UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

MODELO DE REGRESSÃO LOGÍSTICA ORDINAL EM DADOS CATEGÓRICOS NA ÁREA DE ERGONOMIA EXPERIMENTAL

SANTHIAGO GUEDES MONTENEGRO

JOÃO PESSOA
2009

SANTHIAGO GUEDES MONTENEGRO

MODELO DE REGRESSÃO LOGÍSTICA ORDINAL EM DADOS CATEGÓRICOS NA ÁREA DE ERGONOMIA EXPERIMENTAL

Dissertação submetida à apreciação da Banca examinadora do Programa de Pós-graduação em Engenharia de Produção na área de concentração de ergonomia, como requisito parcial à obtenção do título de Mestre em Engenharia de Produção.

Orientador: Luiz Bueno da Silva, Dr.

João Pessoa

2009

M777m Montenegro, Santhiago Guedes.
Modelo de regressão logística ordinal em dados categóricos na área de ergonomia experimental / Santhiago Guedes Montenegro.- João Pessoa, 2009.
85f. : il.
Orientador: Luiz Bueno da Silva
Dissertação (Mestrado) – UFPB/CT
1. Ergonomia (Engenharia de Produção). 2. Regressão Logística Multinomial. 3. Índice de Capacidade para o Trabalho (ICT).

UFPB/BC CDU: 658.3.041(043)

SANTHIAGO GUEDES MONTENEGRO

MODELO DE REGRESSÃO LOGÍSTICA ORDINAL EM DADOS CATEGÓRICOS NA ÁREA DE ERGONOMIA EXPERIMENTAL

Dissertação submetida à apreciação da Banca examinadora do Programa de Pós-graduação em Engenharia de Produção na área de concentração de ergonomia, como requisito parcial à obtenção do título de Mestre em Engenharia de Produção.

Banca Examinadora

_____

Prof. Luiz Bueno da Silva, Dr.
(Orientador – PPGEP - UFPB)

_____

Prof. Eufrásio de Andrade Lima Neto, Dr.
(Departamento de Estatística – UFPB - Examinador Externo)

_____

Prof. Ulisses Umbelino dos Anjos, Dr.
(Departamento de Estatística – UFPB - Examinador Externo)

João Pessoa, 17 de novembro de 2009

Para Cláudia Valéria, mulher forte e corajosa que me ajuda, completa-me, entende-me, faz-me melhor e escreve as minhas dedicatórias...

AGRADECIMENTOS

É ilusão pensar que se pode chegar a algum lugar, atingir algum objetivo, realizar conquistas enfim, sozinho. Assim, tenho muito a agradecer e espero ser perdoado para com aqueles que não terão seus nomes escritos aqui, por obra de minha relapsa memória.

Agradeço primeiramente a Deus, que abriu portas, criou oportunidades, encorajou, deu capacidade e fez dos Seus milagres para tornar possível essa conquista.

À minha esposa, Cláudia Valéria, que suportou desde variações abruptas de humor até dias, semanas, meses de ausência causada pela pesquisa, e ainda me incentivou a continuar quando estava tudo muito difícil;

Aos meus pais, sem os quais não teria subido os degraus anteriores e ainda nesse patamar continuam apoiando e ajudando segundo suas capacidades e às vezes além delas;

Ao professor orientador Luiz Bueno, que acolheu, soube perdoar falhas e esperar que um mestrando pudesse surgir de um aluno imaturo, além de seu vasto apoio. Foi um verdadeiro orientador não apenas na matéria em estudo, mas em disciplinas da vida;

Ao professor Eufrásio Andrade e Ulisses Umbelino, pelas suas preciosas contribuições e disponibilidade em ajudar;

Ao governo brasileiro, que proporcionou uma Universidade com ensino gratuito e de qualidade;

À Rawlla Eriam pela parceria vitoriosa firmada;

À Priscila, por seu auxílio com a coleta de dados;

A Ana Araújo, secretária do programa de pós-graduação em Engenharia de Produção, que sempre lidou com a burocracia e fez fácil lidar com toda a parafernália de documentos necessários, através de seus préstimos;

Ao colega Hilton Freire, por seus conselhos e ajuda;

A colega Ana Maria Braga, por sua amizade e ajuda;

Ao colega Daniel Moura, por sua preciosa ajuda;

Ao colega Ivonaldo Correia, por seu apoio profissional;

Enfim, a todos os que contribuíram direta ou indiretamente com a realização deste trabalho, meu muito obrigado!

“Se vi mais longe foi porque subi nos ombros dos gigantes”

Isaac Newton (1642 – 1727)

RESUMO

Nas análises realizadas em ergonomia experimental, ainda é raro o uso da regressão logística multinomial nominal e ordinal, tendo sido empregada frequentemente a simplificação dessas ferramentas, a regressão logística binária, mesmo onde a Variável Dependente possui mais de duas categorias. A binarização da Variável Dependente leva a prejuízos na análise de dados, pela perda de informação por aglutinação de categorias e desconsideração de ordenação entre as mesmas. Uma análise de dados usando a regressão logística multinomial ordinal foi realizada em um conjunto de dados contendo uma variável categórica, o Índice de Capacidade para o Trabalho (ICT) de enfermeiros de Unidade de Terapia Intensiva (UTI's) de hospitais públicos na cidade de João Pessoa – PB como variável dependente e variáveis termo-ambientais pessoais, e de organização do trabalho como variáveis independentes. Através desta análise, chegou a fatores de risco que levam ao aumento da probabilidade de queda do ICT dos profissionais envolvidos na pesquisa. Características inerentes a VD bem como ao conjunto de dados utilizado levam a conclusão que o uso da Regressão Logística Multinomial Ordinal tornou possível uma análise mais precisa com resultados mais acurados.

Palavras Chaves: Ergonomia; Regressão Logística Multinomial; Modelagem de Dados; Índice de Capacidade para o Trabalho.

ABSTRACT

On analysis performed at experimental ergonomics, still is rare the use of Ordinal and Nominal Multinomial Logistic Regression, having been employed their simplification, Binary Logistic Regression, even on cases where Dependent Variable (DV) have more of two categories. To make the DV becomes binary leads to damages at data analysis, caused due lose of information by category agglutination and ordination disrespect. An analysis using Ordinal Multinomial Logistic Regression was performed on a data set containing a categorical DV, the Work Ability Index (WAI) of Nurses working on João Pessoa city Public Hospital Intensive Care Unit (ICU), and as Independent Variable (ID) thermal comfort variables, environmental variables, personal variables, and work organization variables. Through this analysis, was found out risk factors that lead to increase the probability of the WAI falls on an inferior category. The DV and used data set features allows to conclude that the Ordinal Multinomial Logistic Regression use made possible a more accurate result and analysis.

Keywords: Ergonomics; Multinomial Logistic Regression; Data Modeling; Work Ability Index.

LISTA DE FIGURAS

FIGURA 1:ERGONOMIA APLICADA 25

FIGURA 2: CURVA EM S GERADA POR UMA FUNÇÃO LOGIT 39

FIGURA 3: PROBABILIDADE CUMULATIVA NO MODELO DE *ODDS* PROPORCIONAL 56

FIGURA 4: BOXPLOT DO ICT EM FUNÇÃO DO SEXO 65

FIGURA 5: BOXPLOT DO ICT EM FUNÇÃO DO ESTADO CIVIL DO ENTREVISTADO 66

FIGURA 6: BOXPLOT DO ICT EM FUNÇÃO DA ESCOLARIDADE DO ENTREVISTADO 66

LISTA DE QUADROS

QUADRO 1: TIPOS DE ESCALA DE MENSURAÇÃO 32

QUADRO 2: TIPOS DE VARIÁVEIS..... 32

QUADRO 4: ESCALA SÉTIMA DA SENSAÇÃO TÉRMICA 68

QUADRO 5: ESCALA DE AVALIAÇÃO TÉRMICA DO AMBIENTE 69

QUADRO 6: ESCALA DE MEDIÇÃO DA PREFERÊNCIA TÉRMICA..... 69

LISTA DE TABELAS

TABELA 1: CLASSIFICAÇÃO - ÍNDICE DE CAPACIDADE PARA O TRABALHO 28

TABELA 2: VALORES DO MODELO DE REGRESSÃO LOGÍSTICA QUANDO A VARIÁVEL INDEPENDENTE É DICOTÔMICA 42

TABELA 3: VARIÁVEIS ALTAMENTE CORRELACIONADAS 70

TABELA 4: COEFICIENTES DE REGRESSÃO DO MODELO PROPOSTO 71

TABELA 5: ODDS RATIO PARA AS VARIÁVEIS PRESENTES NO MODELO 73

TABELA 6: PREVISÃO DO MODELO SOBRE OS DADOS COLETADOS 76

TABELA 7: MATRIZ DE CONFUSÃO DO MODELO EM TERMOS PERCENTUAIS 77

TABELA 8: VALORES PREDITOS E OBSERVADOS DO MODELO 77

LISTA DE ABREVIATURAS

ABERGO – Associação Brasileira de Ergonomia

ICT – Índice de Capacidade para o Trabalho

ICU – *Intensive Care Unit*

IEA – *International Ergonomics Association*

MLG – Modelo Linear Generalizado

VA – Variável Aleatória

VD – Variável Dependente

VI – Variável Independente

WAI – *Work Ability Index*

SUMÁRIO

1 INTRODUÇÃO 16

1.1 PROBLEMATIZAÇÃO 16

1.2 OBJETIVOS 18

1.2.1OBJETIVO GERAL 18

1.2.2OBJETIVOS ESPECÍFICOS 18

1.3 ESTRUTURA DO TRABALHO 19

2 ERGONOMIA 20

2.1 HISTÓRICO DA ERGONOMIA 20

2.2 ERGONOMIA – DEFINIÇÃO E ASPECTOS GERAIS 21

2.3 ERGONOMIA – RAMOS DE ATUAÇÃO 25

2.4 ÍNDICE DE CAPACIDADE PARA O TRABALHO (ICT) 26

3 REFERENCIAL TEÓRICO 29

3.1 MODELAGEM 29

3.2 TIPOS DE VARIÁVEIS 31

3.2.1VARIÁVEIS CATEGÓRICAS 33

3.3 ANÁLISE DE REGRESSÃO 34

3.4 MODELOS LINEARES GENERALIZADOS (MLG) 35

3.4.1MODELOS LOGIT PARA DADOS BINÁRIOS 36

3.5 REGRESSÃO LOGÍSTICA 37

3.5.1REGRESSÃO LOGÍSTICA BINÁRIA 37

3.5.2ESTIMAÇÃO DE PARÂMETROS EM REGRESSÃO LOGÍSTICA 40

3.5.3INTERPRETAÇÃO DOS COEFICIENTES 41

3.5.4INFERÊNCIA 43

3.5.5REGRESSÃO LOGÍSTICA MÚLTIPLA 45

3.5.6TESTANDO A SIGNIFICÂNCIA DE UM MODELO LOGÍSTICO MÚLTIPLO 48

3.6 ESTRATÉGIAS DE SELEÇÃO DE MODELOS 49

3.6.1COLINEARIDADE 50

3.6.2MÉTODOS COMPUTACIONAIS – MÉTODOS STEPWISE 51
3.6.3MÉTODO DE SELEÇÃO BASEADO EM CRITÉRIO DE INFORMAÇÃO 52
3.7 REGRESSÃO LOGÍSTICA MULTINOMIAL 52
3.7.1REGRESSÃO LOGÍSTICA MULTINOMIAL ORDINAL 55
3.8 APLICAÇÃO DE ANÁLISE DE REGRESSÃO EM ESTUDOS NA ÁREA DE ERGONOMIA 58
3.9 PROCEDIMENTOS METODOLÓGICOS 62
4 RESULTADOS 64
4.1 VARIÁVEIS TRATADAS 64
4.2 CORRELAÇÃO ENTRE VARIÁVEIS 70
4.3 MODELAGEM 71
4.4 RAZÃO DAS CHANCES (“ODDS RATIO”) 72
4.5 INTERPRETAÇÃO DOS COEFICIENTES 73
4.6 CRÍTICAS AO MODELO 74
5 CONCLUSÕES 78
6 REFERÊNCIAS 79

Capítulo 1

1 INTRODUÇÃO

O primeiro capítulo deste estudo apresenta a introdução ao trabalho, incluindo a problematização do tema e a questão que norteou a pesquisa, além da justificativa da escolha do tema estudado e os objetivos geral e específicos.

1.1 PROBLEMATIZAÇÃO

A ciência tenta conhecer a realidade, interpretando-a através de modelos e teorias de diversos ramos do conhecimento. Assim, ela procura estabelecer relações entre o conhecido e o desconhecido, encontrar ou propor leis explicativas, representar a realidade através de simplificações, recortes ou modelos. Para isso, é necessário controlar, manipular, medir as variáveis consideradas importantes ou relevantes ao entendimento do fenômeno analisado, deve-se coletar informações e dados sobre o fenômeno a ser estudado (FIGUEIRA, 2006). Para Johnson e Wichern (1992), pesquisa científica é um processo de aprendizado interativo. Quando se objetiva a explanação de um fenômeno social ou físico deve-se especificar tal fenômeno a ser observado e então, testar pela coleta de dados. Por sua vez, uma análise de dados coletados em experimento ou observação irá sugerir uma explanação modificada do fenômeno. Através deste processo de aprendizado interativo, variáveis são freqüentemente adicionadas ou retiradas do estudo. Assim, a complexidade de muitos fenômenos requer de um pesquisador a coleta de observações de muitas variáveis, de diferentes tipos.

De acordo com Royston e Sauerbrei (2008), dados são coletados em todas as áreas da vida, e na pesquisa científica, dados em muitas variáveis podem ser coletados para investigar as inter-relações entre elas, ou determinar os fatores que afetam uma saída de interesse.

Uma importante possibilidade advinda de tais análises de dados coletado em experimentos bem controlados ou simples observações, seria a predição. A predição permite a partir do comportamento ou valores de variáveis ditas preditoras ou Variáveis Independentes (VI’s), prever o comportamento da(s) variável(eis) dita(s) dependente(s) (VD’s), através da

criação de modelos lineares ou não-lineares. Tal ação pode ser extremamente útil em situações prática, onde, por exemplo, seja impraticável a medida direta da VD, por motivos econômicos ou técnicos. Nesse contexto, a análise de regressão se insere, como ferramenta poderosa na análise, que permite a implementação de tal ação.

A ergonomia é grande utilizadora desse conjunto de ferramentas, pois procura em dados coletados em seus experimentos, encontrar empiricamente explicações para comportamentos de variáveis de seus interesse controlando outras variáveis, ou procura por variáveis que afetem o comportamento de uma variável de interesse, como o efeito de condições de conforto ambiental sobre a produtividade de um trabalhador nesse ambiente. Silva (2001) por exemplo, utilizou ferramenta de análise de regressão linear para ligar conforto térmico ambiental a produtividade em setores de digitação da Caixa Econômica Federal. A ergonomia enquanto disciplina direcionada para uma abordagem sistêmica de todos os aspectos da atividade humana, desde o seu surgimento, utiliza conhecimento de diversas ciências para adaptar o ambiente ao homem em seus estudos. Um importante aspecto abordado por este campo do conhecimento é a adaptação do trabalho ao homem e suas limitações. Para isso, a ergonomia precisa por vezes, realizar experimentos onde coleta dados em muitas variáveis relacionadas a fenômenos físicos, sociais, ambientais.

Além disso, a ergonomia por vezes precisa tratar com variáveis categóricas, ou ainda mais especificamente, com variáveis dependentes (VD) categóricas. Grande parte das vezes que isso acontece, a análise de regressão logística binária é utilizada através da dicotomização da VD. A dicotomização da variável dependente em casos nos quais a mesma possui mais de duas categorias ocorre não raramente, como por exemplo, no caso do trabalho apresentado por Silva *et al* (2007), onde a VD, o Índice de Capacidade para o trabalho de motoristas de ônibus da cidade de João Pessoa foi dicotomizado, sendo porém esta uma variável categórica com quatro categorias, tendo através do agrupamento de categorias adjacentes, sido transformadas em apenas duas. Tal procedimento tem como objetivo permitir o uso da regressão logística binária.

Tal procedimento porém, descaracteriza a informação contida no conjunto de dados original, pois há uma simplificação da VD. Além da perda de informação por aglutinação de categorias adjacentes, o uso do procedimento comum citado acima, não permite a exploração de outra característica em alguns problemas, como a ordenação entre as

categorias. A desconsideração dessa característica pode levar a modelos com interpretações mais complexas e potencialmente, à perda de poder (AGRESTI, 2007).

Apesar de haver uma grande inserção da regressão logística binária como ferramenta usada em análise de dados na ergonomia, o uso de sua extensão, a regressão logística multinomial, não é freqüente, podendo se dizer que seu uso ainda é raro. Tal limitação de uso gera prejuízos a ergonomia na interpretação dos dados de seus experimentos.

Assim, através de um estudo de caso em dados coletados junto a enfermeiros de UTI's de hospitais públicos na cidade de João Pessoa, procura-se ilustrar o uso dessa ferramenta. Os dados são advindos de um estudo exploratório onde foram coletadas diversas variáveis ambientais, como as temperaturas de bulbo seco, bulbo úmido e de globo, ruído no posto de descanso e junto ao leito do paciente na UTI, iluminação, algumas variáveis pessoais, como escolaridade, tempo de serviço, estado civil, e algumas outras organizacionais, como carteira assinada, trabalho noturno. Assim, a modelagem através da regressão logística multinomial será usada nesse trabalho para procurar os fatores de risco associados com a queda do Índice de Capacidade para o trabalho dos enfermeiros atuantes exclusivamente em UTI's de hospitais públicos na cidade de João Pessoa em uma faixa inaceitável.

1.2 OBJETIVOS

1.2.1 OBJETIVO GERAL

O presente trabalho tem por objetivo contribuir de maneira teórica para a análise de dados na área de ergonomia experimental propondo construir modelos de regressão para estudos envolvendo uma variável dependente não-métrica com características de ordenação entre categorias.

1.2.2 OBJETIVOS ESPECÍFICOS

Realizar uma revisão de literatura na análise de regressão logística, sua inserção em trabalhos anteriores na ergonomia e sua forma de utilização;

Realizar uma revisão de literatura na análise de regressão logística multinomial nominal e ordinal;

Mostrar a regressão logística multinomial aplicada em um estudo exploratório de ergonomia, para estudar o impacto de variáveis métricas e não métricas em uma variável que apresenta singularidades ordinais.

1.3 ESTRUTURA DO TRABALHO

O presente trabalho é divido nas seguintes seções:

O capítulo 2 traz uma revisão sobre ergonomia, um breve histórico de seu surgimento, definição e divisão. O Índice de Capacidade para o Trabalho (ICT) é apresentado nesse capítulo, já que a aplicação desse índice a profissionais do ramo de enfermagem de UTI's resultará na VD do conjunto de dados desse do experimento utilizado nesse estudo.

O capítulo 3 aborda conceitos sobre predição e regressão logística binomial e multinomial. Além disso, apresenta a metodologia utilizada na análise dos dados.

O capítulo 4 apresenta os dados utilizados na modelagem e resultados da análise de regressão logística multinomial ordinal.

O capítulo 5 apresenta as conclusões do trabalho, limitações do modelo e sugestões de trabalhos futuros.

