	39

4. ARTIGO 2: Regressão logística ordinal em estudos clínicos e epidemiológicos.....	46
INTRODUÇÃO.....	51
ANÁLISE UNIVARIADA .....	53
MODELOS DE REGRESSÃO ORDINAL.....	53
Modelo de Chances Proporcionais.....	54
Modelo de Chances Proporcionais Parciais.....	54
Modelo de Razão Contínua.....	55
Modelo Estereótipo.....	56
VERIFICAÇÃO DA QUALIDADE DO AJUSTE DOS MODELOS ORDINAIS....	56
IMPLEMENTAÇÃO COMPUTACIONAL.....	59
EXEMPLOS DE APLICAÇÃO.....	63
Exemplo 1 – Estudo sobre Portadores de Marcapasso.....	63
Exemplo 2 – Levantamento de Nacional de Saúde e Nutrição .....	65
DISCUSSÃO.....	67
REFERÊNCIAS BIBLIOGRÁFICAS.....	70
5. CONSIDERAÇÕES FINAIS.....	82
APÊNDICES.....	84
APÊNDICE A: PROJETO DE PESQUISA .....	85
APÊNDICE B: RESULTADOS DO ARTIGO 1.....	109
APÊNDICE C: RESULTADOS DO ARTIGO 2.....	128
ANEXOS.....	140
ANEXO A: APROVAÇÃO DO DEPARTAMENTO PARA OS BANCOS DE DADOS UTILIZADOS.....	141
ANEXO B: APROVAÇÃO DO COMITÊ DE ÉTICA PARA OS BANCOS DE DADOS UTILIZADOS.....	144
ANEXO C: CERTIFICADO DE QUALIFICAÇÃO.....	147
ANEXO D: ESCALA QLS-BR.....	149
ANEXO E: ESCALA AQUAREL.....	170
ANEXO F: ESCALA SF-36.....	172

## 1. CONSIDERAÇÕES INICIAIS

O tema qualidade de vida (QV) tem ganhado ênfase nos últimos anos e vem sendo utilizado na linguagem cotidiana e no contexto da pesquisa científica.<sup>1</sup>

No passado, o conceito era delegado apenas a filósofos e poetas. Porém, atualmente existe um interesse crescente de pesquisadores da área de saúde nessa definição.<sup>2</sup>

Patrick e Erickson<sup>3</sup> disseram em 1999 que a qualidade de vida é o valor atribuído à duração da vida, modificado pelos prejuízos, estados funcionais e oportunidades sociais que são influenciados por doença, dano, tratamento ou políticas de saúde.

Fica claro que esse conceito é coerente com as mudanças na área de saúde, já que com o aumento das doenças crônicas e degenerativas, a pesquisa médica tem enfatizado os estudos sobre QV considerando a importância não só de uma vida longa, mas uma vida com qualidade. Assim, em diversas áreas da Medicina vem crescendo o interesse de transformar a QV numa medida quantitativa que possa ser comparada entre diferentes populações e até mesmo patologias<sup>1</sup>

Geralmente, o resultado das escalas de QV é mensurado em escalas ordinais. Esse tipo de variável consiste em uma série de categorias com uma determinada ordenação. Além das escalas de qualidade de vida, vários outros eventos de interesse para a área de saúde, são mensurados por meio de instrumentos que resultam em variáveis ordinais, como por exemplo a percepção do indivíduo sobre sua condição de saúde, ou até mesmo diagnósticos médicos sobre gravidade de doenças.

Um exemplo no qual o interesse principal é a investigação da qualidade de vida relacionada à saúde é o estudo de fatores associados com a QV realizado por Cardoso et al<sup>4,5</sup> (2005) em

---

<sup>1</sup> Seidl EMF, Zannon CMLC. Qualidade de vida e saúde: aspectos conceituais e metodológicos. *Cadernos de Saúde Pública*, 2004, 20(2):580-588.

<sup>2</sup> Ciconelli R et al. Tradução para a língua portuguesa e validação do questionário genérico de avaliação de qualidade de vida SF-36. *Revista Brasileira de Reumatologia*, 1999, 39: 143–150.

<sup>3</sup> Ebrahim S. Clinical and public health perspectives and applications of health-related quality of life measurement. *Social Science Medical*, 1995, 41: 1383-94

uma casuística de 273 pacientes com diagnóstico de esquizofrenia originados de dois centros de referência em saúde mental de Belo Horizonte. Nesse estudo, a QV enquanto variável resposta foi mensurada por meio de uma escala com característica ordinal, sendo que os maiores escores refletiam melhor QV. As variáveis independentes eram de natureza clínica e sócio-demográficas.

Participei do estudo de qualidade de vida mencionado como bolsista de iniciação científica durante dois anos, trabalhando com a análise dos dados. O objetivo principal da análise foi a busca dos fatores associados à qualidade de vida na esquizofrenia. Nesse estudo, foi observada a dificuldade de se analisar esse tipo de variável resposta, dada a complexidade da análise estatística que deveria ser empregada, além da escassez da literatura sobre o assunto. Isso serviu com uma motivação para a realização do presente trabalho, como forma de ilustrar melhor a utilização dos modelos para análise de dados ordinais. Uma revisão dessas ferramentas de análise pode facilitar a execução de futuros trabalhos sobre o tema qualidade de vida, que está em constante expansão.

É importante lembrar, que vários autores destacam que as escalas de QV tendem a uma distribuição discreta, assimétrica e limitada. Por isso, métodos de análise usuais como test-t e regressão linear que assumem normalidade podem não ser apropriados. É importante considerar o caráter ordinal original que essas variáveis têm. No entanto, apesar do caráter ordinal das medidas, raramente os dados são analisados como tal. Frequentemente elas são dicotomizadas, tratadas como nominais ou como contínuas. Porém, esse procedimento pode ser inadequado e conduzir a análises estatísticas errôneas.<sup>6</sup>

Assim, um método de análise que vem sendo desenvolvido e apresentado na literatura estatística é a regressão logística ordinal, apropriado para análise de fatores associados quando a variável resposta tem caráter ordinal. Esse tipo de análise proporciona, ainda, o cálculo de

---

<sup>4</sup> Cardoso CS, Caiaffa WT, Bandeira M, Siqueira AL, Abreu MNS, Fonseca JOP. Factores associated with low quality of life in schizophrenia. *Cadernos de Saúde Pública* 2005; 21: 1338-1348.

<sup>5</sup> Cardoso CS, Caiaffa WT, Bandeira M, Siqueira AL, Abreu MNS, Fonseca JOP. Qualidade de vida e dimensão ocupacional na esquizofrenia: uma comparação por sexo. *Cadernos de. Saúde Pública* 2006; 22: 1303-1314.

<sup>6</sup> Lall R, Campbell MJ, Walters SJ, Morgan K. A review of ordinal regression models applied on health-related quality of life assessments. *Statistical Methods in Medical Research*. 2002;11(1):49-67.

risco de ocorrência de um evento adverso, por exemplo, a chance de uma baixa qualidade de vida, que freqüentemente interessa aos epidemiologistas.<sup>7</sup>

O tipo de modelo de regressão ordinal depende da forma como a variável ordinal se apresenta. Em um processo de análise, é possível testar uma variedade desses modelos e então selecionar o apropriado para os dados em estudo. Entretanto, isso aumenta a complexidade da análise e torna a diminuir sua utilização na literatura epidemiológica e biomédica.

Frente a essa discussão, propõem-se comparar o uso de modelos de regressão ordinal, em particular do modelo de chances proporcionais, modelo de razão-contínua, modelo estereótipo e modelo de chances proporcionais parciais, para se avaliar fatores associados quando se utiliza escalas ordinais, utilizando exemplos sobre qualidade de vida.

Optou-se por apresentar o trabalho na forma de artigos conforme regulamento desse programa de pós-graduação.<sup>8</sup> O primeiro artigo, que foi submetido aos Cadernos de Saúde Pública, aborda o planejamento e análise de estudos sobre qualidade de vida, fazendo uma revisão dos principais modelos de regressão logística ordinal, exemplificando a sua aplicação como forma de análise multivariada num estudo sobre fatores associados à qualidade de vida em pacientes com diagnóstico de esquizofrenia. O segundo artigo, que foi submetido à Revista de Saúde Pública, discute a utilização dos modelos de regressão logística ordinal em estudos clínicos e epidemiológicos, considerando a verificação da qualidade do ajuste desses modelos, além de sua implementação computacional em *softwares* como R e STATA.

É importante destacar que algumas informações não puderam ser acrescentadas aos dois artigos, por uma questão de espaço e foram disponibilizadas nos apêndices e anexos, tais como: as estatísticas descritivas e análises univariadas referentes aos bancos de dados utilizados em ambos os artigos; as saídas do programa para todas as análises e as escalas de qualidade de vida consideradas em cada estudo utilizado como exemplo.

---

<sup>7</sup> Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. *Journal Clinical Epidemiological* 1997; 50(1):45-55

<sup>8</sup> Manual de Orientação 2007. Programa de Pós-graduação em Saúde Pública, Departamento de Medicina Preventiva e Social, Faculdade de Medicina - UMG

## **2. OBJETIVOS**

### **2.1. Objetivo geral**

Considerando os diferentes métodos de análise de dados ordinais, o objetivo geral desse estudo é comparar e apresentar recomendações quanto ao uso dos principais modelos de regressão logística ordinal (*modelo de chances proporcionais, modelo de razão-contínua, modelo estereótipo e modelo de chances proporcionais parciais*) como forma de análise multivariada em estudos de qualidade de vida utilizando escala ordinal.

### **2.2. Objetivos específicos**

- a. Investigar o problema de dimensionamento de amostra e/ou cálculo do poder para os modelos de dados ordinais;
- b. Verificar a adequação e ajuste de cada modelo para investigação dos fatores associados a QV considerando bancos de dados secundários;
- c. Verificar a adequação e ajuste de cada modelo considerando diferentes escalas de QV;
- d. Definir o melhor modelo para ajustar os dados em cada situação;
- e. Ilustrar a utilização dos modelos ordinais por meio dos *softwares* R ou STATA.

### **3. ARTIGO 1:**

**Análise de estudos sobre qualidade de vida: aplicação dos modelos de regressão logística ordinal**



**ANÁLISE DE ESTUDOS SOBRE QUALIDADE DE VIDA: APLICAÇÃO DOS  
MODELOS DE REGRESSÃO LOGÍSTICA ORDINAL**

*ANALYSING QUALITY OF LIFE STUDIES: APLICATION OF ORDINAL LOGISTIC  
REGRESSION MODELS*

---

**Mery Natali Silva Abreu<sup>a,b</sup>, Arminda Lucia Siqueira<sup>a,c</sup>, Clareci Silva Cardoso<sup>b</sup>, Waleska  
Teixeira Caiaffa<sup>a,b</sup>**

<sup>a</sup> Departamento de Medicina Preventiva e Social, Programa de Pós-graduação em Saúde Pública,  
Universidade Federal de Minas Gerais – UFMG.

<sup>b</sup> Grupo de Pesquisa em Epidemiologia e Observatório de Saúde Urbana - Universidade Federal de  
Minas Gerais – UFMG.

<sup>c</sup> Departamento de Estatística, Universidade Federal de Minas Gerais – UFMG.

Correspondência: Mery Natali Silva Abreu (A/ C Waleska Teixeira Caiaffa)

Av. Alfredo Balena, 190, 8º andar, sala 8013, Santa Efigênia.

CEP: 31130-100. Belo Horizonte – MG. Telefone/fax: (31) 3248-9949.

E-mail: [merynatali@yahoo.com.br](mailto:merynatali@yahoo.com.br)

## RESUMO

O tema qualidade de vida (QV) tem ganhado ênfase nos últimos anos. Tipicamente, os resultados da QV são mensurados por meio de escalas de caráter ordinal. Procedimentos como a dicotomização da variável resposta e a desconsideração da ordenação geram perda de informação proporcionada pelos dados ou até inferências incorretas. Para análise de dados ordinais, métodos estatísticos específicos são necessários, tais como os modelos de regressão logística ordinal. Precauções na utilização desses modelos devem ser tomadas, devido às premissas por eles exigidas. A proposta deste trabalho é apresentar uma revisão de modelos de regressão logística ordinal, em particular o modelo de chances proporcionais, o modelo de razão-contínua, o modelo estereótipo e duas versões do modelo de chances proporcionais parciais. O ajuste, a inferência estatística e a comparação dos modelos são ilustrados com dados de um estudo transversal sobre qualidade de vida que utilizou a escala QLS-BR, realizado com 273 pacientes com diagnóstico de esquizofrenia. Os resultados indicam que todos os modelos testados mostraram um bom ajuste, mas o modelo de chances proporcionais e o modelo de chances proporcionais parciais foram os mais adequados pelo caráter dos dados utilizados e pela facilidade da interpretação dos resultados.

Palavras-chave: dados ordinais; qualidade de vida; regressão logística ordinal.

## Abstract

The subject quality of life has been increasing emphasis in the last years. Typically, the results of the quality of life are measured by means of ordinal scales. In these situations, specific statistical methods are necessary and procedures as the dichotomization of the response variable, the loss of proportionate information for the data has as consequence, being, therefore not recommended. The models of ordinal logistic regression are appropriate in many situations. However, precautions in the use of these models must be taken, had to the premises for demanded them.. The objective of this work is to present a revision of the models of ordinal logistic regression, in particular of the model the proportional odds model, the partial proportional odds model, the continuation-ratio model and the stereotype model. The adjustment and the comparison of the models are illustrated with data of a transversal study on quality of life, with 273 patients with schizophrenia diagnosis. The results indicate that all the tested models had shown a good adjustment, but the proportional odds model or partial proportional odds model provided a better goodness-of-fit because the character of the used data and the easiness of the interpretation of the results.

Key words: ordinal data; odds ratio (OR); quality of life; ordinal logistic regression.

## INTRODUÇÃO

O interesse pelo tema Qualidade de Vida (QV) tem crescido nos últimos anos, mas ainda é cercado de controvérsias. Falta clareza e consistência quanto ao significado do termo, à mensuração e ainda quanto ao processo de análise de dados. Atualmente, a Organização Mundial de Saúde adotou uma concepção bastante ampla, definindo qualidade de vida como sendo *a percepção do indivíduo a respeito de sua posição na vida dentro do contexto dos valores, da cultura na qual ele vive, e em relação a seus objetivos, expectativas, padrões e preocupações*.<sup>21</sup>

Os novos conceitos de QV são consoantes com as mudanças de paradigmas, que têm influenciado as políticas e as práticas de saúde nas últimas décadas. Além disso, o perfil de morbi-mortalidade indica um aumento da prevalência das doenças crônico-degenerativas e os avanços nos tratamentos têm acarretado aumento também na sobrevivência das pessoas acometidas por esses agravos. Isso faz com que o impacto dessas doenças e de seus tratamentos sejam avaliados em termos de sua influência na QV.<sup>19</sup>

Devido à percepção de que a QV é um fator importante para o estado de saúde, médicos e pesquisadores têm tentado transformá-la numa medida quantitativa que possa ser comparada entre diferentes populações e até mesmo entre diferentes patologias.<sup>9</sup>

Nas duas últimas décadas surgiram vários instrumentos de mensuração da QV, específicos ou genéricos, além do crescente interesse pelo processo de adaptação e validação transcultural. Este acentuado crescimento pela temática mostra os esforços voltados para o amadurecimento conceitual e metodológico de investigações envolvendo a QV. Neste contexto, surgem ainda algumas questões, tais como: Como a QV deve ser medida ou avaliada? Como deveria ser o desenho de estudos sobre QV? Como investigar os fatores associados com uma melhor qualidade de vida dos pacientes?

Em geral a avaliação da QV é feita por meio de questionários desenvolvidos por especialistas da área. São formuladas perguntas que abordam aspectos específicos da vida do paciente e os resultados são mensurados principalmente por meio de escalas ordinais, que consistem em uma série de categorias com determinada ordenação.<sup>15</sup>

Por exemplo, a escala QLS-BR (Quality of Life in Schizophrenia)<sup>5-8</sup>, possui a seguinte pergunta para avaliar o nível de atividade social do paciente: “Você costuma sair com outras pessoas para se divertir?”. As possíveis respostas são cotadas em uma escala tipo likert em sete pontos e correspondem às seguintes opções: nunca, ocasionalmente, às vezes,

freqüentemente. O maior escore indica uma melhor QV. Esse resultado corresponde a uma variável ordinal que possui uma dimensão simples.

Nessa escala há um total de 21 itens e o escore final é definido como a média desses itens, variando entre 0 e 6. Esses escores são divididos em três categorias de uma nova escala ordinal: QV muito comprometida (0-2), QV comprometida (2-5) e QV adequada (5-6).<sup>8,9</sup> Logo, as categorias do resultado final da escala QLS-BR são relacionadas a um *continuum* subjacente, que é o escore variando de 0 a 6 e a variável ordinal pode ser considerada uma variável contínua com dados agrupados.

Existem alguns problemas no que diz respeito à medida de avaliação da qualidade de vida. As escalas de QV tendem a gerar uma distribuição discreta, assimétrica e limitada. Normalmente, essas escalas são tratadas como contínuas devido ao extenso número de categorias, como é o caso da escala SF-36 (*Medical Outcomes Study 36 Item Short Form Healthy Survey*)<sup>9</sup>, cujos escores variam de 0 a 100 (100 indica “ótima” qualidade de vida). Entretanto, métodos tradicionais de análise como test-t e regressão linear que assumem normalidade, pelo menos aproximadamente, podem não ser apropriados, já que se trata de uma distribuição assimétrica. Além disso, para a escala SF-36, por exemplo, os valores finalizam no escore 100 e muitas vezes se concentram nesse valor, o que caracteriza a assimetria dos dados. Portanto, é importante considerar o caráter ordinal original que essas variáveis apresentam.<sup>22</sup>

Considerando a natureza ordinal das escalas utilizadas para avaliar a qualidade de vida e a importância dos estudos sobre esse tema, apresentaremos uma revisão da metodologia para determinação do tamanho da amostra e análise de fatores associados à QV por meio dos modelos de regressão logística ordinal.

## PLANEJAMENTO DE ESTUDOS SOBRE QUALIDADE DE VIDA

A realização de um estudo sobre QV deve ser precedida por um bom planejamento que inclui a escolha dos instrumentos e das variáveis, além do cálculo adequado do tamanho da amostra. Este último aspecto é um passo essencial para que se tenha um poder aceitável para a detecção de diferenças ou efeitos na variável resposta para um nível de significância fixado.<sup>22</sup>

Antes da escolha da fórmula a ser utilizada no dimensionamento de amostra, deve-se definir a medida que sumariza a finalidade principal do estudo, na qual o cálculo deve ser baseado. Em estudos com delineamento do tipo caso-controle, ou mesmo em estudos transversais nos quais a prevalência do evento de interesse é baixa, utiliza-se como medida de risco a razão de chances, tradução para *odds ratio* (OR). Whitehead<sup>24</sup> (1993) sugere a utilização da razão de chances como medida sumário, não só para dados de resposta binária mas também quando se trabalha com dados ordinais.

### Razão de chances (OR) para dados ordinais

Suponha que a resposta de interesse ( $Y$ ) sobre QV tenha  $k$  categorias ordenadas ( $Y_j$  com  $j=1,2,\dots,k$ ) e que dois grupos (A e B) devem ser comparados. Para a categoria  $j$ , OR é dada por:

$$OR_j = \frac{\frac{P(Y \leq Y_j | x^{(A)})}{1 - P(Y \leq Y_j | x^{(A)})}}{\frac{P(Y \leq Y_j | x^{(B)})}{1 - P(Y \leq Y_j | x^{(B)})}} = \frac{\frac{P(Y \leq Y_j | x^{(A)})}{P(Y > Y_j | x^{(A)})}}{\frac{P(Y \leq Y_j | x^{(B)})}{P(Y > Y_j | x^{(B)})}} = \frac{odds^{(A)}}{odds^{(B)}} \quad (1)$$

Como a definição usual, OR é a razão entre duas chances (*odds*), mas agora a chance é definida em termos de probabilidades cumulativas. Para sua interpretação, basta pensar que a resposta tenha sido dicotomizada, sendo que o evento é ser classificado até a categoria  $j$ .

Se A e B representam respectivamente a exposição ou não a um fator de risco, a razão de chances quantifica a chance de um indivíduo do grupo exposto ser classificado até uma determinada categoria comparada com a chance do grupo não exposto.

No contexto dos dados ordinais, de acordo com a suposição de **chances proporcionais**, OR é a mesma para todas as categorias da variável resposta.<sup>24</sup>

### **Cálculo do tamanho da amostra para dados ordinais**

Whitehead<sup>24</sup> (1993) propõe um método não-paramétrico baseado na suposição de OR constante, que foi simplificado por Walters et al<sup>22</sup> (2001), resultando na fórmula (2) para o cálculo do número de sujeitos por grupo para nível de significância  $\alpha$  e poder de  $(1 - \beta)100\%$ .

$$n = \frac{6 \times [(z_{1-\alpha/2} + z_{1-\beta})^2 / (\log OR)^2]}{\left(1 - \sum_{j=1}^k \bar{\pi}_j^3\right)} \quad (2)$$

Em (2), OR é dada pela expressão (1) e  $\bar{\pi}_j$  é a proporção média de sujeitos na categoria  $j$  dos dois grupos comparados (A e B), isto é,  $\bar{\pi}_j = (\pi_{Aj} + \pi_{Bj}) / 2$ .

## **ANÁLISE DE DADOS DE QUALIDADE DE VIDA**

### **Análise descritiva**

Escalas de qualidade de vida com um extenso número de categorias, como as já mencionadas escalas SF-36 e QLS-BR, devem ser sumarizadas através da mediana ao invés da média, pois freqüentemente apresentam distribuição assimétrica dos dados. Deve-se utilizar os percentis para descrever a variabilidade dos dados em substituição ao desvio-padrão.<sup>22</sup>

Outra opção é reagrupar os escores originais das escalas de QV em categorias ordenadas, como é o caso da escala QLS-BR. Dessa forma, há uma redução no número de categorias e podem-se conduzir análises descritivas por meio de distribuição de freqüências, além de tabelas de contingência obtidas pelo cruzamento entre a resposta (QV) e possíveis fatores de interesse.

### **Análise de tabelas de contingência**

O teste qui-quadrado de tendência pode ser uma ferramenta importante quando o objetivo é a comparação entre grupos independentes, cuja variável resposta é ordinal e possui menos de sete categorias.<sup>22</sup>

O teste não paramétrico de Kruskal-Wallis (KW) também pode ser utilizado para avaliar tabelas de contingência em que uma das variáveis possui ordenação, como o caso em que a resposta é uma escala ordinal de QV.

### **Análise multivariada: regressão logística ordinal**

Os testes citados acima analisam a influência de apenas um único fator em relação à variável resposta. Quando é necessário controlar possíveis fatores de confusão, a análise multivariada especial para dados ordinais é a alternativa natural. Existem vários enfoques, tais como uso de modelos mistos ou outra classe de modelos, por exemplo o probito, mas são os modelos de regressão logística ordinal que vêm sendo amplamente divulgados na literatura estatística.<sup>1-4;10;12-18;22</sup>

Consideremos a variável resposta  $Y$  (score de QV) com  $k$  categorias codificadas em  $1, 2, \dots, k$  e  $\underline{x} = (x_1, x_2, \dots, x_p)$  o vetor de variáveis explicativas (covariáveis). As  $k$  categorias de  $Y$  condicionalmente aos valores das covariáveis ocorrem com probabilidades  $p_1, p_2, \dots, p_k$ , isto é,  $p_j = \Pr(Y = j | \underline{x})$  para  $j=1, 2, \dots, k$ . Na modelagem de dados de resposta ordinal podem ser utilizadas as probabilidades individuais  $p_j$  ou as probabilidades acumuladas  $(p_1 + p_2)$ ,  $(p_1 + p_2 + p_3)$ , ...,  $(p_1 + p_2 + p_3 + \dots + p_k)$ . No primeiro caso, a probabilidade de cada categoria é comparada com a probabilidade de uma categoria de referência, ou cada categoria com a categoria anterior, como no modelo de categorias adjacentes. Neste trabalho serão apresentados modelos logísticos com probabilidades acumuladas.

A Tabela 1 apresenta um resumo dos principais modelos de regressão logística para variável resposta com ou sem ordenação, com suas respectivas equações e indicações de uso. A seguir destacamos alguns pontos considerados importantes dos seguintes modelos de regressão logística ordinal: modelo de chances proporcional (MCP), duas versões do modelo de chances proporcionais parciais, sem restrição (MCPN) e com restrição (MCPN-R), modelo de razão-contínua (MRC) e modelo estereótipo (ME).

## Modelos de regressão logística ordinal

### A) Modelo de chances proporcionais (MCP)

O MCP, em inglês *proportional odds model*, também chamado de modelo do logito cumulativo (*cumulative logit model*), produz estimativas de simples compreensão. Seu uso é indicado quando a variável resposta era originalmente uma variável contínua que posteriormente foi agrupada.<sup>1;15</sup>

Como mostrado na Tabela 1, esse modelo compara a probabilidade de uma resposta igual ou menor a uma determinada categoria ( $j = 1, 2, \dots, k-1$ ), com probabilidade de uma resposta maior que esta categoria. Além disso, o modelo é composto por  $k-1$  equações lineares paralelas. No caso particular de apenas duas categorias ( $k=2$ ), o MCP corresponde exatamente ao tradicional modelo de regressão logística binária (ver MB na Tabela 1).

O modelo tem  $(k-1+p)$  parâmetros. O intercepto  $\alpha_j$  do modelo varia para cada uma das equações e satisfaz a condição  $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{k-1}$ ; existem ainda  $p$  coeficientes betas ( $\beta$ ) cujos elementos correspondem aos efeitos das covariáveis na variável resposta. Para uma variável explicativa binária, o coeficiente  $\beta$  representa o *logito* da razão de chances da resposta  $Y$  pela associação com  $x$ , controlado pelas demais covariáveis. Note que  $\beta$  não depende de  $j$ , implicando que a relação entre  $x$  e  $Y$  é independente da categoria. Esse modelo fornece uma única estimativa de OR para todas as categorias comparadas, que pode ser obtida exponenciando o coeficiente  $\beta$ . Essa estimativa é bastante adequada em termos da facilidade de interpretação e da parcimônia do modelo.<sup>15</sup>

Essa característica do modelo resultou na suposição chamada por McCullagh<sup>16</sup> (1980) de *chances proporcionais* que deu nome ao modelo. Esta premissa é assumida para cada covariável incluída no modelo. Antes da construção do modelo, é sempre importante verificar se esta suposição é satisfeita. Para testar a homogeneidade da *razão de chances*, geralmente é utilizado o teste *score*<sup>4</sup>, chamado por Hosmer & Lemeshow<sup>14</sup> (2001) *teste de regressão paralela* e que pode ser usado para se avaliar a evidência de adequação do modelo.

Quando os códigos de  $Y$  são invertidos (isto é,  $Y_1$  é codificado como  $Y_k$ ,  $Y_2$  como  $Y_{k-1}$  e assim por diante), ocorre apenas a inversão do sinal dos parâmetros da regressão. Esse modelo também goza da propriedade de invariância no que se refere à junção das categorias da variável resposta. Esta propriedade implica que quando as categorias de  $Y$  são excluídas ou

reagrupadas, os coeficientes das covariáveis ( $\beta$ ) devem permanecer inalterados, embora os interceptos ( $\alpha$ ) sejam afetados.

## **B) Modelo de chances proporcionais parciais (MCP)**

É rara a chance de todas as covariáveis incluídas no modelo apresentarem a propriedade de chance proporcional. Para contemplar uma situação mais realista, o MCP<sup>17</sup>, em inglês *partial proportional odds model*, permite que algumas covariáveis possam ser modeladas com a suposição de chance proporcional, e para as outras variáveis em que este pressuposto não seja satisfeito, são incluídos no modelo parâmetros específicos que variam para as diversas categorias comparadas. MCP é uma extensão do modelo de chances proporcionais. Existem dois tipos de modelos de chances proporcionais parciais, sem e com restrição, como apresentados a seguir.

### **B1) Modelo de chances proporcionais parciais não-restrito (MCP-NR)**

Como mostrado na Tabela 1, esse modelo considera que para a variável resposta  $Y$  com  $k$  categorias, dentre as  $p$  variáveis preditoras  $\underline{x} = (x_1, x_2, \dots, x_p)$  apenas algumas tenham chances proporcionais. Sem perda de generalidade, vamos assumir que para as  $q$  primeiras covariáveis a propriedade de chances proporcionais não seja válida.<sup>15</sup>

Para uma variável em que não vale a propriedade de chances proporcionais, digamos  $x_1$ ,  $\alpha_j + \beta x_1$  é incrementado pelo coeficiente  $\gamma_{j1}$  que é o efeito associado com cada logito cumulativo, ajustado pelas demais covariáveis.<sup>15</sup> Assim, o coeficiente da covariável é  $\alpha_j + \beta x_1 + \gamma_{j1}$ .