Capítulo 2

2 ERGONOMIA

2.1 HISTÓRICO DA ERGONOMIA

Para Iida (2005), a preocupação em adaptar o ambiente natural e construir objetos artificiais para atender às suas conveniências sempre esteve presente nos seres humanos, desde os tempos mais remotos. Assim, a ergonomia foi precedida de um longo período de gestação, que remonta à pré-história, quando o homem pré-histórico escolheu uma pedra num formato que melhor se adaptasse à forma e movimentos de sua mão, para usá-la como arma, de modo a ter uma ferramenta que lhe facilitasse as tarefas, como caçar, cortar e esmagar.

De acordo com Frias Junior (1999), desde a Antiguidade greco-romana o trabalho já era visto como um fator gerador e modificador das condições de viver, adoecer e morrer dos homens. Estorilio (2003) afirma que estudos sistemáticos do trabalho começaram a aparecer na Renascença (1442 – 1519), onde nesse período, surgem estudos com medidas e observações sistemáticas do homem em atividade de trabalho. Em "*De Re Metallica"*, Georg Bauer de 1556 a.D., faz referência a doenças pulmonares em mineiros e em 1567 a.D., Paracelso, descreve também doenças de mineiros da região da Boêmia e a intoxicação pelo mercúrio. Em 1700 a.D., o médico italiano Bernardino Ramazzini (1633-1714) em sua publicação "*De Morbis Artificum Diatriba*" descreve doenças que ocorriam em mais de cinqüenta profissões (FRIAS JUNIOR, 1999).

No século XIX, Frederick Winslow Taylor lançou seu livro "Administração Científica", com uma abordagem que buscava a melhor maneira de executar um trabalho e suas tarefas. Mediante aumento e redução do tamanho e peso de uma pá de carvão, até que a melhor relação fosse alcançada, Taylor triplicou a quantidade de carvão que os trabalhadores podiam carregar num dia. A partir daí, iniciava-se o estudo do trabalho sob a ótica da "Organização Científica do Trabalho". Nesse período, vários estudos foram realizados com o intuito de responder a questões levantadas por condições de trabalho insatisfatórias

(ESTORILIO, 2003; IIDA, 2005). Na Europa, principalmente na Alemanha, França e países escandinavos começaram então, pesquisas na área de fisiologia do trabalho, na tentativa de transferir para o terreno prático o conhecimento gerado em laboratórios (IIDA, 2005)

Nos primeiros anos do século XX, Frank Bunker Gilbreth e sua esposa Lilian expandiram os métodos de Taylor para desenvolver "Estudos de Tempos e Movimentos" o que ajudou a melhorar a eficiência, eliminando passos e ações desnecessárias. Ao aplicar tal abordagem, Gilbreth reduziu o número de movimentos no assentamento de tijolos de 18 para 4,5 permitindo que os operários aumentassem a taxa de 120 para 350 tijolos por hora.

A Segunda Guerra Mundial (1939 - 1945) marcou o advento de máquinas e armas sofisticadas, criando demandas cognitivas jamais vistas antes por operadores de máquinas, em termos de tomada de decisão, atenção, análise situacional e coordenação entre mãos e olhos. Isso porque, instrumentos bélicos relativamente complexos, como submarinos, tanques, radares e aviões, exigiam dos operadores muita habilidade em condições ambientais bastante desfavoráveis e tensas, no campo de batalha. Foi observado que aeronaves em perfeito estado de funcionamento, conduzidas pelos melhores pilotos, ainda caíam. Em 1943, Alphonse Chapanis, um tenente no exército norte-americano, mostrou que o "erro do piloto" poderia ser muito reduzido quando controles mais lógicos e diferenciáveis substituíssem os confusos projetos das cabines dos aviões (IIDA, 2005).

Embora tenha tido um longo período de gestação como já fora dito, a ergonomia tem uma data formal de nascimento: 12 de julho de 1949. O nome Ergonomia foi proposto no dia 16 de fevereiro de 1950 (IIDA, 2005). Em 1949, K.F.H. Murrel, engenheiro inglês, começou a dar um conteúdo mais preciso a este termo, e fez o reconhecimento desta disciplina científica criando a primeira associação nacional de Ergonomia, a *Ergonomic Research Society*, que reunia fisiologistas, psicólogos e engenheiros que se interessavam pela adaptação do trabalho ao homem. E foi a partir daí, que a Ergonomia se desenvolveu em outros países industrializados e em vias de desenvolvimento.

2.2 ERGONOMIA – DEFINIÇÃO E ASPECTOS GERAIS

A palavra “Ergonomia” tem sua origem etimológica na conjunção de duas palavras gregas: *Ergon* que pode ser traduzido pela palavra trabalho e *nomos* que pode ser traduzido por normas, regras, leis.

O termo ergonomia entrou para o léxico moderno quando Wojciech Jastrzębowski o usou em um artigo em 1857 intitulado *"An outline of ergonomics or science of work"* definindo o termo como "Uso das forças e faculdades humanas com as quais o homem foi dotado por seu criador" (MOURE, 2000). Wisner (1997) *apud* Estorilio (2003) define ergonomia como sendo o conjunto de conhecimentos científicos relacionados ao homem, necessários na concepção de instrumentos, máquinas e dispositivos que possam ser utilizados com o máximo de conforto, segurança e eficiência no trabalho. Ainda Wisner (1997) citado por Estorilio (2003) aborda a ergonomia como sendo a arte que utiliza o saber tecno-científico e o saber dos trabalhadores sobre a sua própria situação de trabalho.

Ergonomia, (ou fatores humanos, como é conhecida nos Estados Unidos), é segundo a definição dada pela Associação Internacional de Ergonomia (IEA) em agosto de 2000, uma disciplina científica relacionada ao entendimento das interações entre os seres humanos e outros elementos ou sistemas, e à aplicação de teorias, princípios, dados e métodos em projetos afim de otimizar o bem estar humano e o desempenho global do sistema (IEA, 2000).

Para a ABERGO, Associação Brasileira de Ergonomia, os ergonomistas contribuem para o planejamento, projeto e a avaliação de tarefas, postos de trabalho, produtos, ambientes e sistemas de modo a torná-los compatíveis com as necessidades, habilidades e limitações das pessoas (ABERGO, 2008).

De acordo com Iida (2005), a *Ergonomics Society*, a associação nacional de ergonomia da Inglaterra, define ergonomia como o estudo do relacionamento entre o homem e seu trabalho, equipamento ambiente e particularmente, a aplicação de conhecimentos de anatomia, fisiologia e psicologia na solução de problemas que surgem desse relacionamento.

Iida (2005) a define como o estudo da adaptação do trabalho ao homem, sendo que o termo trabalho tem nesse caso, uma acepção ampla, tem alcance de toda a situação onde ocorra a relação entre o homem e uma atividade produtiva, envolvendo não apenas aspectos físicos, mas também os organizacionais.

Para a *International Ergononomics Association*, IEA, trata-se de uma disciplina orientada para uma abordagem sistêmica de todos os aspectos da atividade humana. Para darem conta da amplitude dessa dimensão e poderem intervir nas atividades do trabalho é preciso que os ergonomistas tenham uma abordagem holística de todo o campo de ação da

disciplina, tanto em seus aspectos físicos e cognitivos, como sociais, organizacionais, ambientais (IEA, 2000).

Para Slack, Chambers e Johnston (2007) a ergonomia preocupa-se primariamente com os aspectos fisiológicos do projeto de trabalho, ou seja, o corpo humano e como ele se ajusta ao ambiente. Para os referidos autores, isso envolve dois aspectos: primeiro como a pessoa confronta-se com os aspectos físicos de seu local de trabalho (mesas, cadeiras, escrivaninhas, máquinas, computadores); segundo, como uma pessoa relaciona-se com as condições ambientais de sua área de trabalho imediata (temperatura, níveis de iluminação, ruído).

A ergonomia difere de outras áreas do conhecimento pelo seu caráter interdisciplinar e pela sua natureza aplicada, uma vez que uma base múltipla de conhecimentos são abarcados pelo seu escopo de estudo e seu caráter aplicado configura-se na adaptação do posto de trabalho e dos níveis de ambiência às características psicofisiológicas do trabalhador. Silva (2001) corrobora com esse pensamento, afirmando que

> A ergonomia é, fundamentalmente, a aplicação de princípios científicos, métodos e dados subtraídos das diversas disciplinas para o desenvolvimento dos sistemas de engenharia, nos quais o fator humano exerce um importante papel. Ela analisa e estuda as características humanas com o objetivo de viabilizar projetos de ambientes de trabalho mais eficazes e seguros. A psicologia, a ciência cognitiva, a fisiologia, a biomecânica, a antropometria física aplicada e o sistema de engenharia industrial estão dentre as disciplinas básicas que a ergonomia utiliza para tornar esses projetos bem mais sólidos (SILVA, 2001).

Para Iida (2005), a Ergonomia é uma ciência que trata da interação entre os homens e a tecnologia, integra o conhecimento proveniente das ciências humanas para adaptar tarefas, sistemas, produtos e ambientes às habilidades e limitações físicas e mentais dos indivíduos.

Outra definição é dada por Hendrick (2003). Para ele, a ergonomia é o desenvolvimento e aplicação da tecnologia da interface humano-sistema (*hardware*, *software*, ambientes, tarefas e estruturas organizacionais e de processos), incluindo especificações, recomendações, métodos e ferramentas. Ou seja, a arte da ergonomia é a de trabalhar todos os aspectos ligados ao trabalho de modo à melhor adaptá-los ao homem que o exerce.

Para Estorilio (2003) a maioria das definições da ergonomia volta-se para dois objetivos fundamentais: a saúde e a eficiência no trabalho. Iida (2005) relata que a ergonomia procura reduzir as conseqüências nocivas do sistema produtivo sobre o trabalhador, proporcionando saúde, segurança e satisfação ao trabalhador, tendo como conseqüência direta, a eficiência do sistema produtivo, através do estudo dos diversos fatores que influem no desempenho do sistema produtivo.

Para Silva (2001), a ergonomia tem um lado primordial que difere das outras disciplinas: humanização no trabalho. Para o referido autor, na Ergonomia o homem é visto não apenas como uma parte de um sistema, mas como o mais importante componente do sistema tecnológico. A eficácia do projeto, como sua concepção, dependerá principalmente deste componente e, depois, de outros inseridos no sistema, além, é claro, do conhecimento das características individuais, dimensões, capacidades e limitações. Corrobora com esse pensamento Iida (2005), que afirma que a ergonomia inicia-se com o estudo das características do trabalhador, para depois projetar o trabalho que ele consegue executar, preservando sua saúde.

Assim, a Ergonomia tem em vista a transformação do trabalho de modo que o mesmo venha a proporcionar aos trabalhadores um ambiente salutar onde suas atividades possam ser praticadas e contribuam para que as empresas alcancem seus objetivos de desempenho. Conhecer a atividade de trabalho permite auxiliar na concepção dos meios materiais, organizacionais e em formação para que os trabalhadores possam desempenhar as suas funções de maneira eficaz e preservem a sua saúde (GUÉRIN, 2001).

Slack, Chambers e Johnston (2007) afirmam que:

> Deve haver adequação entre pessoas e o trabalho que elas fazem. Para atingir essa adequação, há somente duas alternativas. Ou o trabalho pode ser adequado às pessoas que o fazem ou, alternativamente, as pessoas podem ser adequadas ao trabalho. A ergonomia direciona para a primeira alternativa. (SLACK, CHAMBERS E JOHNSTON, 2007).

A Ergonomia por meio dos seus objetivos, ferramentas e métodos possibilita um forte vínculo entre trabalho e saúde, sendo uma das contribuições mais significativas no que diz respeito à saúde no trabalho. A ação ergonômica além de aplicar métodos, realizar medidas, fazer observações, conduzir entrevistas, deve ajustar os métodos e as suas aplicações

ao contexto em que está inserido e levar em consideração na elaboração da transformação do trabalho os interesses de todos os sujeitos envolvidos (DOPPLER, 2007).

Kroemer *et al* (1994) citado por Silva (2001), afirma que há 2 aspectos distintos na Ergonomia: um relacionado à investigação, pesquisa e experimentação, onde se determinam as particularidades específicas e características humanas, necessárias à elaboração de um projeto de engenharia, e outro relacionado à aplicação da engenharia, onde se projetam ferramentas ou instrumentos, máquinas, ambientes, tarefas e métodos de trabalho para adequar e acomodar o homem.

A figura abaixo mostra o exemplo em um escritório genérico de como a ergonomia tenta intervir em fatores ambientais e instrumentos de trabalho, de modo a haver uma melhor adequação destes ao homem nele presente.

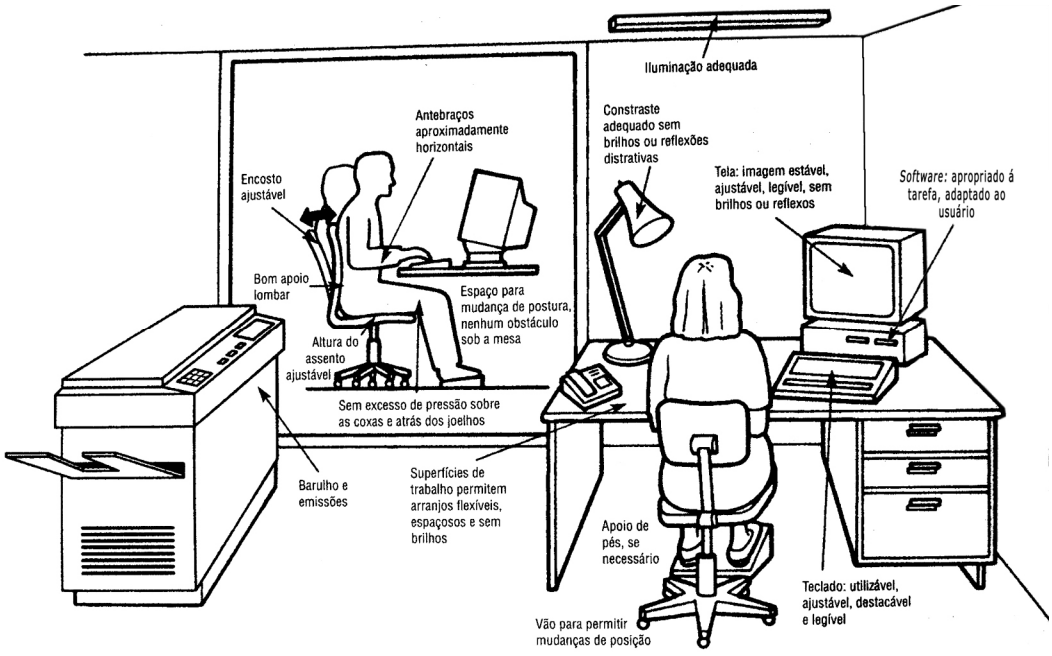

Figura 1:Ergonomia aplicada
Fonte: Adaptado de Slack, Chambers e Jonhston (2007)

2.3 ERGONOMIA – RAMOS DE ATUAÇÃO

A IEA (2000) divide a Ergonomia em três ramos de atuação:

i) Ergonomia Física: que lida com as respostas do corpo humano à carga física e psicológica. Tópicos relevantes incluem manipulação de materiais, arranjo físico de estações de trabalho, demandas do trabalho e fatores tais como

repetição, vibração, força e postura estática, relacionada com lesões músculo-esqueléticas.

ii) Ergonomia Cognitiva: também conhecida como engenharia psicológica, refere-se aos processos mentais, tais como percepção, atenção, cognição, controle motor e armazenamento e recuperação de memória, como eles afetam as interações entre seres humanos e outros elementos de um sistema. Tópicos relevantes incluem carga mental de trabalho, vigilância, tomada de decisão, desempenho de habilidades, erro humano, interação humano-computador e treinamento.

iii) Ergonomia Organizacional: ou macroergonomia, relacionada com a otimização dos sistemas socio-técnicos, incluindo sua estrutura organizacional, políticas e processos. Tópicos relevantes incluem trabalho em turnos, programação de trabalho, satisfação no trabalho, teoria motivacional, supervisão, trabalho em equipe, trabalho à distância e ética.

2.4 ÍNDICE DE CAPACIDADE PARA O TRABALHO (ICT)

A variável dependente nessa análise é o Índice de Capacidade para o Trabalho. Assim, faz-se mister discorrer brevemente sobre o tal índice.

Ferreira (1999) define capacidade como: “qualidade que uma pessoa ou coisa tem de possuir para um determinado fim; habilidade, aptidão”. Assim, capacidade para o trabalho poderia ser entendido como a habilidade ou aptidão de um determinado trabalhador para realizar seu trabalho.

Abordando mais especificamente o assunto, Ilmarien e Rantanen (1999) citado por Chiu *et al.* (2007), afirmam que a capacidade para o trabalho é definida como a habilidade de um trabalhador realizar seu trabalho, levando em conta as demandas específicas do trabalho, condições de saúde individual, recursos mentais e vida no trabalho.

Tuomi *et al.* (1997) *apud* Pereira (2009) definem capacidade para o trabalho como “o quão bem está, ou estará, um trabalhador presentemente, ou num futuro próximo, e quão capaz ele pode executar o seu trabalho, em função das exigências, de seu estado de saúde e de capacidade física e mental”.

Bellusci e Fischer (1999) afirmam que a Organização Mundial de Saúde (OMS), tem demonstrado preocupação com a questão do envelhecimento relacionado ao trabalho e reconhece que modificações nos vários sistemas do corpo humano levam a uma diminuição gradativa na eficácia de cada um deles, com diminuição na capacidade funcional dos indivíduos que pode gerar conflitos entre a capacidade funcional e as exigências do trabalho.

Ainda segundo Bellusci e Fischer (1999), as exigências e os fatores de estresse no trabalho precisam estar equilibrados com a capacidade dos trabalhadores para que eles não envelheçam funcionalmente. Para isso, Bellusci e Fischer (1999, p. 3) pontuam que

> ...há necessidade de uma avaliação contínua dos agentes que desencadeiam sintomas, lesões e doenças e das melhorias das condições de trabalho, procurando soluções para incrementar o equilíbrio da relação entre capacidade e demandas do trabalho. Essas soluções são baseadas em estudos sobre o ambiente de trabalho, as alterações fisiológicas, as mudanças na capacidade para o trabalho, e, em especial, na influência da organização e dos aspectos físicos e ergonômicos no trabalho. (...) na prática, a capacidade para o trabalho precisa ser avaliada para identificar seu declínio em estágio prematuro, acompanhar os efeitos das medidas de prevenção e reabilitação, e para avaliar a incapacidade para o trabalho. (BELLUSCI E FISCHER 1999, p. 3)

Para a avaliação da capacidade para o trabalho um grupo de pesquisadores de um Instituto de saúde ocupacional da Finlândia (Finnish Institute of Occupational Health) desenvolveu uma metodologia conhecida como Índice de Capacidade para o Trabalho (ICT) que foi difundido e atualmente é utilizado por serviços de atenção à saúde de trabalhadores em vários países do mundo sendo considerado um bom indicativo de avaliação e acompanhamento de questões voltadas à saúde no trabalho, tendo em vista, a importância da preservação e manutenção da saúde dos trabalhadores. O estudo utilizando o Índice de Capacidade para o Trabalho ICT - realizado na Finlândia, entre 1981 e 1992, com 6.259 trabalhadores municipais de diferentes ocupações, teve como objetivo a prevenção de doenças e da incapacidade para o trabalho entre trabalhadores em envelhecimento e também a análise de meios para a manutenção da saúde e da capacidade para o trabalho (TUOMI, 2005).

O questionário do ICT é composto por sete itens cuja somatória dos pontos atribuídos a cada um deles define o escore total do índice. Os itens possuem pontuações mínimas e máximas e a equivalência de seus valores são ponderadas conforme as características específicas da atividade realizada no trabalho (TUOMI, 2005).

Analisam-se no Índice da capacidade para o trabalho seus itens e seus valores referenciais:

i) Capacidade para o trabalho atual comparada com a melhor de toda a vida 0-10;
ii) Capacidade para o trabalho em relação às exigências do trabalho 2-10;
iii) Número atual de doenças diagnosticadas por médico 1-7;
iv) Perda estimada para o trabalho devido às doenças 1-6;
v) Faltas ao trabalho por doenças nos últimos 12 meses 1-5;
vi) Prognóstico próprio sobre a capacidade para o trabalho daqui a dois anos 1,4,7;
vii) Recursos mentais 1-4;

Dentro desta proposta temos um intervalo de resultados possíveis entre 7 e 49 pontos, sendo este subdivido em quatro classificações que diagnosticarão o índice de capacidade para o trabalho e definirão os objetivos de quaisquer medidas necessárias a serem tomadas referentes ao avaliado.

Tabela 1: Classificação - Índice de Capacidade para o Trabalho

Pontos	*Classificação*	*Medida*
7 – 27	Baixa	Restaurar a Capacidade para o Trabalho
28 – 36	Moderada	Melhorar a Capacidade para o Trabalho
37 – 43	Boa	Melhorar a Capacidade para o Trabalho
44 – 49	Ótima	Manter a Capacidade para o Trabalho

Fonte: TUOMI *et al*. (2005)

Capítulo 3

3 REFERENCIAL TEÓRICO

A ciência busca conhecer a realidade, e assim, interpretar os acontecimentos e fenômenos baseada no estudo das variáveis intervenientes nestes eventos. Ao fazer isso, a ciência busca encontrar ou propor leis explicativas estabelecendo relações. Há portanto na ciência, a necessidade da transformação de dados coletados nas suas diversas pesquisas em conhecimento, feito através da análise destes dados.

A estatística é, no sentido estrito da palavra, o ramo da ciência que trata com a arte de fazer inferências a respeito de populações, baseado em amostras. No sentido amplo, a estatística inclui compilação, organização e sumarização dos dados; apresentação dos dados na forma de tabelas e gráficos; desenvolvimento de modelos com o propósito de entender fenômeno aleatórios e não aleatórios; uso de modelos para predição; aproximações matemáticas para tomada de decisão e avaliação de riscos e assim por diante (GRAYBILL & IYER, 2006).

3.1 MODELAGEM

O mundo que nos cerca é interpretado através de modelos, já que, seja na vida cotidiana ou em trabalhos científicos, a construção de modelos ajuda a interpretar a complexa realidade. Isso porque, sendo um dado fenômeno de interesse regido por interações de infinitas variáveis, este se faz incompreendido, caso se deseje conhecer toda a realidade sobre ele. O que se faz é tentar compreender tal fato através de modelos. Modelo pode ser rudemente definido como uma representação simplificada e abstrata de fenômeno ou situação concreta, e que serve de referência para a observação, estudo ou análise.

Embora não represente toda a realidade por trás de um dado acontecimento de interesse, os modelos são úteis, pois permitem além da interpretação de tal acontecimento, simulações de mudanças que sofre tal fato, variando-se os parâmetros descritores do modelo que o representa.

Além disso, por se tratar de um “recorte da realidade”, um modelo matemático descritor de um fenômeno pode ser desde o mais simples, considerando um pequeno número de interações com outras variáveis e admitindo que a relação entre as variáveis envolvidas

sejam a mais simples possível, como uma relação linear, como mais complexos, aumentando o número de variáveis e admitindo não-linearidades na relação entre elas.

Para Royston e Sauerbrei (2008), um bom modelo é satisfatório e interpretável do ponto de vista da matéria em estudo, robusto com respeito a mínimas variações dos dados presentes, preditivo em novos dados e, parcimonioso. É preciso ainda, ter em mente os dois principais objetivos da proposição de um modelo, e distingui-los. O primeiro objetivo é o da predição, onde o ajuste do modelo e o erro médio quadrático predito são os principais critérios de adequação do modelo. O segundo objetivo é o da explanação, onde o interesse recai em tentar identificar preditores influentes e ganhar discernimento na relação entre os preditores e a saída.

Um bom modelo precisa ser o mais simples possível, sem tornar-se no entanto, inadequado. Ou seja, deve-se manter o modelo tão simples quanto a complexidade dos dados coletados permitam. Isso porque generalidade e utilidade prática devem ser mantidas em mente quando da proposição de um modelo. Modelos constituídos de exagerado número de preditores, ou com relação entre as variáveis complexa demais, ficam prejudicados quanto à usabilidade. Considere por exemplo, um modelo constituído de muitas variáveis. Todas as variáveis constituintes do modelo precisariam ser medidas de forma idêntica ou parecida, mesmo que seus efeitos sejam bem pequenos. Semelhante modelo é impraticável, e portanto, não é útil e fácil de ser esquecido (WYATT & ALTMAN (1995) *apud* ROYSTON & SAUERBREI (2008)).

Dependendo do papel a ser desempenhado pelo modelo proposto (Predição, explanação, entendimento dos efeitos dos preditores) o julgamento da adequação do modelo muda (ROYSTON & SAUERBREI, 2008 p. 27). Assim, em senso amplo, muitos dos processos de modelagem recaem sobre algum dos casos a seguir:

i) O modelo está predefinido. Tudo o que resta é estimar os parâmetros e checar as principais premissas;

ii) O objetivo é desenvolver um bom preditor. O número de variáveis deve ser pequeno;

iii) O objetivo é desenvolver um bom preditor. Limitar a complexidade do modelo não é importante;

iv) O objetivo é avaliar o efeito de um ou muitos (novos) fatores de interesse, ajustando para alguns fatores estabelecidos em um modelo multivariado;

v) Geração de hipóteses de possíveis efeitos dos fatores em estudos com muitas covariáveis.

3.2 TIPOS DE VARIÁVEIS

Conforme Mayorga (1999) *apud* Vasconcelos (2006), variáveis são quaisquer quantidades ou características que podem possuir diferentes valores numéricos; portanto, podem ser consideradas classificações ou medidas, quantidades que variam, conceitos operacionais que contêm ou apresentam valores, ou aspectos discerníveis em um objeto de estudo e passível de mensuração. Os valores que são adicionados ao conceito operacional para transformá-lo em variável estes podem ser quantidades, qualidades, características, magnitudes, e traços, entre outras, que se alteram em cada caso particular.

De acordo com Pasquali (2003) *apud* Prearo (2008), a legitimidade epistemológica da medida matemática como descritora de fenômenos naturais ocorrem se e somente se, as propriedades estruturais tanto do sistema numérico quanto do fenômeno em estudo forem garantidas.

Segundo ainda esse autor, os axiomas do sistema numérico, são:

i) Identidade: Um número é idêntico a si e somente a si mesmo;

ii) Ordem: Todo número é diferente do outro, não somente em termo qualitativo, mas também em magnitude;

iii) Aditividade: Os números podem ser somados, ou seja, unidos de forma que a soma de dois números resulte em um número diferente;

iv) Razão: O sistema numérico possui um zero absoluto.

Desta forma, o autor já citado considera que quanto maior o acúmulo dessas garantias, maior a aproximação da escala métrica.

Quanto aos tipos de variáveis são apresentados no quadro 2.

Quanto à escala usada para a mensuração de tais variáveis, pode-se generalizar a classificação teórica dessas escalas em dois grandes grupos: variáveis métricas e variáveis não métricas (PREARO, 2008).

i) Dados métricos – também chamados de dados quantitativos, dados intervalares ou dados proporcionais, essas medidas descrevem o indivíduo não apenas pelo atributo, mas pela quantia ou grau em que o indivíduo pode ser caracterizado pelo atributo.

Quadro 1: Tipos de Escala de Mensuração

Escala		Propriedades garantidas	Transformações permitidas	Estatísticas apropriadas
Não Métrica	Nominal	• Identidade	Permutação (troca 1 por 1)	**Freqüências:** Freqüências, percentagens, proporção, moda, qui-quadrado, coeficiente de contingência
	Ordinal	• Identidade • Ordem	Monotônica crescente (isotonia)	**Não paramétricas:** Mediana, correlação de Spearman, Mann-Whitney
Métrica	Intervalar	• Identidade • Ordem • Aditividade	Linear do tipo: Y = a + bX	**Paramétricas:** Média, Desvio-padrão, correlação de Pearson, teste t, teste F
	Razão	• Identidade • Ordem • Aditividade • Razão	Linear do tipo: Y = bX	• Média geométrica • Média Harmônica • Coeficiente de variação • Logarítmos

Fonte: Prearo (2008)

Quadro 2: Tipos de Variáveis

Tipo de Variável	Subtipo	Características	Exemplo
Quantitativa (ou métrica ou numérica)	Discreta	Números inteiros, sem frações, como em contagens. Constituem um conjunto finito de elementos.	Número de empregados numa empresa
	Contínua	Números que podem assumir valores fracionários. Constituem um conjunto infinito de elementos.	Faturamento mensal
Qualitativa (ou não-métrica, ou não-numérica)	Categórica nominal	Categorias, sendo que cada categoria é independente em relação às outras.	Ramo de atividade
	Categórica ordinal	Categorias, sendo que cada categoria mantém uma relação de ordem com as outras.	Percepção Térmica

Fonte: Adaptado de Pereira (2005)

ii) Dados não métricos - também chamados de dados qualitativos, são atributos, características ou propriedades categóricas que identificam ou descrevem o indivíduo. Diferem dos dados métricos no sentido de não indicarem a quantia do atributo que caracteriza o indivíduo.

3.2.1 VARIÁVEIS CATEGÓRICAS

Define-se variável categórica aquela que tem uma escala de medida consistindo de um conjunto de categorias (AGRESTI, 2007). Tal escala pode ser encontrada em ciências sociais para medir atitudes e opiniões, em ciências da saúde para medir por exemplo, a resposta de um paciente a um dado tratamento. Na ergonomia podem ser encontrados dados categóricos por exemplo na medição da percepção térmica de um ambiente ou na ocorrência ou não de uma Lesão por Esforço Repetitivo na execução de uma tarefa.

Como mostrado no quadro 1, as variáveis categóricas tem dois principais tipos de escala de medida: Ordinal ou Nominal. As variáveis categóricas do tipo ordinal são aquelas que suas categorias possuem algum tipo de ordem natural, como por exemplo a percepção térmica do ambiente, que pode ser desde muito frio até muito quente. Já as variáveis categóricas que não possuem ordenação entre suas categorias são chamadas nominais.

Os métodos empregados na análise em variáveis nominais fornecem os mesmos resultados não importa como as categorias são listadas. Já os métodos projetados para a análise de variáveis ordinais utilizam a ordenação das categorias, fornecendo resultados substancialmente iguais no caso de listar-se as categorias da mais alta para a mais baixa ou da mais baixa para a mais alta, porém os resultados das análises deverão apresentar-se diferentes caso as categorias sejam listadas em qualquer outra ordem (AGRESTI, 2007).

Quanto aos métodos usados para a análise em variáveis ordinais, eles não podem ser usados para dados nominais, porém os métodos projetados para a análise com variáveis nominais podem ser utilizados com variáveis ordinais, desde que a única exigência nesse caso é a escala de mensuração categórica.

Embora haja a possibilidade de lidar com variáveis ordinais usando ferramentas para a análise de variáveis nominais, esse emprego implicaria em não levar em consideração a ordem das categorias da variável tratada, o que pode representar uma séria perda de potência na análise (AGRESTI, 2007).

3.3 ANÁLISE DE REGRESSÃO

Uma das buscas da ciência é entender a associação entre variáveis. Isso porque entender tais associações pode ser útil de diversas maneiras, como na predição, ou seja, o conhecimento da associação entre variáveis pode fazer com que o comportamento de uma ou mais variáveis possa ser predito a partir do comportamento das variáveis relacionadas. Ainda é possível com tal conhecimento controlar o valor de uma variável a partir do ajuste das variáveis relacionadas.

De acordo com Royston e Sauerbrei (2008), modelos de regressão realizam muitas tarefas em todas as áreas da ciência onde dados empíricos são analisados, sendo que essas tarefas incluem:

i) Predição de uma saída de interesse;
ii) Identificação de importantes preditores;
iii) Entendimento dos efeitos de preditores;
iv) Ajuste para preditores incontroláveis através de design experimental;
v) Estratificação por risco.

Para Graybill & Iyer (2006), quase todas as decisões que um indivíduo toma são baseadas em predição e muitas dessas predições podem ser feitas através do estudo sistemático de associações e análise de regressão trata do estudo dessas relações. Ainda segundo esses autores, há pelo menos duas razões pelas quais a predição é útil:

i) O valor verdadeiro da variável dependente Y é muito caro ou difícil para ser obtido, porém as variáveis preditoras são mais baratas ou fáceis de serem medidas;
ii) A variável resposta é impossível de ser medida, freqüentemente, por se tratar de valores futuros;

Para que uma predição neste sentido seja realizada, são necessários:

i) As variáveis preditoras, denotadas por X_1, X_2, ... , X_p e os valores observados para essas variáveis;
ii) Uma equação ou formula, para predizer a variável resposta Y usando as variáveis preditoras, X_1, X_2, ... , X_p.

De acordo com Ryan (2009), a Análise de Regressão é uma das técnicas estatísticas mais utilizadas, e seu uso está presente em quase todos os campos de aplicação. Weisberg (2005) afirma que a Análise de Regressão é a parte central de muitos projetos de pesquisa. Já para Johnson & Wichern (1992), análise de regressão é uma metodologia estatística para predizer valores de uma ou mais variável resposta (dependente) a partir de uma coleção de valores de variáveis preditoras (independentes) e que também pode ser utilizada para avaliar os efeitos das variáveis preditoras nas respostas. Graybill & Iyer (2006), concordam com a definição de que a análise de regressão é um método comumente utilizado para obter uma função de predição para predizer valores de uma variável resposta usando as variáveis preditoras.

O termo regressão foi proposto por Francis Galton em 1885, em um estudo onde demonstrou que a altura dos filhos não tende a refletir a altura dos pais, mas tende a regredir para a média da população (MAROCO, 2003 *apud* PREARO, 2008;.JOHNSON & WICHERN, 1992; FIGUEIRA, 2006)[1]. Para Johnson & Wichern (1992), o termo que foi escolhido a partir desse trabalho não reflete a importância nem a amplitude da aplicação desta metodologia.

3.4 MODELOS LINEARES GENERALIZADOS (MLG)

Diversos autores abordam a Regressão Logística como um caso especial de Modelo Linear Generalizado, e até mesmo softwares de análises estatísticas, como o R (ROYSTON & SAUERBREI (2008), FIGUEIRA (2006)). Assim sendo, faz-se necessário expor o conceito de Modelos Lineares Generalizados.

De acordo com Venables e colaboradores (1992), MLG's é um desenvolvimento dos modelos lineares para acomodar não-normalidade na distribuição da variável resposta e transformações para linearidade. Segundo Agresti (2002), MLG's estendem os modelos de regressão ordinários para incluir respostas com distribuições não-normais e modelar funções da média.

Os MLG's contribuem com uma teoria unificada para modelagem que inclui os principais modelos para variáveis discretas e contínuas.

[1] GALTON, F. **Regression Toward Mediocrity in Heridatary Stature.** *Journal of the Anthropological Institute,* 15 (1885), 246 – 263.

Modelos Lineares Generalizados (MLG's) pode ser especificado por três componentes: uma componente aleatória, uma componente sistemática e uma função de ligação. A componente aleatória identifica a distribuição de probabilidade da variável dependente, a componente sistemática especifica uma função linear entre as variáveis independentes e a função de ligação descreve a relação matemática entre a componente sistemática e o valor esperado da componente aleatória (AGRESTI, 2002; FIGUEIRA, 2006).

A componente aleatória de um MLG consiste nas observações da Variável Aleatória (VA) *Y*, com observações independentes (y_1, ..., y_n) de uma distribuição na família exponencial. Essa família tem função densidade de probabilidade na forma:

$$f(y_i, \theta_i) = a(\theta_i)b(y_i)\exp\left[y_i Q(\theta_i)\right] \quad (1)$$

O valor do parâmetro θ_i pode variar de 1 até n, dependendo do valor das variáveis explanatórias; o termo Q(θ) é chamado de parâmetro natural.

A componente sistemática do MLG é definida através de um vetor $\eta = \mathbf{X\beta}$, onde **X** é uma matriz que consiste nas VI's das *n* observações e **β** é um vetor dos parâmetros do modelo.

A função de ligação conecta os valores esperados das observações às variáveis explanatórias. Isso porque, assumindo $\mu_i = E(Y_i|x_i)$, com $i = \{1,...,n\}$ então $\eta_i = g(\mu_i)$. E,

$$g(\mu_i) = \beta_0 + \sum_{j=1}^{n} \beta_j x_j \quad (2)$$

Há duas classes importantes de MLG's: os que constituídos pelos modelos *logit* e os que são constituídos pelos modelos *loglinear*. Nos MLG's constituídos pelos modelos *logit* a VD pode ser associada a uma VA Bernoulli, enquanto que nos MLG's constituídos por modelos *loglinear* a VD pode ser associada a uma VA de Poisson (FIGUEIRA, 2006).

3.4.1 MODELOS LOGIT PARA DADOS BINÁRIOS

Um caso especial dos MLG´s refere-se a VD do tipo binária, ou seja, assumindo apenas dois valores. Geralmente, a VD assume os valores 0 ou 1, denotando fracasso ou sucesso na ocorrência de um evento de interesse.

É importante não confundir sucesso como sendo um evento desejável ou fracasso como um indesejável. Um exemplo seria um estudo na área de ergonomia experimental, onde 1 indique a ocorrência de ICT inaceitável e 0 a não ocorrência deste evento, ou ainda, em um estudo na área de saúde onde 1 indique a ocorrência de câncer e 0 sua não ocorrência.

Em semelhantes casos, diz-se que a VD possui distribuição de Bernoulli, e para semelhante experimento, chamado de experimento de Bernoulli, as probabilidades P(Y=1)=π e P(Y=0) = 1-π, para o qual E(Y)=π. A função densidade de probabilidade nesse caso é:

$$f(y;\pi) = \pi^{y}(1-\pi)^{1-y} = (1-\pi)\left[{}^{\pi}/_{(1-\pi)}\right]^{y} \quad (3)$$

$$= (1-\pi)\exp\left(y\ log\frac{\pi}{1-\pi}\right) \quad (4)$$

Para y=0 e 1. Pode-se fazer a associação da função densidade apresentada na equação (4) com a família exponencial natural apresentada em (1), identificando θ com π, a(π) = 1-π, b(y)=1, e $Q(\pi) = log\,[\pi/(1-\pi)]$. Assim, o parâmetro natural é chamado logit de π. Essa é a função de ligação canônica.

3.5 REGRESSÃO LOGÍSTICA

Por vezes, encontram-se modelos onde há uma necessidade especial quando à variável dependente: ela precisa assumir valores discretos. Geralmente isso ocorre quando a VD é uma variável qualitativa, expressa por duas ou mais categorias (FIGUEIRA, 2006).

As categorias (ou valores), que a VD assume pode ter natureza nominal ou ordinal. Em caso de natureza ordinal, há uma ordem natural entre as possíveis categorias, assim havendo o contexto da Regressão Logística Ordinal. Caso contrário, assume-se o contexto da Regressão Logística Nominal.