Para esse modelo, são estimados  $(k-1)$  interceptos,  $p$  coeficientes betas ( $\beta$ ) que são independentes das categorias comparadas e  $q(k-1)$  parâmetros gama ( $\gamma$ ), os quais são associados com cada covariável e categoria da variável resposta. Se os parâmetros gamas ( $\gamma$ ) são nulos  $\gamma_j = 0$  para todo  $j$ , o modelo se reduz ao modelo de chances proporcionais.

Nesse modelo, para as  $q$  primeiras covariáveis o coeficiente angular depende de  $j$ , implicando que a relação entre  $\underline{x}$  e  $Y$  é dependente da categoria. Conseqüentemente, são estimadas razões de chances para todas as comparações das categorias da variável resposta. Para as demais

covariáveis, os coeficientes angulares ( $\beta$ ) independem de  $j$ , e, portanto, apenas uma razão de chances é estimada.

## **B2) Modelo de chances proporcionais parciais restrito (MCP-P-R)**

Quando a relação entre uma covariável e a variável resposta não é proporcional, freqüentemente, é esperado que haja um tipo de tendência. Peterson e Harrell<sup>17</sup> (1990) propuseram um modelo aplicável quando existe uma relação linear entre o *logito* referente a uma covariável e a variável resposta.<sup>1</sup>

Nesse caso, restrições (representadas pelos parâmetros gamas e que são escalares fixos), podem ser inseridas como parâmetros do modelo de modo a incorporar essa linearidade (ver Tabela 1).

Para uma dada covariável, o coeficiente gama  $\gamma_j$  não depende dos pontos de corte, mas é multiplicado por um coeficiente tau ( $\tau$ ) que é específico para cada *logito*.<sup>15</sup>

A escolha da restrição pode ser decidida de várias maneiras. Idealmente, elas deveriam ser determinadas usando um banco de dados de um estudo piloto ou um valor predefinido a priori.

## **C) Modelo de razão contínua (MRC)**

Feinberg<sup>10</sup> (1980) propôs o modelo MRC, em inglês *continuation ratio logistic model*, que compara a probabilidade de uma resposta igual a uma determinada categoria, digamos  $Y=j$ , com a probabilidade de uma resposta maior,  $Y>j$ , como mostrado na Tabela 1.

Para cada categoria ( $j=1, \dots, k$ ), o intercepto do modelo é  $\alpha_j$  e os coeficientes das covariáveis são os coeficientes betas ( $\beta_j$ ). Esse modelo possui diferentes constantes e coeficientes específicos para cada comparação. Uma vantagem é que o MRC pode ser ajustado através de  $k$  modelos de regressão logística binária. É mais apropriado quando há um interesse intrínseco em uma categoria específica da variável resposta, e não meramente um agrupamento arbitrário de uma variável contínua.<sup>1</sup>

O modelo de razão contínua é afetado pela direção escolhida para modelar a variável, ou seja, a propriedade de invariância da codificação não é válida para esse modelo.<sup>12</sup> A razão de chances obtida quando se modela o crescimento na gravidade não é equivalente ao recíproco

que é obtido quando se modela decréscimo na gravidade. Portanto, não se pode apenas inverter o sinal do coeficiente para trocar o sentido da comparação como ocorre nos modelos de regressão logística binária e modelo de chances proporcionais.<sup>18</sup>

A suposição de heterogeneidade dos pontos de corte pode ser testada incluindo no modelo um termo de interação, entre a exposição de interesse e um fator que indique o ponto de corte usado na comparação. O valor da função de verossimilhança dos modelos com e sem o termo de interação deve ser comparado. Se a heterogeneidade é significativa, o modelo de razão contínua pode ser facilmente adaptado com efeitos para os vários pontos de corte, utilizando o termo de interação incluído no modelo.<sup>20</sup>

#### **D) Modelo *estereótipo***

O modelo *estereótipo* (ME) deve ser utilizado quando a variável resposta é intrinsecamente ordinal e não é uma versão discreta de alguma variável contínua, como por exemplo, as possíveis respostas do item da escala de QV QLS-BR<sup>5-8</sup> (nunca, raramente, ocasionalmente, às vezes e freqüentemente), mencionadas na introdução.

Esse modelo foi proposto por Anderson<sup>2</sup> (1984) que afirma que os diagnósticos médicos tendem a ser fixos e invariáveis (estereotipados) baseando-se na classificação da gravidade da doença, tais como, média, moderada e grave. Neste caso, o modelo deve ser flexível o bastante para capturar a multidimensionalidade natural dessas respostas.<sup>12</sup>

Esse é o modelo mais flexível para análise de respostas ordinais e pode ser considerado uma extensão do modelo de regressão multinomial (ver MM na Tabela 1).<sup>12</sup>

Devido ao caráter ordinal dos dados é imposta uma estrutura linear ao *logito* desse modelo. Ou seja, são atribuídos pesos aos coeficientes dados por  $\beta_{jl} = \omega_j \beta_l$  com  $j = 1, \dots, k$  e  $l = 1, \dots, p$ . (ver Tabela 1).

Além dos pesos ( $\omega_k$ ) para a variável dependente  $Y$ , há um parâmetro beta para cada variável explicativa. Esses pesos são diretamente relacionados com o efeito das covariáveis. Então, a razão de chances formada terá uma tendência de crescimento, já que os pesos normalmente são construídos com ordenação ( $0 = \omega_1 \leq \omega_2 \leq \dots \leq \omega_j$ ). Assim, o efeito das covariáveis na primeira razão de chances é menor que o efeito na segunda e assim sucessivamente.<sup>15</sup>

A maior dificuldade dessa modelagem é a determinação desses pesos, mas existem algumas possibilidades. Greenland<sup>12</sup> (1994) sugere que os pesos podem ser decididos a priori, ou seja, valores apropriadamente escolhidos ou estimados a partir de dados de um estudo piloto, ou ainda utilizar modelos lineares generalizados que estimam os pesos como parâmetros adicionais do modelo.

### **Implementação computacional**

O modelo de chances proporcionais (MCP) encontra-se implementado na maioria dos pacotes estatísticos comerciais, incluindo SPSS<sup>®</sup>, Minitab<sup>®</sup>, S-plus<sup>®</sup>, SAS<sup>®</sup> e Stata<sup>®</sup>, e também o *software* R que é de livre acesso.

O Modelo de Chances Proporcionais parciais (MCP) pode ser ajustado no SAS<sup>®</sup> usando o procedimento PROC CATMOD<sup>15</sup> ou no Stata<sup>®</sup> utilizando o comando *gologit2* desenvolvido por Willians<sup>23</sup> (2006).

Os pacotes SAS<sup>®</sup>, Stata<sup>®</sup>, S-plus<sup>®</sup> e R possuem comandos que programam uma reestruturação dos dados a partir da variável resposta  $Y$  para execução do modelo de razão contínua. O artifício consiste em criar uma variável binária que representa as comparações propostas pelo modelo ( $Y=j; Y>j$ ), um vetor que indica o ponto de corte ( $k$ ) que foi aplicado, ou seja, em qual categoria a variável resposta foi dividida e um vetor que é utilizado para replicar as outras variáveis explicativas da mesma forma que a resposta foi replicada. O modelo pode ser obtido ajustando uma regressão logística binária nos dados reestruturados com a nova resposta dicotômica como variável dependente, incluindo como covariável a variável criada que indica o nível do ponto de corte e as covariáveis reestruturadas.<sup>20</sup>

O Modelo Estereótipo (ME) pode ser ajustado através do SAS, quando os conjuntos de pesos forem predeterminados. Alternativamente os pesos podem ser considerados como parâmetros adicionais do modelo, como nas macros para SAS<sup>®</sup> e STATA<sup>®</sup> desenvolvidas por Hendricks<sup>13</sup> (2000). Além disso, o modelo pode ser ajustado através dos *softwares* R e S-plus<sup>®</sup>, por meio de comandos no contexto de modelos lineares generalizados, sendo que os pesos são estimados como parâmetros adicionais do modelo.<sup>20</sup>

## EXEMPLOS DE APLICAÇÃO

Para exemplificar a análise de dados de qualidade de vida por meio dos modelos mencionados anteriormente foram utilizados dados obtidos a partir de um estudo transversal, que teve a participação de 273 pacientes com diagnóstico de esquizofrenia, registrados em dois serviços de referência em saúde mental.<sup>5-8</sup> As entrevistas foram conduzidas por profissionais de saúde previamente treinados. Foi aplicado um questionário que incluía informações clínicas e sociodemográficas dos pacientes. A qualidade de vida foi mensurada por meio da QLS-BR, instrumento adaptado e validado para o contexto brasileiro, apresentando boas características de validade e confiabilidade.<sup>5;6</sup>

A QLS-BR possui em sua estrutura um total de 21 itens distribuídos em três domínios específicos: (1) social, (2) ocupacional e (3) intrapsíquico e relações interpessoais. O resultado da avaliação é pontuado em uma escala do tipo *likert*, cujos escores variam de zero a seis pontos, sendo que um escore maior representa melhor QV. Todos os itens e domínios da escala, além da escala global, podem ser analisados categorizando os escores em (0 + 2) como QV muito comprometida, (2 + 5) como QV comprometida e (5 + 6) como QV inalterada.<sup>5;6</sup>

Foram criados dois exemplos de aplicação e em ambos foram utilizadas duas covariáveis categóricas, “sexo (feminino/masculino)” e “estado civil (casado/solteiro)”, pela importância relatada na literatura<sup>7;8</sup> como associadas à QV na esquizofrenia.

Além da análise descritiva dos dados por meio de tabelas de contingência, foi testada a associação entre sexo e estado civil com a qualidade de vida dos pacientes, por meio do teste de Kruskal-Wallis (KW) implementado no pacote StatXact versão 6. Dos quatro modelos apresentados anteriormente, em três foi utilizado o *software* R versão 2.2.1<sup>20</sup>. Apenas o MCPP foi ajustado por meio do *software* STATA versão 9.0.<sup>23</sup>

A suposição de chances proporcionais foi testada para cada covariável e no modelo final, foi utilizado o teste escore. O ajuste de cada modelo foi avaliado através do teste de *deviance*.<sup>14</sup> Para comparar os modelos multinomial e estereótipo foi utilizado o teste da razão de velhossimilhanças.<sup>14</sup> Por fim, foi calculado o poder estatístico por meio do método de Whitehead<sup>24</sup> (1993).

Foi adotado o nível de significância de 5% e a probabilidade de significância foi denotada por p.

### **Exemplo 1**

Para este exemplo foi escolhida a QV no domínio ocupacional da escala QLS-BR devido à sua distribuição mais homogênea entre as categorias, facilitando as comparações. A análise descritiva dos dados por meio de tabela de contingência mostrou que a maioria (70,8%) dos pacientes encontra-se na categoria com QV comprometida independente do sexo ou estado civil.

Para QV muito comprometida, nota-se uma tendência de pior qualidade de vida para pacientes solteiros e do sexo masculino (39,2%), enquanto que no grupo de casados do sexo feminino foram encontradas apenas duas pacientes (8,3%). Nos dois casos, a associação entre sexo ( $p < 0,01$ ) e estado civil ( $p = 0,04$ ) e a baixa QV mostrou-se significativa.

### **Modelo de chances proporcionais**

O MCP foi calculado de forma uni e multivariada e verificou-se que as estimativas em ambas as análises foram bastante semelhantes, sugerindo minimização do efeito de confusão. De acordo com a Tabela 2, a suposição de regressão paralela não foi violada ( $p = 0,36$  do teste do escore), indicando homogeneidade das razões de chances para todas as categorias comparadas. Portanto, nesse caso, não houve necessidade de ajustar o modelo de chances proporcionais parciais. O teste de *deviance* indica que o modelo está bem ajustado e ambas as variáveis (sexo e estado civil) mostraram-se estatisticamente associadas ao desfecho. Um exemplo de interpretação é que homens apresentam duas vezes mais chance de estarem em uma categoria de pior qualidade de vida que mulheres.

Também foram mostrados na Tabela 2 os resultados do modelo logístico de resposta binária após reagrupar as categorias de QV comprometida e QV inalterada e compara-las à QV muito comprometida. Entretanto, nesse modelo a variável estado civil não se mostrou significativa ( $p > 0,05$ ), diferentemente do modelo ordinal. Isso demonstra que não se deve dicotomizar a variável ordinal, pois isso pode levar a conclusões incorretas, como nesse exemplo.

### **Modelo de razão-contínua**

Também na Tabela 2 encontram-se os resultados do MRC sem os termos de interação, que foram testados e não se mostraram significativos ( $p = 0,86$ ), evidenciando que as razões de chances são homogêneas, como observado pelo modelo de chances proporcionais. Esse

modelo também se mostrou adequado pelo teste de *deviance*. A razão de chances de 1,88 pode ser interpretada como a chance relativa dos pacientes solteiros terem uma pior QV em uma categoria específica, comparado aos pacientes casados.

### **Modelo multinomial**

Antes de ajustar o modelo estereótipo, foi ajustado o modelo multinomial, calculadas suas razões de chances e erros-padrão foram analisados para uma posterior comparação. Em ambos os modelos, a última categoria (QV inalterada) foi considerada como referência. Os resultados estão na Tabela 3.

Segundo a análise multivariada, por exemplo, a chance de homens terem QV muito comprometida é 2,46 vezes maior que das mulheres, quando comparada à QV inalterada.

Ressalta-se que nas comparações da QV comprometida e QV inalterada ambas covariáveis não se mostraram significantes, explicado possivelmente pela proximidade entre essas duas categorias.

### **Modelo estereótipo**

O modelo estereótipo também foi ajustado e os pesos foram estimados como parâmetros do modelo e apresentados na Tabela 3. Na análise multivariada a variável estado civil perde sua significância. A chance dos homens terem QV muito comprometida é três vezes maior que das mulheres, quando comparada à QV inalterada. Essa mesma chance é de 1,38 vezes quando se compara QV comprometida e QV inalterada.

Por meio do teste da razão de verossimilhança, percebe-se que não há diferença significativa entre os modelos multinomial e estereótipo ( $p=0,77$ ), e ambos apresentam um bom ajuste ( $p=0,60$  e  $p=0,58$ , respectivamente).

### **Exemplo 2**

Foi escolhido para este exemplo, um dos 21 itens da escala QLS-BR, componente do domínio ocupacional chamado de “Funcionamento Ocupacional”. É importante ressaltar que, esse foi o único item da escala em que a suposição de chances proporcionais foi violada para uma das variáveis explicativas.

### **Modelo de chances proporcionais**

Na Tabela 4 encontram-se os resultados do Modelo de Chances Proporcionais calculados de forma bivariada e multivariada para o item Funcionamento Ocupacional. A suposição de regressão paralela foi violada para a covariável sexo ( $p < 0,05$  do teste do escore), indicando heterogeneidade da razão de chances nas categorias comparadas. Isso também pode ser observado na Tabela 5 contendo os resultados dos modelos de regressão logística binária tendo com resposta a QV dicotomizada. Para a covariável estado civil, os coeficientes estimados para as duas diferentes comparações variam pouco (0,39 e 0,32). Ao contrário, confirmando a violação da suposição de chances proporcionais, os coeficientes da covariável sexo variam substancialmente (1,53 para 0,59), fazendo com que a razão de chances passe de 4,64 na primeira comparação para apenas 1,81 na segunda. Nesse caso, o MCP não é adequado e deve-se ajustar o MCP.

### **Modelo de chances proporcionais parciais**

Na Tabela 5 encontram-se os resultados do MCP sem restrição. Note que foram estimados dois coeficientes para a variável sexo (sem chances proporcionais) e apenas um para a variável estado civil.

Esse modelo indica que homens têm quase cinco vezes mais chance que mulheres de estarem na categoria de QV muito comprometida quando comparada a QV comprometida ou inalterada. Por outro lado, homens têm aproximadamente duas vezes mais chance que mulheres de estar na categoria de QV muito comprometida ou comprometida quando comparadas com a categoria de QV inalterada. Quanto ao estado civil, solteiros têm 1,44 vezes mais chance do que casados de estarem numa categoria de pior QV.

### **Cálculo do poder**

Como a amostra do estudo utilizado no exemplo foi pré-determinada, pode-se calcular o poder associado a este tamanho de amostra ( $n = 273$ ). Outras metodologias já foram empregadas,<sup>7</sup> mas aqui o poder será calculado de acordo com Whitehead<sup>24</sup> (1993) e Walters et al<sup>22</sup> (2001), já que os dados utilizados como exemplo são provenientes de um estudo transversal e utilizou-se a razão de chances como medida sumário.

Assim, de acordo com a fórmula (2), o poder é calculado como a probabilidade acumulada da distribuição normal padrão do seguinte percentil:

$$z_{1-\beta} = \left[ (\log OR) \times \sqrt{\frac{n \left( 1 - \sum_{i=1}^k \bar{\pi}_i^3 \right)}{6}} \right] - z_{1-\alpha/2}$$

Foi considerada como covariável principal o estado civil. De acordo com os dados do exemplo 1,  $\bar{\pi}_1 = 0,157$ ;  $\bar{\pi}_2 = 0,291$ ;  $\bar{\pi}_3 = 0,051$ , onde  $\bar{\pi}_1$  representa a proporção média de casados e solteiros que estão na categoria de QV muito comprometida ( $Y_1$ ). Logo,  $\sum_{i=1}^k \bar{\pi}_i^3 = 0,029$ . Pela análise univariada, a razão de chances para a variável estado civil é de 1,982 ( $\log OR = 0,684$ ). Considerando o tamanho de amostra de 136 para cada grupo (aproximadamente a metade de 273),  $z_{1-\beta} = 1,249$  e portanto  $1 - \beta = 0,894$ , correspondendo ao poder de quase 90% de identificação de fator de risco para QV.

## DISCUSSÃO

De forma geral, os modelos de regressão logística ordinal demonstram ser apropriados para análise de dados que têm como respostas medidas de qualidade de vida. A modelagem difere quanto à forma dessas escalas: elas podem ter categorias ordenadas agrupadas a partir de uma variável latente contínua, ou categorias discretas, mas com ordenação.

Para os dados utilizados no primeiro exemplo, cuja variável resposta era QV no domínio ocupacional, não houve diferença significativa entre o modelo estereótipo e o modelo multinomial. Apesar disso, o modelo estereótipo deve ser preferido já que ele considera o caráter ordinal dos dados e estima um menor número de parâmetros. Deve-se destacar, entretanto, que esse tipo de modelo é mais apropriado para situações em que a variável resposta ordinal tem categorias discretas.

Entretanto, considerando que a variável resposta utilizada no exemplo apresentava categorias ordenadas agrupadas a partir de uma variável latente contínua, o modelo de chances proporcionais ou modelo de razão contínua seriam os mais indicados. Apesar de esses dois modelos concordarem no critério de homogeneidade da razão de chances das categorias comparadas, eles discordam quanto às magnitudes de risco, já que consideram comparações

distintas. Não se recomendaria a utilização do modelo de razão contínua, apesar de ter apresentado um bom ajuste, uma vez que ele é indicado quando se tem interesse pela comparação de uma categoria específica da resposta, o que não ocorre nesse caso.

Outro ponto que deve ser considerado é o fato de que comparações baseadas em várias regressões logísticas binárias podem levar a inferências incorretas, como mostrado no exemplo analisado. Por isso, os modelos de regressão ordinal fornecem estimativas mais confiáveis para se analisar dados ordinais, podendo destacar o MCP devido a sua parcimônia. Considerando que sua principal restrição – a suposição de chances proporcionais – não foi violada para os dados do exemplo 1, o MCP foi considerado o mais apropriado.

Para o segundo exemplo, entretanto, o MCP foi inadequado devido a violação da suposição. Nesse caso, um modelo de regressão ordinal alternativo é o modelo de chances proporcionais parciais (MCP). Esse modelo apresenta-se como uma opção interessante, já que a suposição de chances proporcionais nem sempre é encontrada. O MCP permite que algumas variáveis tenham apenas uma razão de chances para todas as categorias, e outras tenham razão de chances para comparações em cada categoria da variável resposta, como ocorreu no segundo exemplo. Uma dificuldade é que esse modelo não está implementado em muitos *softwares* comumente utilizados para análise estatística.

Apesar de algumas diferenças nos resultados dos modelos ajustados, todos os ajustes foram razoáveis. Os MCP e MCP mostraram-se adequados para a análise de dados do estudo de QV em pacientes com esquizofrenia, devido ao caráter da variável resposta sobre QV (variável contínua agrupada), além da parcimônia e facilidade de interpretação dos resultados desses modelos.

No que se refere à interpretação dos resultados das análises envolvendo dados ordinais, deve-se ter alguns cuidados para que ele não seja feito da mesma forma que nos modelos binários. É importante considerar que a interpretação deve sempre levar em conta a proporcionalidade das chances, ou seja, a chance de estar em uma categoria de melhor ou pior qualidade de vida, dependendo do caso. Tal procedimento quase sempre é negligenciado na maioria dos estudos envolvendo dados ordinais.

Da mesma forma, deve-se ressaltar que para se ter uma boa análise estatística, o planejamento e dimensionamento da amostra devem considerar o caráter ordinal dos dados sobre QV. Além disso, é importante lembrar que os modelos ordinais não são os únicos métodos de análise para verificação de fatores associados à qualidade de vida. Existem várias outras formas de

análise que não foram discutidas nesse artigo, como os modelos lineares generalizados ou as árvores de decisão.

Outro ponto a ser mencionado é que existem vários tipos de variáveis ordinais, além dos escores de QV, que são utilizadas no contexto de saúde pública e que também podem ser analisadas pelos modelos discutidos nesse artigo.

Finalmente, conclui-se que os modelos de regressão logística ordinal são ferramentas apropriadas para analisar dados de QV, além de mostrar-se com grande potencial para uso em outras investigações envolvendo dados de natureza ordinal. Recomenda-se que sejam evitados procedimentos simples, tais como a dicotomização da variável resposta e a desconsideração da ordenação, que têm como conseqüências a perda de informação contida nos dados e provavelmente inferências incorretas ou menos apropriadas.

#### **AGRADECIMENTOS**

As autoras agradecem a FAPEMIG, pelo financiamento do Projeto “Qualidade de Vida na Esquizofrenia”, processo nº CDS-301/02 e também ao CNPQ e CAPES pelas bolsas às pesquisadoras Waleska Teixeira Caiaffa e Mery Natali Silva Abreu.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. *Int. J. Epidemiol.* 1997; 26:1323-33.
2. Anderson JA. Regression and ordered categorical variables. *J. R. Statist. Soc.* 1984, 16: 1-30.
3. Bender R, Grouven U. Ordinal logistic regression in medical research. *J. R. Coll. Physicians Lond.* 1997; 31: 546-51.
4. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 1990; 46:1171-1178
5. Cardoso CS, Bandeira M, Caiaffa WT, Fonseca JOP. Escala de qualidade de vida para pacientes com esquizofrenia - QLS-BR, Adaptação transcultural para o Brasil. *J. Bras. Psiquiatr.* 2002; 51: 31-38.
6. Cardoso CS, Bandeira M, Caiaffa WT, Siqueira AL, Fonseca IK, Fonseca JOP. Qualidades psicométricas da escala de qualidade de vida para pacientes com esquizofrenia: Escala QLS-BR. *J. Bras. Psiquiatr.* 2003; 52: 211-222.
7. Cardoso CS, Caiaffa WT, Bandeira M, Siqueira AL, Abreu MNS, Fonseca JOP. Factors associated with low quality of life in schizophrenia. *Cad. Saúde Pública* 2005; 21: 1338-1348.
8. Cardoso CS, Caiaffa WT, Bandeira M, Siqueira AL, Abreu MNS, Fonseca JOP. Qualidade de vida e dimensão ocupacional na esquizofrenia: uma comparação por sexo. *Cad. Saúde Pública* 2006; 22: 1303-1314.
9. Ciconelli RM, Ferraz MB, Santos W, Meinão I, Quaresma MR. Tradução para a língua portuguesa e validação do questionário genérico de avaliação de qualidade de vida SF-36. *Ver. Bras. Reumatol.* 1999; 39:143-150.
10. Fienberg SE. *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA; MIT Press; 1980. p. 110-116.
11. Fleck et al. Desenvolvimento da versão em português do instrumento de avaliação da qualidade de vida da OMS (WHOQOL-100). *Rer. Bras. Psiquiatr.* 1999; 21: 19-28.
12. Greenland S. Alternative models for ordinal logistic regression. *Stat. Med.* 1994;13(16):1665-77.

13. Hendrickx J. Special restrictions in multinomial logistic regression. *Stata Technical Bulletin* 2000; STB-56: 18 - 26
14. Hosmer WD, Lemeshow S. *Applied Logistic Regression*. Wiley; 2000.
15. Lall R, Campbell MJ, Walters SJ, Morgan K. A review of ordinal regression models applied on health-related quality of life assessments. *Stat. Methods Med. Res.* 2002; 11: 49-67.
16. McCullagh P. Regression models for ordinal data. *J. R. Stat. Soc. Ser. B. Methodol.* 1980; 42: 109-142.
17. Peterson BL, Hanrrel FE. Partial proportional odds models for ordinal response variables. *Appl. Statistic.* 1990; 39:205-217.
18. Scott SC, Goldberg MS, Mayo NE. Statistical assessment of ordinal outcomes in comparative studies. *J. Clin. Epidemiol.* 1997; 50:45-55
19. Seidl EMF, Zannon CMLC. Qualidade de vida e saúde: aspectos conceituais e metodológicos. *Cad. Saúde Pública* 2004; 20: 580-588.
20. *The R Project for Statistical Computing* <http://www.r-project.org/about.html> (acessado em 19 de outubro de 2006).
21. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol. Med.* 1998; 28: 551-558.
22. Walters SJ, Campbell MJ, Lall R. Design and analysis of trials with quality of life as an outcome: a practical guide. *J. Biopharm. Stat.* 2001, 11(3): 155-176.
23. Williams R. gologit2: Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *Forthcoming in Stata Journal*, 2006.
24. Whitehead J. Sample size calculations for ordered categorical data. *Stat. Med.* 1993, 12: 2257-2271.