3.5.1 REGRESSÃO LOGÍSTICA BINÁRIA

A modelagem usando regressão logística é muito útil para situações nas quais deseja-se estar apto a predizer a presença ou ausência de uma característica ou resultado, baseado num conjunto de variáveis preditoras, sendo a VD do tipo dicotômica. A VD Y nesse caso, geralmente é codificada pelos valores 0 e 1, como sendo a ausência ou presença da característica em estudo (FIGUEIRA, 2006).

É habitual fazer $E(Yi|Xi) = \pi_i$, que é $P(Yi = 1)$. O comportamento da relação entre Xi e πi tem comportamento curvilinear em valores muito pequenos ou muito grandes de Xi, bem como, tem um comportamento aproximadamente linear em valores intermediários de Xi. Essa relação pode ser expressa por uma curva em forma de S, conforme mostrado na figura 2.

A relação entre Xi e πi é dada por:

$$\pi_i(X) = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}} \tag{5}$$

O modelo dado equação (5) satisfaz a exigência de $0 \leq \pi i \leq 1$. O modelo em termos da VD, Y, seria escrito como :

$$Y = \pi(X) + \varepsilon \tag{6}$$

Da equação (5), pode-se algebricamente chegar a:

$$ln\left(\frac{\pi}{1-\pi}\right) = g(x) \tag{7}$$

Onde:

$$g(x) = \beta_0 + \beta_1 X \tag{8}$$

O modelo acima, é chamado Modelo de Regressão Logística, já que vem de uma transformação logística, também conhecida como transformação *logit*.

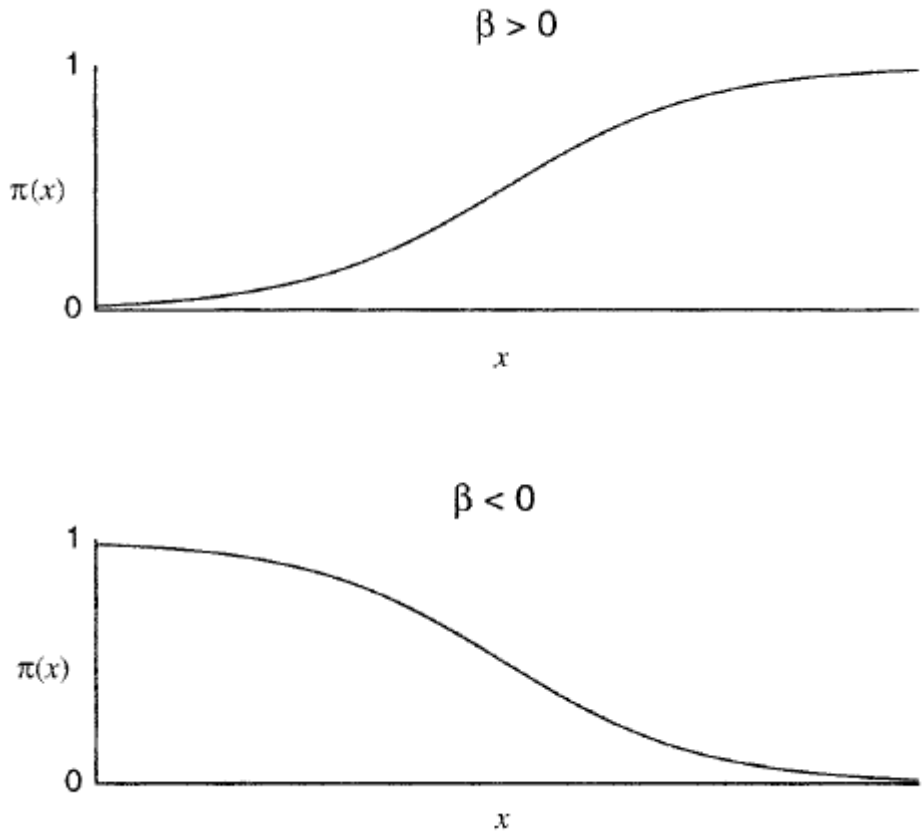

Figura 2: Curva em S gerada por uma função logit
Fonte: Agresti (2007)

Em (5), quando x tende a infinito, π (X) tende a zero se β_1 for negativo e tende a 1 caso β_1 seja positivo, como ilustrado na figura 2. Caso β_1 seja zero, a variável *Y* é independente da variável *X*.

Percebe-se a ausência do termo ε no modelo apresentado, já que o lado esquerdo do modelo é uma função de *E(Y|X)*, em substituição de *Y*, o que serve para remover o termo erro do modelo.

No caso logístico binário, ε pode assumir dois valores: se $y = 1$, então $\varepsilon = 1-\pi(x)$ com probabilidade $\pi(x)$, e se $y = 0$, então $\varepsilon = -\pi(x)$ com probabilidade $1 - \pi(x)$. Assim, a V.A. ε tem média zero e variância $\pi(x)\,[1-\pi(x)]$. Essa proposição indica que independente dos erros serem grandes ou pequenos, pode-se esperar que sua média seja nula (FIGUEIRA, 2006, pg. 72).

Assim, tem-se que:

$$y_i = \frac{\exp(gx_i)}{1 + \exp(gx_{i)}} + \varepsilon_i \qquad (9)$$

Onde ε_i segue as suposições para todo i,l={1,2,...,n)

i) $E(\varepsilon_i|x_i) = 0$

ii) $Var\ (\varepsilon_i|x_i) = \pi(x_i)[1-\pi(x_i)]$

iii) $Cov\ (\varepsilon_i,\varepsilon_l) = 0\ se\ i \neq l.$

3.5.2 ESTIMAÇÃO DE PARÂMETROS EM REGRESSÃO LOGÍSTICA

O método mais comumente utilizado para a estimação dos parâmetros de um Modelo de Regressão Logística é o método da Máxima Verossimilhança (RYAN, 2009).

O método da máxima verossimilhança é ilustrado a seguir. A função de distribuição de probabilidade de Y_i para a regressão logística binária simples é:

$$f(y_i, \pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{10}$$

As observações são independentes. Logo, a função distribuição de probabilidade conjunta de y_1, y_2, ... , y_n será:

$$\prod_{i=1}^{n} f(y_i, \pi_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1-y_i)^{1-y_i} \tag{11}$$

Com $y_i = \{0,1\}$.

Então, a função de verossimilhança será dada por:

$$L(\beta) = \prod_{i=1}^{n} \pi_i^{y_i}(1-y_i)^{1-y_i} \tag{12}$$

O princípio da máxima verossimilhança é estimar o valor de β que maximiza a função de verossimilhança. A aplicação do logaritmo natural ajuda no processo de manipulação algébrica.

$$l(\beta) = \ln[L(\beta)] = ln\left[\prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}\right] \tag{13}$$

$$= \sum_{i=1}^{n} [y_i \ln(\pi_i)) + (1-y_i)\ln(1-\pi_i)] \tag{14}$$

$$= \sum_{i=1}^{n} [y_i \ln(\pi_i) + \ln(1-\pi_i) - y_i \ln(1-\pi_i)] \tag{15}$$

$$= \sum_{i=1}^{n} [y_i \ln \left(\frac{\pi_i}{1-\pi_i}\right) + \ln (1-\pi_i)] \quad (16)$$

Fazendo as devidas substituições de valores na equações acima, temos:

$$l(\beta) = \sum_{i=1}^{n} \left[y_i(\beta_0 + \beta_1 x_i) + \ln \left(\frac{1}{1+\exp (\beta_0 + \beta_1 x_i)}\right)\right] \quad (17)$$

$$= \sum_{i=1}^{n} [y_i(\beta_0 + \beta_1 x_i) - \ln (1+\exp (\beta_0 + \beta_1 x_i)] \quad (18)$$

O valor de β que maximiza l(β) é encontrado após derivar-se l(β) em relação aos parâmetros (β_0, β_1):

$$\frac{\partial l(\beta)}{\partial \beta_0} = \sum_{i=1}^{n} \left[y_i - \frac{1}{1+\exp (\beta_0 + \beta_1 x_i)} \exp (\beta_0 + \beta_1 x_i)\right] \quad (19)$$

$$\frac{\partial l(\beta)}{\partial \beta_1} = \sum_{i=1}^{n} \left[y_i x_i - \frac{1}{1+\exp (\beta_0 + \beta_1 x_i)} \exp (\beta_0 + \beta_1 x_i) x_i\right] \quad (20)$$

Essas equações igualadas a zero, geram o seguinte sistema de equações:

$$\sum_{i=1}^{n} (y_i - \pi_i) = 0 \quad (21)$$

$$\sum_{i=1}^{n} x_i (y_i - \pi_i) = 0 \quad (22)$$

Tais equações são não-lineares nos parâmetros e requerem o emprego de processo interativo na sua solução.

3.5.3 INTERPRETAÇÃO DOS COEFICIENTES

Considere a situação em que a VI e a VD são dicotômicas. Nesta situação há dois valores possíveis para π(x) e dois equivalentes para 1- π(x).

A chance da resposta quando x=1 é definida como π(1)/[1-π(1)]. De maneira análoga, a chance da resposta quando x = 0 é definida como π(0))/[1-π(0)]. O logaritmo da razão é dado por:

$$g(1) = ln \left[\frac{\pi(1)}{1-\pi(1)}\right] \quad (23)$$

e

$$g(0) = \ ln\left[\frac{\pi(0)}{1-\pi(0)}\right] \tag{24}$$

Os valores obtidos na VD relacionados com o valor da VI são mostrados na tabela (2).

Tabela 2: Valores do modelo de regressão logística quando a variável independente é dicotômica

	x=1	x=0
Y = 1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
Y = 0	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$
Total	1,0	1,0

A razão das chances ("*Odds ratio*"), denotada por ψ, é definida por:

$$\psi = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} \tag{25}$$

Já o logaritmo da "Odds ratio", "log odds", é:

$$\ln(\psi) = ln\left\{\frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}\right\} = g(1) - \ g(0) \tag{26}$$

Usando manipulação algébrica e a expressão para o modelo de regressão logística na tabela 1, temos:

$$\psi = \frac{\frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \Big/ \frac{1}{1 + \exp(\beta_0 + \beta_1)}}{\frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \Big/ \frac{1}{1 + \exp(\beta_0)}} \tag{27}$$

$$= \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) \tag{28}$$

Assim, o logaritmo da razão das chances sendo dado por:

$$\ln(\psi) = \ln[\exp(\beta_1)] = \beta_1 \tag{29}$$

A razão das chances como definida acima é uma medida de associação muito utilizada em muitas áreas, como epidemiologia, saúde e ergonomia (SILVA *et al.*, 2007).

Devido a sua fácil interpretação, a razão das chances é uma medida muito utilizada em regressão logística. É definida como a chance de ocorrência de um evento entre indivíduos que têm um fator de risco, comparados com indivíduos não expostos, sujeitos ao evento. Por exemplo, se Y denota a ocorrência de câncer de pulmão e X representa se a pessoa é fumante ou não, um valor $\psi = 2$ indica que a chance de uma pessoa que fuma adquirir câncer no pulmão é duas vezes maior que uma pessoa que não fuma também o adquirir (SOUZA, 2006).

O intervalo de confiança para a razão das chances de 100(1-α)% é obtido calculando o intervalo para de confiança para β_1 e aplicando a exponencial. Tem-se:

$$\exp\left[\hat{\beta}_1 \pm z_{1-\alpha/2} SE\left(\hat{\beta}_1\right)\right] \tag{30}$$

Onde $SE\left(\hat{\beta}_1\right)$ é o erro padrão de $\hat{\beta}_1$.

3.5.4 INFERÊNCIA

Após estimar os coeficientes de regressão, a significância da variável no modelo geralmente é o primeiro aspecto que o analista observa. Usualmente, isso envolve a formulação de testes de hipóteses para saber se a variável é ou não significantemente correlata com a saída. Segundo Hosmer e Lemeshow (2002), uma aproximação para testar a significância do coeficiente de uma variável em um modelo é relacionada com a seguinte questão: “O modelo que inclui a variável em questão nos diz mais a respeito da saída do que o modelo que não a inclui?”

Na regressão logística, a comparação dos valores observados e preditos é baseado na função logaritmo da verossimilhança, apresentada na equação (18).

Conceitualmente, utiliza-se o modelo saturado para essa comparação, onde esse trata-se de um modelo em que há tantos parâmetros quando dados. Um exemplo simples de modelo saturado é usar uma amostra com n=2 para ajustar um modelo logístico simples, onde também se tem dois parâmetros.

A comparação então usa a seguinte expressão:

$$D = -2ln\left[\frac{Verossimilhança\ do\ modelo\ ajustado}{Verossimilhança\ do\ modelo\ saturado}\right] \tag{31}$$

A quantia dentro dos colchetes é chamada de razão de verossimilhança e o uso do sinal negativo multiplicado por dois é necessário para obter uma quantia a qual possui distribuição conhecida, podendo assim, ser usada no contexto do teste de hipóteses.

Usando a equação (14) na equação (31), esta se torna:

$$D = -2\sum_{i=1}^{n}\left[y_i\, ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i) ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \tag{32}$$

A estatística D é chamada *deviance*, e desempenha um papel fundamental em algumas aproximações para verificar a bondade do ajuste (HOSMER e LEMESHOW, 2000).

No modelo saturado, no entanto, $\hat{\pi}_i = y_i$, e a verossimilhança nesse caso vale 1. Assim, a equação (29) se torna:

$$D = -2\ln(Verossimilhaça\ do\ modelo\ ajustado) \tag{33}$$

No contexto da análise de significância de uma variável em um modelo ajustado, a comparação do modelo com e sem a variável em questão se dá através da comparação da estatística *deviance* com e sem a variável no modelo. Assim, define-se:

$$G = D(modelo\ sem\ a\ variáve) - D(modelo\ com\ a\ variável) \tag{34}$$

Uma vez que a verossimilhança do modelo saturado está presente no modelo com e sem a variável, pode-se escrever G da seguinte forma:

$$G = -2\left[\frac{(verossimilhança\ sem\ a\ variável)}{(verossimilhança\ com\ a\ variável)}\right] \tag{35}$$

No caso de regressão logística simples, seria testado se a inclusão de uma variável independente x melhoraria o ajuste do modelo sem a variável, ou seja, se o modelo apenas com o intercepto β_0 descreveria melhor o comportamento dos dados observados. Isso pode ser encarado como fazer $\beta_1 = 0$ na equação de regressão. Sob a hipótese de $\beta_1 = 0$, a estatística G segue uma distribuição qui-quadrado (χ^2) com um grau de liberdade. Semelhante formulação de teste de hipóteses permite afirmar se uma variável é ou não significante no modelo de regressão, além de permitir calcular o p valor de tal variável.

Já o teste de Wald, compara o valor de $\hat{\beta}_1$ obtido da estimação de máxima verossimilhança e o seu erro padrão, $\widehat{SE}(\hat{\beta}_1)$.

$$W = \frac{\hat{\beta}_1}{\widehat{SE}(\hat{\beta}_1)} \tag{36}$$

Sob a hipótese de que $\beta_1 = 0$, W segue a distribuição normal padrão.

Já o teste de *Score* tem como principal vantagem o uso de pequeno esforço computacional no seu cálculo. Esse teste é baseado na teoria da distribuição das derivadas do log da máxima verossimilhança.

O teste de *Score* é dado por:

$$ST = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y})\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{37}$$

Sob a hipótese de que $\beta_1 = 0$, a estatística de *Score* tem distribuição normal padrão.

3.5.5 REGRESSÃO LOGÍSTICA MÚLTIPLA

Pode ser visto como uma extensão do caso simples, onde tem-se agora em vez de um preditor X, um conjunto com p preditores. Hosmer e Lemeshow (1989) estabelecem esta generalização da seguinte forma: Seja um conjunto com p variáveis independentes, denotadas por $x_i^T = (x_{i0}, x_{i1}, \ldots, x_{ip})$, o vetor da i-ésima linha da matriz (X) das variáveis explicativas; denota-se por $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$, o vetor de parâmetros desconhecidos e β_j o j-ésimo parâmetro associado a variável explicativa x_j. No modelo de regressão múltipla, a probabilidade de sucesso é dada por:

$$\pi_i(x_i) = P(Y_i = 1|X = x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})} \tag{38}$$

$$= \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} \tag{39}$$

E a probabilidade de fracasso se torna:

$$1 - \pi_i(x_i) = P(Y_i = 0|X = x_i) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})} \tag{40}$$

Assim, a função g(.) toma a forma:

$$g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \tag{41}$$

Os erros seguem as mesma suposições do caso simples. O modelo logístico múltiplo é dado por:

$$y_i = \frac{\exp(g_i)}{1 + \exp(g_i)} + \varepsilon_i \tag{42}$$

Onde

$$g_i(x) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_n x_{ip} \tag{43}$$

Observação: No modelo apresentado acima, pode-se ter várias variáveis discretas, do tipo escala nominal, cujos diversos números usados para representar níveis dessas escalas e não possuem significado numérico. Estas são as variáveis *dummies* (FIGUEIRA, 2006). Nesse caso, temos:

$$g(x) = \beta_0 + \beta_1 x_1 + \cdots + \sum_{l=1}^{k_j - 1} \beta_{jl} x_{jl} + \cdots + \beta_p x_p \tag{44}$$

Quando temos uma variável na escala nominal com k possíveis valores. Foram introduzidas k-1 variáveis *dummies*, onde a *j*-ésima variável está na escala nominal com k_j níveis; cada uma das k_j - 1 variáveis *dummies* é denotada por x_{jl} e seu coeficiente β_{jl}, com l = $\{1,...,k_j\text{-}1\}$.

As matrizes serão usadas no contexto da Regressão Logística Múltipla:

Y = $(y_1, ..., y_n)'_{1xn}$

Π = $(\pi_1,..., \pi_n)'_{1xn}$

B = $(\beta_0,...,\beta_p)'_{1x(p+1)}$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{n1} \\ 1 & x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \tag{45}$$

$$\mathbf{V} = \begin{bmatrix} \pi_1(1-\pi_1) & 0 & \cdots & 0 \\ 0 & \pi_2(1-\pi_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_n(1-\pi_n) \end{bmatrix} \tag{46}$$

A função de verossimilhança é dada por:

$$L(\mathrm{B}) = \prod_{i=1}^{n} \pi_i^{yi} (1-\pi_i)^{1-yi} \tag{47}$$

Com $y_i = \{0,1\}$. O estimador de máxima verossimilhança é a solução para:

$$\sum_{i=1}^{n}(y_i - \pi_i) = 0 \tag{48}$$

E,

$$\sum_{i=1}^{n} x_{ij}(y_i - \pi_i) = 0 \tag{49}$$

j = {1, ..., p}

de forma compacta, pode-se representar as p + 1 equações de verossimilhança, em notação matricial, como:

$$\frac{\partial L(\beta)}{\partial \beta}\, \boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{\Pi}) = \mathbf{0} \tag{50}$$

Para encontrar o valor que maximiza l(B), utiliza-se o processo interativo de Newton – Raphson, para isso necessitando derivar l(B) em relação a cada parâmetro

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n}\left[y_i x_{ij} - \frac{\exp(x_i^T\beta)}{1 + \exp(x_i^T\beta)}\, x_{ij}\right] \tag{51}$$

$$= \sum_{i=1}^{n}[y_i - \pi_i]x_{ij} \tag{52}$$

A estimação da variância e covariância dos coeficientes advém da matriz de informação. A obtenção dessa matriz segue a teoria de que os estimadores são obtidos da matriz da segunda derivada parcial da função de verossimilhança. Essas derivadas parciais têm a seguinte forma:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = -\sum_{i=1}^{n} x_{ij}^2 \pi_i(1 - \pi_i) \tag{53}$$

e

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = -\sum_{i=1}^{n} x_{ij} x_{il} \pi_i(1 - \pi_i) \tag{54}$$

para *j,l = 1,2,...,p*. A matriz *(p+1) x (p+1)* contendo o negativo dos termos das equações (53) e (54) é conhecida como a matriz de informação observada, denotada como ***I(β)*** (HOSMER & LEMESHOW, 2002). As variâncias são assim, obtidas de $\boldsymbol{\Gamma^{-1}(\beta)}$. Desta forma, $\boldsymbol{Var\ (\beta) = \Gamma^{-1}(\beta)}$**.** O elemento na *j-ésima* diagonal desta matriz se refere à variância do

parâmetro $\hat{\beta}_j$, enquanto que um elemento qualquer da matriz I_{jl} refere-se à covariância dos parâmetros $\hat{\beta}_j$ e $\hat{\beta}_l$. Comumente, usa-se a notação $\widehat{Var}(\hat{\beta}_j)$ *e* $\widehat{Cov}\,(\hat{\beta}_j, \hat{\beta}_l), j, l = 0,1,2, \ldots, p$ para denotar os valores desta matriz. Ainda desta matriz, advém o erro padrão estimado dos coeficientes estimados, denotado por:

$$\widehat{SE}(\hat{\beta}_j) = \left[\widehat{Var}(\hat{\beta}_j)\right]^{1/2} \tag{55}$$

O vetor escore U(B) pode ser escrito como:

$$U(\beta) = X^T y - X^T \pi = X^T (y - \pi) \tag{56}$$

A matriz de informação de Fischer é dada por:

$$\hat{I}(\hat{\beta}) = \boldsymbol{X^T V X} \tag{57}$$

Onde V já foi mostrada anteriormente na matriz (46) e **X** a matriz de dados, mostrada na matriz (45).

3.5.6 TESTANDO A SIGNIFICÂNCIA DE UM MODELO LOGÍSTICO MÚLTIPLO

Como no caso do modelo logístico simples, no caso múltiplo após a estimação dos parâmetros do modelo inicia-se a fase de verificação do ajuste do modelo. O primeiro passo nessa fase usualmente é conferir a significância das variáveis presentes no modelo. O teste da razão de verossimilhança no caso múltiplo se processa de maneira análoga ao caso simples, com a diferença de que os valores ajustados $\hat{\pi}$ são baseados em um vetor contendo p+1 parâmetros, já que o modelo apresenta p VI's. Sob a hipótese de que os p parâmetros coeficientes das covariáveis serem iguais a zero, a distribuição de G será qui-quadrado com p graus de liberdade.

Diante da conclusão de que algum ou todos os coeficientes das covariáveis são não nulos, em geral observa-se o teste estatístico de Wald univariado, ou seja, testa-se a significância do parâmetro β_j, para j = 1, ..., p. O teste de Wald é dado por:

$$W = \frac{\hat{\beta}_j}{\widehat{SE}\,(\hat{\beta}_j)} \tag{58}$$

Sob a hipótese de que o parâmetro β_j é nulo, essa estatística segue a distribuição normal padrão.

Já o análogo do teste de Wald com mais de uma variável é obtido da seguinte maneira:

$$W = \hat{\beta}'\left[\widehat{Var}(\hat{\beta})\right]^{-1}\hat{\beta} \quad (59)$$

$$W = \hat{\beta}'(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})\hat{\beta} \quad (60)$$

Neste caso, W segue a distribuição qui-quadrado com p+1 graus de liberdade, sob a hipótese de cada um dos p+1 coeficientes ser nulo.

A estimação dos intervalos de confiança para o intercepto e os demais parâmetros do modelo de regressão logística são baseados nos seus respectivos testes de Wald. Os limites de um intervalo de confiança de 100(1-α)% para um parâmetro $\hat{\beta}_j$ é:

$$\hat{\beta}_j \pm z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_j) \quad (61)$$

Para o caso de regressão logística simples, temos $g(x) = \beta_0 + \beta_1 x$. Nesse caso, a variancia do estimador do logit exige a obtenção de uma soma de variâncias.

$$\widehat{Var}[\hat{g}(x)] = \widehat{Var}(\widehat{\beta_0}) + x^2\widehat{Var}(\widehat{\beta_1}) + 2\widehat{Cov}(\widehat{\beta_0}, \widehat{\beta_1}) \quad (62)$$

Os limites para o intervalo de confiança de um intervalo de confiança de 100(1-α)% do logit para o caso de regressão logística simples é:

$$\hat{g}(x) + z_{1-\alpha/2}\widehat{SE}[\hat{g}(x)] \quad (63)$$

3.6 ESTRATÉGIAS DE SELEÇÃO DE MODELOS

No estágio inicial de uma pesquisa envolvendo a proposição de modelo, há sempre a suspeita de que muitas variáveis são potenciais preditores de uma ou mais VD's de interesse. Após a coleta de tais dados, cabe ao analista verificar se cada uma daquelas variáveis afetam ou não a saída estudada, bem como o grau em que isso ocorre, sendo aquele um importante preditor ou não. Wiesberg (2005) relata que há problemas em que o pesquisador depara-se com centenas e algumas vezes milhares até, de variáveis potencialmente preditoras.

Hosmer & Lemeshow (2007) defendem o uso de métodos concebidos para lidar com a situação em que hajam muitas variáveis a serem incluídas ou não como covariáveis em um modelo. Para os referidos autores, o objetivo de qualquer método gerado para esse fim é selecionar aquelas variáveis que resultem em uma melhor modelo dentro de um contexto científico. Para alcançar isso, deve-se ter:

i) Um plano básico para selecionar variáveis para o modelo;

ii) Um conjunto de métodos para verificar a adequação do modelo em termos de suas variáveis individuais e seu ajuste como um todo.

A busca estatística por um modelo, envolve a eleição de um tal que, sendo o mais simples possível, ainda seja capaz de explicar os dados. Modelos contendo um excessivo número de covariáveis além de não serem práticos do ponto de vista de sua utilização, possui uma maior instabilidade numérica e dependência dos dados originais (HOSMER & LEMESHOW, 2007).

A estratégia de eleição de um modelo muda com o tipo de estudo em que este seja empregado. Alguns estudos são elaborados para responder a determinadas questões, e assim, estas questões devem guiar a escolha dos termos no modelo. É o caso de analise confirmatória, onde há um conjunto restrito de modelos a serem utilizados, por exemplo, para checar a hipótese de um sobre um efeito, comparando modelos com e sem aquele efeito. Em estudos exploratórios, a busca entre possíveis modelos devem buscar indícios sobre a estrutura de dependência e levantar questões para futuras pesquisas (AGRESTI, 2002).

É útil em ambos os casos, primeiro estudar o efeito em Y de cada preditor, graficamente para preditores contínuos e através de tabelas de contingências, para preditores discretos.

Verificar a correlação entre as VI's, de modo a não permitir que variáveis altamente correlacionadas participem do mesmo modelo são tidas como primárias nesse caso. Agresti (2002) relata que o problema da multicolinearidade é um problema pertencente a não apenas regressão linear, mas a qualquer modelo de regressão abarcado por MLG's.

3.6.1 COLINEARIDADE

Dois termos X_1 e X_2 são ditos colineares, ou linearmente dependentes, se é possível obter uma equação linear do tipo:

$$c_1X_1 + c_2X_2 = c_0 \tag{64}$$

Ou seja, se a partir de uma combinação linear uma variável possa ser obtida através de outra.

A colinearidade entre variáveis é medida através da correlação da amostra, $r_{1,2}^2$. A colinearidade perfeita entre duas variáveis ocorre quando , $r_{1,2}^2 = 1$, e a não colinearidade perfeita advém quando , $r_{1,2}^2 = 0$.

Quando duas variáveis são colineares, uma precisa ser deletada do modelo. A parte difícil reside em escolher qual a variável a ser deletada (WEISBERG, 2005, pg.216). O conhecimento do analista sobre as variáveis em relação ao assunto em estudo pesa significativamente na hora de tomar a decisão de qual variável deve deixar o modelo.

3.6.2 MÉTODOS COMPUTACIONAIS – MÉTODOS *STEPWISE*

Estes métodos fornecem compromisso computacional , já que nesse conjunto de algoritmos, apenas algumas possibilidades de combinações das variáveis são verificadas, entre todas as possibilidades. Por causa disso, os métodos Stepwise não garantem achar o subconjunto de variáveis dito ótimo, porém seus resultados são úteis (WEISBERG, 2005). De acordo com Agresti (2002), Goodman (1971) propôs métodos análogos, para regressão logística, da seleção *forward* e da eliminação *backward*, usados vastamente em regressão linear.

Seleção *Forward* – Esse método consiste em ir adicionando variáveis ao modelo até que adições futuras não melhorem o ajuste ou não haja mais variáveis a serem adicionadas.

Esse algoritmo facilita muito na seleção de variáveis quando é grande o número das mesmas. Por exemplo, quando no estudo estão presente 10 variáveis, há 2^k possibilidades de combinações entre variáveis que resultem em um modelo, ou seja, o modelo final precisaria ser eleito entre 1024 possibilidades, através do exame um a um. Com o uso desse algoritmo, o número de possibilidades seria $k + (k-1) + ... + 1 = k(k-1)/2 = 45$ das 1024 possibilidades iniciais.

O Algoritmo de eliminação *Backward* vai na direção oposta ao algoritmo apresentado anteriormente, começando com um modelo mais complexo e sequencialmente ir removendo termos. Em cada estágio, elimina-se o termo que a sua eliminação cause o menor dano ao ajuste do modelo, como por exemplo, aquela que apresente o maior p-valor. O processo termina quando a eliminação de qualquer variável leve a um ajuste significantemente mais pobre.

Deve haver um cuidado especial com variáveis *dummies* presentes no modelo, devendo ser toda ela adicionada ou excluída do modelo, não podendo permanecer apenas algumas categorias da mesma ou ser retirada (AGRESTI, 2002).

Novamente, esse algoritmo observa $k(k-1)/2$ possibilidades das 2^k iniciais.

Uma outra variante *stepwise* também é utilizada, sendo uma combinação dos algoritmos *forward* e *backward*, onde em cada passo pode ser adicionada ou deletada uma variável de tal forma que o subconjunto candidato minimize o critério de informação observado. Esse algoritmo tem a vantagem de fazer uma varredura em um número maior de possibilidades de subconjuntos, sem varrer os 2^k candidatos possíveis.

3.6.3 MÉTODO DE SELEÇÃO BASEADO EM CRITÉRIO DE INFORMAÇÃO

O critério para comparar os vários modelos gerados pela combinação de covariáveis é baseado na falta de ajuste de um modelo e sua complexidade. Isso porque, se por um lado é de interesse que o modelo final possua uma melhor sensibilidade aos dados, errando menos, é de igual modo que o modelo seja parcimonioso, ou seja, que seja o mais simples possível. De modo a manter o compromisso entre esses interesses conflitantes, Akaike propôs um critério que penaliza o modelo por ter muitas variáveis e por apresentar falta de ajuste. O critério de informação de Akaike é então dado por:

$$\text{AIC} = -2\,(\textit{log verossimilhança maximizada} - \textit{número de parâmetros no modelo}) \qquad (65)$$

Esse critério pode ser observado em métodos *stepwise* para indicar a superioridade ou não de um modelo em relação a uma tentativa anterior. O modelo que apresentar menor índice AIC apresenta superioridade em termos do critério de informação e este deve ser preferível em relação a outro que apresente índice menor.

3.7 REGRESSÃO LOGÍSTICA MULTINOMIAL

No tratamento dado até aqui, a VD assume sempre dois valores. A generalização dessa situação modela respostas categóricas com mais de duas categorias.

De acordo com Agresti (2007), modelos *logit* com multicategorias usa todos os pares de categorias para especificar a “odds” de a saída recair sobre uma categoria em relação a outra categoria. Nesse tipo de modelagem, a ordem entre as categorias é considerada irrelevante.

Dentre as categorias da VD, uma é eleita para ser a categoria de referência. Assim, se a última categoria (J) da VD é usada para esse fim, o logit para essa modelagem é:

$$\ln\left(\frac{\pi_j}{\pi_J}\right), j = 1, \dots, J-1 \tag{62}$$

Por exemplo, se J = 3, para a modelagem nesse caso, calcula-se log(π1/π3) e log(π2/π3), dado que nesse modelo, ou a resposta recai em j, j=1,2 ou em J, J=3.

O modelo logit com categoria de referência com um preditor x, é:

$$\ln\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x \quad , j = 1, \dots, J-1 \tag{63}$$

Esse modelo apresenta J-1 equações, com parâmetros distintos para cada uma delas. Quando J = 2, essa modelagem se torna o caso logístico binário.

Por uma questão de simplicidade de notação, por vezes faz-se:

$$\eta_j = \alpha_j + \beta x \tag{64}$$

Então,

$$\text{logit}\left(\frac{\pi_j}{\pi_J}\right) = \eta_j + \varepsilon \quad , j = 1, \dots, J-1 \tag{65}$$

E o modelo multinomial expresso em termos de probabilidades de ocorrência das categorias da VD, toma a forma de:

$$\hat{\pi}_j = \frac{\exp(\hat{\eta}_j)}{\sum_{h=1}^{J-1} \exp(\hat{\eta}_h)} \qquad , j = 1, \dots, J-1 \tag{66}$$

Por exemplo, se J=3, ter-se-ia:

$$\hat{\pi}_1 = \frac{\exp(\hat{\eta}_1)}{\exp(\hat{\eta}_1) + \exp(\hat{\eta}_2)} \tag{67}$$

$$\hat{\pi}_2 = \frac{\exp(\hat{\eta}_2)}{\exp(\hat{\eta}_1) + \exp(\hat{\eta}_2)} \tag{68}$$

$$\hat{\pi}_3 = \frac{1}{\exp(\hat{\eta}_1) + \exp(\hat{\eta}_2)} \tag{69}$$

Onde, em $\hat{\pi}_3$, o numerador igual a 1 representa α3 = β3 = 0 com relação a categoria de referência.

Estendendo a modelagem para o caso de termos um conjunto de p variáveis explicativas, o modelo de regressão para a categoria j se torna:

$$logit\left(\frac{\pi_j}{\pi_J}\right) = \beta_{oj} + \beta_{1j}x_1 + \cdots + \beta_{pj}x_p + \varepsilon \quad , j = 1, \ldots, J-1 \tag{70}$$

Para construir a função de verossimilhança, Hosmer & Lemeshow (2002) ilustram o processo para o caso em que a VD possui três categorias, com o auxílio de três variáveis binárias, que têm por objetivo ilustrar a que categoria pertence uma observação; ditas Y0, Y1 e Y2. Caso um valor observado da VD esteja na categoria 2, por exemplo, faz-se então Y0=0; Y1=0 e Y2=1.

Assim, usando essa notação, a função de verossimilhança condicional para uma amostra de n observações independentes é:

$$l(\beta) = \prod_{i=1}^{n}[\pi_0(x_i)^{y_o}\pi_1(x_i)^{y_1}\pi_2(x_i)^{y_2}] \tag{71}$$

Tomando o logaritmo da equação () e usando o fato que $\sum y_{ji} = 1$ para cada i, o logaritmo da função de verossimilhança é:

$$L(\beta) = \sum_{i=1}^{n} y_{1i}g_1(x_i) + y_{2i}g_2(x_i) - \ln\{1 + exp[g_1(x_i)] + exp[g_2(x_i)]\} \tag{72}$$

As equações de verossimilhança são achadas tomando as derivadas parciais primeiras de L(β) com respeito a cada um parâmetros desconhecidos. A forma geral dessas equações, dadas por Hosmer & Lemeshow (2002) é:

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^{n} x_{ki}\left(y_{ji} - \pi_{ji}\right) \tag{73}$$

Para j= 1, 2, ..., J-1 e k = 0,1,2,..., p, com x0i = 1 para cada objeto.

O estimador de máxima verossimilhança $\hat{\beta}$ é obtido igualando essas equações a zero e as solucionando para β.

A matriz da derivada parcial segunda é necessária para a obtenção da matriz de informação e estimação da matriz de covariância dos estimadores de máxima verossimilhança. A forma geral dos elementos na matriz da derivada parcial segunda é:

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk}\partial \beta_{jk'}} = -\sum_{i=1}^{n} x_{k'i}x_{ki}\pi_{ji}\left(1 - \pi_{ji}\right) \tag{74}$$

e

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{jk'}} = \sum_{i=1}^{n} x_{k'i} x_{ki} \pi_{ji} \pi_{j'i} \tag{75}$$

Para j e j'= 1,2,..., J-1 e k e k'=0,1,2,...,p. A matriz de informação observada é a matriz 2(p+1) por 2(p+1) cujos elementos são os valores negativos das encontrados das equações (74) e (75), avaliadas em $\hat{\beta}$. O estimador da matriz covariância dos estimadores de máxima verossismilhança é o inverso da matriz de informação observada.

A "odds ratio" de um modelo multinomial, supondo a categoria Y=0 como a referência, é dada por:

$$OR_j(a,b) = \frac{P(Y = j|x = a)/P\,(Y = 0|x = a)}{P(Y = j|x = b)/P(Y = 0|x = b)} \tag{76}$$

Representando a razão das chances ("*odds ratio*") da saída Y=j versus a saída Y=0, para os valores da covariável em x=a versus x=b.

A indicação preliminar da importância de uma variável independente no modelo, pode ser obtida a partir da estatística do teste de Wald. Porém, deve-se usar o teste da razão da verossimilhança para avaliar significância. Por exemplo, para testar a significância de uma variável independente em um modelo, compara-se o logaritmo da verossimilhança do modelo contendo a VI com o logaritmo da verossimilhança do modelo apenas contendo os interceptos. De acordo com Hosmer & Lemeshow (2007), sob a hipótese nula de que todos os coeficientes de regressão sejam nulos no modelo, o negativo do dobro da mudança no logaritmo da verossimilhança segue uma distribuição qui-quadrado com dois graus de libredade.

3.7.1 REGRESSÃO LOGÍSTICA MULTINOMIAL ORDINAL

Quando a VD possui uma ordenação entre as suas categorias, o uso do modelo logístico para respostas ordinais tem interpretações mais simples e potencialmente maior poder (AGRESTI, 2007).

A regressão logística para respostas ordinais se baseia no uso da probabilidade acumulada de Y. Assim, a probabilidade considerada agora, é a de que o valor de Y recai numa faixa de interesse, j, ou em categorias que se situem em faixas inferiores a ela. Então, dada uma categoria j de interesse:

$$P(Y \leq j) = \pi_1 + \ \ldots + \pi_j, \quad j = 1, \ldots, J \tag{77}$$

A probabilidade acumulada reflete a ordenação entre as categorias da VD. Decorre que $P(Y \leq 1) \leq P(Y \leq 2) \leq \cdots \leq P(Y \leq J) = 1$.

Os logits para probabilidade acumulada são:

$$logit[P(Y \leq j)] = \ln\left[\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right] = \ln\left[\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right], j = 1, \ldots, J-1 \qquad (78)$$

Para J = 3, por exemplo, o modelo usa logit [P(Y≥3)] = logit [$\pi_3/(\pi_2 + \pi_1)$] e logit [P(Y≥2)] = logit [$(\pi_3 + \pi_2)/ \pi_1$]. Assim, cada logit cumulativo usa todas as categorias da resposta.