Tabela 1: Informações sobre os principais modelos de regressão logística (com categorias ordinais ou não)

Modelo	Forma funcional do modelo	Indicação de uso
Modelo binário (MB)	$\lambda(x) = \ln \left\{ \frac{\Pr(Y = 1   \underline{x})}{\Pr(Y = 0   \underline{x})} \right\} = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$	Variável resposta com duas categorias (Y=0,1)
Modelo de Chances Proporcionais (MCP)	$\lambda_j(x) = \ln \left\{ \frac{\Pr(Y = 1   \underline{x}) + \dots + \Pr(Y = j   \underline{x})}{\Pr(Y = j+1   \underline{x}) + \dots + \Pr(Y = k   \underline{x})} \right\} = \ln \left\{ \frac{\sum_{l=1}^j \Pr(Y = l   \underline{x})}{\sum_{l=j+1}^k \Pr(Y = l   \underline{x})} \right\}$ $\lambda_j(x) = \alpha_j + (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p), \quad j = 1, \dots, k-1$	Variável resposta original contínua e posteriormente agrupada e a suposição de chances proporcionais é válida
Modelo de Chances Proporcionais Parciais Não-Restrito (MCP-PR)	$\lambda_j(x) = \ln \left\{ \frac{\Pr(Y = 1   \underline{x}) + \dots + \Pr(Y = j   \underline{x})}{\Pr(Y = j+1   \underline{x}) + \dots + \Pr(Y = k   \underline{x})} \right\} = \ln \left\{ \frac{\sum_{l=1}^j \Pr(Y = l   \underline{x})}{\sum_{l=j+1}^k \Pr(Y = l   \underline{x})} \right\}$ $\lambda_j(x) = \alpha_j + [(\beta_1 + \gamma_{j1})x_1 + \dots + (\beta_q + \gamma_{jq})x_q + (\beta_{q+1} x_{q+1}) + \dots + (\beta_p x_p)], \quad j = 1, \dots, k-1$	Quando a suposição de chances proporcionais não é válida
Modelo de Chances Proporcionais Parciais Restrito (MCP-PR)	$\lambda_j(x) = \ln \left\{ \frac{\Pr(Y = 1   \underline{x}) + \dots + \Pr(Y = j   \underline{x})}{\Pr(Y = j+1   \underline{x}) + \dots + \Pr(Y = k   \underline{x})} \right\} = \ln \left\{ \frac{\sum_{l=1}^j \Pr(Y = l   \underline{x})}{\sum_{l=j+1}^k \Pr(Y = l   \underline{x})} \right\}$ $\lambda_j(x) = \alpha_j + \{\tau_j [(\beta_1 + \gamma_1)x_1 + \dots + (\beta_q + \gamma_q)x_q] + (\beta_{q+1} x_{q+1}) + \dots + (\beta_p x_p)\}, \quad j = 1, \dots, k-1$	Suposição de chances proporcionais não é válida e existe relação linear entre OR de uma covariável e a variável resposta
Modelo de Razão Contínua (MRC)	$\lambda_j(x) = \ln \left\{ \frac{\Pr(Y = j   \underline{x})}{\Pr(Y = j+1   \underline{x}) + \dots + \Pr(Y = k   \underline{x})} \right\} = \ln \left\{ \frac{\Pr(Y = j   \underline{x})}{\sum_{l=j+1}^k \Pr(Y = l   \underline{x})} \right\}$ $\lambda_j(x) = \alpha_j + (\beta_{j1} x_1 + \beta_{j2} x_2 + \dots + \beta_{jp} x_p), \quad j = 1, \dots, k$	Há um interesse intrínseco em uma categoria específica da variável resposta
Modelo Multinomial (MM)	$\lambda_j(x) = \ln \left\{ \frac{\Pr(Y = j   \underline{x})}{\Pr(Y = 0   \underline{x})} \right\}$ $\lambda_j(x) = \alpha_j + (\beta_{j1} x_1 + \dots + \beta_{jp} x_p), \quad j = 1, \dots, k$	Variável resposta nominal com três ou mais categorias sem ordenação
Modelo Estereótipo (ME)	$\lambda_j(x) = \ln \left\{ \frac{\Pr(Y = j   \underline{x})}{\Pr(Y = 0   \underline{x})} \right\}$ $\lambda_j(x) = \alpha_j + \omega_j (\beta_1 x_1 + \dots + \beta_p x_p), \quad j = 1, \dots, k$	Variável resposta ordinal discreta que não provem de alguma variável contínua agrupada

Tabela 2: Resultados dos modelos de regressão logística binária, de chances proporcionais e de razão contínua tendo como resposta a qualidade de vida no domínio ocupacional da escala QLS-BR

Tipo de modelo	Covariável (referência)	$\beta$	$EP(\beta)$	$OR$	Teste <i>Wald</i> (p)	Teste escore (p)
Binário* <sup>1</sup>	Sexo (feminino)	0,46	0,38	1,58	1,49 (0,22)	--
	Estado civil (casado)	0,90	0,29	2,45	9,41 (<0,01)	--
Chances Proporcionais <sup>2</sup>	Sexo (feminino)	0,71	0,25	2,03	2,79 (<0,01)	2,02 (0,36)
	Estado civil (casado)	0,68	0,33	1,97	2,10 (0,04)	2,02 (0,36)
Razão Contínua <sup>3</sup>	Sexo (feminino)	0,60	0,23	1,82	2,56 (0,01)	0,18 (0,86)
	Estado civil (casado)	0,63	0,29	1,88	2,14 (0,03)	0,18 (0,86)

\*As categorias QV comprometida e QV inalterada foram reagrupas e comparadas à QV muito comprometida

<sup>1</sup>Teste de Hosmer-Lemeshow (p = 0,47); <sup>2</sup>Teste de *deviance* (p=0,55);

<sup>3</sup>Teste de *deviance* (p=0,26);

EP= erro-padrão; OR = *odds ratio* (razão de chances)

Tabela 3: Resultados dos modelos de regressão logística multinomial e estereótipo tendo como resposta a qualidade de vida no domínio ocupacional da escala QLS-BR

Tipo de modelo	Covariável (referência)	Comparações					
		QV inalterada <i>versus</i> QV muito comprometida			QV inalterada <i>versus</i> QV comprometida		
		$\beta_1$	$EP(\beta_1)$	OR <sub>1</sub> (p)	$\beta_2$	$EP(\beta_2)$	OR <sub>2</sub> (p)
Multinomial*	Sexo (feminino)	0,90	0,46	2,46 (0,05)	0,02	0,41	1,02 (0,97)
	Estado civil (casado)	1,19	0,53	3,29 (0,02)	0,71	0,45	2,03 (0,12)
Estereótipo**	Sexo (feminino)	1,10	0,40	3,00 (<0,01)	0,32	0,40	1,38 (<0,01)
	Estado civil (casado)	0,95	0,58	2,59 (0,10)	0,28	0,58	1,32 (0,10)

Peso  $\omega$  do Modelo Estereótipo = 0,29; EP= erro-padrão; OR = *odds ratio* (razão de chances)

\*Teste de *deviance* (p=0,60); \*\*Teste de *deviance* (p=0,58);

Tabela 4: Resultados dos modelos de chances proporcionais tendo como resposta a qualidade de vida no item funcionamento ocupacional da escala QLS-BR

Tipo de análise	Covariável (referência)	$\beta$	EP( $\beta$ )	OR	Teste Wald (p)	Teste score (p)
Univariada	Sexo (feminino)	1,09	0,26	2,99	4,25 (<0,01)	4,88 (0,03)
	Estado civil (casado)	0,46	0,31	1,59	1,49 (0,14)	0,08 (0,78)
Multivariada*	Sexo (feminino)	1,07	0,26	2,92	4,15 (<0,01)	4,88 (0,09)
	Estado civil (casado)	0,33	0,31	1,44	1,15 (0,25)	

\* Teste de *deviance* (p=0,20); EP= erro-padrão; OR = *odds ratio* (razão de chances)

Tabela 5: Resultados dos modelos de regressão logística binária e de chances proporcionais parciais não restrito (MCPN-NR) tendo como resposta a qualidade de vida no item funcionamento ocupacional da escala QLS-BR

Tipo de modelo	Covariável (referência)	Comparações							
		QV muito comprometida <i>versus</i> QV comprometida + QV inalterada				QV muito comprometida + QV comprometida <i>versus</i> QV inalterada			
		$\beta$	$EP(\beta)$	OR	Teste Wald (p)	$\beta$	$EP(\beta)$	OR	Teste Wald (p)
Regressão binária	Sexo (feminino)	1,53	0,36	4,64	4,28 (<0,01)	0,59	0,34	1,81	1,76 (0,08)
	Estado civil (casado)	0,39	0,41	1,48	0,95 (0,34)	0,32	0,41	1,38	0,79 (0,43)
Chances proporcionais parciais não restrito	Sexo (feminino)	1,54	0,36	4,65	4,29 (<0,01)	0,60	0,34	1,81	1,77 (0,09)
	Estado civil (casado)	0,36	0,32	1,44	1,15 (0,25)	0,36	0,32	1,44	1,15 (0,25)

EP= erro-padrão; OR = *odds ratio* (razão de chances)

#### **4. ARTIGO 2:**

**Uso de modelos de regressão logística ordinal em estudos clínicos e epidemiológicos**



**USO DE REGRESSÃO LOGÍSTICA ORDINAL  
EM ESTUDOS CLÍNICOS E EPIDEMIOLÓGICOS**

*ORDINAL LOGISTIC REGRESSION*

*IN CLINICAL AND EPIDEMIOLOGICAL STUDIES*

---

**Mery Natali Silva Abreu<sup>a,b</sup>, Arminda Lucia Siqueira<sup>a,c</sup>, Bruna Guimarães Oliveira<sup>d</sup>, Waleska Teixeira Caiaffa<sup>a,b</sup>.**

<sup>a</sup> Programa de Pós-graduação em Saúde Pública, Faculdade de Medicina, Universidade Federal de Minas Gerais – UFMG.

<sup>b</sup> Grupo de Pesquisa em Epidemiologia GPE e Observatório de Saúde Urbana, Faculdade de Medicina, Universidade Federal de Minas Gerais – UFMG.

<sup>c</sup> Departamento de Estatística, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais – UFMG.

<sup>d</sup> Serviço de Cardiologia, Hospital das Clínicas da Universidade Federal de Minas Gerais – UFMG.

Correspondência: Mery Natali Silva Abreu (A/C Dra. Waleska Teixeira Caiaffa)

Av. Alfredo Balena, 190, 8º andar, sala 8013, Santa Efigênia CEP: 31130-100. Belo Horizonte – MG. Tel/fax: 31 3248-9949 E-mail: [merynatali@yahoo.com.br](mailto:merynatali@yahoo.com.br)

## RESUMO

Os modelos de regressão logística ordinal vêm sendo desenvolvidos para análise de estudos clínicos e epidemiológicos. Entretanto, a verificação da adequação de cada modelo tem recebido atenção limitada. Usando os softwares R ou STATA foram feitas análises formal ou gráfica para comparar os modelos ordinais por meio de dois bancos de dados: o primeiro refere-se à qualidade de vida em portadores de marcapasso e o segundo à condição de saúde obtida pelo “Segundo Levantamento Nacional de Saúde e Nutrição (NHANES II)”. Os resultados mostraram que, para o primeiro conjunto de dados, os modelos de chances proporcionais e chances proporcionais parciais apresentaram melhor ajuste. No segundo banco de dados, o modelo estereótipo se adequou melhor às premissas do modelo. O uso de modelos logísticos ordinais depende da categorização da variável resposta, da adequação às suposições, da qualidade do ajuste e da parcimônia do modelo.

Palavras-chave: dados ordinais; estudos clínicos; estudos epidemiológicos; qualidade de vida; regressão logística ordinal.

## ABSTRACT

The ordinal logistic regression models have been developed for analysis of clinical and epidemiological studies. However, the adequacy of such models for adjustment has received limited attention. Using R or STATA we performed formal and graphical analyses to compare ordinal models using two data sets: one on quality of life in pacemaker patients and other on health condition within the “National Health and Nutrition Examination Survey (NHANES II)”. Results show that for the first data set, the proportional odds model and the partial proportional odds model showed better fit, but the second data set, the stereotype model better fulfilled models assumptions. The ordinal logistic models perform differently depending on categorization of outcome (e.g. refinement of measure), adequacy to the assumptions, goodness-of-fit and parsimony of the model.

Key words: ordinal data; clinical studies; epidemiological studies; quality of life; ordinal logistic regression.

## INTRODUÇÃO

Os modelos de regressão logística ordinal vêm sendo desenvolvidos e descritos na literatura nos últimos anos para análise de dados cuja resposta apresente categorias com ordenação. Por sua vez, a informação ordenada, na forma de escore, tem sido utilizada em estudos clínicos e epidemiológicos, tais como, a qualidade de vida em escalas intervalares (adequada, comprometida ou muito comprometida), os indicadores de condição de saúde (boa, regular, ruim) e mesmo de gravidade das doenças (leve, moderada, grave).<sup>1</sup>

Estes modelos permitem também calcular a razão de chances (OR) ou risco de ocorrência de um evento, de grande interesse em estudos da área médica.<sup>1</sup>

Brevemente apresentaremos alguns dos vários trabalhos produzidos sobre o assunto. Walker e Duncan<sup>20</sup> propuseram inicialmente o modelo do logito cumulativo em 1967, depois chamado de modelo de chances proporcionais por McCullagh<sup>13</sup>, em 1980. Esse modelo permite analisar variáveis ordinais provenientes de uma variável contínua, fornecendo um único coeficiente, ou seja, uma única estimativa da razão de chances para todas as categorias comparadas.

O modelo de razão contínua, proposto por Fienberg<sup>7</sup>, também em 1980, é bastante utilizado e compara a chance de estar em uma dada categoria da variável resposta ordinal à chance da exposição ser maior que essa categoria. Já o modelo estereótipo proposto por Anderson<sup>2</sup> (1984) é considerado uma extensão do modelo de regressão multinomial<sup>11</sup> e recomenda-se seu uso na presença de variáveis qualitativas ordinais. Mais recentemente, em 1990, Peterson e Harrell<sup>16</sup> criaram uma variação do modelo de chances proporcionais, chamado de modelos de chances proporcionais parciais, que permite que algumas covariáveis possam ser modeladas com vários coeficientes para cada categoria comparada.

Apesar da diversidade dos modelos ordinais e da grande variedade de estudos sobre o assunto, a utilização na área de saúde pública tem sido escassa.<sup>1</sup> Esta constatação pode ser atribuída não só à complexidade de alguns modelos, mas ao fato desses modelos exigirem alguns cuidados na análise, como por exemplo, a validação de seus pressupostos.<sup>12</sup> Outro fator que poderia estar relacionado à pequena utilização destes modelos refere-se ao fato de existirem poucas opções de modelagem em pacotes

estatísticos comerciais utilizados na área de saúde pública, tais como SPSS<sup>®</sup> e MINITAB<sup>®</sup>. Mesmo utilizando *softwares* mais complexos como SAS<sup>®</sup> e STATA<sup>®</sup>, é freqüente a dificuldade em selecionar os comandos apropriados e, normalmente, ocorrem problemas na interpretação dos resultados.<sup>3</sup> Adiciona-se o custo elevado, tendo em vista licenças altamente restritas da maioria dos pacotes estatísticos comerciais .

Um pacote estatístico que vem se tornando cada vez mais popular é o *software* livre R, distribuído sob licença pública geral. Possui uma ampla variedade de técnicas estatísticas, incluindo vários modelos de regressão logística ordinal, como o modelo de chances proporcionais, modelo de razão contínua e modelo estereótipo, permitindo que eles sejam testados e que o ajuste possa ser comparado.<sup>18</sup>

Esse artigo tem como objetivo discutir o ajuste e a adequação dos principais modelos de regressão ordinal e mostrar sua implementação no *software* R versão 2.2.1. Para complementar, o ajuste do modelo de chances proporcionais parciais que, por ainda não se encontrar implementado no R, será feito através do *software* STATA<sup>®</sup> versão 9.1.

A metodologia é ilustrada com dois bancos de dados. No primeiro exemplo, serão utilizadas informações produzidas no processo de validação e adaptação transcultural do AQUAREL (Assessment of QUALity of life and RELated events), instrumento específico para avaliação da qualidade de vida em portadores de marcapasso cardíaco. Foram avaliados 139 pacientes portadores de marcapasso, nos quais foram aplicadas as escalas AQUAREL e SF-36 (*Medical Outcomes Study 36 Item Short Form Healthy Survey*), além de coletadas informações clínicas e sócio-demográficas.<sup>15</sup>

No segundo exemplo, será utilizado o banco de dados do “Segundo Levantamento Nacional de Saúde e Nutrição (NHANES II)” encontrável no endereço [http://www.cdc.gov/nchs/products/elec\\_prods/subject/nhanesii.htm](http://www.cdc.gov/nchs/products/elec_prods/subject/nhanesii.htm)<sup>14</sup>. Estas informações são amplamente utilizadas como exemplos em estudos estatísticos e epidemiológicos. Este inquérito inclui informações socioeconômicas como idade, sexo, cor de pele, estado civil; dados antropométricos; história nutricional e de saúde; e hematologia (hemoglobina, albumina, vitaminas), dentre outros. Para este estudo, excluídas as informações a respeito das crianças, o banco de dados totalizou 10337 entrevistados com idades entre 20 e 74 anos. Este mesmo banco de dados já foi utilizado por Richard

Willians<sup>21</sup>, como exemplo na implementação de um comando no *software* Stata<sup>®</sup>, encontrando-se disponível no *site* do autor (<http://www.nd.edu/~rwilliam/gologit2>).<sup>8</sup>

O artigo inicia com uma breve revisão dos principais modelos de regressão logística ordinal. A seguir, são apresentadas as principais estratégias para o ajuste desses modelos, as técnicas de verificação de qualidade do ajuste bem como a implementação computacional através dos *softwares* R e Stata<sup>®</sup>. A metodologia é ilustrada detalhadamente utilizando os bancos de dados supracitados, seguida por uma discussão sobre utilização destes modelos contextualizando nas principais dificuldades.

## ANÁLISE UNIVARIADA

Como em qualquer procedimento analítico que utiliza modelos de regressão, a análise múltipla por meio dos modelos ordinais deve sempre ser precedida pelo cruzamento de cada covariável com o evento de interesse. As tabelas de contingência são instrumentos importantes na análise descritiva dos dados ora em estudo.

Através dessa análise preliminar, conhecida como uni ou bivariada, pode-se selecionar os fatores mais importantes que serão introduzidos no modelo de regressão, além de se ter uma idéia inicial da razão de chances a ser estimada. O teste qui-quadrado de tendência pode ser um dos testes apropriados para seleção dos efeitos principais, já que este considera o caráter ordinal da variável resposta. Normalmente utiliza-se um nível de significância pouco conservador (geralmente entre 10 e 25%) para entrada das covariáveis no modelo.<sup>11</sup>

## MODELOS DE REGRESSÃO ORDINAL

Sendo o evento de interesse ordinal, após a análise univariada, deve-se construir o modelo final de regressão logística ordinal, para controlar possíveis fatores de confusão.

Considere  $Y$  a variável resposta com  $k$  categorias codificadas em  $1, 2, \dots, k$  e  $\underline{x} = (x_1, x_2, \dots, x_p)$  o vetor de variáveis explicativas ou covariáveis. As  $k$  categorias de  $Y$  condicionalmente aos valores de  $\underline{x}$  ocorrem com probabilidades  $p_1, p_2, \dots, p_k$ , isto é,  $p_j = \Pr(Y = j)$ , para  $j = 1, 2, \dots, k$ .

### Modelo de Chances Proporcionais

No modelo de chances proporcional (MCP) são considerados  $(k - 1)$  pontos de corte das categorias sendo que o  $j$ -ésimo ( $j=1, \dots, k-1$ ) ponto de corte é baseado na seguinte comparação de probabilidades acumuladas:

$$\lambda_j(\underline{x}) = \ln \left[ \frac{\Pr(Y \leq j | \underline{x})}{\Pr(Y > j | \underline{x})} \right] = \ln \left[ \frac{\Pr(Y = 1 | \underline{x}) + \dots + \Pr(Y = j | \underline{x})}{\Pr(Y = j+1 | \underline{x}) + \dots + \Pr(Y = k | \underline{x})} \right], \quad j=1, \dots, k-1 \quad (1)$$

Levando-se em consideração as  $p$  covariáveis, a forma do modelo MCP é:

$$\lambda_j(\underline{x}) = \alpha_j + (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) = \alpha_j + \beta^T \underline{x}, \quad j=1, \dots, k-1 \quad (2)$$

O termo  $\alpha_j$ , intercepto do modelo, varia para cada uma das  $k$  categorias e  $\beta$  é um vetor cujos elementos correspondem aos efeitos das covariáveis na variável resposta. O vetor  $\beta$  não depende do índice  $j$ , implicando que a relação entre  $\underline{x}$  e  $Y$  é independente da categoria. Para uma variável explicativa binária  $\underline{x}$ , esse vetor  $\beta$  representa o *logito* da razão de chances da resposta  $Y$  pela associação com  $\underline{x}$ , controlado pelas demais covariáveis.<sup>12</sup>

Logo, o modelo possui uma suposição de chances proporcionais acerca dos  $(k-1)$  pontos de corte, também chamada de suposição de regressão paralela, que é assumida para cada covariável incluída no modelo. Essa suposição deve ser testada para cada covariável e no modelo final, utilizando, por exemplo, o teste escore.<sup>11</sup>

Esse modelo é apropriado para analisar variáveis ordinais provenientes de uma variável contínua que foi agrupada.

### Modelo de Chances Proporcionais Parciais

Como a suposição de chances proporcionais é difícil de ser encontrada na prática, alternativamente pode ser utilizado o modelo de chances proporcionais parciais.<sup>16</sup> Este modelo permite que algumas covariáveis possam ser modeladas com a suposição de chance proporcional, mas, para as variáveis em que essa suposição não é satisfeita, é

incrementado por um coeficiente ( $\gamma$ ) que é o efeito associado com cada  $j$ -ésimo logito cumulativo, ajustado pelas demais covariáveis.<sup>11</sup>

A forma geral do modelo é a mesma anterior, isto é,  $\lambda_j(\underline{x}) = \ln \left[ \frac{\Pr(Y \leq j | \underline{x})}{\Pr(Y > j | \underline{x})} \right]$ , mas

agora:

$$\lambda_j(\underline{x}) = \alpha_j + [(\beta_1 + \gamma_{j1})x_1 + \dots + (\beta_q + \gamma_{jq})x_q + (\beta_{q+1}x_{q+1}) + \dots + (\beta_p x_p)], j = 1, \dots, k - 1 \quad (3)$$

Normalmente é esperado que haja um tipo de tendência linear entre cada razão de chance dos pontos de corte específicos e a variável resposta.<sup>1</sup> Nesse caso, um conjunto de restrições ( $\gamma_{kl}$ ) podem ser incluídas no modelo, para esclarecer essa linearidade. Quando essas restrições são incorporadas, esse modelo é chamado de modelo de chances proporcionais parciais restrito, cujo logito passa a ter a seguinte forma:

$$\lambda_j(\underline{x}) = \alpha_j + \{\tau_j [(\beta_1 + \gamma_1)x_1 + \dots + (\beta_q + \gamma_q)x_q] + (\beta_{q+1}x_{q+1}) + \dots + (\beta_p x_p)\}, j = 1, \dots, k - 1 \quad (4)$$

Os parâmetros  $\tau_j$  são escalares fixos que tomam a forma de restrições alocadas nos parâmetros. Nesse caso, para uma dada covariável  $X_m$ ,  $\gamma_m$  não depende dos pontos de corte, mas é multiplicado por  $\tau_j$  para cada  $j$ -ésimo logito.<sup>12</sup>

### Modelo de Razão Contínua

Esse modelo compara a probabilidade de uma resposta igual à categoria com determinado escore, digamos  $y_j$ ,  $Y = y_j$ , com a probabilidade de uma resposta maior,  $Y >$

$$y_j, \text{ isto é: } \lambda_j(\underline{x}) = \ln \left[ \frac{\Pr(Y = j | \underline{x})}{\Pr(Y > j | \underline{x})} \right] = \ln \left[ \frac{\Pr(Y = j | \underline{x})}{\Pr(Y = j+1 | \underline{x}) + \dots + \Pr(Y = k | \underline{x})} \right].$$

A forma do modelo é:

$$\lambda_j(\underline{x}) = \alpha_j + \beta_j^T \underline{x}, j = 1, \dots, k \quad (5)$$

Aqui  $\alpha_j$  é o intercepto do modelo e  $\beta_j$  é o vetor de coeficientes das covariáveis  $\underline{x}$ . Esse modelo possui diferentes constantes e coeficientes para cada comparação e pode ser ajustado através de  $k$  modelos de regressão logística binária.<sup>11</sup>

Esse tipo de modelo é mais apropriado quando há um interesse intrínseco em uma categoria específica da variável resposta.<sup>1</sup>

### **Modelo *Estereótipo***

O modelo *estereótipo* pode ser considerado uma extensão do modelo de regressão multinomial<sup>11</sup> e compara cada categoria da variável resposta com uma categoria de referência, que normalmente é a primeira categoria ou a última. Entretanto, devido ao caráter ordinal dos dados é imposta uma estrutura linear ao *logito da chance*  $\beta_{jl}$  ( $j=1,\dots,k$  e  $l=1,\dots,p$ ), ou seja, são atribuídos pesos ( $\omega_j$ ) aos coeficientes<sup>12</sup> tal que a forma do modelo é:

$$\lambda_j(\underline{x}) = \ln \left[ \frac{\Pr(Y = j | \underline{x})}{\Pr(Y = 0 | \underline{x})} \right] = \alpha_j + \omega_j (\beta_1 x_1 + \dots + \beta_p x_p), \quad j = 1, \dots, k \quad (6)$$

Os pesos do modelo são diretamente relacionados com o efeito das covariáveis. Então, a razão de chances formada terá uma tendência de crescimento, já que os pesos normalmente são construídos com ordenação ( $0 = \omega_1 \leq \omega_2 \leq \dots \leq \omega_k$ ).

Esse modelo deve ser utilizado quando a variável resposta é uma variável ordinal com categorias discretas.

Em todos os modelos ordinais mencionados a significância dos coeficientes deve ser testada por meio do teste de Wald<sup>11</sup> e nesse artigo ele foi calculado utilizando a aproximação pela distribuição normal padronizada.

### **VERIFICAÇÃO DA QUALIDADE DO AJUSTE DOS MODELOS ORDINAIS**

Assim como em qualquer tipo de análise de regressão, é importante avaliar a qualidade do ajuste dos modelos de regressão logística ordinal, pois a falta de ajuste pode, por exemplo, levar à viés de estimação de efeitos importantes. A avaliação do ajuste pode detectar: covariáveis importantes; interações omitidas; casos em que a função de ligação (logito) não foi apropriada; casos em que a forma funcional da modelagem das

covariáveis não está correta; e, finalmente, casos em que a suposição de chances proporcional foi violada.<sup>4</sup>

Embora muitos métodos tenham sido desenvolvidos para avaliar o ajuste de modelos de regressão logística binária, poucos desses métodos foram estendidos para dados de resposta ordinal.<sup>11</sup>

Normalmente, a qualidade do ajuste dos modelos ordinais é verificada usando os testes de Pearson ou *deviance*. Esses testes envolvem a criação de uma tabela de contingência na qual as linhas consistem de todas as possíveis configurações das covariáveis do modelo e as colunas são as categorias da resposta ordinal.<sup>17</sup> As contagens esperadas dessa tabela são expressas por  $E_{lj} = \sum_{l=1}^{N_l} \hat{p}_{lj}$ , onde  $N_l$  é o número total de indivíduos classificados na linha  $l$  e  $\hat{p}_{lj}$  representa a probabilidade de um indivíduo na linha  $l$  ter a resposta  $j$  calculada para a partir do modelo adotado.<sup>17</sup>

O teste de Pearson para avaliar a adequação do ajuste compara essas contagens esperadas com as observadas pela fórmula:

$$\chi^2 = \sum_{l=1}^L \sum_{j=1}^k \frac{(O_{lj} - E_{lj})^2}{E_{lj}} \quad (7)$$

A estatística *deviance* também compara contagens observadas e esperadas, mas através da fórmula:

$$D^2 = 2 \sum_{l=1}^L \sum_{j=1}^k O_{lj} \log \frac{O_{lj}}{E_{lj}} \quad (8)$$

Os testes para avaliar a qualidade do ajuste do modelo são baseadas nas aproximações das estatísticas (6) e (7) para a distribuição qui-quadrado com  $(L-1)(k-1)p$  graus de liberdade ( $L$  é o número de linhas,  $k$  o número de colunas da tabela de contingência e  $p$  o número de covariáveis do modelo). Diferenças significativas levam à conclusão de falta de ajuste do modelo aos dados estudados.<sup>17</sup>

Pulkstenis e Robinson<sup>17</sup> (2004) relatam que as estatísticas (7) e (8) não fornecem uma boa aproximação da distribuição qui-quadrado quando são ajustadas covariáveis contínuas e propõem pequenas modificações neste caso.

Nesse estudo, em todos os modelos considerados, foram utilizados os testes de qualidade do ajuste de Pearson ou *deviance*, uma vez que se encontram implementados nos *softwares* estatísticos usuais.

Como a literatura sobre ferramentas de diagnóstico ou avaliação de ajuste para modelos ordinais é relativamente escassa, Hosmer e Lemeshow<sup>11</sup> (2000) sugerem que devem ser utilizadas regressões binárias separadas para cada ponto de corte para que sejam criadas as estatísticas de diagnóstico para os modelos ordinais. Normalmente são construídos gráficos dos resíduos para os modelos de chances proporcionais usando o ajuste desses modelos para prever uma série de eventos binários  $Y \geq j$ ,  $j=1,2,\dots,k$ . Assim, para a variável indicadora  $[Y \geq j]$ , o resíduo score, para o caso  $i$  e covariável  $p$  é dados por:<sup>10</sup>

$$U_{ip} = X_{ip}([Y_i \geq j] - \hat{P}_{ij}), \hat{P}_{ij} = \frac{1}{1 + \exp[-(\hat{\alpha}_j + \underline{x}\hat{\beta})]} \quad (9)$$

Seja  $\underline{x}$  o vetor de covariáveis associadas ao caso  $i$ . Nos gráficos do resíduo score são colocados no eixo vertical a média  $\bar{U}_{.p}$  e os respectivos intervalos de confiança, com as categorias da variável resposta no eixo horizontal. Se a suposição de chances proporcionais for válida, para cada covariável, os intervalos de confiança para cada categoria da variável resposta devem ter uma aparência semelhante.<sup>10</sup>

Existem ainda os resíduos parciais que são muito usados para verificar se todas as covariáveis do modelo se comportam de forma linear. No contexto de regressão ordinal, é necessário calcular modelos de regressão logística binária para todos os pontos de corte da variável resposta  $Y$ , sendo o resíduo parcial para cada caso  $i$  e covariável  $p$  definido da seguinte forma:<sup>10</sup>

$$r_{ip} = \hat{\beta}_p X_{ip} + \frac{[Y_i \geq j] - \hat{P}_{ij}}{\hat{P}_{ij}(1 - \hat{P}_{ij})}, \hat{P}_{ij} = \frac{1}{1 + \exp[-(\hat{\alpha}_j + \underline{x}\hat{\beta})]} \quad (10)$$

Os gráficos de resíduos parciais geram estimativas de como  $X_p$  se relaciona com cada ponto de corte de  $Y$ .<sup>10</sup>

Assim, os resíduos parciais são usados para verificar a necessidade de transformações nas covariáveis (linearidade) ou mesmo a validade da suposição de chances proporcionais (paralelismo das curvas).<sup>10</sup>

## IMPLEMENTAÇÃO COMPUTACIONAL

Nessa seção serão mostrados todos os passos para ajuste dos modelos mostrados na Seção 2 nos *softwares* R ou STATA®.