De acordo com Agresti (2007), um modelo para logit acumulativo se parece com um modelo de regressão logística binária, na qual as categorias de 1 a j se combinam para formar uma única categoria e as outras j+1 a J formam uma segunda categoria.

Para apenas um preditor x, tal modelo de logit cumulativo pode ser escrito da seguinte forma:

$$\text{logit}[P(Y \leq j)] = \beta_{0j} + \beta_1 x \ , \qquad j = 1, \ldots, J-1 \qquad (79)$$

Na equação acima, β não possui um índice j, indicando que o efeito da variável x é descrito por apenas um parâmetro para todas as categorias.

A figura 4 ilustra o caso de um modelo de regressão logística multinomial ordinal, onde a propriedade de Odds proporcional vale, com quatro categorias e uma variável explicativa. As curvas se mostram parecidas com o caso binomial. Neste modelo, o intercepto é o parâmetro que diferencia o modelo para uma categoria de outra categoria, como se pode perceber na equação acima o índice j no parâmetro β_0.

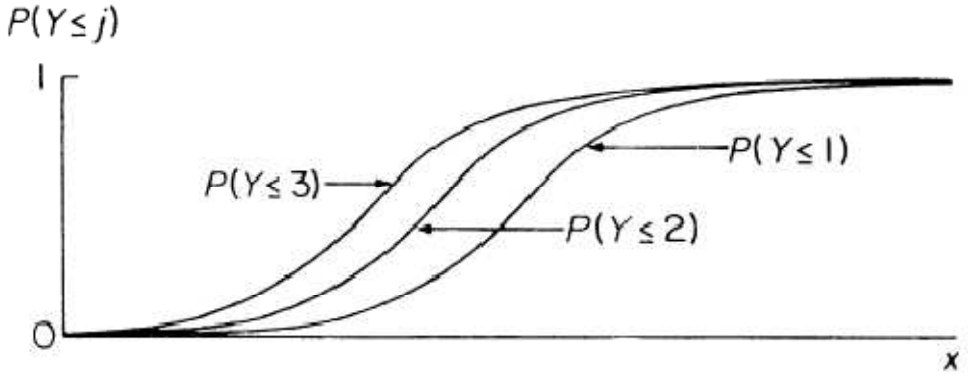

Figura 3: Probabilidade cumulativa no modelo de *odds* proporcional
FONTE: AGRESTI (2007)

Segundo Ananth e Klaeinbaum (1997), o modelo é invariante quando a codificação das categorias é invertida (a J-ésima categoria passa a ser a primeira, a primeira passa a ser a J-ésima, a segunda passa a ser a penúltima e assim por diante). Agresti (2007), afirma que nesse caso porém, os sinais dos $\hat{\beta}$'s ficam invertidos.

Ananth e Kleinbaum (1997) afirmam que existem vários modelos de regressão logística usados quando a resposta possui ordenação, tais como o modelo de *odds* proporcional, modelo de *odds* proporcional parcial, modelo de razão contínua e modelo esteriótipo. Apresenta-se aqui apenas o modelo de Odds Proporcionais.

3.7.1.1 MODELO DE *ODDS* PROPORCIONAIS

Considere uma variável resposta Y multinomial com saídas categóricas, denotadas por 1,2,..., *k* e seja x_i um vetor *p*-dimensional das covariáveis. A dependência de Y em x para o modelo de *odds* proporcional tem a seguinte representação:

$$\Pr(Y \leq y_j|x) = \frac{\exp(\alpha_j - \boldsymbol{x'\beta})}{1 + \exp(\alpha_j - \boldsymbol{x'\beta})}, j = 1,2, \dots, k \tag{80}$$

Ou, na forma de logit:

$$logit(\Pi_j) = ln\left[\frac{\Pi_j}{1 - \Pi_j}\right] \tag{81}$$

$$ln\left[\frac{\Pr\,(Y \leq y_j|x}{\Pr\,(Y > y_j)}\right] = \alpha_j - \boldsymbol{x'\beta} \tag{82}$$

Onde $\Pi_j = Pr\ (Y \leq y_j)$ é a probabilidade cumulativa do evento $Y \leq y_j$. α_j são os intercepto desconhecidos, satisfazendo a condição $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_k$, e $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$'é um vetor dos coeficientes de regressão correspondente ao vetor das covariáveis, **x**.

Percebe-se que nesse caso, os termos de $\boldsymbol{\beta}$ não dependem de *j*, ou seja, no modelo de odds proporcional, não importa em que faixa da VD esteja o valor, os β's permanecem os mesmos. Em outras palavras, a relação entre Y e **x** permanece inalterada ao se percorrer toda a extensão de Y. Apenas muda nesse caso, para cada faixa de valor da VD, os $\alpha_{j's}$. Essa premissa faz com que esse modelo receba o nome de modelo de odds proporcionais, pois assume-se haver uma idêntica odds ratio nos *k* pontos de corte, ou suposição de regressão paralela (ABREU, 2009).

3.8 APLICAÇÃO DE ANÁLISE DE REGRESSÃO EM ESTUDOS NA ÁREA DE ERGONOMIA

Na ergonomia acontece de lidar-se com variáveis de diversas naturezas, não sendo rara a manipulação com variáveis categóricas. Essa seção traz alguns exemplos de como a análise de regressão logística tem sido utilizada no campo da ergonomia nos últimos anos:

Evans e Patterson (2000) realizaram um estudo de campo epidemiológico para determinar a incidência de dor do pescoço e ombros em uma população de não secretários usuários de computador e testar a hipotese de que baixa habilidade de digitação, horas de uso de computador, *score* de tensão e estação de trabalho mau montada são associados com enfermidades no pescoço e ombros. Participaram do estudo 170 sujeitos de sete locais de trabalho da cidade de Hong Kong que responderam o número de horas de uso de computador, dores no pescoço e ombros e níveis de tensão. Estação de trabalho e fatores posturais foram observados pelos pesquisadores. Estatística descritiva e matriz de correlação foram usadas para revelar a natureza das variáveis e suas correlações. ANOVA univariada, teste de Kruskall Wallis e teste do qui-quadrado foram usados para testar diferenças de variáveis entre locais de trabalho. Análise de Regressão Múltipla foi usada para identificar as variáveis que são preditoras das dores no pescoço e ombros. Devido a natureza categórica da variável dependente, dores no pescoço ou ombros, regressão logística foi usada na análise. Apenas *score* de tensão e gênero foram achados como preditivas de dores no pescoço e ombros.

Fogleman e Lewis (2002), coletaram dados junto a 373 pessoas, com o intuito de identificar fatores de risco junto a pessoas que usam terminais de vídeo no trabalho. Os respondentes foram questionados acerca de sintomas em seis regiões do corpo, informações demográficas. Dois métodos multivariados foram utilizados: A análise fatorial exploratória, com o intuito de obter informações descritivas a partir dos dado; a regressão logística, que foi usada para estimar os riscos. Os resultados indicaram, com significância estatística, que os riscos de desconforto em cada região do corpo crescem com o número de horas de uso do teclado. Posicionamento de monitor e teclado impróprias também foram significativamente associados com o desconforto na cabeça/olho e ombros/costas respectivamente. Assim, esses resultados permitiram concluir que a ergonomia da estação de

trabalho é importante, bem como a necessidade de limitar o número de horas de trabalho ininterruptas no teclado para reduzir sintomas músculo-esqueléticos.

Pennathur, Sivasubramanian e Contreras (2003), investigaram os efeitos da idade e do gênero de mexico-americanos idosos no nível de dificuldade na realização de tarefas de manutenção da casa (preparação de alimentos, compras de mercearia, limpeza da casa e lavanderia), tarefas pessoais (vestir-se, banhar-se, arrumar-se), tarefas de transferência (subir e descer da cama, sentar-se e levantar-se de cadeiras, entrar e sair do banho, uso de escadas) e tarefas de gerenciamento (usar o telefone, acessar o e-mail, operar fechaduras). Um questionário foi administrado para 62 mexico-americanos idosos (31 homens e 31 mulheres), com idade variando entre 65 e 84 anos (idade média 74 anos com desvio padrão 6.2 anos). Os sujeitos da pesquisa quantificaram suas respostas em 1-quase impossível de realizar a tarefa, 2- possível com ajuda, 3- fácil e possível sem ajuda. Foi realizada uma regressão logística com a idade (variável contínua) e gênero como variáveis preditoras, e as respostas as questões como variáveis resposta categóricas ordinais. Os resultados mostraram que idade e gênero tem efeitos significativos em tarefas diárias envolvendo alcance considerável, giro e inclino.

Shuval e Donchin (2005) realizaram um estudo para examinar a relação entre fatores de risco ergonômicos e sintomas músculo-esqueletal da extremidade superior em trabalhadores que usam terminais de vídeo em uma companhia de alta tecnologia em Israel. A população do estudo foram 84 trabalhadores, compostos de programadores, gerentes, administradores e especialistas de *marketing*. Dados dos sintomas músculo-esqueletal, fatores individuais e organizacionais e estresse foram obtidos através de questionários, enquanto que os dados ergonômicos foram obtidos através de observação direta, através do método conhecido como RULA. A estatística analítica e descritiva foi realizada através dos testes de qui-quadrado, (que foi usado para comparar variáveis categóricas), ANOVA (que foi usada para comparar médias) e análise de regressão logística, com intervalo de confiança de 95%. A análise de regressão logística foi realizada usando o método de seleção de variável *backward* de modo a detectar tanto quanto possível, covariáveis independentemente associadas com os sintomas músculo-esqueletais. As variáveis que continuaram no modelo final foram analisadas novamente usando o método *Enter*. Estresse no trabalho entrou no modelo de modo a atender um importante conceito na literatura. A correlação entre as variáveis contínuas foram analisadas através do coeficiente de correlação de Person, com significância bi-caudal e $\alpha = 0,01$, usando SPSS© 11. Os resultados indicam a necessidade de implementar um programa de intervenção focando na postura dos braços/pulsos e levando em consideração

necessidades especiais de subgrupos: gênero, trabalhando 10h por dia, trabalhando 7,1 – 9h por dia com uma VDT, e funcionários experimentando desconforto nas estações de trabalho.

Subramanian, Silva e Coutinho (2007) utilizaram técnicas multivariadas, quando ao usar regressão linear e a técnica exploratória da análise de descriminante determinaram quais características termo-ambientais representam melhor a sensação térmica declarada por bancários. Constataram que a temperatura de bulbo seco foi a variável que representou melhor a sensação térmica, em situação de conforto, e encontraram a temperatura de 23,79 ^{o}C como sendo a ideal para que as atividades bancárias sejam exercidas com satisfação térmica.

Kahya (2008) estudou a respeito dos efeitos do desempenho no trabalho na efetividade. Para isso, 143 funcionários de uma companhia de médio porte participaram. Considerou-se que desempenho da tarefa e desempenho contextual como sendo duas dimensões distintas do comportamento no trabalho que contribuem independentemente com resultado. Assim, 7 itens para performance na tarefa, 12 itens para performance contextual e 3 itens de efetividade foram usados. A analise de regressão múltipla foi realizada com o intuito de verificar a relativa contribuição de cada variável e as dimensões da performance no trabalho para a predição da efetividade. Os resultados demonstraram que dois itens, “atenção a detalhes importantes” e “criatividade para resolver problemas” foram os mais eficazes para contribuir com a efetividade. Níveis de educação e experiência no trabalho tiveram menor efeito na efetividade.

Bellusci (1999) relata o uso de modelos de regressão logística para avaliar as respostas ao questionário do Índice de Capacidade para o Trabalho (ICT) de 807 servidores de uma instituição judiciária federal. Caracterizou as variáveis como: variável dependente Índice de Capacidade para o Trabalho – ICT; variáveis independentes idade, sexo, estado conjugal, escolaridade, tempo de serviço no tribunal, cargo que ocupa, função e local de trabalho (lotação). Como resultado, obteve que as mulheres com maior tempo de trabalho na instituição e com cargo de auxiliar operacional de serviços diversos têm maiores chances de apresentarem o índice baixo ou moderado.

McFADDEN (1997) construiu um modelo a partir dos dados de uma população de 70164 pilotos de aeronaves obtidos da administração federal de aviação norte-americana, onde 475 homens e 22 mulheres passaram por incidentes entre os anos de 1986 e 1992. De modo a controlar por idade, experiência (horas totais de vôo), exposição ao risco e empregador (linha aérea maior ou menor) simultaneamente, a autora construiu um modelo de incidentes por erro do piloto para homens usando regressão logística. A regressão indicou que

juventude, inexperiência e empregador linhas aéreas menores foram contribuintes independentes para aumentar o risco de incidentes por erro do piloto. Os resultados também dão suporte a literatura para dar conta de que a performance do piloto não difere significativamente entre homens e mulheres.

Horn e Salvendy (2009), realizaram dois estudos para refinar e validar um modelo previamente testado para medir a percepção do consumidor quanto à criatividade de um produto. Um estudo com amostra de n = 208 amostras, realizado pela web, avaliou cadeiras e lâmpadas, enquanto outro, com amostra de n = 105, feitos em papel, avaliou produtos individualmente selecionados. A Análise Fatorial Exploratória indicou três principais fatores: Afeto, Importância e Inovação, que respondia por 72% da variância comum. Resultados da Regressão indica que o fator Afeto significantemente prediz o desejo de aquisição de consumidores criativos (65% da variância explicada). Uma contribuição importante desse estudo foi descobrir que afeto (R^2 = 0,28) é igualmente influente a inovação (R^2 = 0,25) na percepção do consumidor da criatividade do produto.

Mathiassem e Åhsberg (1999) com o propósito primário de aumentar a base de dados para diretrizes ergonômicas, investigaram a resistência dos ombros a flexão isométrica em 20 homens e 20 mulheres saudáveis, com idade variando entre 20 e 55 anos, altura variando entre 1.53 m e 1.90 m e peso variando entre 48 kgs e 106 kgs. Os participantes foram instruídos a manter o braço dominante reto na posição horizontal em frente ao corpo até a exaustão. A análise de regressão mostrou que o tempo de resistência (T_{lim}) é significantemente correlacionado com o torque nos ombros com relação a capacidade máxima (%MVC), porém não correlacionados significantemente com o torque absoluto, gênero ou idade. A distribuição da idade, GTA (torque glenohumeral correspondente a gravidade na horizontal), %MVC e T_{lim} foram examinadas quanto à normalidade, através da curtose e simetria. Apenas a variável T_{lim} foi considerada crítica, o que foi resolvido com a aplicação do logaritmo natural à variável. Assim, ln (T_{lim}) passou a ser utilizado no modelo em substituição a T_{lim}. Numa primeira análise, as variáveis preditoras foram sendo adicionadas a um modelo de regressão múltipla, na seguinte ordem: (1) Gênero, (2) idade, gênero-idade, (3) GTA, gênero-GTA, idade-GTA, (4) %MVC, gênero - %MVC, idade-%MVC. As variáveis no passo (3) foram substituídas pelas variáveis do passo (4), uma vez que dados em GTA estão inclusos em %MVC. Após isso, numa segunda análise, foi utilizada a técnica de análise de discriminante *backward* de modo a determinar quais as variáveis melhor explicavam o ln(T_{lim}). No passo seguinte, as variáveis menos significantes de acordo com o teste-*F* foram

removidas (se P > 0.10). e um novo ajuste de regressão calculado. A habilidade de predição do modelo resultante da extração era comparado com a capacidade do modelo completo e aceito caso a diferença fosse considerada aceitável de acordo com a estatística F_p. Assim, provou-se que o modelo baseado apenas na variável %MVC como regressora não continha erros significativos quando comparados com o modelo completo. No entanto, o modelo explica apenas 30% da variação, sendo que os autores concluem que os outros aproximadamente 50% da variação pode ser explicada por fatores psicológicos, como tolerância a dor, motivação e humor.

Gaspary e colaboradores (2008) propuseram um modelo para explicar o ICT dos policiais rodoviários federais em função em termos do tempo de serviço na corporação, autonomia no trabalho e a possibilidade de promoção. O modelo encontrado através da técnica de Regressão Linear Múltipla foi significativo (F = 9,899; *p-value* = 0,000) com coeficiente de determinação (R^2) igual a 0,452, mostrando que 45,2% da variabilidade total do ICT pode ser explicada pelos fatores tempo de serviço, satisfação com autonomia no trabalho e possibilidade de promoção. Ainda encontrou-se que o acréscimo de uma unidade na escala de satisfação com a autonomia no trabalho aumentaria o ICT em 0,495 unidades; e ainda, o aumento de uma unidade na escala de satisfação com a possibilidade de promoção, acresceria 0,362 unidades ao ICT. Já um ano de serviço passado na corporação diminui 0,337 unidades no ICT.

3.9 PROCEDIMENTOS METODOLÓGICOS

Tratar-se á da caracterização dos dados e variáveis medidas, bem como sua caracterização e a caracterização de sua escala de mensuração conforme o quadro 2 apresentado anteriormente. Especial destaque merecerá a Variável Dependente, já que sua natureza justificará a técnica a ser utilizada para a modelagem, conforme o quadro 3 também apresentado anteriormente.

As Variáveis Independentes por sua vez, também serão analisadas sob o ponto de vista da possibilidade de se agruparem em um modelo ou de sua mútua exclusão, causada por colinearidade. Para tal serão tratadas duas a duas todo o conjunto de covariáveis medidas, verificando quais pares não podem coexistir no mesmo modelo.

Verificada as limitações acima, será iniciada a fase de geração/ajuste/seleção de modelos.

A geração de modelos se dará através da combinação das covariáveis candidatas a regressores; por sua vez, não serão testadas todas as possibilidades combinatória

advindas de todas as covariáveis presentes no subconjunto de variáveis candidatas, que como visto são de 2^k possibilidades, onde k é o número de covariáveis medidas, porém o método *stepwise backward* será utilizado para diminuir o número de possibilidades combinatórias.

O ajuste de modelos se dará através do método da máxima verossimilhança, usando o pacote estatístico livre R (R DEVELOPMENT CORE TEAM, 2009).

Casos nos quais todas as covariáveis presentes tenham sido aprovadas no modelo pelo teste de Wald, serão comparados usando o critério de informação de Akaike, de modo que o se possa equilibrar entre a complexidade do modelo e o ajuste do mesmo.

Após a fase de ajustar o modelo, será a vez de criticar o modelo. Para isso, será testado se o modelo usado é adequado aos dados, bem como a verificação do erro na predição do modelo, para verificar a qualidade do ajuste do modelo eleito.

Capítulo 4

4 RESULTADOS

4.1 VARIÁVEIS TRATADAS

Nesta seção busca-se descrever quais as variáveis serão tratadas neste estudo. Como o estudo que deu origem à coleta dos dados apresentados aqui foi um estudo exploratório, diversas variáveis de diversas naturezas foram coletadas. Todas são brevemente apresentadas aqui, mesmo que não se achem significativas no modelo final.

Sexo do enfermeiro, de natureza dicotômica, apresentada como SEXO no programa. Arbitrariamente para fins de representação no R, usou-se 0 para representar os participantes do sexo feminino e 1 para o sexo masculino. Na amostra tratada, a maioria são mulheres, o que é um reflexo de uma profissão historicamente exercida por mulheres. Aqui, apenas 14,89% dos respondentes são homens. A figura ? nos dá uma idéia de como a mediana do ICT em varia conforme o sexo do entrevistado, onde percebe-se que pela mediana do ICT dos entrevistados feminino que seu ICT teve um valor muito concentrado em 2, “BOM”, enquanto que o ICT dos entrevistados do sexo masculino variou mais, embora a mediana tenha como valor 3 “ÓTIMO”.