O Quadro 1 mostra uma lista com alguns dos principais comandos do R e quando devem ser utilizados. O Quadro 2 mostra uma comparação dos principais comandos para ajuste dos modelos ordinais tanto no *software* R, quanto no STATA. Esses comandos serão detalhados no texto para uma melhor compreensão.

A maioria dos comandos será ilustrada com o banco de dados do segundo exemplo, mencionado na introdução, denominado *NHANES.dta* e cujas variáveis foram nomeadas da seguinte forma:

- Variável resposta *y*: *saude*
- Covariáveis *X*: *idade, insuf\_cardiaca, diabetes, cor\_pele*

### *Software R*

Primeiramente, será dada uma noção geral de como trabalhar no R e em seguida será feito um breve resumo dos comandos principais. Mais informações podem ser obtidas no próprio *site* do Projeto R (<http://www.r-project.org/about.html>) ou por meio de apostilas introdutórias como a desenvolvida por Colin<sup>6</sup> (2004).

Como o R fornece uma interface de entrada por linha de comando, todos os comandos são digitados. Assim, os comandos devem ser digitados após o sinalizador de comando “>”. É importante saber que há uma distinção entre letras minúsculas e maiúsculas e que as funções em R sempre aparecem entre parênteses. Além disso, pode-se designar nomes ou atributos a objetos R usando a combinação “<-” ou “=”.

Por exemplo:

```
> x<-1+2 ou
```

```
> x=1+2
```

```
> x
```

```
[1] 3
```

Agora, considere o banco de dados *NHANES.dta*. Esse banco tem a extensão *.dta*, já que é proveniente do pacote estatístico STATA. Para leitura dos dados no R deve-se

adicionar o módulo *foreign*, que lê dados em diversos formatos. Deve-se executar os seguintes comandos:

```
> library(foreign)
> nhanes=read.dta ("c:/NHANES.dta")
```

Dessa forma, o banco de dados que estava dentro do diretório *c* foi lido no pacote R com o nome *nhanes*, com todas as suas variáveis.

### **Análise por meio de tabelas de contingência**

O comando *table* integrante dos módulos de instalação do programa R, possibilita o cruzamento entre variáveis categóricas e o comando *prop.trend.test* retorna o cálculo do teste qui-quadrado de tendência para as variáveis em questão.

```
> table(nhanes$diabetes, nhanes$saude)
> prop.trend.test(nhanes$diabetes, nhanes$saude)
```

### **Ajuste dos Modelos no Software R**

#### **A) Modelo de Chances Proporcionais (MCP)**

No *software* R o modelo de chances proporcionais pode ser ajustado por meio do comando *lrm*, desenvolvido por Frank Harrell e integrante do pacote *Design*. Esse comando ajusta modelos de regressão logística binária e ordinal de chances proporcionais usando o método de máxima verossimilhança ou alternativamente, máxima verossimilhança penalizada.<sup>18</sup>

Os argumentos utilizados são: fórmula, ou seja, os termos a serem incluídos no modelo (variável resposta e covariáveis), nome do banco de dados a ser utilizado, entre outros.

O modelo pode ser especificado com os comandos:

```
> library(Design)
> mcp=lrm(saude~idade+insuf_cardiaca+diabetes+cor_pele, nhanes)
> mcp
```

As saídas mostradas após execução dos comandos são: expressão utilizada, tabela de frequências para a resposta, vetor com algumas estatísticas importantes, estimativas dos

coeficientes, vetor das primeiras derivadas do log da função de verossimilhança e *deviance* do modelo.<sup>18</sup>

## B) Modelo de Razão Contínua (MRC)

O modelo de razão contínua (MRC) pode ser implementado no *software* R por meio de uma reestruturação dos dados que é feita com o comando *cr.setup*, integrante do pacote *Design*. Esse comando cria várias novas variáveis a partir da variável resposta *y* que serão usadas no ajuste do modelo de razão contínua<sup>1;18</sup> da seguinte forma:

```
> cr=cr.setup(nhanes$saude)
```

Quatro novas variáveis são adicionadas através desse comando:

- 1- *y* → nova variável binária que será usada como resposta no ajuste do modelo de regressão logística binária;
- 2- *cohort* → vetor indicando qual ponto de corte (das comparações do MRC) foi aplicado;
- 3- *subs* → vetor utilizado para replicar as outras variáveis (explicativas) da mesma forma que *y* foi replicado;
- 4- *reps* → variável que especifica quantas vezes cada observação original foi replicada.

O modelo é obtido ajustando uma regressão logística binária nos dados reestruturados com uma nova resposta dicotômica (*y*) como variável dependente, incluindo como uma covariável a variável criada (*cohort*) que indica o nível do ponto de corte e reestruturando as covariáveis através do vetor (*subs*).<sup>1;18</sup> Os comandos são:

```
> saude.cr=cr$y
> cohort=cr$cohort
> idade.cr=idade[cr$subs]
> insuf.cr=insuf_cardiaca[cr$subs]
> diabetes.cr=diabetes[cr$subs]
> cor_pele.cr=cor_pele[cr$subs]
> mrc=lm(saude.cr~idade.cr+insuf.cr+diabetes.cr+cor_pele.cr+cohort)
> mrc
```

A suposição de heterogeneidade dos pontos de corte pode ser testada incluindo no modelo um termo de interação entre a exposição de interesse e a variável indicadora do ponto de corte (cohort), o que é chamado de modelo saturado, e comparando o valor do log da função de verossimilhança dos modelos com e sem o termo de interação. Se a heterogeneidade é significativa, o modelo de razão contínua pode ser facilmente adaptado com efeitos para os vários pontos de corte<sup>1;18</sup> através dos comandos:

```
>mracs=lm(saude.cr~(idade.cr+insuf.cr+diabetes.cr+cor_pele.cr)*cohort)
> mracs
```

### C) Modelo Estereótipo (MRC)

O modelo estereótipo pode ser ajustado por meio de modelos lineares generalizados que possuem matrizes de restrições estimadas. Os pesos (restrições) são estimados como parâmetros adicionais do modelo, utilizando a família multinomial para o ajuste.

No *software* R, o modelo estereótipo pode ser ajustado por meio do comando *rrvglm* desenvolvido por Thomas W. Yee e integrante do pacote VGAM.<sup>19;22</sup> O pacote VGAM não pode ser instalado diretamente dentro do *software* R e encontra-se disponível no *site* do autor.<sup>19</sup> Dessa forma, com os comandos listados abaixo, o modelo estereótipo pode ser ajustado:

```
> install.packages
("VGAM", repos="http://www.stat.auckland.ac.nz/~yee")
> library(VGAM)
> ms=rrvglm(saude~idade+insuf_cardiaca+diabetes+cor_pele,nhanes,multinomial)
> summary(ms)
```

### Análise de Resíduos

A função *residuals.lrm* do pacote *Design* é usada para construir gráficos dos resíduos (escore ou parcial) após ajuste do MCP no *software* R, por meio dos seguintes comandos:<sup>18</sup>

```
> mcp=lm(saude~idade+insuf_cardiaca+diabete+cor_pele,nhanes,x=TRUE,y=TRUE)
> par(mfrow=c(1,5))
```

```
> residuals (mcp, type='score.binary', pl=TRUE)
> residuals (mcp, type='partial', pl=TRUE)
```

No gráfico do resíduo escore (*score.binary*), se a suposição de chances proporcionais for válida, é esperado que para cada covariável a tendência em torno das categorias da variável resposta tenha um comportamento horizontal constante. Já no gráfico do resíduo parcial (*partial*), para um modelo bem ajustado, espera-se que as curvas sejam lineares e paralelas.<sup>10</sup>

## **Software STATA®**

### **Ajuste do Modelo de Chances Proporcionais Parciais (MCP)**

O modelo de chances proporcionais parciais não está implementado no *software* R, mas pode ser ajustado no STATA® 9.0 utilizando o comando *gologit2* desenvolvido por Willians<sup>8;21</sup> (2006). Esse comando testa a suposição de chances proporcionais por meio da opção *autofit* e ajusta coeficientes para as várias categorias das variáveis em que essa suposição é violada. Existe ainda a alternativa de parametrização do modelo através da opção *gamma*. De uma forma geral, para os dados do segundo exemplo, o MCP pode ser executado no *software* STATA® por meio dos comandos:

```
. use c:/NHANES.dta
. gologit2 saude idade cor_pele diabetes insuf_cardiaca, autofit lrforce
```

## **EXEMPLOS DE APLICAÇÃO**

### **Exemplo 1 - Estudo dos Portadores de Marcapasso**

Esse exemplo foi dividido em duas partes. Na primeira parte foi utilizada como variável resposta o domínio arritmia da escala AQUAREL (Assessment of QUALity of life and RELated events). Em seguida, a variável resposta foi o domínio aspecto emocional da escala SF-36 (*Medical Outcomes Study 36 Item Short Form Healthy Survey*).<sup>15</sup>

### Escala AQUAREL

O domínio arritmia da escala AQUAREL foi escolhido arbitrariamente entre os três domínios da escala por resultar em um exemplo didático. Os escores desse domínio podem variar entre zero e 100. Para o exemplo, esses escores foram categorizados utilizando como pontos de corte seus quartis (1º quartil=70, 2º quartil=85 e 3º quartil=100). Assim, foi criada uma variável ordinal com quatro categorias: QV ruim (escores de 0 a 69,9), QV razoável (escores de 70 a 84,9), QV boa (escores de 85 a 99,9) e QV muito boa (escore 100).

Após análise univariada, foram selecionadas dentre todas as informações clínicas e sociodemográficas do estudo as covariáveis significativas. Para facilitar a construção do exemplo e considerar poucas covariáveis no modelo, utilizou-se um nível de significância conservador de 5%, diferente dos níveis usados habitualmente na análise univariada. Como mostrado na Tabela 1, as variáveis selecionadas (valor- $p \leq 0,05$ ) foram: sexo (feminino/masculino), uso do medicamento digoxina (sim/não), uso de medicamento antiarrítmico (sim/não), portador de doença de chagas (sim/não), classificação funcional (escala cujos resultados podem ser pontuados de 1 a 4, de acordo com consumo metabólico) e união conjugal (com ou sem parceiro).<sup>15</sup> A análise univariada mostrou que os portadores de marcapasso que utilizam medicamentos como digoxina ou antiarrítmicos, que possuem doença de Chagas, ou que pertencem a classificação funcional III têm maiores percentuais de QV ruim. Já os pacientes do sexo masculino ou que possuem parceiros apresentaram maiores percentuais de QV muito boa.

Optou-se por ajustar o modelo de chances proporcionais já que se trata de uma variável ordinal proveniente de uma variável contínua que foi categorizada. Iniciou-se o modelo com todas as variáveis supracitadas. As que permaneceram no modelo final foram: sexo, uso de digoxina e classificação funcional. De acordo com os resultados da Tabela 2, o teste de *deviance* (8) mostrou que o modelo apresenta bom ajuste e o teste do escore mostrou que a suposição de chances proporcionais era válida.

Os gráficos de resíduos (escore e parcial), mostrados na Figura 1, foram construídos para avaliar a adequação desse modelo aos dados analisados. A Figura 1a reforça a conclusão do teste escore, pois no gráfico do resíduo escore as covariáveis mostram um

formato horizontal razoavelmente constante em torno de zero para todas as categorias da variável resposta.<sup>10</sup> Nos gráficos de resíduos parciais (Figura 1b), para todas as variáveis explicativas o aspecto é linear e as retas são aproximadamente paralelas, mostrando que a suposição de chances proporcionais parece ser válida.

### **Escala SF-36**

O domínio aspectos emocionais da escala SF-36 foi escolhido para esse exemplo, por possuir escores que resultam numa variável ordinal com as seguintes categorias: 0; 33; 67 e 100. Após análise univariada (Tabela 1), por meio do teste qui-quadrado de tendência foram selecionadas, dentre todas as informações clínicas e sociodemográficas, as seguintes covariáveis: sexo (masculino/feminino), uso do medicamento digoxina (sim/não), uso de diurético (sim/não), uso de medicamento do sistema nervoso central (sim/não), classificação funcional (I/II/III) e união conjugal (com e sem parceiro).<sup>15</sup>

Também foi utilizado inicialmente o modelo de chances proporcionais e os resultados do modelo final ajustado estão apresentados na Tabela 2. Ao contrário do exemplo anterior, nenhuma das covariáveis adequa-se à suposição de regressão paralela como pode ser observado não só pelo teste escore (valor- $p < 0,05$ ), mas também pelo gráfico de resíduos parciais (Figura 2). Nestes gráficos as retas se cruzam, não apresentando o paralelismo esperado pelo MCP para todas as covariáveis analisadas.

Logo, o próximo passo foi utilizar o modelo de chances proporcionais parciais. Os resultados, apresentados na Tabela 3, mostram que os coeficientes de três covariáveis foram significativos apenas para a primeira comparação, que se refere aos níveis extremos (0 versus 33 a 100). Isso indica que quem tem parceiro, não usa diurético nem medicamento do sistema nervoso central tem, respectivamente, oito, três e 17 vezes mais chance de não ter uma péssima qualidade de vida (nível zero).

### **Exemplo 2 - Levantamento Nacional de Saúde e Nutrição**

Neste exemplo, a variável dependente foi a condição de saúde classificada em cinco categorias (1= ruim, 2=regular, 3=média, 4=boa e 5=excelente),<sup>14</sup> aceita como um bom indicador de qualidade de vida relacionado à saúde. Existe um grande número de

possíveis variáveis explicativas nesse banco de dados, sendo quase todas relacionadas ao evento de interesse. Usar todas essas variáveis impossibilitaria a construção de um exemplo didático. Por isso, foi utilizada a metodologia de árvore de decisão através do algoritmo CART<sup>5</sup> para seleção das variáveis mais significantes em relação ao evento condição de saúde.

O CART, abreviação de *Classification And Regression Tree* (Árvore de Classificação e Regressão) é um método baseado em árvores de decisão que predizem ou classificam observações futuras, por meio de sucessivas divisões binárias no conjunto de dados, de modo a tornar os subconjuntos resultantes cada vez mais homogêneos em relação à variável resposta. Esse método é recomendado quando há grandes bancos de dados e/ou para seleção das variáveis com os efeitos principais como é o caso desse exemplo.<sup>5</sup>

As covariáveis selecionadas pelo algoritmo CART foram idade (em anos), insuficiência cardíaca (não/sim), diabetes (não/sim) e cor de pele (branca/negra/outras). Para a variável cor de pele foram criadas variáveis indicadoras, considerando como referência a cor negra.

Estão apresentados na Tabela 4 os resultados das análises univariada e múltipla do modelo de chances proporcionais. Observa-se que as estimativas dos coeficientes e da razão de chances da análise univariada diferem muito da análise múltipla, sugerindo um efeito de confusão entre as variáveis explicativas (principalmente diabetes, insuficiência cardíaca e idade), ajustadas no modelo multivariado.

Na análise univariada, o teste escore para as variáveis insuficiência cardíaca e diabetes não sugere violação da suposição de chances proporcionais. Entretanto, o mesmo não ocorre para as variáveis cor de pele e idade, cujos testes de regressão paralela foram altamente significantes. Também para o modelo múltiplo, a suposição de chances proporcionais foi violada. Além disso, o teste de *deviance* indicou falta de ajuste para o modelo de chances proporcionais.

Os gráficos de resíduos (escore e parcial) para avaliar a adequação do modelo de chances proporcionais estão apresentados na Figura 3. O formato dos gráficos da Figura 3a (resíduo escore) reforça a conclusão do teste escore, pois as variáveis insuficiência cardíaca e diabetes mostram um formato horizontal próximo do zero. Entretanto, a variável idade apresenta uma oscilação no comportamento para as categorias de

condição de saúde boa e média estando bem abaixo da linha do resíduo zero. O mesmo ocorre para a covariável cor de pele. Entretanto, nesse caso, a oscilação maior foi observada nas categorias de condição de saúde razoável e ruim para a classificação outras e na categoria média para a cor branca.

Nos gráficos do resíduo parcial (Figura 3b), a suposição de regressão paralela é bastante razoável (aspecto linear e retas aproximadamente paralelas) para as variáveis explicativas insuficiência cardíaca e diabetes. Já no gráfico para a categoria outras da variável cor de pele, apesar do comportamento linear, observa-se cruzamento das curvas, violando, portanto, a suposição do paralelismo. A covariável idade não apresenta um comportamento linear, o que poderia contribuir para a falta de ajuste do modelo. Mesmo incluindo termos de graus mais elevados para a idade, o teste de *deviance* continuou mostrando um ajuste ruim.

Considerando o ajuste ruim, foi avaliado neste conjunto de dados o modelo estereótipo, apresentado na Tabela 5. Em todos os casos, o efeito das covariáveis foi significativo (valor- $p < 0,01$ ), e o teste *deviance* indicou um bom ajuste do modelo. A interpretação das estimativas da razão de chances é de que pessoas com insuficiência cardíaca têm quase 15 vezes mais chance de ter uma condição de saúde ruim comparada à condição excelente. Essa razão de chances diminui à medida que a condição de saúde se aproxima da excelente, chegando a 1,56 na comparação de saúde boa e excelente.

## DISCUSSÃO

Como discutido por alguns autores<sup>1;3;9;11;12</sup>, os modelos de regressão logística ordinal são recomendados para análise de dados ordinais em várias ocasiões. Ananth & Kleinbaum<sup>1</sup> (1997) relataram que os modelos de chances proporcionais e de razão contínua são os mais utilizados em aplicações epidemiológicas e biomédicas em relação aos demais modelos. Entretanto, esses modelos possuem fortes suposições que, se não válidas, podem levar a conclusões incorretas.<sup>1</sup> Como foi observado no exemplo 1 deste artigo, em que a resposta era o domínio aspectos emocionais da escala SF-36, os gráficos de resíduos mostraram retas não paralelas, indicando violação da premissa principal do modelo de chances proporcionais. Inferências baseadas nesse modelo poderiam ter levado a conclusões incorretas, dado que apenas na primeira comparação

foram encontradas diferenças estatisticamente significativas. Nesse caso, uma alternativa foi o ajuste do modelo de chances proporcionais parciais, com estimativa da razão de chances para as comparações.

Lall et al<sup>12</sup> (2002) também discutem a utilização dos modelos de regressão logística ordinal. Estes autores afirmam que o tipo de modelo a ser utilizado depende do caráter da variável resposta ordinal, ou seja, quando as categorias ordenadas são provenientes de uma variável latente subjacente ou quando as categorias ordenadas são de origem discretas. O exemplo 1 pode ser utilizado para estas considerações. A variável resposta referente à QV no domínio arritmia da escala AQUAREL era uma variável ordinal proveniente de uma variável contínua que foi agrupada. Neste caso, o modelo de chances proporcionais foi o mais indicado e a premissa de retas paralelas mostrou-se válida. Já no exemplo 2, que tinha como variável resposta a condição de saúde avaliada no estudo NHANES II, o modelo estereótipo foi o mais indicado por se tratar de uma variável ordinal com categorias discretas. Esse modelo, apesar de estimar um número maior de parâmetros, sendo menos parcimonioso, foi o mais adequado por apresentar melhor ajuste que os demais de acordo com o teste *deviance*.

Na presença de uma resposta ordinal, outras opções de análises multivariadas devem ser consideradas. Um caminho, aqui adotado, foi utilizar a metodologia de árvores de decisão. Dentre os algoritmos de árvores de decisão existentes, o CART<sup>5</sup> (*Classification and Regression Trees*) foi utilizado para seleção dos efeitos principais e posterior inclusão no modelo multivariado. Esse tipo de análise também pode ser útil para verificação de fatores associados, assim como a regressão logística ordinal. Entretanto, é mais descritivo e considera um número maior de variáveis no modelo final. Em contraponto, a regressão é uma técnica paramétrica, que apesar de impor uma estrutura rígida do modelo, é mais conservadora e parcimoniosa, além de permitir a estimação de intervalos de confiança e testes de hipóteses para os parâmetros.

Existem outras formas alternativas de modelagem para dados ordinais, por exemplo a utilização de outras funções de ligação, como probito e complementar log-log. No entanto, tais abordagens merecem uma análise e discussão à parte deste artigo.

Na construção dos modelos ordinais, Hosmer & Lemeshow<sup>11</sup> (2000) propõem estratégias como as adotadas nos exemplos 1 e 2 deste artigo. Recomendam

inicialmente fazer uma análise univariada para seleção dos efeitos principais e incluir no modelo apenas as variáveis significativas com um nível de significância pré-fixado. Em seguida, ajustar o modelo, verificar sua adequação por meio dos testes adequados e gráficos de resíduos e, por fim, interpretar o modelo por meio da estimativa da razão de chances.

Deve-se destacar, nesta última parte da recomendação dos autores acima, a escassez de métodos para verificação do ajuste dos modelos ordinais. Até o momento não nos foi possível encontrar na literatura nenhuma técnica de verificação do ajuste do modelo estereótipo. As estatísticas de diagnóstico existentes, propostas por Harrel<sup>10</sup> (2002) e aplicáveis ao modelo de chances proporcionais, nada mais são do que gráficos traçados a partir de regressões binárias separadas para os pontos de corte da variável ordinal. Embora incompletas até o momento, essas técnicas são de grande importância para se ter uma indicação da qualidade de ajuste dos modelos ordinais. A análise dos resíduos parciais, mesmo que gráfica, é considerada muito útil para modelos ordinais, pois eles simultaneamente checam linearidade, indicando eventuais transformações que devem ser utilizadas, bem como o pressuposto de chances proporcionais.

Por outro lado, deve-se ter cuidado na interpretação dos resíduos, principalmente considerando a escassa existência de informações de como fazê-la. Os gráficos por vezes podem sugerir informações confusas e dificultar a tomada de decisões quanto à violação ou não da suposição de chances proporcionais. Uma alternativa é utilizar a análise dos resíduos conjuntamente com o teste do escore, pois quando houver alguma dúvida quanto ao formato do gráfico, esse teste pode contribuir para a conclusão final.

Por fim, deve-se ressaltar que uma boa implementação computacional e o domínio dos comandos para execução dos modelos ordinais são essenciais até mesmo para que esses modelos sejam comparados e a escolha do modelo mais adequado seja feita. Para isso, o programa R torna-se uma ferramenta importante, trazendo vários modelos, além dos gráficos de diagnóstico propostos por Harrell<sup>10</sup>. Um ponto a ser acrescentando é, ainda, a falta do modelo de chances proporcionais parciais no R, embora sua implementação já esteja em andamento.

Finalmente, os modelos de regressão logística ordinal têm-se mostrado apropriados para análise de dados com resposta ordinal e a escolha do melhor modelo depende do caráter

da variável ordinal, adequação do modelo às suposições, qualidade do ajuste e capacidade de boa explicação com reduzido número de parâmetros a serem estimados.

## AGRADECIMENTOS

As autoras agradecem a FAPEMIG, pelo financiamento do Projeto “Qualidade de Vida na Esquizofrenia”, processo nº CDS-301/02 e também ao CNPQ pela bolsa de produtividade em pesquisa da Dra. Waleska Teixeira Caiaffa e a CAPES pela bolsa de mestrado de Mery Natali Silva Abreu.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. *Int. J. Epidemiol.* 1997; 26:1323-33.
2. Anderson JA. Regression and ordered categorical variables. *J. R. Statist. Soc.* 1984, 16: 1-30.
3. Bender R, Benner A. Calculating ordinal regression models in SAS and S-Plus. *Biometrical Journal.* 2000; 6: 677-699.
4. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 1990; 46:1171-1178
5. Breiman L et al. *Classification and regression trees*. Pacific Grove, Califórnia: Wadsworth and Brooks, 1984.
6. Colin RB. *Bioestatística usando R - apostila para biólogos*. Bragança, 2004.
7. Fienberg SE. *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA; MIT Press; 1980. p. 110-116.
8. gologit2: Generalized ordered logit/ partial proportional odds models for ordinal dependent variables. <http://www.nd.edu/~rwilliam/gologit2> (acessado em 10 de novembro de 2006).
9. Greenland S. Alternative models for ordinal logistic regression. *Stat. Med.* 1994;13(16):1665-77.

10. Harrell, Jr., FE. *Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer New York, 2002.
11. Hosmer WD, Lemeshow S. *Applied Logistic Regression*. Wiley; 2000.
12. Lall R, Campbell MJ, Walters SJ, Morgan K. A review of ordinal regression models applied on health-related quality of life assessments. *Stat. Methods Med. Res.* 2002; 11: 49-67.
13. McCullagh P. Regression models for ordinal data. *J. R. Stat. Soc. Ser. B. Methodol.* 1980; 42: 109-142.
14. NHANES II Public-use Data Files. [http:// www.cdc.gov/ nchs/ products/ elec\\_prods/ subject/ nhanesii.htm](http://www.cdc.gov/nchs/products/elec_prods/subject/nhanesii.htm) (acessado em 10 de novembro de 2006).
15. Oliveira BG et al. Versão em Português, Adaptação Transcultural e Validação de Questionário para Avaliação da Qualidade de Vida para Pacientes Portadores de Marcapasso: AQUAREL. *Arq. Bras. Cardiol.* 2006; 87: 75-83.
16. Peterson BL, Hanrrel FE. Partial proportional odds models for ordinal response variables. *Appl. Statistic.* 1990; 39:205-217.
17. Pulkstenis E, Robinson TJ. Goodness-of-fit tests for ordinal response regression models. *Stat. Med.* 2004; 23(6):999-1014.
18. The R Project for Statistical Computing <http://www.r-project.org/about.html> (acessado em 10 de novembro de 2006).
19. Thomas Yee's Personal Home Page. <http://www.stat.auckland.ac.nz/~yee> (acessado em 10 de novembro de 2006).
20. Walker SH, Ducan DB. Estimation of the probability of an event as a function of several independent variables. *Biometrika* 1967; 54:167-179.
21. Williams R. gologit2: Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *Forthcoming in Stata Journal*, 2006.
22. Yee TW, Hastie TJ. Reduced-rank vector generalized linear models. *Statistical Modelling* 2003; 3: 15-41.