Idade do enfermeiro, de natureza discreta, já que foram computados apenas anos inteiros de vida do profissional participante. Apresentada ao programa como IDADE. A idade média dos participantes foi de 33 anos, com desvio padrão de 8,82 anos.

IMC, o índice de massa corpórea do participante, uma variável física do participante, medida em kg/m^2. Apresentada no programa como IMC. Essa variável é obtida através da divisão do peso do participante pela sua altura ao quadrado. Com isso é possível detectar se o mesmo está abaixo de seu peso ideal, no peso ideal, acima ou muito acima. Embora seja uma variável contínua, pode ser categorizada. Valores abaixo de 20,0 kg/m^2 demonstram estar-se abaixo do peso ideal; valores entre 20,1 e 25,0 kg/m^2 é a faixa considerada ideal; entre 25,1 e 30,0 kg/m^2 considera-se estar acima do peso ideal, enquanto que acima desse patamar é considerado obesidade mórbida. O IMC médio dos participantes foi de 25,20 kg/m^2, demonstrando que em média os participantes estão um pouco acima do peso. O desvio padrão da amostra foi de 4,36 kg/m^2.

Estado Conjugal, de natureza matemática categórica, uma das variáveis pessoais do profissional amostrado. Apresentada no programa como E_CIVIL. Para fins de codificação no programa, usou-se: 1- solteiro (a); 2- casado (a); 3- União estável; 4- Separado (a) /divorciado (a); 5- Viúvo (a). A maioria dos entrevistados é solteira 46,81%, seguidos de casados 36,17%, União estável 8,51% e Separados/Divorciados também com 8,51%. Para efeito de análise, considerou-se que o estado de União Estável e Casado são muito

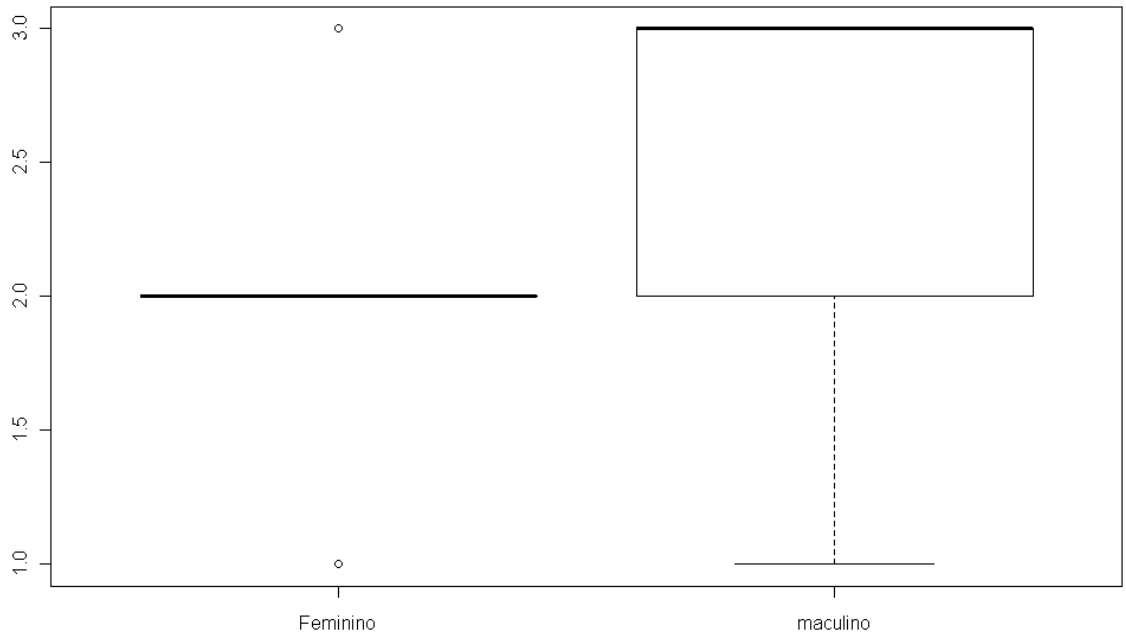

Figura 4: Boxplot do ICT em função do sexo

semelhantes, e portanto foram aglutinados numa única categoria, 3. A figura 5 mostra que o ICT dos solteiros teve mediana 2, “BOM”, porém teve alguns representantes na categoria 1, “REGULAR” e poucos na categoria 3, enquanto que os casados/união estável tiveram sua concentração maciça no valor 2, “BOM”, e os divorciados tiveram como mediana o valor 2, porém tiveram alguns representantes na categoria 3, “EXCELENTE”.

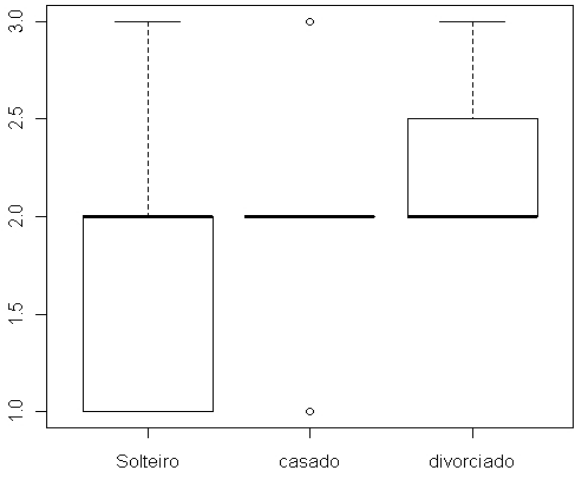

Figura 5: Boxplot do ICT em função do estado civil do entrevistado

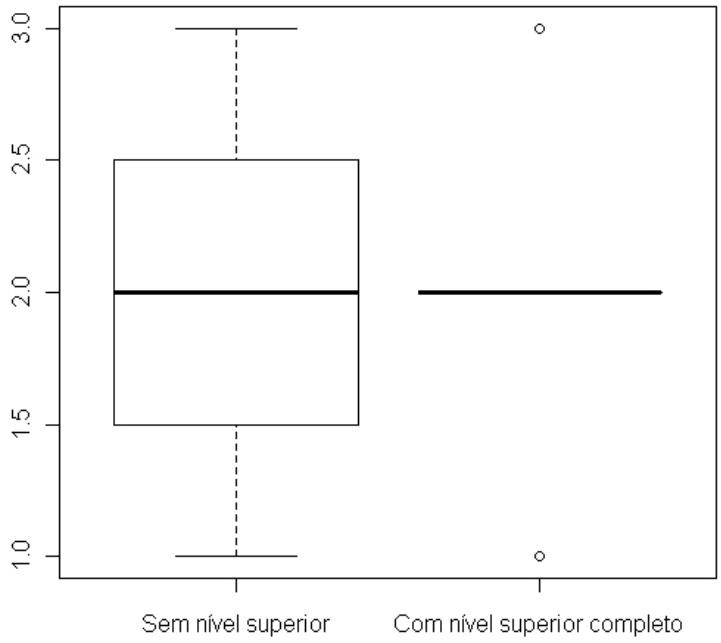

Figura 6: Boxplot do ICT em função da escolaridade do entrevistado

Escolaridade, de natureza categórica, uma variável que mede o preparo profissional do entrevistado. Foram considerados nove níveis na entrevista, a saber: 1- Ensino fundamental Incompleto; 2- Ensino fundamental completo; 3- Curso técnico de primeiro grau completo; 4 – Ensino médio incompleto; 5- Ensino médio completo; 6-Curso técnico de segundo grau completo; 7- Faculdade incompleta; 8- Faculdade completa; 9 – Pós-graduação completa/incompleta. Dos respondentes, 23,4% possuem pós graduação completa ou incompleta, 17,02% possuem curso superior completo, 8,51% possuem curso superior incompleto, 40,42% possuem curso técnico e 10,64% ensino médio completo. As demais categorias não foram contempladas na amostra devido ao requisito para investidura na profissão de enfermeiro. Apresenta-se essa variável no programa por ESC. Durante a analise, a aglutinação das categorias de nível superior completo, (os entrevistados que possuem nível superior completo combinados com os que além disso possuem pós graduação), em contrates com os que não possuem nível superior (nível médio, nível técnico ou nível superior incompleto), se mostrou uma medida necessária para que a variável pudesse ter significância no modelo. Assim, a categoria 1 representa os que possuem nível superior completo, contra o nível 0, os que não possuem nível superior completo. A figura 6 mostra como o ICT varia com a escolaridade. Percebe-se que há um maior espalhamento dos dados no fator “nível superior incompleto”em relação ao fator “nível superior completo”, embora as medianas possam graficamente ser percebidas como iguais.

Tempo de Serviço na função que exercia quando foi entrevistado, dado em anos ou em frações do mesmo, como meses, possui natureza contínua. Apresenta-se no programa como a variável TSERV. Os entrevistados em média possuíam 6,93 anos na função, ou seja, 6 anos e 11 meses aproximadamente. O desvio-padrão da amostra foi de 7,66 anos ou 7 anos e 8 meses aproximadamente.

Terceirização, contabiliza os funcionário presentes na amostra que são terceirizados. De natureza dicotômica (sim ou não). Apresenta-se no programa como a variável TER. Para fins de codificação da variável, usou-se 0 para funcionário não terceirizados e 1 para terceirizados. 23,40% da amostra é composta por terceirizados, sendo os outros 76,60% funcionários das instituições visitadas.

Carteira de trabalho contabiliza se o funcionário entrevistado possui registro em carteira de trabalho. Natureza dicotômica. 29,78% dos entrevistados dizem possuir o registro. Para representação neste estudo, usa-se a notação CAR para referir-se a essa variável. Foi codificada como 0 não possui registro e 1 possui registro na carteira de trabalho.

Insalubridade que identifica os entrevistados que recebem adicional de insalubridade ou periculosidade. Possui natureza dicotômica, sendo que 0 codifica não receber e 1 codifica receber. A variável é referida como INS. 63,83% dos entrevistados dizem não receber o beneficio.

A variável NOT representa se o entrevistado trabalho durante a noite ou não, possuindo assim, natureza dicotômica. 39,30% dos entrevistados não trabalham a noite.

A variável ATV representa há quanto tempo o profissional realiza a mesma tarefa no trabalho, sendo apresentados quatro opções para resposta: 1- menos de 6 meses; 2- entre 6 meses e um ano; 3- um a dois anos; 4- mais de dois anos. Assim, possui natureza categórica ordinal. 65,96% dos entrevistados disse realizar a mesma tarefa há mais de 2 anos, enquanto que 8,51% dos entrevistados disseram realizar a mesma tarefa entre 1 ano e 2 anos; 14,89% entre 6 meses e um ano e 10,64% disseram realizar a mesma tarefa havia menos de seis meses.

I-0 trata da idade inicial com que o profissional iniciou-se no mercado de trabalho, não necessariamente no cargo de enfermeiro. Possui natureza discreta (apenas anos inteiro foram contabilizados). A idade inicial no mercado de trabalho média dos enfermeiros amostrados foi 24,70 anos ou 24 anos e 8 meses, aproximadamente. O desvio-padrão da amostra foi de 7,60 anos ou 7 anos e 7 meses aproximadamente.

As variáveis seguintes são as variáveis ambientais. Expressam a opinião dos profissionais quanto ao ambiente de trabalho e medições realizadas nas UTI's em que trabalham.

ACE – registra a opinião do enfermeiro sobre o ambiente de trabalho. Contém a expressão do profissional se ele aceitaria mais do que rejeitaria o ambiente da UTI em que se encontrava. Assim, a variável possui natureza binária (Aceitaria-1 rejeitaria-0). 14,89% dos entrevistados rejeitaram mais que aceitaram o ambiente da UTI.

PER – variável que contém a percepção térmica dos entrevistados. Foi requisitado a cada participante que durante um plantão inteiro de doze horas registrasse sua percepção térmica, conforme o quadro abaixo:

Quadro 3: Escala sétima da sensação térmica

Valor	Percepção Térmica
+3	Com muito calor
+2	Com calor
+1	Levemente com calor
0	Neutro
-1	Levemente com frio

-2	Com frio
-3	Com muito Frio

Fonte: ASHRAE, 2004

Foi feita a média aritmética das respostas de cada respondente, obtendo-se sua sensação média ao longo de seu plantão de trabalho. A percepção média dos participantes foi de -0,76, o que se colocaria como em média, a sensação dos profissionais quanto a seu ambiente de trabalho entre neutro e levemente frio.

AVA é a variável designada para expressar a avaliação térmica do ambiente segundo os entrevistados. Nela estão contidas as avaliações feitas de hora em hora durante um plantão inteiro de doze horas de um enfermeiro entrevistado. A escala de avaliação usada para isso está mostrada abaixo. A variável AVA teve como média 0,66, ou seja, a maioria dos entrevistados considera o ambiente entre confortável e levemente inconfortável.

Quadro 4: Escala de avaliação térmica do ambiente

Valor	Avaliação
0	Confortável
1	Levemente Inconfortável
2	Inconfortável
3	Muito Inconfortável

Fonte: ASHRAE, 2004

A variável PREF contém a preferência térmica de cada entrevistado avaliada pelo próprio a cada hora de seu plantão de 12 horas. Em outras palavras, o entrevistado preenchia como queria termicamente estar se sentindo a cada hora. A escala de medição usada é mostrada no quadro abaixo.

Quadro 5: Escala de medição da preferência térmica

Valor	Preferência Térmica
+3	Bem mais aquecido
+2	Mais Aquecido
+1	Um pouco mais Aquecido
0	Como está
-1	Um pouco mais refrescado
-2	Mais refrescado
-3	Bem mais refrescado

Fonte: ASHRAE, 2004

TG contém a temperatura de globo medida na UTI em que trabalham os participantes. É uma variável contínua. A temperatura de globo média variou entre a mínima de 21,68 °C e a máxima de 24,61 °C.

TBS designa a temperatura de bulbo seco medida nas UTI's. Variável do tipo contínua. A média variou entre a mínima de 21,61 °C e a máxima de 24,84 °C.

TBU designa a temperatura de bulbo úmido medida nas UTI's. Também é uma variável contínua. Variou entre a mínima de 18,88 °C e a máxima de 19,56 °C.

RUI1 designa o ruído médio medido em dB no posto 1, que foi considerado o posto de enfermagem onde o profissional fica. Variou entre a mínima de 61,71 dB e a máxima de 73,46 dB. Variável contínua.

RUI2 designa o ruído médio medido em dB no posto 2, que foi considerado próximo ao leito do paciente. Variável contínua. Variou entre 60,58 dB e 75,17 dB.

ILM1 designa a iluminância média medida em luxes do posto 1, o posto de enfermagem. Variável contínua. A variável possui valores entre 91,78 luxes e 250,58 luxes. Variável contínua.

ILM2 designa a iluminância média medida em luxes do posto 2, próximo ao leito do paciente. Variável contínua. A variável possui valores entre 90,13 luxes e 211,49 luxes. Variável contínua.

4.2 CORRELAÇÃO ENTRE VARIÁVEIS

Uma das primeiras etapas na seleção de variáveis é a verificação da correlação entre VI's. A correlação de Pearson foi verificada para os pares de VI's contínuas. As seguintes variáveis se acharam altamente correlacionadas:

Tabela 3: Variáveis altamente correlacionadas

Variáveis	Coeficiente de correlação de Pearson
ILM1 e ILM2	0,74
RUI1 e RUI2	0,99
TBU e RUI2	-0,83
TBU e RUI1	-0,76
TBS e ILM2	-0,86
TBS e ILM1	-0,72
TG e ILM2	-0,89
TG e TBS	0,995

Para fins de avaliação, computou-se como coeficiente acima de 0,70 como altamente correlacionados. As demais variáveis ficaram bem abaixo desse patamar. Percebe-se que as variáveis ambientais medidas nas UTI's são altamente correlacionadas entre si.

Ruido e temperatura possuem uma dependência inversa, já que a refrigeração desses ambientes envolve o uso de condicionadores de ar.

O cálculo da correlação entre variáveis no R se dá através do comando cor (Variável1, Variável2).

4.3 MODELAGEM

A eleição de um modelo, conforme visto anteriormente envolve compromisso entre a complexidade do modelo, que nesse caso se observa pelo número de variáveis envolvidas, e pelo erro observado (no presente caso, deviance).

O método utilizado para a escolha desse modelo será o método *stepwise backward*. Assim, inicia-se a análise com o total dos regressores e após o ajuste do modelo, observa-se o índice AIC. Retira-se então uma variável, a que possuir o maior p-valor, e novamente ajusta-se o modelo. Caso o AIC observado na segunda situação for menor que o AIC observado no passo anterior, a variável que foi excluída deve permanecer excluída. Caso contrário, adiciona-se a variável de volta ao modelo e termina-se o processo. O processo continua até não haverem mais variáveis a serem excluídas ou quando o AIC não melhorar mais.

Após o processo de seleção de variáveis, as seguintes VI's se encontraram no modelo: SEXO, TSERV, INS, NOT, AVA, TBS, RUI1, ESC, E_CIVIL, ILM1.

Tabela 4: Coeficientes de regressão do modelo proposto

	Coeficiente	Erro padrão	Wald Z	P valor
SEXO	4,36	1,46	2,98	0,0029
TSERV	-0,22	0,07	-3,00	0,0027
INS	1,31	0,78	1,68	0,0930
NOT	-2,56	0,99	-2,58	0,0099
AVA	-1,43	0,63	-2,28	0,0228
TBS	2,31	0,80	2,86	0,0043
RUI1	0,32	0,12	2,66	0,0079
ILM1	0,04	0,02	2,44	0,0148
ESC	2,93	1,00	2,91	0,0036
E_CIVIL-3	1,35	0,77	1,74	0,0820
E_CIVIL-4	4,54	1,68	2,70	0,0069
Intercepto 2	-77,01	26,95	-2,86	0,0043

Intercepto 3	-81,20	27,36	-2,97	0,0030

O modelo então é:

$$\text{logit}[P(Y \geq 2)] = -77{,}01 + 4{,}36SEXO - 0{,}22TSERV + 1{,}31INS - 2{,}56NOT - 1{,}43AVA + 2{,}31TBS + 0{,}32RUI + 0{,}04ILM1 + 2{,}93ESC + \sum_{l=3}^{4}(1{,}35E\_CIVIL_3 + 4{,}54E\_CIVIL_4) \quad (83)$$

$$\text{logit}[P(Y \geq 3)] = -81{,}20 + 4{,}36SEXO - 0{,}22TSERV + 1{,}31INS - 2{,}56NOT - 1{,}43AVA + 2{,}31TBS + 0{,}32RUI + 0{,}04ILM1 + 2{,}93ESC + \sum_{l=3}^{4}(1{,}35E\_CIVIL_3 + 4{,}54E\_CIVIL_4) \quad (84)$$

Alternativamente, podemos escrever:

$$\hat{\pi}_2 = \frac{e^{-(-77{,}01+4{,}36SEXO-0{,}22TSERV+1{,}31INS-2{,}56NOT-1{,}43AVA+2{,}31TBS+0{,}32RUI1+0{,}04ILM1+2{,}93ESC+\sum_{l=3}^{4}(1{,}35E_{CIVIL_3}+4{,}54E\_CIVIL_4)}}{1 + e^{-(77{,}01+4{,}36SEXO-0{,}22TSERV+1{,}31INS-2{,}56NOT-1{,}43AVA+2{,}31TBS+0{,}32RUI1+0{,}04ILM1+2{,}93ESC+\sum_{l=3}^{4}(1{,}35E_{CIVIL_3}+4{,}54E\_CIVIL_4)}} \quad (85)$$

$$\hat{\pi}_{3=} \frac{e^{-(-81{,}20+4{,}36SEXO-0{,}22TSERV+1{,}31INS-2{,}56NOT-1{,}43AVA+2{,}31TBS+0{,}32RUI+0{,}04ILM1+2{,}93ESC+\sum_{l=3}^{4}(1{,}35E\_CIVIL_3+4{,}54E\_CIVIL_4)}}{1 + e^{(-81{,}20+4{,}36SEXO-0{,}22TSERV+1{,}31INS-2{,}56NOT-1{,}43AVA+2{,}31TBS+0{,}32RUI+0{,}04ILM1+2{,}93ESC+\sum_{l=3}^{4}(1{,}35E\_CIVIL_3+4{,}54E\_CIVIL_4)}} \quad (86)$$

Como não houve representantes na categoria ICT ruim, que seria a categoria 0, o modelo não contempla essa categoria. Já a categoria 1, ICT “moderado” foi usado como referência, conforme a teoria vista anteriormente.