Tabela 1: Distribuição de frequência de alguns fatores segundo a classificação da QV no domínio Arritmia da escala AQUAREL e resultados do teste qui-quadrado de tendência

	Domínio Arritmia (AQUAREL)					Valor-p <sup>1</sup>
	Total (n)	QV ruim (%)	QV razoável (%)	QV boa (%)	QV muito boa (%)	
Sexo						
Feminino	84	29,8	23,8	26,2	20,2	0,005
Masculino	55	16,4	16,4	25,5	41,8	
União Conjugal						
Sem parceiro	68	30,9	23,5	23,5	22,1	0,026
Com parceiro	71	18,3	18,3	28,2	35,2	
Uso de Digoxina						
Presente	17	41,2	23,5	29,4	5,9	0,018
Ausente	111	21,6	19,8	25,2	33,3	
Classificação funcional <sup>2</sup>						
I	85	12,9	22,4	24,7	40,0	<0,001
II	31	38,7	12,9	35,5	12,9	
III	23	47,8	26,1	17,4	8,7	
Doença de Chagas						
Não	31	9,7	16,1	29,0	45,2	0,004
Sim	77	27,3	23,4	28,6	20,8	

	Domínio Aspectos Emocionais (SF-36)					Valor-p <sup>1</sup>
	Total (n)	Escore 0 (%)	Escore 33 (%)	Escore 67 (%)	Escore 100 (%)	
Sexo						
Feminino	84	51,2	8,3	8,3	32,1	0,029
Masculino	55	27,3	18,2	12,7	41,8	
União conjugal						
Sem parceiro	68	54,4	5,9	8,8	30,9	0,024
Com parceiro	71	29,6	18,3	11,3	40,8	
Uso de Digoxina						
Presente	17	64,7	5,9	11,8	17,6	0,004
Ausente	111	37,8	12,6	9,9	39,6	
Uso de Diurético						
Presente	57	52,6	10,5	8,8	28,1	0,023
Ausente	71	32,4	12,7	11,3	43,7	
Uso de MSNC						
Presente	23	73,9	4,3	4,3	17,4	0,002
Ausente	105	34,3	13,3	11,4	41,0	
Classificação funcional <sup>2</sup>						
I	85	31,8	12,9	11,8	43,5	0,011
II	31	54,8	16,1	3,2	25,8	
III	23	60,9	4,3	13,0	21,7	

<sup>1</sup>Teste qui-quadrado de tendência

<sup>2</sup>Classificação funcional: I = consumo metabólico  $\geq 7$  mets; II = consumo metabólico  $\geq 5$  e  $< 7$  mets; III = consumo metabólico  $\geq 2$  e  $< 5$  mets MSNC=Medicamento do Sistema Nervoso Central

Tabela 2: Modelos finais de chances proporcionais tendo como resposta a qualidade de vida no domínio arritmia da escala AQUAREL e o domínio aspetos emocionais da escala SF-36

Variável resposta	Covariável	Categorias	$\beta$	$EP(\beta)$	Valor-p		OR	
					Teste Wald	Teste escore	Estimativa	IC (95%)
Domínio Arritmia (AQUAREL) <sup>1</sup>	Sexo	Feminino					1,00	--
		Masculino	0,85	0,35	0,01		2,34	(1,19; 4,60)
	Uso de Digoxina	Presente					1,00	--
		Ausente	1,00	0,50	0,05	0,37	2,72	(1,01; 7,27)
	Classificação Funcional <sup>3</sup>	I	1,74	0,52	<0,01		5,67	(2,07; 15,57)
		II	0,64	0,57	0,26		1,90	(0,62; 5,82)
III						1,00	--	
Domínio Emocional (SF-36) <sup>2</sup>	União conjugal	Sem parceiro					1,00	--
		Com parceiro	0,79	0,35	0,02	<0,01	2,21	(1,12; 4,37)
Uso de Diurético (SF-36) <sup>2</sup>	Uso de Diurético	Presente					1,00	--
		Ausente	1,03	0,36	<0,01	0,01	2,80	(1,39; 5,64)
	Uso de MSNC	Presente					1,00	--
Ausente		1,71	0,52	<0,01	<0,01	5,53	(1,99; 15,34)	

<sup>1</sup>Teste de deviance (valor-p = 0,565) <sup>2</sup> Teste de deviance (valor-p = 0,302)

<sup>3</sup> Classificação funcional: I = consumo metabólico  $\geq 7$  mets; II = consumo metabólico  $\geq 5$  e  $< 7$  mets; III = consumo metabólico  $\geq 2$  e  $< 5$  mets  
EP= erro-padrão; OR = odds ratio (razão de chances) MSNC = Medicamento do Sistema Nervoso Central

Tabela 3: Modelos finais de chances proporcionais parciais tendo como resposta a qualidade de vida no domínio aspetos emocionais da escala SF-36

Covariável	Categorias	Comparações								
		0 versus (33; 67; 100)			(0 + 33) versus (67 + 100)			(0 + 33 + 67) versus 100		
		$\beta_1$	OR <sub>1</sub>	Valor-p	$\beta_2$	OR <sub>2</sub>	Valor-p	$\beta_3$	OR <sub>3</sub>	Valor-p
União conjugal	Sem parceiro		1,00	<0,01		1,00	0,71		1,00	0,97
	Com parceiro	2,09	8,08		-0,16	0,85		-0,02	0,98	
Uso de Diurético	Presente		1,00	<0,01		1,00	0,62		1,00	0,52
	Ausente	2,24	9,39		0,22	1,25		0,28	1,32	
Uso de MSNC	Presente		1,00	<0,01		1,00	0,38		1,00	0,58
	Ausente	2,85	17,29		0,57	1,77		0,38	1,46	

OR = *odds ratio* (razão de chances) MSNC = Medicamento do Sistema Nervoso Central

Tabela 4: Modelos finais de chances proporcionais tendo como resposta a condição de saúde (NHANES II)

Tipo análise	Covariável	Categorias	n (%)	$\beta$	$EP(\beta)$	OR	Valor-p	
							Teste <i>Wald</i>	Teste <i>escore</i>
Univariada	Insuficiência cardíaca	Não	9862 (95%)			1,00	<0,01	0,94
		Sim	473 (5%)	2,02	0,09	7,58		
	Diabetes	Não	9836 (65%)			1,00	<0,01	0,60
		Sim	499 (5%)	1,74	0,08	5,68		
	Cor de Pele	Negra	1086 (10%)			1,00	<0,01	0,01
		Branca	9049 (88%)	-0,75	0,06	0,47		
		Outras	200 (2%)	-0,66	0,13	0,51		
	Idade	(em anos)	48 ± 17	0,04	0,01	1,04	<0,01	<0,01
Multivariada*	Insuficiência cardíaca	Não	9862 (95%)			1,00	<0,01	<0,01
		Sim	473 (5%)	1,50	0,09	4,46		
	Diabetes	Não	9836 (65%)			1,00	<0,01	
		Sim	499 (5%)	1,21	0,09	3,36		
	Cor de Pele	Negra	1086 (10%)			1,00	<0,01	
		Branca	9049 (88%)	-0,87	0,14	0,42		
		Outras	200 (2%)	-0,64	0,06	0,53		
	Idade	(em anos)	48 ± 17	0,04	0,01	1,04	<0,01	

\*Teste de *deviance* (valor-p = 0,053)

EP= erro-padrão; OR = *odds ratio* (razão de chances)

Tabela 5: Modelo de regressão logística ordinal (estereótipo)\* tendo como resposta a condição de saúde - referência saúde excelente (NHANES II)

Covariável	Categorias	Comparações da condição de saúde							
		Ruim versus excelente		Regular versus excelente		Média versus excelente		Boa versus excelente	
		$\beta_1$	OR <sub>1</sub>	$\beta_2$	OR <sub>2</sub>	$\beta_3$	OR <sub>3</sub>	$\beta_4$	OR <sub>4</sub>
Insuficiência cardíaca	Não		1,00		1,00		1,00		1,00
	Sim	2,69	14,76	2,02	7,55	1,19	3,29	0,45	1,56
Diabetes	Não		1,00		1,00		1,00		1,00
	Sim	2,23	9,35	1,68	5,36	0,99	2,69	0,37	1,45
Cor da pele	Negra		1,00		1,00		1,00		1,00
	Branca	-1,69	0,18	-1,27	0,28	-0,75	0,47	-0,28	0,75
	Outras	-1,29	0,28	-0,97	0,38	-0,57	0,57	-0,21	0,81
Idade (em anos)		0,07	1,08	0,06	1,06	0,03	1,03	0,01	1,01

\*Teste de deviance (valor-p = 0,432)

OR = odds ratio (razão de chances) Todas as variáveis foram significativas ao nível de significância de 1%

Quadro 1: Principais comandos do *software R*

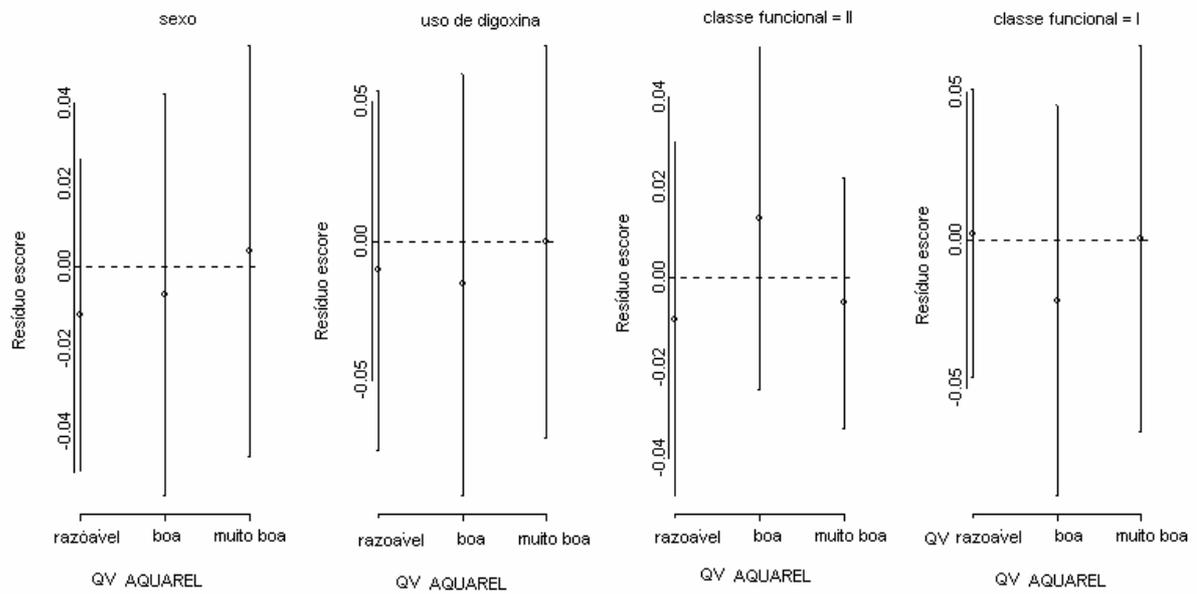
<b>Comando</b>	<b>Pacote</b>	<b>Descrição</b>
read.spss read.dta read.dbf	foreign	Lê arquivos de outras bases de dados (por exemplo, SPSS, STATA ou dbase) e importa para o programa R.
summary	base	Produz estatísticas descritivas (média, mediana, etc.) ou resultados sumarizados de ajuste de modelos de regressão.
mean	base	Calcula a média aritmética.
sd	base	Calcula o desvio-padrão.
table	base	Constrói tabelas de frequência simples ou tabelas de contingência com o cruzamento de dois fatores.
t.test	stats	Calcula o teste t.Student para uma ou duas amostras independentes.
chisq.test	stats	Calcula o teste qui-quadrado para tabelas de contingência.
prop.trend.test	stats	Calcula o teste qui-quadrado de tendência
lrm	Design	Ajusta regressão logística binária ou modelo ordinal de chances proporcionais.
cr.setup	Design	Cria várias novas variáveis para ajuste do modelo de razão contínua.
vglm	VGAM	Ajusta modelos lineares generalizados, utilizando diferentes famílias de dados, como por exemplo a multinomial.
rrvglm	VGAM	Ajusta o modelo ordinal estereótipo.
residuals	stats	Extraí os resíduos dos modelos.

Quadro 2: Comandos para execução dos modelos de regressão logística ordinal nos softwares R e STATA®, exemplificando com o banco de dados do estudo NHANES

	R	STATA®
Leitura do banco de dados	<pre>&gt; library(foreign) &gt; nhanes=read.dta ("c:/NHANES.dta") &gt; library(Design) &gt; mcp=lrn(saude~idade+insuf_cardiaca +diabetes+cor_pele,nhanes) &gt; mcp</pre>	<pre>. use c:/NHANES.dta</pre>
Modelo de chances proporcionais	<pre>&gt; mcp=lrn(saude~idade+insuf_cardiaca +diabetes+cor_pele,nhanes) &gt; mcp</pre>	<pre>. ologit saude idade insuf_cardiaca diabetes cor_pele</pre>
Modelo de chances proporcionais parciais	---	<pre>. gologit2 saude idade cor_pele diabetes insuf_cardiaca, autofit lrforce</pre>
Modelo de razão contínua	<pre>&gt; cr=cr.setup(nhanes\$saude) &gt; saude.cr=cr\$y &gt; cohort=cr\$cohort &gt; idade.cr=idade[cr\$subs] &gt; insuf.cr=insuf_cardiaca[cr\$subs] &gt; diabetes.cr=diabetes[cr\$subs] &gt; cor_pele.cr=cor_pele[cr\$subs] &gt; mrc=lrn(saude.cr~idade.cr+insuf.cr+ diabetes.cr+cor_pele.cr+cohort) &gt; mrc</pre>	<pre>. ocratio saude idade insuf_cardiaca diabetes cor_pele</pre>
Modelo estereótipo	<pre>&gt; install.packages ("VGAM", repos="http://www.stat.auckland.ac.nz/~yee") &gt; library(VGAM) &gt; ms=rrvglm(saude~idade+insuf_cardiaca+diabetes +cor_pele,nhanes,multinomial) &gt; summary(ms)</pre>	<pre>. slogit saude idade insuf_cardiaca diabetes cor_pele</pre>

\* Banco de dados: NHANES; variável resposta: saude; covariáveis: idade, insuf\_cardiaca, diabetes e cor\_pele

## (a) Resíduo escore



## (b) Resíduo parcial

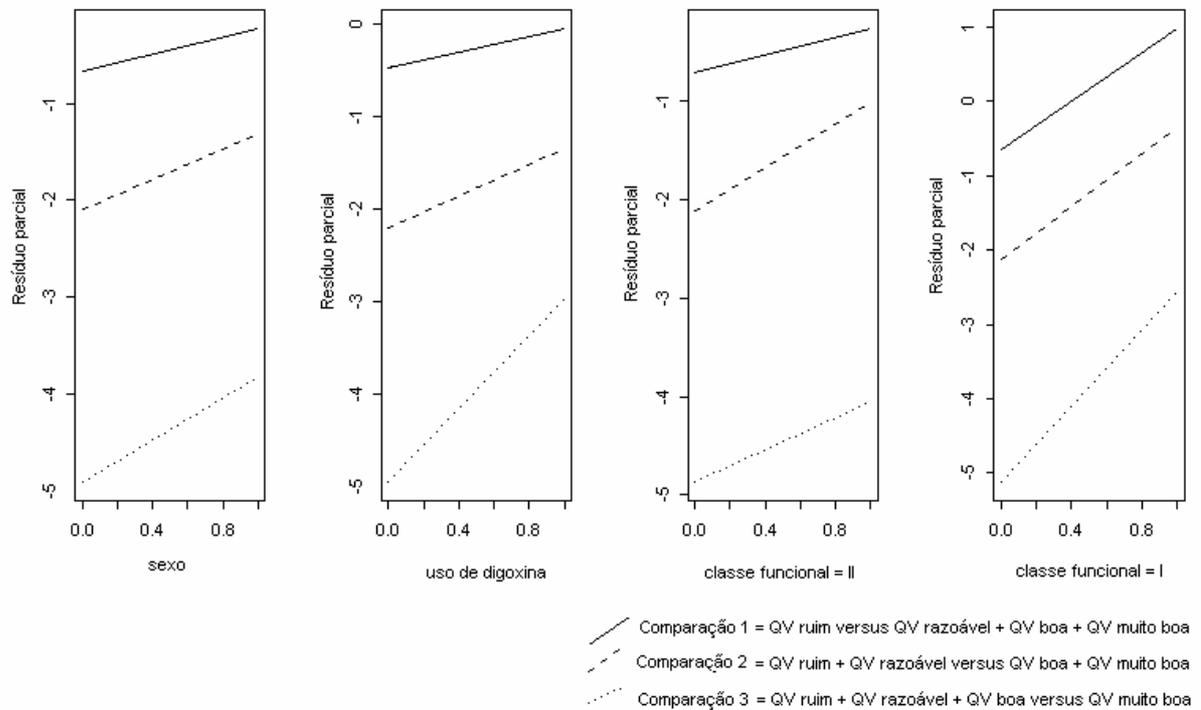


Figura 1: Gráficos de resíduos (escore e parcial) para as covariáveis incluídas no modelo tendo como resposta o domínio arritmia da escala AQUAREL

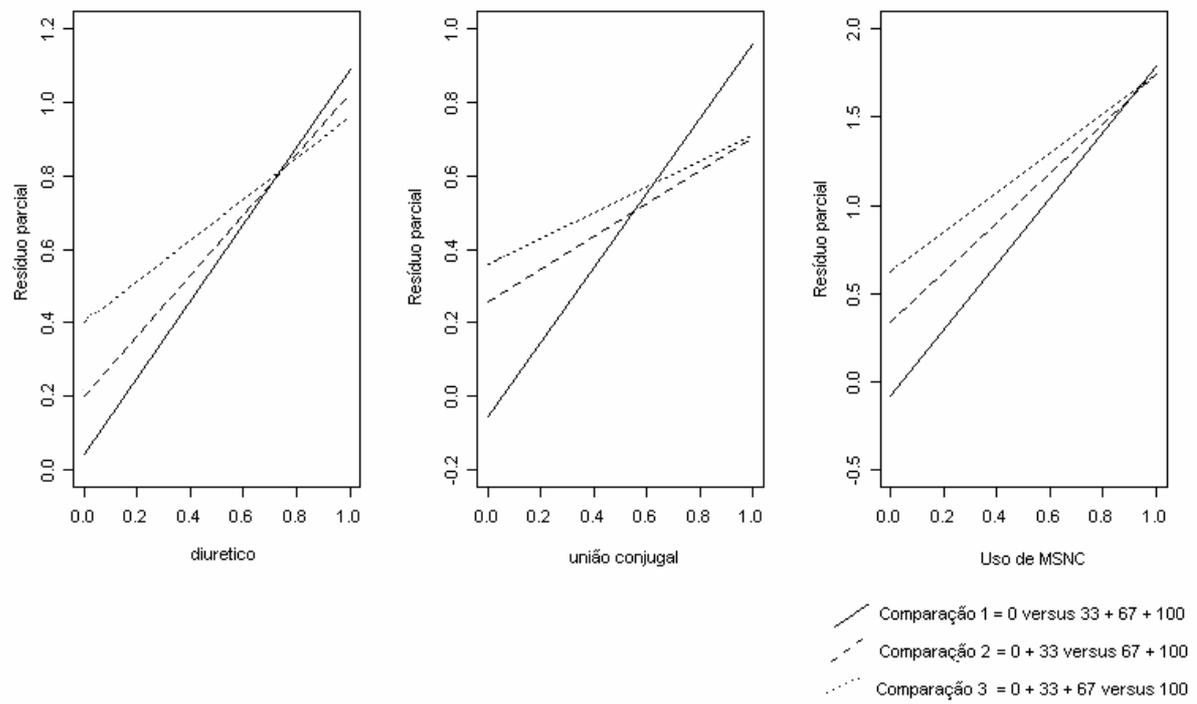
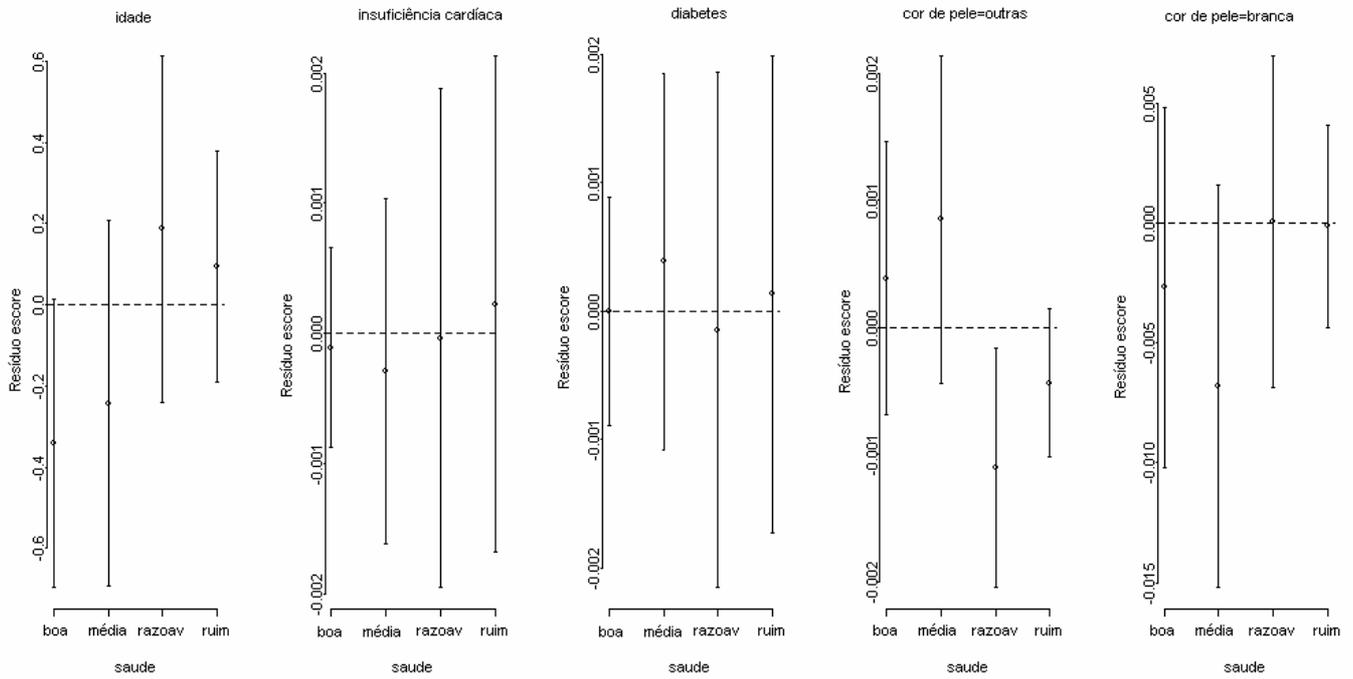


Figura 2: Gráficos de resíduos parciais as covariáveis incluídas no modelo tendo como resposta o domínio aspectos emocionais da escala SF-36

## (a) Resíduo escore



## (b) Resíduo parcial

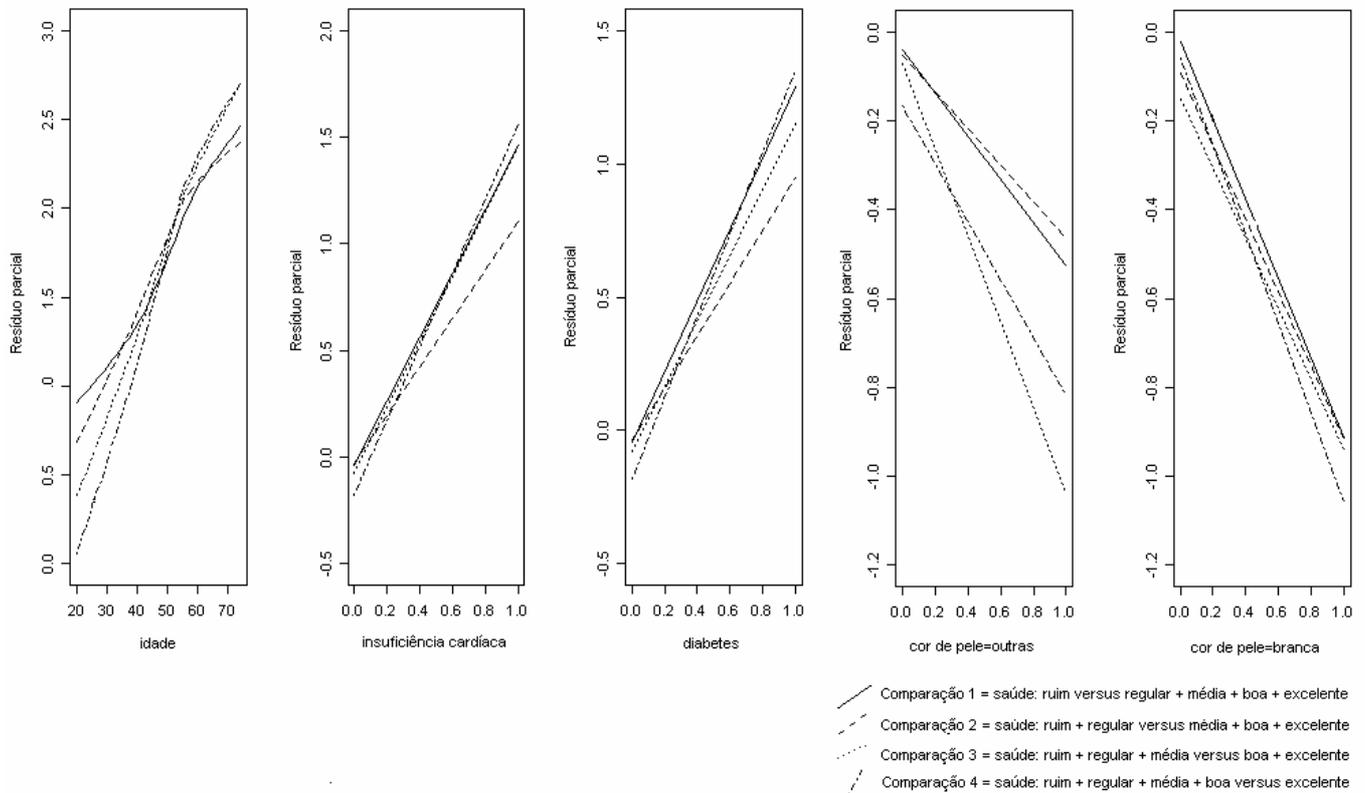


Figura 3: Grficos de resduos (escore e parcial) para as covariveis includas no modelo tendo como resposta a condiço de sade

## 5. CONSIDERAÇÕES FINAIS

O presente estudo pretende destacar que para se ter uma boa análise estatística em estudos sobre qualidade de vida, primeiramente deve-se considerar o planejamento e cálculo amostral, sempre levando em conta o caráter ordinal dos dados.

No que diz respeito à análise de fatores associados à QV, os modelos de regressão logística ordinal devem ser empregados, sendo que cada modelo investigado apresentou melhor adequação dependendo da forma como a variável resposta é apresentada.

Diante disso, uma boa implementação computacional e o domínio dos comandos para execução dos modelos ordinais é essencial até mesmo para que esses modelos sejam comparados e a escolha do mais adequado à situação seja feita.

A verificação do ajuste dos modelos ordinais, de acordo com as estatísticas apropriadas, e a utilização de gráficos de diagnósticos são outros pontos importantes nesse tipo de análise, além da avaliação da validade das suposições associadas a cada modelo.

Em resumo, a escolha do melhor modelo ordinal depende do caráter da variável resposta ordinal, adequação do modelo às suposições, qualidade do ajuste e reduzido número de parâmetros estimados.

Nesse estudo foram discutidos, como métodos de análise de fatores associados à QV, apenas os modelos de regressão logística ordinal. Entretanto, esses não são adequados quando a variável resposta ordinal tem um extenso número de categorias que não podem ser reagrupadas, como por exemplo alguns domínios da escala SF-36. Nesse caso, deve-se optar por outras formas de análise.

Outra limitação é a falta de técnicas gráficas para verificação do ajuste do modelo estereótipo. Tais técnicas ainda não se encontram disponíveis na literatura e deveriam ser estudadas.

Além disso, o modelo de chances proporcionais parciais não está implementado no *software* R, utilizado para analisar os dados do presente estudo. Entretanto, a implementação desse modelo já está em andamento e será apresentada em trabalhos futuros.

Por fim, os modelos logísticos ordinais mostraram-se apropriados para analisar os dados de QV e recomenda-se que sejam evitados procedimentos mais simples, tais como a dicotomização da variável resposta e a desconsideração da ordenação, que têm como consequência a perda de informação proporcionada pelos dados.

**APÊNDICES**

## **APÊNDICE A: PROJETO DE PESQUISA**

**Mery Natali Silva Abreu**

**USO DE MODELOS DE REGRESSÃO LOGÍSTICA ORDINAL  
EM ESTUDOS SOBRE QUALIDADE DE VIDA**

Projeto apresentado à Banca de Qualificação do Programa de Pós-Graduação em Saúde Pública, nível de mestrado, na Área de Concentração em Epidemiologia da Faculdade de Medicina da UFMG.

Orientadora: Arminda Lucia Siqueira

Co-Orientadora: Waleska Teixeira Caiaffa

Colaboradora: Clareci Silva Cardoso

**Belo Horizonte**  
**Universidade Federal de Minas Gerais**  
**Setembro/2005**

## 1. INTRODUÇÃO

O tema qualidade de vida (QV) tem ganhado ênfase nos últimos anos, mas ainda é cercado de controvérsias. O termo qualidade de vida é utilizado na linguagem cotidiana e no contexto da pesquisa científica.<sup>1</sup> Já na década de 70, Campbell disse que “QV é uma entidade vaga e etérea, algo sobre o qual muita gente fala, mas ninguém sabe o que é”.<sup>2</sup>

No passado, o conceito era delegado apenas a filósofos e poetas. Porém, atualmente existe um interesse crescente de pesquisadores da área de saúde nessa definição.<sup>3</sup>

Patrick e Erickson<sup>4</sup> disseram em 1999 que a qualidade de vida *é o valor atribuído à duração da vida, modificado pelos prejuízos, estados funcionais e oportunidades sociais que são influenciados por doença, dano, tratamento ou políticas de saúde.*

Ainda em 1999, a Organização Mundial de Saúde (WHOQOL GROUP)<sup>5</sup> definiu a qualidade de vida como sendo *a percepção do indivíduo a respeito de sua posição na vida dentro do contexto dos valores, da cultura na qual ele vive, e em relação a seus objetivos, expectativas, padrões e preocupações.*

Esses conceitos são consoantes com as mudanças de paradigmas que têm influenciado as políticas e as práticas de saúde nas últimas décadas. Além disso, o perfil de morbimortalidade indica um aumento da prevalência das doenças crônico-degenerativas, e os avanços nos tratamentos têm acarretado aumento também na sobrevivência das pessoas acometidas por esses agravos. Isso faz com que o impacto dessas doenças e de seus tratamentos sejam avaliados em termos de sua influência na QV.<sup>1</sup>

Assim, a avaliação da QV foi incluída principalmente nos ensaios clínicos como uma terceira dimensão a ser avaliada, além da eficácia e segurança dos medicamentos.<sup>6</sup>

Devido à percepção de que a QV é um fator importante para o estado de saúde, médicos e pesquisadores têm tentado transformá-la numa medida quantitativa que possa ser comparada entre diferentes populações e até mesmo patologias.<sup>3</sup>

Normalmente, a avaliação da QV é feita por meio de questionários desenvolvidos por especialistas da área que está sendo estudada. São utilizadas perguntas que abordam

aspectos específicos da vida do paciente, e os resultados são mensurados principalmente por meio de escalas.<sup>7</sup>

Na última década houve uma proliferação de instrumentos para medir a QV, alguns mais específicos, outros mais genéricos, além do crescente interesse pelo processo de adaptação e validação transcultural. Diante do desenvolvimento desses instrumentos, isto é, do crescimento dos estudos que avaliam a QV, surge ainda uma questão: Quais fatores poderão estar associados com uma melhor qualidade de vida?

Geralmente, o resultado das escalas de QV é mensurado em escalas ordinais. Esse tipo de variável, consiste em uma série de categorias com uma determinada ordenação. Para a análise dos dados ordinais, métodos estatísticos mais complexos são necessários como os modelos de regressão ordinal, ainda pouco divulgados e utilizados.<sup>7</sup>

Lall, Campbell e Walters vêm mostrando técnicas de análises estatísticas na construção de modelos que têm como resposta a avaliação da QV.<sup>7-11</sup> Em dois artigos, esses autores falam dos métodos *Bootstrap* para analisar esse tipo de variável. Os métodos *Bootstrap* envolvem basicamente amostragem aleatória com reposição para os dados originais, produzindo sub-amostras aleatórias de mesmo tamanho da amostra original, cada qual sendo conhecida como uma amostra *Bootstrap*. A partir dessas sub-amostras é possível obter uma estimativa do parâmetro de interesse através de sua média nessas sub-amostras.<sup>9</sup>

No primeiro artigo, Walters<sup>10</sup> compara diferentes métodos, dentre eles o estimador *Bootstrap*, para estimação do tamanho de amostra e poder, em estudos cuja resposta primária é a medida de QV. Ele verifica que para cada situação um método diferente mostra-se mais adequado. Se há um piloto ou banco de dados histórico confiável, por exemplo, o método de simulação *Bootstrap* baseado nesses dados, produz estimação mais acurada e confiável do tamanho de amostra.<sup>10</sup>

Num segundo artigo<sup>9</sup>, esses autores, utilizando dados da escala SF-36, fazem uma comparação entre os métodos estatísticos convencionais (como teste-t, ANOVA, regressão linear), utilizados quando há distribuição normal dos dados e as estimativas pelo método *Bootstrap*. Entretanto, no conjunto de dados estudado, ambos métodos produzem resultados semelhantes. Isso pode ser explicado, pela suposição de que existe

uma variável latente contínua envolvendo o construto QV, e que a medida ordinal dessas variáveis, refletem intervalos contíguos desse *continuum*.<sup>9</sup>

Apesar desses resultados, os autores destacam que as escalas de QV tendem a uma distribuição discreta, assimétrica e limitada. Por isso, métodos de análise como test-t e regressão linear que assumem normalidade podem não ser apropriados. É importante considerar o caráter ordinal original que essas variáveis tem.<sup>9</sup> No entanto, os resultados das escalas de QV, apesar do caráter ordinal, raramente são analisados como tal. Frequentemente essas escalas são dicotomizadas, tratadas como nominais ou como contínuas. Porém, isso pode ser inconveniente e conduzir a análises estatísticas errôneas.<sup>12</sup>

Assim, um método de análise mais adequado e que vem sendo desenvolvido e apresentado na literatura estatística para análise desse tipo de dado é a regressão logística ordinal.<sup>12</sup> Esse tipo de análise proporciona, ainda, o cálculo de risco de ocorrência de um evento adverso, por exemplo, uma baixa qualidade de vida, que frequentemente interessa aos epidemiologistas.<sup>13</sup>

Lall et al.<sup>7</sup> discutem os modelos de regressão ordinal como principais técnicas de análise multivariada, que deveriam ser utilizadas em estudos sobre QV. Esse tipo de modelo requer alguns cuidados na análise dos dados e são cercados de pressupostos. O tipo de modelo de regressão ordinal depende do processo como os dados foram processados e gerados. E em um processo de análise, é possível testar uma variedade desses modelos e então selecionar o apropriado para os dados em estudo.<sup>7</sup>

Apesar de vários modelos para resposta ordinal terem sido propostos, a utilização na literatura epidemiológica e biomédica tem sido mínima.<sup>13</sup>

Frente a essa discussão, e considerando os diferentes métodos de análise de dados sobre QV mensurados por meio de escalas ordinais, neste trabalho propomos as comparações do uso dos modelos de regressão ordinal, em particular do modelo de chances proporcionais, modelo de razão-contínua, modelo estereótipo e modelo de chances proporcionais parciais, para se avaliar os fatores associados à QV quando se utiliza escalas ordinais.

A seguir será apresentada uma revisão da utilização dos instrumentos de avaliação da QV, mais especificamente de três desses instrumentos: WHOQOL, SF-36 e QLS. Além

disso, serão ilustradas técnicas de análise estatística utilizadas em estudos que tem como resposta primária a QV, como os modelos de regressão logística ordinal.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1. Instrumentos para avaliar a QV na área médica

Percebe-se que os instrumentos para avaliar a QV vêm sendo amplamente utilizados na área médica. Foram selecionados três desses instrumentos para essa revisão: WHOQOL, SF-36 e QLS. Dois desses instrumentos são mais genéricos e podem ser utilizados em diferentes populações, com diversas patologias. O terceiro, entretanto, é um instrumento mais específico, utilizado para uma população restrita.

#### 2.1.1. WHOQOL (*World Health Organization Quality of Life*)

O WHOQOL é um instrumento que foi desenvolvido para avaliar a QV, podendo ser aplicado por diferentes culturas. Nesse instrumento, a QV foi definida como uma percepção individual, oscilando por um caminho complexo, entre saúde física, estado psicológico, nível de independência, relacionamento social e nível de envolvimento.<sup>14</sup>

É uma escala com 100 itens, 24 facetas e 6 domínios que refletem a multidimensionalidade do construto: domínio físico, domínio psicológico, nível de independência, relações sociais, meio ambiente como também espiritualidade/religião/crenças pessoais.<sup>6</sup> A escala de resposta é do tipo *likert* com uma escala de intensidade (variando de nada a extremamente), capacidade (variando de nada a completamente), frequência (variando de nunca a sempre) e avaliação (variando de muito insatisfeito a muito satisfeito).<sup>14</sup>

Existe ainda uma versão reduzida da escala, a WHOQOL-BREF com 26 itens<sup>5</sup>, além da versão traduzida e validada para a língua portuguesa.<sup>6</sup>

Em pesquisa bibliográfica realizada na base indexadora de produção científica MedLine, apenas no ano de 2005, foram localizados 35 artigos que utilizaram esse instrumento para avaliar a QV de pacientes com os mais diversos tipos de doenças.

#### 2.1.2. SF – 36 (*Medical Outcomes Study 36 Item Short Form Healthy Survey*)

Outro instrumento que vem sendo amplamente utilizado para avaliar a QV é o SF-36. É um instrumento genérico, de fácil administração e compreensão, porém não tão extenso como os anteriores.<sup>3</sup>

É um questionário auto-aplicável que leva cerca de cinco minutos para ser completado. É multidimensional formado por 36 itens, englobados em oito escalas ou componentes: capacidade funcional, aspectos físicos, dor, estado geral da saúde, vitalidade, aspectos sociais, aspectos emocionais e saúde mental. Apresenta um escore final de 0 a 100, sendo que zero corresponde ao pior estado geral de saúde.<sup>15</sup>

Esse instrumento também já foi traduzido e validado para a língua portuguesa.<sup>3</sup> Chama atenção o quanto essa escala vem sendo utilizada em todo o mundo. Só em 2005 foram localizados 495 artigos, quando se faz uma busca utilizando “SF-36” como palavra chave.

A Tabela 1 mostra uma relação do percentual de artigos nas principais áreas de aplicação, que foram encontradas: psiquiatria, cardiologia, reumatologia e oncologia, para as escalas WHOQOL e SF-36.

Tabela 1: Relação dos artigos sobre os instrumentos WHOQOL e SF-36, publicados em 2005, por área de aplicação.

Áreas	WHOQOL (35 artigos) %	SF-36 (495 artigos) %
Psiquiatria		
Depressão	13%	12%
Outros	39%	5%
Cardiologia	13%	12%
Reumatologia	--	9%
Oncologia	--	42%
Outras áreas (diversas)	35%	20%

### 2.1.3. QLS (*Quality of life in schizophrenia*)

Diante de instrumentos mais genéricos, há ainda o surgimento de instrumentos para medir a QV de populações restritas, como é o caso da escala QLS que é uma escala desenvolvida especificamente para pacientes com esquizofrenia. Possui em sua estrutura um total de 21 itens distribuídos em três domínios específicos: Domínio social, domínio ocupacional e domínio intrapsíquico e interpessoal. Cada item aborda um aspecto específico da vida do paciente, considerando informações sobre sintomatologia e funcionamento dos pacientes nas três semanas precedentes à entrevista. É uma escala do tipo *likert*, cujos escores variam de zero a seis, sendo que um escore maior representa uma melhor QV. O escore global da escala é formado pela média dos 21 itens que constituem uma variável discreta ordinal, considerada na maioria das vezes como contínua, devido ao grande número de categorias.<sup>16,17</sup>

Este instrumento foi adaptado e validado para o Brasil (QLS-BR) apresentando boas características de medida no que se refere à validade e confiabilidade.<sup>16,17</sup>

Apesar de ser uma escala aplicável a uma população bastante específica, só em 2005 foram localizados 15 artigos, utilizando esse instrumento.

## 2.2. Regressão logística ordinal

Como pode ser verificado pelos instrumentos citados, geralmente, o resultado das escalas de QV são mensurados em escalas ordinais. Existem vários tipos de modelos de regressão logística ordinal que podem ser utilizados. Vários trabalhos sobre esse assunto vêm sendo desenvolvidos e resultados muito satisfatórios estão sendo encontrados.<sup>7,12,13;18-30</sup> A seguir, há uma breve explicação sobre cada um dos modelos: *modelo de chances proporcionais*, *modelo de razão-contínua*, *modelo estereótipo* e *modelo de chances proporcionais parciais*.

### 2.2.1. Modelo de chances proporcionais (*Proportional odds model*)

O modelo de chances proporcionais, também chamado de modelo do logito cumulativo (*cumulative logit model*) é mais apropriado e com interpretação mais fácil, quando a variável resposta a ser considerada é uma variável contínua que foi categorizada.<sup>7,13</sup>

Sejam  $Y_1, Y_2, \dots, Y_k$  as categorias da variável resposta  $Y$  e  $(X_1, X_2, \dots, X_p)$  o conjunto de variáveis explicativas ou covariáveis. Esse modelo compara a probabilidade de uma resposta igual ou menor a determinado valor,  $Y \leq k$ , com a probabilidade de uma resposta maior,  $Y > k$ , sendo expresso pela equação<sup>32</sup>:

$$c_k(x) = \log \left[ \frac{P(Y \leq y_k | x)}{P(Y > y_k | x)} \right] = \log \left[ \frac{\phi_0(x) + \phi_1(x) + \dots + \phi_k(x)}{\phi_{k+1}(x) + \phi_{k+2}(x) + \dots + \phi_k(x)} \right]$$

$$c_k(x) = \tau_k - x' \beta \quad \text{para } k=0, 1, \dots, k-1. \quad (1)$$

O termo  $\tau_k$  representa o intercepto do modelo, que varia para cada uma das  $k$  categorias e  $\beta$  é o vetor dos coeficientes do modelo.

Portanto, este modelo é definido pelo *logito* das probabilidades cumulativas e foi originalmente proposto por Walker e Duncan<sup>31</sup> e depois chamado de *modelo de chances proporcionais* por McCullagh<sup>18</sup>, por descrever a sua principal característica.

Pela expressão 1, percebe-se que o vetor de coeficientes da equação,  $\beta$ , não depende de  $k$ , implicando que o relacionamento entre  $X$  e  $Y$  é independente de  $k$ .<sup>13</sup> Logo, esse modelo fornece uma única estimativa da razão de chances para todas as categorias comparadas, dada pelo exponencial dos termos do vetor  $\beta$ . Esta estimativa não é uma média ponderada das chances de cada comparação, e sim uma estimativa obtida usando o método de máxima verossimilhança ou o método dos mínimos quadrados. Essa é uma estimativa ideal em termos da facilidade de interpretação e em termos da parcimônia do modelo.<sup>7</sup>

Por outro lado, o modelo carrega uma suposição, chamada por McCullagh<sup>18</sup> de razão de chances proporcionais, a cerca dos  $k$ -pontos de corte, ou seja, razão de chances semelhante para todas as categorias comparadas. Essa premissa é assumida para cada covariável incluída no modelo, e é sempre importante verificar antes da construção do modelo se essa suposição é satisfeita.<sup>7</sup> Normalmente é utilizado para testar a homogeneidade da razão de chances, um teste usando uma aproximação da distribuição qui-quadrado<sup>23</sup>, chamado por Hosmer & Lemeshow de teste de regressão paralela.<sup>32</sup>

Apesar deste teste ser anticonservativo ele pode ser usado para se ter alguma evidencia da adequação do modelo.<sup>32</sup>

Esse modelo encontra-se implementado na maioria dos pacotes estatísticos, inclusive aqueles muito utilizados como SPSS e Minitab.

### 2.2.2. Modelo de razão contínua (*Continuation-ratio logistic model*)

Feinberg<sup>19</sup> propôs um método alternativo para a análise de dados com resposta ordenada. Nesse modelo, a razão de chances representa a chance da exposição ter um nível menor da variável resposta, em comparação à chance da exposição ser igual que essa categoria.<sup>12</sup>

Assim, o modelo compara a probabilidade de uma resposta igual à categoria com determinado score, digamos  $k$ ,  $Y = k$ , com a probabilidade de uma resposta maior,  $Y \leq k$ , sendo expresso pela equação<sup>32</sup>:

$$c_k(x) = \ln \left[ \frac{P(Y = k | x)}{P(Y < k | x)} \right] = \ln \left[ \frac{\phi_k(x)}{\phi_{k+1}(x) + \phi_{k+2}(x) + \dots + \phi_{k-1}(x)} \right]$$

$$c_k(x) = \theta_k - x' \beta \quad \text{para } k=0, 1, \dots, k-1. \quad (2)$$

Esse modelo possui diferentes constantes e um coeficiente para cada comparação. Uma vantagem desse tipo de modelo é que ele pode ser ajustado através de  $k$  modelos de regressão logística binária.<sup>32</sup>

Mas é importante lembrar que o modelo de razão contínua é afetado pela direção escolhida para modelar a variável. Portanto, a razão de chances obtida quando se modela o crescimento na gravidade, não é equivalente ao recíproco que é obtido quando se modela decrescimento na gravidade.<sup>12</sup>

Existe também um teste para heterogeneidade dos pontos de corte que é específico para testar o relacionamento entre a resposta e a exposição.<sup>12</sup>

Esse tipo de modelo é mais apropriado quando há um interesse específico em uma categoria da variável resposta e não meramente um agrupamento arbitrário de uma variável contínua subjacente.<sup>13</sup>

O modelo de razão contínua é análogo ao modelo de riscos proporcionais de Cox<sup>22</sup>, quando se utiliza tempos discretos,<sup>13</sup> e pode ser desenvolvido pelo *software* STATA, através de um comando desenvolvido por Wolfe (1998).<sup>32</sup>

### 2.2.3. Modelo *estereótipo* (*Stereotype model*)

O modelo *estereótipo* deve ser utilizado quando a variável resposta é uma variável ordinal que não é uma versão discreta de alguma variável contínua, como por exemplo, uma escala de QV que tem como categorias (leve, moderado, grave).<sup>20</sup>

Esse modelo acessa o caráter ordinal da variável resposta através de uma ordenação da razão de chances das categorias. Ele pode ser considerado uma extensão do modelo de regressão politômica ou multinomial. Compara cada categoria da variável resposta com uma categoria de referência, que normalmente é a primeira categoria, e é dado pela seguinte equação<sup>7</sup>:

$$\log \left[ \frac{P(Y = y_k | x_1 \dots x_p)}{P(Y = y_1 | x_1 \dots x_p)} \right] = \alpha_k + \beta_{k1}x_1 + \dots + \beta_{kp}x_p \quad K = 2, \dots, c \quad (3)$$

Entretanto, devido o caráter ordinal dos dados é imposta uma estrutura ao *log* da chance  $\beta_{kp}$  desse modelo, ou seja, são atribuídos pesos aos coeficientes.<sup>7</sup>

$$\beta_{kj} = \phi_k \beta_j \quad k=2, \dots, c$$

$$j = 1, \dots, p \quad (4)$$

$$\log \left[ \frac{P(Y = y_k / x_1 \dots x_p)}{P(Y = y_1 / x_1 \dots x_p)} \right] = \alpha_k + \phi_k (\beta_1 x_1 + \dots + \beta_p x_p)$$

$$k = 2, \dots, c \quad (5)$$

Então, a razão de chances formada terá uma tendência de crescimento, já que os pesos podem ser construídos pela ordenação. Assim, o efeito das covariáveis na primeira razão de chances é menor que o efeito no segundo e assim sucessivamente.<sup>7</sup>

A maior dificuldade dessa modelagem é a determinação desses pesos, mas existem várias possibilidades. Greenland<sup>26</sup> sugere que os pesos podem ser decididos a priori,

sendo estimados por algum estudo piloto ou conjunto de valores apropriadamente escolhidos.

O modelo estereótipo pode ser ajustado através do *software* SAS, quando os conjuntos de pesos foram predeterminados. Entretanto, Hendricks<sup>28</sup> desenvolveu macros (para SAS e STATA) que ajustam a estimativa dos pesos como conjuntos dos parâmetros do modelo.

Lall et al.<sup>7</sup> propõem ainda o uso de técnicas *Bootstrap* para obtenção de erros-padrão e testes corretos, quando são ajustados esses tipos de modelo.

#### 2.2.4. Modelo de chances proporcionais parciais (*Partial proportional odds models*)

Há geralmente um consenso de que a suposição de chances proporcionais é bastante restrita. Principalmente quando se considera mais que uma covariável e na prática, a chance de todas as covariáveis no modelo terem chances proporcionais é muito rara. Por isso, o modelo de chances proporcionais parciais<sup>24</sup> permite que algumas covariáveis possam ser modeladas com a suposição de chance proporcional, e ainda que as variáveis em que essa suposição não seja satisfeita, tenham estimativas da razão de chances para as várias categorias comparadas.<sup>7</sup>

Existem dois tipos de modelos de chances proporcionais parciais: com ou sem restrição. Esses modelos são uma extensão do modelo de chances proporcionais.

##### 2.2.4.1. Modelo de chances proporcionais parciais não-restrito (*Unconstrained partial proportional odds models*)

Suponha que dentre as  $p$  variáveis preditoras ( $X_1, X_2, \dots, X_p$ ), algumas tenham chances proporcionais e outras, digamos  $q$  delas, não tenham. O modelo é dado por<sup>7</sup>:

$$\log \left[ \frac{P(Y \leq y_k | x_1 \dots x_p)}{P(Y > y_k | x_1 \dots x_p)} \right] = \alpha_k + [(\beta_1 x_1 + \gamma_{k1} T_1) + (\beta_2 x_2 + \gamma_{k2} T_2) + \dots + (\beta_q x_q + \gamma_{kq} T_q) + \dots + (\beta_p x_p)]$$

$$k=1,2,\dots,c-1 \quad (6)$$

Os parâmetros  $\beta_1, \dots, \beta_p$  são as componentes de cada covariável, para as quais a proporcionalidade acerca dos pontos de corte pode ser assumida. Os  $T_1, \dots, T_q$  ( $=X_1 \dots X_q$ ) existem apenas para as  $q$  variáveis que não tem chances proporcionais. Para esse modelo, são estimados  $c-1$  interceptos,  $p$  parâmetros  $\beta$  que são independentes dos pontos de corte, e um vetor  $(c-1) \times q$  dos  $\gamma$  parâmetros associados com cada covariável e pontos de corte.<sup>7</sup>

Logo, para esse modelo, alguns variáveis possuem apenas uma razão de chances para todas as categorias, e outras possuem razão de chances para comparações em cada categoria da variável resposta.

#### 2.2.4.2. Modelo de chances proporcionais com restrições parciais (*Constrained partial proportional odds models*)

Peterson e Harrell<sup>24</sup> propuseram ainda um segundo modelo de chances proporcionais parciais, com restrição. Esse modelo é aplicável quando existe uma relação linear entre o logito da razão de chances de uma covariável e a variável resposta.<sup>13</sup>

Nesse caso, um conjunto de restrições ( $\gamma_{kl}$ ) podem ser inseridas nos parâmetros do modelo, para esclarecer essa linearidade. Quando essas restrições são incorporadas no modelo, ele passa a ter a seguinte forma<sup>7</sup>:

$$\log \left[ \frac{P(Y \leq y_k | x_1 \dots x_p)}{P(Y > y_k | x_1 \dots x_p)} \right] = \alpha_k - [(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) + \tau_k (\gamma_1 T_1 + \gamma_2 T_2 + \dots + \gamma_q T_q)]$$

$$k=1,2,\dots,c-1 \quad (6)$$

Os  $\tau_k$  são escalas fixas que tomam a forma de restrições alocadas nos parâmetros. Nesse caso,  $X_m, \gamma_m$  não dependem dos pontos de corte, mas são multiplicados por  $\tau_k$  para cada k-ésimo logito<sup>7</sup>.

A escolha da restrição pode ser decidida de várias maneiras, idealmente elas deveriam ser determinados usando um banco de dados piloto ou um valor predefinido a priori. Entretanto, alguns autores<sup>13</sup> têm examinado o logito das chances obtido pelo modelo de chances proporcionais parciais não-restritos, para determinar o conjunto de restrições para esse modelo.

### **2.2.3. Exemplo de aplicação: Estudo de QV utilizando a escala QLS-BR**

Foi realizado um estudo de fatores associados com a QV em uma amostra de 123 pacientes com diagnóstico de esquizofrenia. A variável resposta (QV) foi mensurada por meio da escala QLS validada para o Brasil. Este instrumento possui característica ordinal, sendo que os maiores escores refletiam melhor QV. As variáveis independentes eram de natureza clínica e sócio-demográficas.<sup>33</sup>

Esse estudo tinha como objetivo investigar a QV desses pacientes por sexo, explorando os fatores associados com a baixa QV no domínio ocupacional. Para isso, foi utilizado o escore categorizado do domínio ocupacional da escala QLS-BR, cujas categorias são: QV muito comprometida (<2), QV comprometida (2 – 5) e QV inalterada (>5).<sup>16,17,33</sup>

Pode-se dizer que essa escala possui uma variável latente contínua em sua estrutura, que foi agrupada, tendo como pontos de corte os escores 2 e 5. Por isso, os fatores associados a baixa QV foram analisados utilizando o modelo de chances proporcionais, baseado em probabilidades cumulativas.<sup>33</sup>

O modelo de chances proporcional faz duas comparações nesse caso. Compara a probabilidade de uma resposta igual ou menor que 2, com a probabilidade de uma resposta maior que 2, ou seja, QV muito comprometida versus QV comprometida e QV inalterada. E compara ainda, a probabilidade de uma resposta igual ou menor que 5, com a probabilidade de uma resposta maior que 5, ou seja, QV muito comprometida e QV comprometida versus QV inalterada.

Entretanto, quando se utiliza esse modelo é importante ressaltar que se devem ter alguns cuidados na interpretação da razão de chances, já que é oferecida uma estimativa única desse parâmetro. A interpretação é feita, em termos do aumento proporcional da chance em cada categoria.

Nesse estudo foi verificado, ainda, que para todas as variáveis, a suposição de chances proporcionais era válida, e por isso não seria o caso do uso do modelo de chances proporcionais parciais.

Considerando a escala QLS-BR, para exemplificar as comparações feitas caso fosse utilizado o modelo de razão contínua, temos as seguintes comparações: QV comprometida versus QV muito comprometida, e QV inalterada versus QV comprometida e QV muito comprometida. Entretanto, esse modelo deve ser utilizado apenas quando se tem um interesse específico por uma das categorias da variável resposta, o que não era o caso do estudo.

Ainda exemplificando para a escala QLS-BR, o modelo estereótipo faria as seguintes comparações nesse caso: QV muito comprometida (categoria de referência) versus QV comprometida e QV muito comprometida versus QV inalterada, fornecendo duas constantes e dois coeficientes. Entretanto, esse modelo também não seria apropriado nesse caso, já que ele deve ser utilizado apenas quando a variável resposta é uma variável ordinal que não é uma versão de alguma variável contínua.

Apesar das restrições à utilização de cada um dos modelos no estudo da escala de QV QLS-BR, apenas a título de exemplificação, a Tabela 2 faz uma comparação entre o modelo de chances proporcionais, o modelo de razão contínua e o modelo estereótipo, quanto às categorias comparadas.

Tabela 2: Comparação das categorias definidas pelos pontos de corte entre os três modelos ordinais para a escala QLS-BR.

Pontos de corte	<i>Modelo de chances proporcionais</i>		<i>Modelo de razão contínua</i>		<i>Modelo estereótipo</i>
2	QV muito comprometida	$P(Y \leq 2)$	QV muito comprometida	$P(Y=2)$	QV muito comprometida
	vs. QV comprometida e QV inalterada	vs. $P(Y > 2)$	vs. QV comprometida e QV inalterada	vs. $P(Y > 2)$	vs. QV comprometida
5	QV muito comprometida e QV comprometida	$P(Y \leq 5)$	QV comprometida	$P(Y=5)$	QV muito comprometida
	vs. QV inalterada	vs. $P(Y > 5)$	vs. QV inalterada	vs. $P(Y > 5)$	vs. QV inalterada

### 2.3. Dimensionamento de amostra / Cálculo de poder

Não se pode esquecer de que, antes que um estudo sobre QV seja realizado, deve-se ter um planejamento que inclui a escolha dos instrumentos e o cálculo do tamanho da amostra.

Além disso, sabe-se que o cálculo do tamanho de amostra é obrigatório em protocolos de pesquisa e são essenciais para que os artigos de ensaios clínicos sejam aceitos em periódicos. Entretanto, em algumas situações comuns esses cálculos não são facilmente acessíveis.<sup>34</sup>

Além disso, um bom dimensionamento amostral é essencial para o sucesso da análise estatística. Alguns trabalhos recentes sobre cálculo amostral são encontrados, até mesmo no contexto de QV. Segundo Walters et al (2001)<sup>35</sup> as medidas das escalas de QV podem levar a vários problemas na determinação do tamanho de amostra, devido à distribuição dessas variáveis. Esses autores fazem uma revisão de estudos sobre QV, para verificar o que tem sido usado na prática quanto aos métodos de dimensionamento amostral e argumentam que a utilização de dados sobre QV, como variáveis contínuas e normalmente distribuídas não é adequada.

Portanto, quando se fala do dimensionamento de amostra no planejamento de estudos sobre QV, antes da escolha da fórmula a ser utilizada, deve-se ater a várias questões

com: qual é a finalidade principal do estudo, qual é a principal medida do resultado, se os dados pretendem detectar diferenças nos tratamentos, o quão grande é essa diferenças, entre outros.<sup>10</sup> Logo, para cada situação um caminho diferente para estimação amostral pode ser seguido.

### **3. JUSTIFICATIVA**

Como mencionado durante várias situações, a preocupação com a qualidade de vida vem crescendo nos últimos anos. Além disso, nos estudos em que se avalia a QV através de escalas ordinais sempre surgem dificuldades na análise dos dados. Portanto, esses estudos sempre remetem ao seguinte questionamento: Qual método de análise é mais apropriado?

As técnicas de análises multivariadas quando se considera uma variável resposta ordinal, ainda são pouco divulgadas e utilizadas. Modelos de regressão linear ou de regressão logística binária não devem ser utilizados quando se considera esse tipo de variável, já que não levam em consideração o caráter de ordenação da variável resposta, além de não serem apropriados quando se considera o tipo de distribuição das escalas de QV. Por outro lado, os modelos de regressão ordinal são complexos e exigem um grande conhecimento dos dados e da forma como foram gerados. Existem vários modelos de regressão logística ordinal e em cada situação um modelo específico deve ser utilizado.

Vale ressaltar que, no estudo mencionado na Seção 2.2.3, foi observada a dificuldade de se analisar esse tipo de variável resposta, dada a escassez da literatura e complexidade da análise que deveria ser utilizada.

Além disso, pacotes estatísticos utilizados na saúde pública ou epidemiologia tais como SPSS<sup>®</sup> e Minitab<sup>®</sup> não possuem muitas opções de modelos ordinais implementados, e os que existem, muitas vezes não são aplicáveis à distribuição dos dados, por exemplo, quando não existe a proporcionalidade das chances.

Levando em consideração esses fatores, torna-se importante um estudo ilustrando os modelos ordinais e comparando a aplicabilidade desses métodos em cada situação.

## 4. OBJETIVOS

### 4.1. Objetivo geral

Considerando os diferentes métodos de análise de dados ordinais, o objetivo geral desse estudo é comparar e apresentar recomendações quanto ao uso dos modelos de regressão logística ordinal: *modelo de chances proporcionais*, *modelo de razão-contínua*, *modelo estereótipo* e *modelo de chances proporcionais parciais* como forma de análise multivariada em estudos de qualidade de vida utilizando escala ordinal.

### 4.2. Objetivos específicos

- f. Ilustrar o planejamento de estudo sobre QV, iniciando com a escolha dos instrumentos e cálculo amostral;
- g. Investigar o problema de dimensionamento de amostra e/ou cálculo do poder para os modelos de dados ordinais;
- h. Verificar a adequação e ajuste de cada modelo para investigação dos fatores associados a QV considerando bancos de dados secundários;
- i. Verificar a adequação e ajuste de cada modelo considerando diferentes escalas de QV;
- j. Definir o melhor modelo para ajustar os dados em cada situação;

## 5. MÉTODO

O estudo será realizado por meio de um banco de dados secundário, obtido a partir de um estudo transversal, realizado com 273 pacientes com diagnóstico de esquizofrenia. Esses pacientes são originários do SERSAM Divinópolis (Serviço de Referência em Saúde Mental) e CERSAM Pampulha (Centro de Referência em Saúde Mental). A qualidade de vida foi mensurada por meio da escala QLS-BR em entrevista conduzida tendo o próprio paciente como informante.<sup>16</sup> As entrevistas foram conduzidas por profissional de saúde previamente treinado para aplicação de instrumentos. Além disso, foi preenchido um questionário com informações clínicas e sócio-demográficas dos pacientes.

Avalia-se ainda a possibilidade de utilização de um segundo banco, com a medida da QV feita através da escala SF-36 e de outros bancos de dados, para comparação das metodologias diante de diferentes escalas.

### **Análise estatística**

Será avaliado o método de dimensionamento de amostra e cálculo de poder, mais apropriado para cada modelo e escala utilizados.

Serão ajustados modelos de regressão logística ordinal, através dos quatro modelos mencionados, para se avaliar os fatores associados à QV. Nesses modelos a variável resposta será definida como o escore de QV categorizado. As variáveis explicativas serão de natureza clínica e sócio-demográfica.

Para esta comparação metodológica e construção dos modelos, serão consideradas as variáveis de relevância encontradas nos estudos anteriores realizados para verificação dos fatores associados à QV na esquizofrenia.<sup>17</sup> A análise univariada precederá toda investigação para detecção das associações significativas.

Serão verificadas todas as suposições dos modelos de regressão ordinal e feitas simulações de dados, para testar situações em que as variáveis não se enquadram nos padrões de exigência dos modelos.

Por fim, haverá a verificação do ajuste dos modelos, através das estatísticas adequadas buscando os modelos que melhor representam os conjunto de dados estudados.

Pretende-se ainda, verificar o efeito da dicotomização das escalas de QV, e construção de modelos binários, comparados aos modelos ordinais. Serão utilizados os softwares R<sup>®</sup>, SPSS<sup>®</sup>, MINITAB<sup>®</sup>, STATA<sup>®</sup>, para ajuste dos modelos de regressão logística ordinal.

## 6. PRODUÇÃO CIENTÍFICA

A proposta do estudo inclui a preparação de dois artigos científicos, sendo um deles uma revisão bibliográfica do método de regressão logística ordinal, considerando ser essa uma metodologia pouco encontrada na literatura. Esse artigo já se encontra em andamento. O segundo artigo será sobre o detalhamento da análise estatística, comparação dos modelos propriamente ditos e dimensionamento da amostra.

## 7. CONSIDERAÇÕES ÉTICAS

Os dados utilizados neste trabalho são provenientes de pesquisas que já foram aprovadas por comitês de ética.

## 8. CRONOGRAMA

Atividades	Período			
	Mar/05 a Jul/05	Ago/05 a Dez/05	Jan/06 a Jul/06	Ago/06 a Dez/06
1. Revisão bibliográfica				
2. Preparação da versão final do projeto				
3. Qualificação				
4. Análise dos dados				
5. Preparação dos resultados				
6. Preparação dos artigos				
7. Apresentação dos resultados				

## 9. REFERÊNCIAS BIBLIOGRÁFICAS

- 1- Seidl EMF, Zannon CMLC. *Qualidade de vida e saúde: aspectos conceituais e metodológicos*, Cadernos de Saúde Pública, 2004, 20(2):580-588.
- 2- Awad G, Voruganti LN. *Intervention research in psychosis: issues related to the assessment of quality of life*. Schizophr Bull, 2000, 20: 557-64
- 3- Cionelli, Ferraz, Santos, Meinão, Quaresma. *Tradução para a língua portuguesa e validação do questionário genérico de avaliação de qualidade de vida SF-36*. Rev Bras Reumatol, 1999, 39: 143–150.
- 4- Ebrahim S. *Clinical and public health perspectives and applications of health-related quality of life measurement*. Soc Sci Med, 1995, 41: 1383-94
- 5- The WHOQOL Group. *Development of the World Health Organization WHOQOL-BREF quality of life assessment*. The WHOQOL Group. Psychol Med. 1998;28(3):551-8.
- 6- Fleck et al. *Desenvolvimento da versão em português do instrumento de avaliação da QV da OMS (WHOWOL-100)*. Rev Bras Psiquiatr, 1999, 21(1):19-28.
- 7- Lall R, Campbell MJ, Walters SJ, Morgan K. *A review of ordinal regression models applied on health-related quality of life assessments*. Stat Methods Med Res. 2002;11(1):49-67.
- 8- Walters SJ, Campbell MJ, Lall R. *Design and analysis of trials with quality of life as an outcome: a practical guide*. J Biopharm Stat. 2001;11(3):155-76.
- 9- Walters SJ, Campbell MJ. *The use of bootstrap methods for estimating sample size and analysing health-related quality of life outcomes*. Stat Med. 2005; 15;24(7):1075-102.
- 10- Walters SJ. *Sample size and power estimation for studies with health related quality of life outcomes: a comparison of four methods using the SF-36*. Health Qual Life Outcomes. 2004; 25;2(1):26.

- 11- Walters SJ, Campbell MJ. *The use of bootstrap methods for analysing Health-Related Quality of Life outcomes (particularly the SF-36)*. Health Qual Life Outcomes. 2004; 9;2(1):70.
- 12- Scott SC, Goldberg MS, Mayo NE. *Statistical assessment of ordinal outcomes in comparative studie*. J Clin Epidemiol. 1997; 50(1):45-55
- 13- Ananth CV, Kleinbaum DG. *Regression models for ordinal responses: a review of methods and applications*. Int J Epidemiol. 1997;26(6):1323-33.
- 14- The WHOQOL Group. *The world health organization quality of life assessment (WHOQOL): Development and general psychometric propertie*. Soc. Sci. Med., 1998, 46(12): 1569-1585.
- 15- Brazier JE. et al. *Validating the SF-36 health survey questionnaire: new outcome measure for primary car*. BMJ. 1992; 305(6846):160-4.
- 16- Cardoso CS, Bandeira M, Caiaffa WT, Fonseca JOP. *Escala de qualidade de vida para pacientes com esquizofrenia-QLS-BR, Adaptação transcultural para o Brasil*. J brasil Psiq, 2002; (51):31-38.
- 17- Cardoso CS, Caiaffa WT, Bandeira M, Siqueira AL, Abreu MNS, Fonseca JOP. *Qualidade de vida na esquizofrenia: Fatores associados*. Cadernos de Saúde Pública. (no prelo)
- 18- McCullagh P. *Regression models for ordinal data*. J R Statist Soc [B]. 1980; 42:109-142.
- 19- Fienberg SE. *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA: MIT Press; 1980: 110-116.
- 20- Anderson JA. *Regression and ordered categorical variables*. J R Statisti Soc; 1984, 16: 1-30.
- 21- Cox C, Chuang C. *A comparison of chi-square partitioning and two logit analyses of ordinal pain data from a pharmaceutical study*. Stat Med; 1984, 3:273-285
- 22- Cox C. *Multinomial regression models based on continuation ratios*. Stat Med; 1988, 7:435-441

- 23- Brant R. *Assessing proportionality in the proportional odds model for ordinal logistic regression*. *Biometrics*. 1990; 46(4):1171-8
- 24- Peterson BL, Hanrrel FE. *Partial proportional odds models for ordinal response variables*. *Appl Stat*. 1990; 36:205-217.
- 25- Lee J. *Cumulative logit modelling for ordinal response variables: applications to biomedical research*. *Comput Appl Biosci*. 1992; 8(6):555-62.
- 26- Greenland S. *Alternative models for ordinal logistic regression*. *Stat Med*. 1994;13(16):1665-77.
- 27- Bender R, Grouven U. *Ordinal logistic regression in medical research*. *J R Coll Physicians Lond*. 1997; 31(5):546-51.
- 28- Hendrickx J. *Special restrictions in multinomial logistic regression*. *Stata Technical Bulletin*; 2000, STB-56: 18 - 26
- 29- Yee, Thomas W. and Hastie, Trevor J. *Reduced-rank vector generalized linear models*. *Statistical Modelling*, 2003, 3:15-41.
- 30- Pulkstenis E, Robinson TJ. *Goodness-of-fit tests for ordinal response regression models*. *Stat Med*. 2004; 23(6):999-1014.
- 31- Walker SH, Duncan DB. *Estimation of the probability of an event as a function of several independent variables*. *Biometrika*, 1967; 54:167-179.
- 32- HOSMER WD.; LEMESHOW S. *Applied Logistic Regression*, 2. ed. , Wiley, 2000.
- 33- Cardoso CS, Caiaffa WT, Bandeira M, Siqueira AL, Abreu MNS, Fonseca JOP. *Qualidade de vida e dimensão ocupacional: Uma comparação por sexo na esquizofrenia*. *Cadernos de Saúde Pública*. (submetido)
- 34- Campbell MJ, Julious SA, Altaman DG. *Estimation sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons*. *BMI*, 1995; 311: 1145-1148.
- 35- Walters SJ, Campbell MJ, Paisley S. *Methods for Determining Sample Sizes for Studies Involving Health-Related Quality of Life Measures: A Tutorial*. *Health Services & Outcomes Research Methodology* 2: 83–99, 2001.

- 36- Cardoso, C.S.; Bandeira, M.; Caiaffa, W.T.; Siqueira, A.L.; Fonseca I.K.; Fonseca J.O.P. 2003. Qualidades psicométricas da escala de qualidade de vida para pacientes com esquizofrenia: Escala QLS-BR. *Jornal Brasileiro de Psiquiatria*, 52:211-222.

**APÊNDICE B**  
**RESULTADOS DO ARTIGO 1**

## Estatísticas Descritivas dos Dados Utilizados no Artigo 1

### Exemplo 1 – Variável resposta Domínio ocupacional

Tabela 1: Tabela de frequência dos dados sobre qualidade de vida no domínio ocupacional da escala QLS-BR, por estado civil e sexo.

Estado civil	Sexo	QV - Domínio ocupacional			Total
		QV muito comprometida	QV comprometida	QV inalterada	
solteiro	masculino	56 (39,2%)	76 (53,1%)	11 (7,7%)	143
solteiro	feminino	19 (23,5%)	54 (66,7%)	8 (9,9%)	81
casado	masculino	9 (36,0%)	12 (48,0%)	4 (16,0%)	25
casado	feminino	2 (8,3%)	17 (70,8%)	5 (20,8%)	24

Tabela 2: Resultados dos modelos de regressão logística binária com as comparações feitas pelos modelos de chances proporcionais e modelo de razão contínua tendo como resposta o Domínio Ocupacional da Escala QLS-BR.

Covariáveis (categoria de referência)	Pontos de corte								
	QV muito comprometida X QV comprometida + QV inalterada			QV muito comprometida + QV comprometida X QV inalterada			QV comprometida X QV inalterada		
	$\hat{\beta}$	$EP(\hat{\beta})$	OR	$\hat{\beta}$	$EP(\hat{\beta})$	OR	$\hat{\beta}$	$EP(\hat{\beta})$	OR
	Sexo (masc.)	0,897	0,292	2,45	0,289	0,407	1,34	-0,020	0,419
Estado civil (solt.)	0,460	0,378	1,59	0,851	0,443	2,34	0,756	0,459	2,13

EP= erro-padrão; OR= *odds ratio* (razão de chances)

### Exemplo 2 – Variável resposta Item Funcionamento ocupacional

Tabela 3: Tabela de frequência dos dados sobre qualidade de vida no Item Funcionamento Ocupacional da escala QLS-BR, por estado civil e sexo.

Estado civil	Sexo	QV - Domínio ocupacional			Total
		QV muito comprometida	QV comprometida	QV inalterada	
solteiro	masculino	54 (37,8%)	72 (50,3%)	17 (11,9%)	143
solteiro	feminino	8 (9,9%)	57 (70,4%)	16 (19,8%)	81
casado	masculino	62 (27,7%)	129 (57,6%)	33 (14,7%)	25
casado	feminino	6 (24,0%)	15 (60,0%)	4 (16,0%)	24

## SAÍDAS DO PROGRAMA - SOFTWARE R

### Exemplo 1 – Variável resposta Domínio ocupacional

#### {Leitura dados e criação de vetores para análise}

```
> library(foreign)
> banco=read.spss("c:/bancoqlsr.sav")
> summary(banco)
      Length Class  Mode
SEXO    273   -none- numeric
ESTCIV  273   -none- numeric
F2_012  273   -none- numeric
F2_0X12 273   -none- numeric
F2_01X2 273   -none- numeric
F2_1X2  273   -none- numeric

> y.ord=banco$F2_012
> y.bin1=banco$F2_0X12
> y.bin2=banco$F2_01X2
> y.bin3=banco$F2_1X2
> sexo=banco$SEXO
> ec=banco$ESTCIV
```

#### {Ajuste dos modelos de regressão logística binária}

```
> library(Design)
```

{Comparação QV muito comprometida X QV comprometida e QV inalterada}

#### Regressão logística binária univariada

```
> regbl_1=lrn(y.bin1~sexo)
> regbl_1

Logistic Regression Model
lrn(formula = y.bin1 ~ sexo)

Frequencies of Responses
  0   1
86 187

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273   4e-12   10.88      1    0.001   0.603   0.205
Gamma   Tau-a      R2      Brier
0.433   0.089   0.055   0.208

      Coef  S.E.  Wald Z P
Intercept 0.4603 0.1584 2.91 0.0037
Sexo      0.9260 0.2909 3.18 0.0015
```

```

> regb1_2=lrn(y.bin1~ec)
> regb1_2

Logistic Regression Model
lrn(formula = y.bin1 ~ ec)

Frequencies of Responses
  0  1
86 187

              Obs  Max Deriv Model L.R.          d.f.          P          C          Dxy
              273    6e-07    2.38            1      0.1227      0.538      0.075
      Gamma      Tau-a          R2      Brier
      0.27      0.033      0.012      0.214

              Coef  S.E.  Wald Z P
Intercept  0.6865  0.1416  4.85  0.0000
ec         0.5532  0.3705  1.49  0.1354

```

### Regressão logística binária multivariada

```

> regb1=lrn(y.bin1~sexo+ec)
> regb1

Logistic Regression Model
lrn(formula = y.bin1 ~ sexo + ec)

Frequencies of Responses
  0  1
86 187

              Obs  Max Deriv Model L.R.          d.f.          P          C          Dxy
              273    3e-11    12.44            2      0.002      0.615      0.23
      Gamma      Tau-a          R2      Brier
      0.376      0.1      0.063      0.207

              Coef  S.E.  Wald Z P
Intercept  0.3954  0.1665  2.37  0.0176
sexo       0.8967  0.2922  3.07  0.0022
ec         0.4605  0.3775  1.22  0.2225

```

{ Comparação QV muito comprometida e QV comprometida X QV inalterada }

### Regressão logística binária univariada

```

> regb2_1=lrn(y.bin2~sexo)
> regb2_1

Logistic Regression Model
lrn(formula = y.bin2 ~ sexo)

Frequencies of Responses
  0  1
245 28

              Obs  Max Deriv Model L.R.          d.f.          P          C          Dxy
              273    8e-08    0.82            1      0.3648      0.544      0.089
      Gamma      Tau-a          R2      Brier
      0.181      0.016      0.006      0.092

              Coef  S.E.  Wald Z P
Intercept -2.3224  0.2706 -8.58  0.0000
sexo       0.3655  0.4012  0.91  0.3623
> regb2_2=lrn(y.bin2~ec)

```

```

> regb2_2

Logistic Regression Model

lrm(formula = y.bin2 ~ ec)

Frequencies of Responses
  0  1
245 28

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273    4e-09    3.72      1    0.0538    0.579    0.158
Gamma   Tau-a      R2    Brier
0.417   0.029    0.028    0.091

      Coef  S.E.  Wald Z P
Intercept -2.379 0.2398 -9.92 0.0000
ec         0.887 0.4400  2.02 0.0438

```

### Regressão logística binária multivariada

```

> regb2=lrm(y.bin2~sexo+ec)
> regb2

Logistic Regression Model

lrm(formula = y.bin2 ~ sexo + ec)

Frequencies of Responses
  0  1
245 28

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273    5e-09    4.22      2    0.1213    0.6    0.199
Gamma   Tau-a      R2    Brier
0.294   0.037    0.032    0.09

      Coef  S.E.  Wald Z P
Intercept -2.4911 0.2942 -8.47 0.0000
sexo      0.2888 0.4067  0.71 0.4777
ec        0.8514 0.4432  1.92 0.0547

```

{Comparação QV comprometida X QV inalterada}

### Regressão logística binária univariada

```
> regb3_1=lrn(y.bin3~sexo)
> regb3_1

Logistic Regression Model
lrn(formula = y.bin3 ~ sexo)

Frequencies of Responses
  0  1
159 28

Frequencies of Missing Values Due to Each Variable
y.bin3  sexo
   86    0

              Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
              187   7e-14      0.03      1      0.8619   0.509   0.018
Gamma      Tau-a      R2      Brier
0.036      0.005      0      0.127

              Coef      S.E.      Wald Z P
Intercept -1.76929 0.2793 -6.33 0.0000
sexo      0.07156 0.4111  0.17 0.8618
```

```
> regb3_2=lrn(y.bin3~ec)
> regb3_2

Logistic Regression Model
lrn(formula = y.bin3 ~ ec)

Frequencies of Responses
  0  1
159 28

Frequencies of Missing Values Due to Each Variable
y.bin3  ec
   86    0

              Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
              187   2e-11      2.59      1      0.1077   0.57   0.139
Gamma      Tau-a      R2      Brier
0.36      0.036      0.024   0.125

              Coef      S.E.      Wald Z P
Intercept -1.923 0.2456 -7.83 0.000
ec      0.753 0.4538  1.66 0.097
```

## Regressão logística binária multivariada

```

> regb3=lrn(y.bin3~sexo+ec)
> regb3

Logistic Regression Model

lrn(formula = y.bin3 ~ sexo + ec)

Frequencies of Responses
  0  1
159 28

Frequencies of Missing Values Due to Each Variable
y.bin3  sexo  ec
   86    0    0

              Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
              187  2e-11  2.59      2      0.2739  0.569  0.138
              Gamma  Tau-a      R2      Brier
              0.201  0.035  0.024  0.125

              Coef      S.E.  Wald Z P
Intercept -1.91462 0.3003 -6.38 0.0000
sexo      -0.02046 0.4191 -0.05 0.9611
ec        0.75637 0.4590  1.65 0.0993

```

## {Ajuste do Modelo de chances proporcionais}

### Univariada

```

> mcp1=lrn(y.ord~sexo)
> mcp1

Logistic Regression Model

lrn(formula = y.ord ~ sexo)

Frequencies of Responses
      QV inalterada      QV comprometida QV muito comprometida
              28              159              86

              Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
              273  9e-13  9.29      1      0.0023  0.582  0.164
              Gamma  Tau-a      R2      Brier
              0.343  0.091  0.04  0.092

              Coef      S.E.  Wald Z P
y>=QV comprometida  1.7650 0.2356  7.49 0.0000
y>=QV muito comprometida -1.2633 0.2138 -5.91 0.0000
sexo2=masculino     0.7597 0.2529  3.00 0.0027

```

```

> mcp2=lrn(y.ord~ec)
> mcp2

Logistic Regression Model

lrn(formula = y.ord ~ ec)

Frequencies of Responses
      QV inalterada      QV comprometida QV muito comprometida
              28                159                86

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273   6e-14   5.82      1      0.0158   0.551   0.101
Gamma   Tau-a      R2      Brier
0.336   0.056   0.025   0.091

      Coef      S.E.      Wald Z P
y>=QV comprometida      1.5700 0.3131   5.01 0.0000
y>=QV muito comprometida -1.4312 0.3088  -4.63 0.0000
ec2=solteiro            0.7736 0.3248   2.38 0.0172

```

## Multivariada

```

> mcp=lrn(y.ord~sexo+ec)
> mcp

Logistic Regression Model

lrn(formula = y.ord2 ~ sexo2 + ec2)

Frequencies of Responses
      QV inalterada      QV comprometida QV muito comprometida
              28                159                86

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273   3e-11  13.79      2      0.001   0.603   0.207
Gamma   Tau-a      R2      Brier
0.329   0.114   0.059   0.091

      Coef      S.E.      Wald Z P
y>=QV comprometida      1.2646 0.3291   3.84 0.0001
y>=QV muito comprometida -1.8086 0.3406  -5.31 0.0000
sexo2=masculino          0.7085 0.2542   2.79 0.0053
ec2=solteiro            0.6842 0.3259   2.10 0.0358

```

## {Ajuste do Modelo de razão contínua}

### Reestruturação dos dados

```

> cr=cr.setup(y.ord)
> y.mrc=cr$y
> cohort=cr$cohort
> sexo.cr=sexo[cr$subs]
> ec.cr=ec[cr$subs]

```

## Univariada

```
> mrc1=lrn(y.mrc~sexo.cr+cohort)
> mrc1

Logistic Regression Model

lrn(formula = y.mrc ~ sexo.cr + cohort)

Frequencies of Responses
  0  1
215 245

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      460      2e-13      145.63      2      0      0.791      0.583
Gamma      Tau-a      R2      Brier
0.731      0.291      0.362      0.176

      Coef      S.E.      Wald Z P
Intercept      -1.1948 0.2051 -5.82 0.0000
sexo.cr=masculino      0.6482 0.2329 2.78 0.0054
cohort=y.ord>=QV comprometida      2.6110 0.2509 10.41 0.0000
```

```
> mrc2=lrn(y.mrc ~ec.cr+cohort)
> mrc2

Logistic Regression Model

lrn(formula = y.mrc ~ ec.cr + cohort)

Frequencies of Responses
  0  1
215 245

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      460      2e-13      143.59      2      0      0.779      0.558
Gamma      Tau-a      R2      Brier
0.767      0.278      0.358      0.178

      Coef      S.E.      Wald Z P
Intercept      -1.3670 0.2843 -4.81 0.0000
ec.cr=solteiro      0.7015 0.2921 2.40 0.0163
cohort=y.ord>=QV comprometida      2.5780 0.2487 10.36 0.0000
```

## Multivariada

```
> mrc=lrn(y.mrc ~sexo.cr+ec.cr+cohort)
> mrc

Logistic Regression Model

lrn(formula = y.mrc ~ sexo.cr + ec.cr + cohort)

Frequencies of Responses
  0  1
215 245

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      460      9e-13      150.34      3      0      0.798      0.597
      Gamma      Tau-a      R2      Brier
      0.698      0.298      0.372      0.174

              Coef      S.E.      Wald Z P
Intercept      -1.6959  0.3177  -5.34  0.0000
sexo.cr=masculino      0.6017  0.2348  2.56  0.0104
ec.cr=solteiro      0.6312  0.2948  2.14  0.0322
cohort=y.ord>=QV comprometida      2.6629  0.2557  10.41  0.0000
```

## Modelo saturado

```
> mcrcs=lrn(y.mrc ~(sexo.cr+ec.cr)*cohort)
> mcrcs

Logistic Regression Model

lrn(formula = y.mrc ~ (sexo.cr + ec.cr) * cohort)

Frequencies of Responses
  0  1
215 245

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      460      1e-12      153.47      5      0      0.802      0.603
      Gamma      Tau-a      R2      Brier
      0.708      0.301      0.379      0.173

              Coef      S.E.      Wald Z P
Intercept      -1.8736  0.4058  -4.62  0.0000
sexo.cr=masculino      0.8902  0.2926  3.04  0.0023
ec.cr=solteiro      0.6069  0.3881  1.56  0.1178
cohort=y.ord>=QV comprometida      3.0804  0.5852  5.26  0.0000
sexo.cr=masculino * cohort=y.ord>=QV comprometida      -0.8965  0.5100  -1.76  0.0788
ec.cr=solteiro * cohort=y.ord>=QV comprometida      0.1054  0.5993  0.18  0.8604
```

**{Ajuste do Modelo multinomial}****Univariada**

```

> mult1=vglm(y.ord~sexo,multinomial)
> summary(mult1)

Call:
vglm(formula = y.ord ~ sexo2, family = multinomial)

Pearson Residuals:
           Min           1Q   Median           3Q          Max
log(mu[,1]/mu[,3]) -2.1782 -0.44508 -0.25044  1.1052  1.8119
log(mu[,2]/mu[,3]) -2.3357 -0.60257  0.64510  0.8432  0.8432

Coefficients:
           Value Std. Error t value
(Intercept):1  0.479573   0.35291  1.35893
(Intercept):2  1.697731   0.30167  5.62769
sexo2masculino:1 0.986764   0.45453  2.17098
sexo2masculino:2 0.071556   0.41114  0.17404

Number of linear predictors: 2

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])

Dispersion Parameter for multinomial family: 1

Residual Deviance: 487.2013 on 542 degrees of freedom

Log-likelihood: -243.6006 on 542 degrees of freedom

Number of Iterations: 5

```

```

> mult2=vglm(y.ord ~ec2,multinomial)
> summary(mult2)

Call:
vglm(formula = y.ord ~ ec2, family = multinomial)

Pearson Residuals:
           Min           1Q   Median           3Q          Max
log(mu[,1]/mu[,3]) -2.1796 -0.39823 -0.39823  1.22085  1.83484
log(mu[,2]/mu[,3]) -2.4574 -0.67594  0.75768  0.76016  0.76016

Coefficients:
           Value Std. Error t value
(Intercept):1 0.10536   0.45947  0.22931
(Intercept):2 1.20397   0.38006  3.16786
ec2solteiro:1 1.28093   0.52621  2.43425
ec2solteiro:2 0.71140   0.45258  1.57188

Number of linear predictors: 2

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])

Dispersion Parameter for multinomial family: 1

Residual Deviance: 492.1505 on 542 degrees of freedom

Log-likelihood: -246.0752 on 542 degrees of freedom

Number of Iterations: 5

```

## Multivariada

```
> mult=vglm(y.ord ~sexo2+ec2,multinomial)
> summary(mult)

Call:
vglm(formula = y.ord ~ sexo2 + ec2, family = multinomial)

Pearson Residuals:
             Min           1Q       Median           3Q          Max
log(mu[,1]/mu[,3]) -2.4057 -0.47193 -0.27591  1.04757  2.40506
log(mu[,2]/mu[,3]) -2.5337 -0.59997  0.62934  0.84275  0.84275

Coefficients:
             Value Std. Error  t value
(Intercept):1  -0.413443   0.52874 -0.781941
(Intercept):2   1.197066   0.42622  2.808588
sexo2masculino:1  0.902881   0.46043  1.960950
sexo2masculino:2  0.014838   0.41537  0.035723
ec2solteiro:1    1.192253   0.53198  2.241177
ec2solteiro:2    0.709825   0.45471  1.561038

Number of linear predictors: 2

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])

Dispersion Parameter for multinomial family: 1

Residual Deviance: 482.2711 on 540 degrees of freedom

Log-likelihood: -241.1355 on 540 degrees of freedom

Number of Iterations: 5
```

**{Ajuste do Modelo estereótipo}****Univariada**

```

> st1=rrvglm(y.ord~sexo2,multinomial)
> summary(st1)

Call:
rrvglm(formula = y.ord ~ sexo2, family = multinomial)

Pearson Residuals:
           Min          1Q      Median          3Q          Max
log(mu[,1]/mu[,3]) -2.1782 -0.44508 -0.25044  1.1052  1.8119
log(mu[,2]/mu[,3]) -2.3357 -0.60257  0.64510  0.8432  0.8432

Coefficients:
           Value Std. Error t value
I(lv.mat)    0.072516   0.39167  0.18515
(Intercept):1 0.479573   0.35295  1.35876
(Intercept):2 1.697731   0.30172  5.62693
sexo2masculino 0.986764   0.45463  2.17046

Number of linear predictors: 2

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])

Dispersion Parameter for multinomial family: 1

Residual Deviance: 487.2013 on 542 degrees of freedom

Log-likelihood: -243.6006 on 542 degrees of freedom

Number of Iterations: 4
> st2=rrvglm(y.ord ~ec2,multinomial)
> summary(st2)

Call:
rrvglm(formula = y.ord ~ ec2, family = multinomial)

Pearson Residuals:
           Min          1Q      Median          3Q          Max
log(mu[,1]/mu[,3]) -2.1796 -0.39823 -0.39823  1.22085  1.83484
log(mu[,2]/mu[,3]) -2.4574 -0.67594  0.75768  0.76016  0.76016

Coefficients:
           Value Std. Error t value
I(lv.mat)    0.55538   0.25698  2.16120
(Intercept):1 0.10536   0.45948  0.22931
(Intercept):2 1.20397   0.37999  3.16843
ec2solteiro  1.28093   0.52620  2.43432

Number of linear predictors: 2

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])

Dispersion Parameter for multinomial family: 1

Residual Deviance: 492.1505 on 542 degrees of freedom

Log-likelihood: -246.0752 on 542 degrees of freedom

Number of Iterations: 4

```

## Multivariada

```

> st=rrvglm(y.ord ~sexo2+ec2,multinomial)
> summary(st)

Call:
rrvglm(formula = y.ord ~ sexo2 + ec2, family = multinomial)

Pearson Residuals:
           Min          1Q      Median          3Q          Max
log(mu[,1]/mu[,3]) -2.4396 -0.47230 -0.27861  1.04669  2.47106
log(mu[,2]/mu[,3]) -2.5715 -0.60414  0.68859  0.83697  0.83697

Coefficients:
           Value Std. Error t value
I(lv.mat)    0.29007   0.23418  1.23869
(Intercept):1 -0.35398   0.58128 -0.60897
(Intercept):2  1.35540   0.44825  3.02373
sexo2masculino 1.10699   0.39703  2.78816
ec2solteiro   0.95406   0.57979  1.64553

Number of linear predictors: 2

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])

Dispersion Parameter for multinomial family: 1

Residual Deviance: 483.6975 on 541 degrees of freedom

Log-likelihood: -241.8488 on 541 degrees of freedom

Number of Iterations: 4

```

## Exemplo 2 – Variável resposta Item Funcionamento ocupacional

### {Leitura dados e criação de vetores para análise}

```
library(foreign)
banco=read.spss("c:/bancoqls_itensR.sav")
y.ord=banco$FUNC_OCU
y.bin1=banco$FUNC_O_A
y.bin2=banco$FUNC_O_B
sexo=banco$SEXO
ec=banco$ESTCIV
```

### {Ajuste dos modelos de regressão logística binária}

```
> library(Design)
```

### {Comparação QV muito comprometida X QV comprometida e QV inalterada}

```
> mb=lrn(y.bin1~sexo2+ec2)
> mb
```

Logistic Regression Model

```
lrn(formula = y.bin1 ~ sexo2 + ec2)
```

Frequencies of Responses

	QV muito comprometida	QV comp + QV inalterada
	71	202

	Obs	Max	Deriv	Model L.R.	d.f.	P	C	Dxy
	273		6e-07	24.44	2	0	0.671	0.341
Gamma	0.558		Tau-a	R2	Brier			
			0.132	0.126	0.176			

	Coef	S.E.	Wald	Z	P
Intercept	2.4571	0.4663	5.27	0.0000	
sexo2=masculino	1.5336	0.3580	4.28	0.0000	
ec2=solteiro	0.3908	0.4135	0.95	0.3446	

```
> mb2=lrn(y.bin2~sexo2+ec2)
> mb2
```

Logistic Regression Model

```
lrn(formula = y.bin2 ~ sexo2 + ec2)
```

Frequencies of Responses

	QV muito comp. + QV compr.	QV inalterada
	230	43

	Obs	Max	Deriv	Model L.R.	d.f.	P	C	Dxy
	273		2e-13	4	2	0.135	0.588	0.176
Gamma	0.267		Tau-a	R2	Brier			
			0.047	0.025	0.131			

	Coef	S.E.	Wald	Z	P
Intercept	-1.0848	0.3843	-2.82	0.0048	
sexo2=masculino	0.5922	0.3366	1.76	0.0785	
ec2=solteiro	0.3219	0.4061	0.79	0.4279	

## {Ajuste do Modelo de chances proporcionais}

### Univariada

```
> mcp1=lrn(y.ord~sexo)
> mcp1

Logistic Regression Model

lrn(formula = y2.ord ~ sexo2)

Frequencies of Responses
      QV inalterada      QV comprometida QV muito comprometida
              43                159                71

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273   4e-10  19.09      1          0      0.615  0.229
Gamma   Tau-a      R2      Brier
0.473   0.131      0.079  0.132

      Coef  S.E.  Wald Z P
y>=QV comprometida  1.123 0.2047  5.49 0
y>=QV muito comprometida -1.770 0.2295 -7.71 0
sexo2=masculino      1.095 0.2578  4.25 0
```

```
> mcp2=lrn(y.ord~ec)
> mcp2

Logistic Regression Model

lrn(formula = y2.ord ~ ec2)

Frequencies of Responses
      QV inalterada      QV comprometida QV muito comprometida
              43                159                71

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273   2e-08  2.24      1          0.1343  0.531  0.062
Gamma   Tau-a      R2      Brier
0.213   0.036      0.01  0.132

      Coef  S.E.  Wald Z P
y>=QV comprometida  1.308 0.2939  4.45 0.0000
y>=QV muito comprometida -1.433 0.2969 -4.83 0.0000
ec2=solteiro      0.464 0.3109  1.49 0.1356
```

## Multivariada

```

> mcp3=lrn(y.ord~sexo+ec)
> mcp3=lrn(y.ord~sexo+ec)
> mcp3

Logistic Regression Model

lrn(formula = y2.ord ~ sexo2 + ec2)

Frequencies of Responses
      QV inalterada      QV comprometida QV muito comprometida
              43                159                71

      Obs  Max Deriv Model L.R.      d.f.      P      C      Dxy
      273   6e-10   20.43      2      0      0.628  0.256
      Gamma  Tau-a      R2      Brier
      0.407   0.146   0.085   0.132

      Coef      S.E.      Wald Z P
y>=QV comprometida      0.8452 0.3140  2.69 0.0071
y>=QV muito comprometida -2.0593 0.3425 -6.01 0.0000
sexo2=masculino      1.0733 0.2585  4.15 0.0000
ec2=solteiro      0.3627 0.3146  1.15 0.2489

```



## Modelo com parametrização gamma

```
. gologit2 func_ocup sexo estciv, autofit lrforce gamma
```

---

Testing parallel-lines assumption using the .05 level of significance...

Step 1: Constraints for parallel lines imposed for estciv (P Value = 0.8926)  
Step 2: Constraints for parallel lines are not imposed for  
sexo (P Value = 0.03201)

Wald test of parallel-lines assumption for the final model:

( 1) [0]estciv - [1]estciv = 0

chi2( 1) = 0.02  
Prob > chi2 = 0.8926

An insignificant test statistic indicates that the final model does not violate the proportional odds/parallel-lines assumption

If you refit this exact same model with gologit2, instead of autofit, you can save time by using the parameter

```
pl(estciv)
```

---

Generalized Ordered Logit Estimates

Number of obs	=	273
LR chi2(3)	=	25.29
Prob > chi2	=	0.0000
Pseudo R2	=	0.0484

Log likelihood = -248.40394

( 1) [0]estciv - [1]estciv = 0

---

	func_ocup	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
0						
	sexo	1.535714	.3577206	4.29	0.000	.8345943 2.236833
	estciv	.3644684	.3175926	1.15	0.251	-.2580016 .9869384
	_cons	.5355293	.1672278	3.20	0.001	.207769 .8632897
1						
	sexo	.5955824	.3356668	1.77	0.076	-.0623124 1.253477
	estciv	.3644684	.3175926	1.15	0.251	-.2580016 .9869384
	_cons	-2.009638	.240845	-8.34	0.000	-2.481686 -1.537591

---

Alternative parameterization: Gammas are deviations from proportionality

---

	func_ocup	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Beta						
	sexo	1.535714	.3577206	4.29	0.000	.8345943 2.236833
	estciv	.3644684	.3175926	1.15	0.251	-.2580016 .9869384
Gamma_2						
	sexo	-.9401314	.43844	-2.14	0.032	-1.799458 -.0808049
Alpha						
	_cons_1	.5355293	.1672278	3.20	0.001	.207769 .8632897
	_cons_2	-2.009638	.240845	-8.34	0.000	-2.481686 -1.537591

---

**APÊNDICE C**  
**RESULTADOS DO ARTIGO 2**

## Estatísticas Descritivas dos Dados Utilizados no Artigo 2

### Exemplo 1A – Estudo sobre Portadores de Marcapasso (Escala AQUAREL)

Tabela 4: Estatísticas descritivas do domínio Arritmia da escala AQUAREL (contínuo)

<b>AQUAREL</b>	
<b>Arritmia</b>	
Média	80,3
Mediana	85,0
Desvio-padrão	20,3
Mínimo	20,0
Máximo	100,0
Percentil 25	70,0
50	85,0
75	100,0

Tabela 5: Estatísticas descritivas do domínio Arritmia da escala AQUAREL após Codificação (de acordo com percentis)

<b>AQUAREL Arritmia</b>	<b>Frequência</b>	<b>Percentual</b>
QV ruim	34	24,5
QV razoável	29	20,9
QV boa	36	25,9
QV muito boa	40	28,8
<b>Total</b>	<b>139</b>	<b>100,0</b>

### Exemplo 1B – Estudo sobre Portadores de Marcapasso (Escala SF-36)

Tabela 6: Estatísticas descritivas do domínio Aspectos Emocionais da escala SF-36

<b>Aspectos emocionais SF-36</b>	<b>Frequência</b>	<b>Percentual</b>
0	58	41,7
33,33	17	12,2
66,67	14	10,1
100,00	50	36,0
<b>Total</b>	<b>139</b>	<b>100,0</b>

**Exemplo 2 – Levantamento Nacional de Saúde e Nutrição (NHANES II)**

Tabela 7: Estatísticas descritivas da Condição de Saúde, de acordo com o estudo

## NHANES II

<b>Aspectos emocionais SF-36</b>	<b>Frequência</b>	<b>Percentual</b>
excelente	2407	23,3
bom	2591	25,1
médio	2938	28,4
razoável	1670	16,2
ruim	729	7,1
<b>Total</b>	<b>10335</b>	<b>100,0</b>

## Análises Univariadas do Artigo 2

### Exemplo 1A – Estudo sobre Portadores de Marcapasso (Escala AQUAREL)

Tabela 7: Associação entre os fatores clínicos e sócio-demográficos e a QV no Domínio Arritmia da Escala AQUAREL com resultado do teste Qui-quadrado de Tendência.

	AQUAREL Arritmia				Valor-p*
	QV ruim	QV razoável	QV boa	QV muito boa	
<b>Sexo</b>					
Feminino	25 29,8%	20 23,8%	22 26,2%	17 20,2%	<b>0,005</b>
Masculino	9 16,4%	9 16,4%	14 25,5%	23 41,8%	
<b>Estado civil</b>					
com parceiro	13 18,3%	13 18,3%	20 28,2%	25 35,2%	<b>0,026</b>
sem parceiro	21 30,9%	16 23,5%	16 23,5%	15 22,1%	
<b>Escolaridade</b>					
> 4 anos	6 26,1%	4 17,4%	5 21,7%	8 34,8%	0,749
4 anos ou menos	28 24,1%	25 21,6%	31 26,7%	32 27,6%	
<b>Trabalha</b>					
sim	2 11,8%	6 35,3%	3 17,6%	6 35,3%	0,504
não	32 26,2%	23 18,9%	33 27,0%	34 27,9%	
<b>Medicação</b>					
Ausente	8 25,8%	8 25,8%	4 12,9%	11 35,5%	0,937
Presente	23 23,2%	19 19,2%	30 30,3%	27 27,3%	
<b>Antiarrítmico</b>					
Ausente	24 21,8%	23 20,9%	27 24,5%	36 32,7%	<b>0,073</b>
Presente	7 38,9%	3 16,7%	6 33,3%	2 11,1%	
<b>IECA</b>					
Ausente	19 27,1%	13 18,6%	14 20,0%	24 34,3%	0,876
Presente	12 20,7%	13 22,4%	19 32,8%	14 24,1%	
<b>Diogoxina</b>					
Ausente	24 21,6%	22 19,8%	28 25,2%	37 33,3%	<b>0,018</b>
Presente	7 41,2%	4 23,5%	5 29,4%	1 5,9%	
<b>Diurético</b>					
Ausente	15 21,1%	15 21,1%	15 21,1%	26 36,6%	0,157
Presente	16 28,1%	11 19,3%	18 31,6%	12 21,1%	

<b>Beta bloqueador</b>					
Ausente	30	23	28	36	0,799
	25,6%	19,7%	23,9%	30,8%	
Presente	1	3	5	2	
	9,1%	27,3%	45,5%	18,2%	
<b>Med. Sistema Nervoso</b>					
Ausente	24	20	28	33	0,230
	22,9%	19,0%	26,7%	31,4%	
Presente	7	6	5	5	
	30,4%	26,1%	21,7%	21,7%	
<b>Outros medicamentos</b>					
Ausente	20	14	13	25	0,994
	27,8%	19,4%	18,1%	34,7%	
Presente	11	12	20	13	
	19,6%	21,4%	35,7%	23,2%	
<b>Classe funcional</b>					
I	11	19	21	34	<0,001
	12,9%	22,4%	24,7%	40,0%	
II	12	4	11	4	
	38,7%	12,9%	35,5%	12,9%	
III	11	6	4	2	
	47,8%	26,1%	17,4%	8,7%	
<b>Indicação do transplante</b>					
BAV	4	7	13	10	0,518
	11,8%	20,6%	38,2%	29,4%	
DNS	0	2	0	1	
	,0%	66,7%	,0%	33,3%	
Outros	0	0	0	1	
	,0%	,0%	,0%	100,0%	
<b>Doença de chagas</b>					
não	3	5	9	14	0,004
	9,7%	16,1%	29,0%	45,2%	
sim	21	18	22	16	
	27,3%	23,4%	28,6%	20,8%	
<b>Pressão arterial diastólica</b>					
Normal	7	10	18	14	0,780
	14,3%	20,4%	36,7%	28,6%	
Alterada	6	5	5	9	
	24,0%	20,0%	20,0%	36,0%	
<b>Pressão arterial sistólica</b>					
Normal	11	13	15	16	0,179
	20,0%	23,6%	27,3%	29,1%	
Alterada	2	2	8	7	
	10,5%	10,5%	42,1%	36,8%	
<b>Idade</b>					
23 a 40 anos	5	4	3	8	0,194
	25,0%	20,0%	15,0%	40,0%	
41 a 60 anos	15	11	11	11	
	31,3%	22,9%	22,9%	22,9%	
61 a 70 anos	8	10	9	8	
	22,9%	28,6%	25,7%	22,9%	
71 anos ou mais	6	4	13	13	
	16,7%	11,1%	36,1%	36,1%	

\* Teste Qui-quadrado de Tendência

### Exemplo 1B – Estudo sobre Portadores de Marcapasso (Escala SF-36)

Tabela 8: Associação entre os fatores clínicos e sócio-demográficos e a QV no domínio Aspectos Emocionais da escala SF36 com resultado do teste Qui-quadrado de Tendência.

	SF36 Aspecto emocional				Valor-p*
	0	33	67	100	
<b>Sexo</b>					
Feminino	43 51,2%	7 8,3%	7 8,3%	27 32,1%	<b>0,029</b>
Masculino	15 27,3%	10 18,2%	7 12,7%	23 41,8%	
<b>Estado civil</b>					
com parceiro	21 29,6%	13 18,3%	8 11,3%	29 40,8%	<b>0,024</b>
sem parceiro	37 54,4%	4 5,9%	6 8,8%	21 30,9%	
<b>Escolaridade</b>					
> 4 anos	4 17,4%	8 34,8%	3 13,0%	8 34,8%	0,210
4 anos ou menos	54 46,6%	9 7,8%	11 9,5%	42 36,2%	
<b>Trabalha</b>					
sim	5 29,4%	3 17,6%	3 17,6%	6 35,3%	0,525
não	53 43,4%	14 11,5%	11 9,0%	44 36,1%	
<b>Medicação</b>					
Ausente	9 29,0%	5 16,1%	6 19,4%	11 35,5%	0,367
Presente	45 45,5%	10 10,1%	7 7,1%	37 37,4%	
<b>Antiarrítmico</b>					
Ausente	44 40,0%	12 10,9%	12 10,9%	42 38,2%	0,323
Presente	9 50,0%	3 16,7%	1 5,6%	5 27,8%	
<b>IECA</b>					
Ausente	28 40,0%	11 15,7%	7 10,0%	24 34,3%	0,830
Presente	25 43,1%	4 6,9%	6 10,3%	23 39,7%	
<b>Diogoxina</b>					
Ausente	42 37,8%	14 12,6%	11 9,9%	44 39,6%	<b>0,004</b>
Presente	11 64,7%	1 5,9%	2 11,8%	3 17,6%	
<b>Diurético</b>					
Ausente	23 32,4%	9 12,7%	8 11,3%	31 43,7%	<b>0,023</b>
Presente	30 52,6%	6 10,5%	5 8,8%	16 28,1%	

<b>Beta bloqueador</b>					
Ausente	49	14	12	42	
	41,9%	12,0%	10,3%	35,9%	
Presente	4	1	1	5	0,619
	36,4%	9,1%	9,1%	45,5%	
<b>Med. Sistema Nervoso</b>					
Ausente	36	14	12	43	
	34,3%	13,3%	11,4%	41,0%	
Presente	17	1	1	4	0,002
	73,9%	4,3%	4,3%	17,4%	
<b>Outros medicamentos</b>					
Ausente	25	9	8	30	
	34,7%	12,5%	11,1%	41,7%	
Presente	28	6	5	17	0,100
	50,0%	10,7%	8,9%	30,4%	
<b>Classe funcional</b>					
I	27	11	10	37	
	31,8%	12,9%	11,8%	43,5%	
II	17	5	1	8	
	54,8%	16,1%	3,2%	25,8%	0,011
III	14	1	3	5	
	60,9%	4,3%	13,0%	21,7%	
<b>Doença de chagas</b>					
não	10	8	1	12	
	32,3%	25,8%	3,2%	38,7%	
sim	39	2	7	29	0,394
	50,6%	2,6%	9,1%	37,7%	
<b>Pressão arterial diastólica</b>					
Normal	15	6	7	21	
	30,6%	12,2%	14,3%	42,9%	
Alterada	14	1	1	9	0,155
	56,0%	4,0%	4,0%	36,0%	
<b>Pressão arterial sistólica</b>					
Normal	22	4	8	21	
	40,0%	7,3%	14,5%	38,2%	
Alterada	7	3	0	9	0,734
	36,8%	15,8%	,0%	47,4%	
<b>Idade</b>					
23 a 40 anos	6	4	5	5	
	30,0%	20,0%	25,0%	25,0%	
41 a 60 anos	22	4	2	20	
	45,8%	8,3%	4,2%	41,7%	
61 a 70 anos	16	3	4	12	0,984
	45,7%	8,6%	11,4%	34,3%	
71 anos ou mais	14	6	3	13	
	38,9%	16,7%	8,3%	36,1%	

\* Teste Qui-quadrado de Tendência

## Exemplo 2 – Levantamento Nacional de Saúde e Nutrição (NNHANES II)

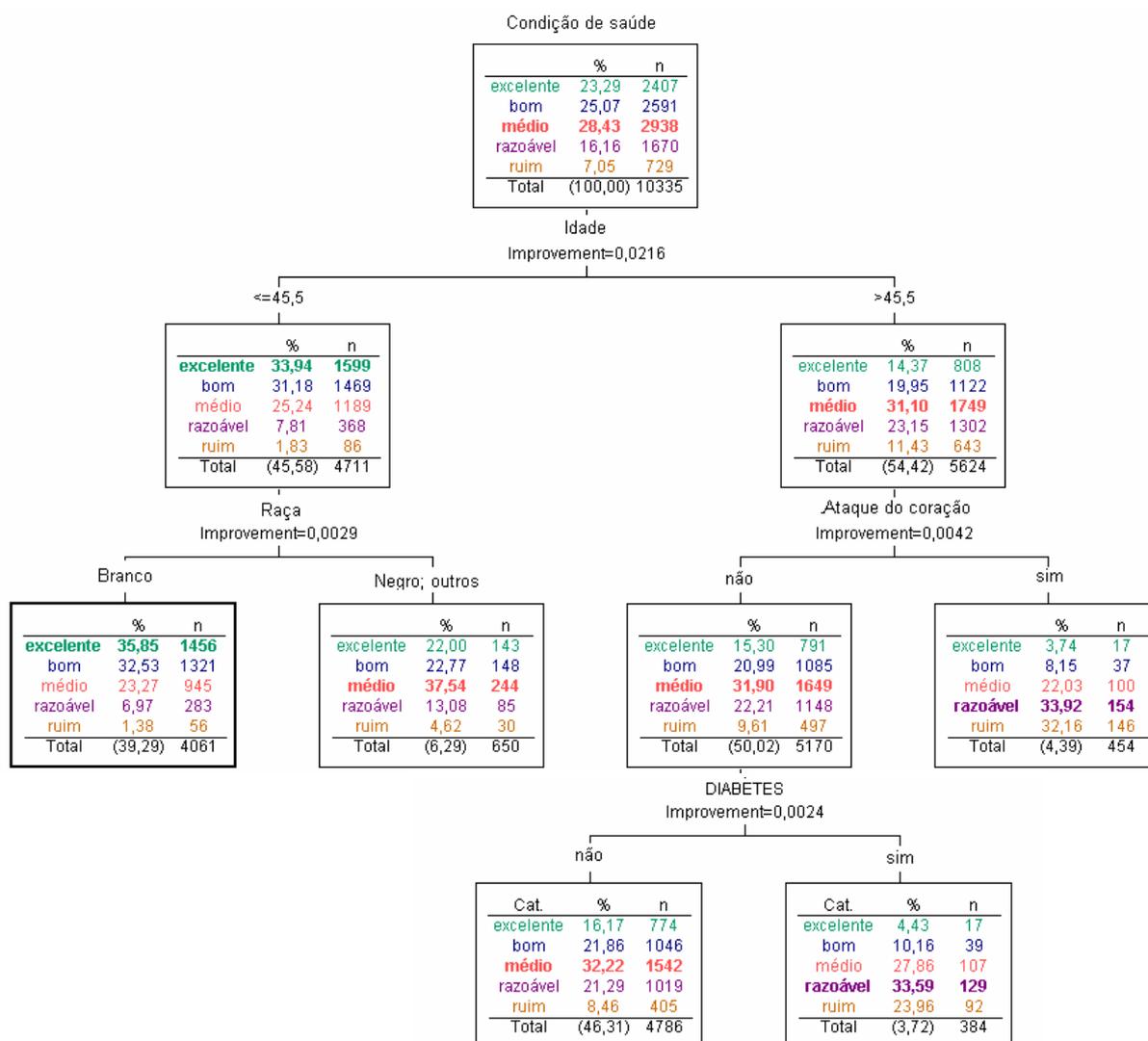


Figura 1: Árvore de decisão de acordo com o algoritmo CART para seleção dos efeitos principais que devem ser incluídos no modelo logístico.

## PROGRAMA R PARA EXECUÇÃO DOS MODELOS - ARTIGO 2

### Exemplo 1A – Estudo sobre Portadores de Marcapasso (Escala AQUAREL)

#### {Leitura dados e criação de vetores para análise}

```
library(foreign)
bancob=read.spss("c:/banco Bruna.sav")
summary(bancob)
y.ar=bancob$AQUA_A
sexo=bancob$SEXO
antiar=bancob$ANTARRIT
dig=bancob$DIGOXINA
clasfun=bancob$CLASSFUN
chagas=bancob$CHAGAS
estciv=bancob$ESTCIV
```

#### {Ajuste do modelo de chances proporcionais}

##### Univariada

```
library(Design)
mcp1sexo=lrn(y.ar~sexo)
mcp1sexo
mcp1estciv=lrn(y.ar~estciv)
mcp1estciv
mcp1antiar=lrn(y.ar~ antiar)
mcp1antiar
mcp1dig=lrn(y.ar~dig)
mcp1dig
mcp1clasfun=lrn(y.ar~clasfun)
mcp1clasfun
mcp1chagas=lrn(y.ar~chagas)
mcp1chagas
```

##### Multivariada

```
mcp1=lrn(y.ar~sexo+estciv+antiar+dig+clasfun+chagas)
mcp1
mcp2=lrn(y.ar~sexo+antiar+dig+clasfun+chagas)
mcp2
mcp3=lrn(y.ar~sexo+antiar+dig+clasfun)
mcp3
mcp4=lrn(y.ar~sexo+dig+clasfun)
mcp4

mcp=lrn(AQUAREL~sexo+dig+clasfun, x=TRUE, y=TRUE)
mcp
```

#### {Análise resíduos}

```
par(mfrow=c(1,4))
resid(mcp, 'score.binary', pl=TRUE)
resid(mcp, 'partial', pl=TRUE)
```

## Exemplo 1B – Estudo sobre Portadores de Marcapasso (Escala SF-36)

### {Leitura dados e criação de vetores para análise}

```
library(foreign)
bancob=read.spss("c:/banco Bruna reduzido mod.sav")
summary(bancob)
y.sf=bancob$SFLMEMO
sexo=bancob$SEXO
dig=bancob$DIGOXINA
diur=bancob$DIURETIC
medsnc=bancob$MEDSNC
imccod=bancob$IMCCOD
estciv=bancob$ESTCIV
clasfun=bancob$CLASSF_A
```

### {Ajuste do modelo de chances proporcionais}

#### Univariada

```
mcp2dig=lm(y.sf~dig)
mcp2dig
mcp2clasfun=lm(y.sf~clasfun)
mcp2clasfun
mcp2sexo=lm(y.sf~sexo)
mcp2sexo
mcp2diur=lm(y.sf~diur)
mcp2diur
mcp2estciv=lm(y.sf~estciv)
mcp2estciv
mcp2medsnc=lm(y.sf~medsnc)
mcp2medsnc
```

#### Multivariada

```
mcp2=lm(y.sf~dig+clasfun+sexo+diur+estciv+medsnc)
mcp2
mcp2_1=lm(y.sf~dig+sexo+diur+estciv+medsnc)
mcp2_1
mcp2_2=lm(y.sf~dig+diur+estciv+medsnc)
mcp2_2
mcp2_3=lm(y.sf~diur+estciv+medsnc)
mcp2_3

mcp2=lm(y.sf~diur+estciv+medsnc, x=TRUE, y=TRUE)
mcp2
```

### {Análise de resíduos}

```
par(mfrow=c(1,3))
resid(mcp2, 'score.binary', pl=TRUE)
resid(mcp2, 'partial', pl=TRUE)
```

## {Ajuste do modelo de chances proporcionais parciais} - STATA

```
. use "C:\Mery\Mestrado\artigo ajuste\análises Banco Bruna\banco Bruna
Stata.dta"
. gologit2 sf sexo diago diure medsnc classfun2 estciv, autofit lrforce
```

## Exemplo 2 – Levantamento Nacional de Saúde e Nutrição (NHANES II)

### {Leitura dados e criação de vetores para análise}

```
library(foreign)
library(Design)
banco2=read.spss("c:/nhanesR.sav")
saude=banco2$HEALTH
idade=banco2$AGE
ataque_coracao=banco2$HEARTATK
diabetes=banco2$DIABETES
raca=banco2$RACE2
```

### {Ajuste do modelo de chances proporcionais}

```
mcp1=lm(saude~idade)
mcp1
mcp2=lm(saude~ataque_coracao)
mcp2
mcp3=lm(saude~diabetes)
mcp3
mcp4=lm(saude~raca)
mcp4
mcp=lm(saude~idade+ataque_coracao+diabetes+raca, x=TRUE, y=TRUE)
mcp
```

### {Análise de resíduos}

```
par(mfrow=c(1,5))
resid(mcp, 'score.binary', pl=TRUE)
resid(mcp, 'partial', pl=TRUE)
```

### {Ajuste do modelo de razão contínua}

#### Reestruturação dos dados

```
mcr=cr.setup(saude2)
ymrc=mcr$y
idademrc=idade[mcr$subs]
atcormrc=ataque_coracao[mcr$subs]
diabmrc=diabetes[mcr$subs]
racamrc=raca[mcr$subs]
coh=mcr$cohort
```

## Ajuste dos modelos

```
mrc1=lm(ymrc~idademrc+coh)
mrc1
mrc2=lm(ymrc~atcormrc+coh)
mrc2
mrc3=lm(ymrc~diabmrc+coh)
mrc3
mrc4=lm(ymrc~racamrc+coh)
mrc4
mrc=lm(ymrc~idademrc+atcormrc+diabmrc+racamrc+coh, x=TRUE, y=TRUE)
mrc
```

## {Ajuste do modelo multinomial}

```
library(VGAM)
mmu1=vglm(saude2~idade,multinomial)
summary(mmu1)
mmu2=vglm(saude2~ ataque_coracao,multinomial)
summary(mmu2)
mmu3=vglm(saude2~diabetes,multinomial)
summary(mmu3)
mmu4=vglm(saude2~raca,multinomial)
summary(mmu4)

mmu=vglm(saude2~idade+ataque_coracao+diabetes+raca,multinomial)
summary(mmu)
```

## {Ajuste do modelo estereótipo}

```
s1=rrvglm(saude2~idade,multinomial)
summary(s1)
s2=rrvglm(saude2~ataque_coracao,multinomial)
summary(s2)
s3=rrvglm(y saude2~diabetes,multinomial)
summary(s3)
s4=rrvglm(saude2~raca,multinomial)
summary(s4)

s=rrvglm(saude2~idade+ataque_coracao+diabetes+raca,multinomial)
summary(s)
```