O pseudo-R^2 obtido foi de 0,52, ou seja, 52% da variação do ICT pode ser explicada pelos regressores.

4.4 RAZÃO DAS CHANCES (“ODDS RATIO”)

Um dos aspectos interessantes na regressão logística em geral é saber como a mudança em um fator ou variável explicativa pode afetar a variável dependente. Esse papel é desempenhado observando-se a razão das chances ou “odds ratio”.

Tabela 5: Odds Ratio para as variáveis presentes no modelo

Variável	BETA	"Odds Ratio"
SEXO	4,36	78,26
TSERV	-0,22	0,8025
INS	1,31	3,70
NOT	-2,56	0,077
AVA	-1,43	0,2393
TBS	2,31	10,07
RUI1	0,32	1,377
ILM1	0,04	1,041
ESC	2,93	18,72
E_CIVIL-3	1,35	3,86
E_CIVIL-4	4,54	93,69

4.5 INTERPRETAÇÃO DOS COEFICIENTES

Por se tratar de um estudo exploratório, pode-se adquirir algum discernimento a cerca da matéria em estudo através da interpretação da "*Odds ratio*" do modelo proposto. É importante perceber que o modelo proposto trata de $Pr(Y \geq j)$, ou seja, a probabilidade de o ICT observado seja maior ou igual a uma categoria de interesse. Pode-se assim interpretar que as categorias inferiores a j se combinam numa única, e assim, a razão das chances se dá entre a chance de o evento recair sobre uma categoria inferior a j e o evento recair sobre uma categoria superior. Deve-se também perceber que no modelo proposto fez-se a suposição das odds proporcionais, conforme visto anteriormente no capítulo 3. Conclui-se disso, que as razões das chances obtidas irão confrontar o aumento da chance de o ICT estar recair numa categoria superior a atual ao se mudar o valor da variável em uma unidade.

A variável SEXO, sendo que ao passar do sexo feminino para o masculino, há um aumento na chance de o ICT sair da faixa "Moderado" (1) para a faixa "Bom" ou "Ótimo" da ordem de 78,41 vezes .

A variável Tempo de Serviço (TSERV) ao se passar uma ano na profissão de enfermeiro em UTI, o profissional tem 0,8025 de chance de passar de um ICT numa categoria inferior para uma categoria superior. Isso quer dizer que a cada ano, o profissional na ativa numa UTI tem 24,61% de chances de seu ICT cair numa faixa inferior.

A odds ratio da variável Insalubridade (INS) mostra que ao se passar da situação de não se receber o benefício para a situação de recebê-lo, a chance de o profissional ter seu ICT numa faixa superior à atual é de 3,71 vezes maior que se não o recebesse.

A variável NOT, que simboliza o trabalho noturno dos enfermeiros, mostra que ao se passar da situação de não trabalhar a noite para trabalhar a noite, a chance de o profissional ter seu ICT numa faixa superior à atual é de 0,077, ou seja, a chance de o ICT do profissional ir para uma faixa inferior à atual 12,99 vezes maior.

A avaliação térmica no modelo mostra que ao se distanciar na escala da situação de conforto térmico em direção ao desconforto térmico em uma unidade, a chance de o ICT do profissional cair é 4,17 vezes maior que a chance de aumentar.

A temperatura de bulbo seco, variável representante da temperatura ambiental da UTI, mostra que ao se aumentar um grau numa UTI, que em geral possui um ambiente frio, o ICT do enfermeiro tem a chance de aumentar em 10,07 vezes.

Ao se aumentar o ruído na posto de enfermagem em 1 dB, a chance de o enfermeiro ter seu ICT atual aumentado é de 37,71%.

Ao se aumentar a iluminação em 1 lux no posto de enfermagem, aumenta-se a chance de seu ICT atual subir de nível em 4,08%.

A escolaridade também influi no ICT dos enfermeiro, uma vezes que ao se passar da situação em que o profissional não possui nível superior completo para a situação em que o mesmo o possui, aumenta-se a chance de se ter um ICT numa categoria superior em 18,73 vezes.

Os profissionais casados ou com união estável têm 3,86 vezes mais chances de ter um ICT numa categoria superior aos profissionais solteiros, enquanto que os divorciados têm 93,69 vezes mais chances de ter um ICT numa categoria superior que um profissional solteiro.

4.6 CRÍTICAS AO MODELO

Com o modelo ajustado, o mesmo foi utilizado para tentar prever os mesmo dados que o geraram. Constatou-se que através do modelo acertou-se as previsões em 33 das 47 entradas de dados. Assim, em 70,21% dos pontos o modelo acerta. A tabela 6 mostra dado por dado onde o modelo acerta e erra e, graças à característica ordinal da VD, a diferença entre o predito e o medido pôde ser computada. Dos 14 pontos onde o modelo erra na

predição, 9 previram uma categoria inferior à categoria medida. Neste caso, em cerca de 64% das vezes em que o modelo erra, esse erro é pessimista.

A matriz de confusão, que mostra como o modelo se comporta em predizer os dados presentes é mostrado na tabela 7, em termos percentuais e em termos quantitativos na tabela 7.

Tabela 6: Previsão do modelo sobre os dados coletados

Dado	ICT medido pelo questionário	ICT previsto pelo modelo	Previsto é igual ao real?	Diferença
1	ÓTIMA	BOA	NÃO	-1
2	BOA	BOA	SIM	0
3	ÓTIMA	BOA	NÃO	-1
4	BOA	BOA	SIM	0
5	MODERADA	MODERADA	SIM	0
6	BOA	BOA	SIM	0
7	BOA	BOA	SIM	0
8	MODERADA	BOA	NÃO	+1
9	BOA	ÓTIMA	NÃO	+1
10	BOA	BOA	SIM	0
11	ÓTIMA	BOA	NÃO	-1
12	BOA	BOA	SIM	0
13	BOA	BOA	SIM	0
14	ÓTIMA	BOA	NÃO	-1
15	ÓTIMA	BOA	NÃO	-1
16	BOA	BOA	SIM	0
17	ÓTIMA	ÓTIMA	SIM	0
18	BOA	BOA	SIM	0
19	BOA	BOA	SIM	0
20	BOA	BOA	SIM	0
21	MODERADA	MODERADA	SIM	0
22	BOA	ÓTIMA	NÃO	+1
23	BOA	MODERADA	NÃO	-1
24	ÓTIMA	BOA	NÃO	-1
25	BOA	BOA	SIM	0
26	BOA	BOA	SIM	0
27	BOA	BOA	SIM	0
28	ÓTIMA	BOA	NÃO	-1
29	MODERADA	MODERADA	SIM	0
30	BOA	BOA	SIM	0
31	BOA	BOA	SIM	0
32	MODERADA	MODERADA	SIM	0
33	BOA	BOA	SIM	0
34	BOA	BOA	SIM	0
35	BOA	BOA	SIM	0
36	BOA	BOA	SIM	0
37	ÓTIMA	ÓTIMA	SIM	0
38	MODERADA	MODERADA	SIM	0
39	MODERADA	MODERADA	SIM	0
40	BOA	BOA	SIM	0
41	MODERADA	BOA	NÃO	+1
42	ÓTIMA	ÓTIMA	SIM	0
43	MODERADA	BOA	NÃO	+1
44	BOA	BOA	SIM	0
45	BOA	BOA	SIM	0
46	BOA	BOA	SIM	0
47	ÓTIMA	BOA	NÃO	-1

Tabela 7: Matriz de confusão do modelo em termos percentuais

	Predito		
Real	1	2	3
1	66,67	33,33	0,00
2	3,70	88,89	7,41
3	0,00	72,72	27,28

Tabela 8: Valores Preditos e Observados do modelo

OBSERVADO	PREDITO	QUANTIDADE
1	1	6
1	2	3
1	3	0
2	1	1
2	2	24
2	3	2
3	1	0
3	2	8
3	3	3

A matriz de confusão mostrada na tabela 7 mostra que o modelo é muito tendencioso para o valor ICT “Bom”, havendo uma tendência natural no modelo em predizer bem este valor, onde ele acerta em 88,89% dos casos. Em cerca 74,46% das saídas do modelo (35 das 47 saídas no presente caso), o modelo prever ICT “BOM”. Esse fato vem da amostra usada na construção do modelo, que possui tendência para esse valor.

O modelo mostra-se com certa precisão no valor 2, ICT “Bom”, que não é acompanhada nas demais faixas, onde o modelo se mostra errar mais. Isso se deve em parte à tendência natural dos dados apresentados, onde 57,44% dos dados pertenciam a categoria 2 “BOM”, 23,40% pertenciam a categoria 3 “EXCELENTE”, 19,14% dos dados pertenciam à categoria 1 “REGULAR” e nenhum exemplar da amostra pertencia à categoria 0 “RUIM”.

Na categoria 1 “REGULAR”, o modelo acerta em cerca de 67% dos pontos, enquanto que a categoria 3 “EXCELENTE”, o modelo acerta em apenas 27,28% dos casos.

Capítulo 5

5 Conclusões

O teste de adequação global do modelo apresenta p-valor de 0,0035. Isso implica que existe um modelo com os dados presente relacionando o ICT categorizado e as VI's apresentadas. No entanto, o valor do Pseudo-R^2 apresentado, 0,52 mostra que o ajuste ainda deixa a desejar. Os fatores que contribuem para essa baixa aderência do modelo aos dados são a amostra pequena (quarenta e sete respondentes) para muitas variáveis coletadas (vinte e três coletadas, dez variáveis independentes no modelo final) além de ser uma amostra tendenciosa para uma das categorias e ainda, por variáveis explicativas latentes ainda existentes, fatores de risco que influem no valor do ICT que não foram contemplados neste conjunto de dados. Uma amostra mais significativa e mais bem distribuída traria um melhor ajuste do modelo, além da inclusão de variáveis adicionais.

Apesar de tudo isso, o modelo atual proposto fornece certo discernimento acerca da matéria em questão, mostrando que fatores de conforto ambiental, associados a fatores pessoais interferem na Capacidade para o Trabalho de enfermeiros em UTI's públicas na cidade de João Pessoa.

No modelo apresentado, o fator de risco associado à variável Sexo não pode ser totalmente confiável, pois na amostra o número de representantes do sexo feminino é muito maior que o número de representantes do sexo oposto.

Apesar de ter o modelo se mostrar tendencioso a predizer o valor 2, "BOM", devido à tendência dos dados, essa situação seria ainda pior,caso se utilizasse a regressão logística binária, já que a aglutinação dos valores 2 e 3 traria uma variável dependente ainda mais viciada. Daí, conclui-se que o uso da ferramenta multinomial trouxe benefício para essa análise de dados, e uma contribuição para a ergonomia, podendo o seu uso trazer ainda mais benesses futuras.

Para trabalhos futuros, uma análise confirmatória com uma amostra maior, novas variáveis inclusas além das atuais presentes no modelo, a variável ICT mais bem distribuída e anonimato garantido com maior ênfase poderiam levar a um modelo preditivo mais preciso.

6 REFERÊNCIAS

ABREU, Mery Natali Silva; SIQUEIRA, Arminda Lucia e CAIAFFA, Waleska Teixeira. **Regressão logística ordinal em estudos epidemiológicos**. *Rev. Saúde Pública* [online]. 2009, vol.43, n.1, pp. 183-194. ISSN 0034-8910.

AGRESTI, Alan. **Categorical Data analysis.** Nova Jersey, 2 ed. 2002. Editora John Wiley and Sons.

_____. **An Introduction to Categorical Data Analysis.** Nova Jersey, 2 ed. 2007. Editora John Wiley and Sons.

ANANTH, C.V.; KLEINBAUM D.G. **Regression models for ordinal responses: a review of methods and applications.** International Journal of Epidemiology. 1997;26(6):1323-33. DOI: 10.1093/ije/26.6.1323

ANDERSON, J.A. **Regression and Ordered Categorical Variables**. Journal of Royal Statistical Society. Series B (Metodological), vol. 46, n 1, 1984, PP.1-30

ASHRAE STANDARD, **Thermal Envolvimental Conditions for Human Occupaney**. American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2004.

ASSOCIAÇÃO BRASILEIRA DE ERGONOMIA **O que é Ergonomia?** 2008. Disponível em <www.abergo.org.br> Acesso em 5 de fev de 2009

BELLUSCI, Silvia Meirelles, FISCHER, Frida Marina **Envelhecimento funcional e condições de trabalho em servidores forenses** Revista Saúde Pública [*on-line*] dezembro de 1999 p.602-9

CHAMBERS, John **Software for data analysis: Programming with R.** Ed. Springer, 2008.

CHENG, Hung-Yuan; CHANG, Yu-Ming **Extraction of product form features critical to determining consumers' perceptions of product image using a numerical definition-based systematic approach** International Journal of Industrial Ergonomics (2008), doi:10.1016/j.ergon.2008.04.

CHIANG, Hua-Cheng; LAI, Hsin-Hsi; CHANG, Yu-Ming **A measurement scale for evaluating the attractiveness of a passenger car form aimed at young consumers** International Journal of Industrial Ergonomics (2006), doi:10.1016/j.ergon.2006.09.014

CHIU, Min-Chi ; WANG, Mao-Jiun J; LU, Chih-Wei; PAN, Shung-Mei, KUMASHIRO, Masaharu e ILMARIEN, Juhani **Evaluating work ability and quality of life for clinical nurses in Taiwan** Revista Nursing Outlook Volume 55, Issue 6, Novembro-Dezembro de 2007, pp 318-326

DEMIDENKO, Eugene **Sample size determination for logistic regression revisited** Revista *Statistics in Medicine*. 2007; vol 26:pgs 3385–3397

DOPPLER, F. Trabalho e saúde. In: FALZON, P. **Ergonomia**. São Paulo: Blucher, 2007. p. 47-58

ESTORILIO, C.C.A. **O trabalho dos engenheiros em situações de projeto de produto:** uma análise de processo baseada na ergonomia**.** São Paulo, 2003. Tese (Doutorado em engenharia) Departamento de Engenharia de Produção, Escola Politécnica, Universidade de São Paulo.

EVANS, Owen; PATTERSON, Kim **Predictors of neck and shoulder pain in non-secretarial computer users.** Rev. International Journal of Industrial Ergonomics Vol. 26, 2000 p.357-365

FIGUEIRA, Cleonis Viater **Modelos de Regressão Logística.** Porto Alegre, 2006. Dissertação (mestrado em matemática) Programa de Pós-graduação em matemática do Instituto de Matemática Universidade Federal do Rio Grande do Sul.

FOGLEMAN, Maxwell; LEWIS, R.J. **Factors associated with self-reported musculoskeletal discomfort in video display terminal (VDT) users** International Journal of Industrial Ergonomics vol 29 (2002) p.311–318

FRIAS JÚNIOR, Carlos Alberto da Silva. **A saúde do trabalhador no Maranhão:** uma visão atual e proposta de atuação. Dissertação [Mestrado] Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública; 1999. 135 p.

GASPARY, L. T., SELAU, L. P. R., AMARAL, F. G. **Análise das condições de trabalho da polícia rodoviária Federal e sua influência na capacidade para trabalhar** Revista Gestão Industrial (*on-line*) 2008, v. 04, n. 02: p. 48-64, 2008

GRAYBILL, Franklin A.; IYER, Hariharan K. **Regression analysis:Concepts and Applications.** California, Duxbury Press 2006

GUÉRIN, F. *et al.* **Compreender o trabalho para Transformá-lo**. São Paulo: ABDR, 2001

GUITIÉRREZ, J.L.G.; JIMÉNEZ, B.M.; HERNÁNDEZ, E.G.; LÓPEZ, A.L. **Spanish version of the Swedish Occupational Fatigue Inventory (SOFI): Factorial replication, reliability and validity** International Journal of Industrial Ergonomics vol 35 (2005) pgs. 737–746

HENDRICK, Hal W. **Boa ergonomia é boa economia.** Tradução Stephania Padovani; Revisão Marcelo M. Soares. ABERGO, jan 2003 20 p.

HORN, Diaina; SALVENDY, **Gavriel Measuring consumer perception of product creativity: Impact on satisfaction and purchasability.** Human Factor and Ergonomics in Manufacturing, vol 19 Issue 3, pp. 223-240 fev 2009

HOSMER, D.W.; LEMESHOW, S. **Applied Logistic Regression** 2 ed. 2000. Editora John Wiley and Sons.

IIDA, I. **Ergonomia** – Projeto e Produção, 2 ed., São Paulo: Edgard Blucher Ltda, 2005.

ILMARIEN, J; TUOMI, K.; SEITSAMO, J. **New dimensions of work ability** International Congress Series Volume 1280, Junho de 2005, Pags 3-7

INTERNATIONAL ERGONOMICS ASSOCIATION, The **What is Ergonomics?** 2000 disponivel em <www.iea.cc>. Acesso em 05 de fev de 2009.

JOHNSON, Richard A. ; WICHERN, Dean W. **Applied multivariate statistical analysis** 3 ed. Nova Jersey, 1992 Prentice Hall, 3 edição 600 p.

MATHIASSEN, Svend Erik ; ÅHSBERG, Elizabeth **Prediction of shoulder flexion endurance from personal factors** International Journal of Industrial Ergonomics vol. 24 (1999) pgs 315-329

McFADDEM, K.L. **Predicting pilot error incidents of US airlines pilots using logistic regression.** Applied Ergonomics Vol 28. No. 3. p.209-212, 1997.

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística Aplicada e probabilidade para engenheiros.** 2 ed. Rio de Janeiro, 2003 LTC editora.463 p.

MOURE, M.L. **Utilização da análise ergonômica do trabalho para concepção e aplicação de uma metodologia para avaliação da exposição ao ruído em canteiros de obras.** São Paulo, 2000. Tese (Doutorado em engenharia) Departamento de Engenharia de Produção, Escola Politécnica, Universidade de São Paulo.

PENNATHUR, A.; SIVASUBRAMANIAN, S.; CONTRERAS, L.R. **Functional limitations in Mexican American elderly.** International Journal of Industrial Ergonomics vol.31 (2003) p.41–50

PEREIRA, Daniel Augusto de Moura. **Análise da capacidade de trabalho e das condições de conforto térmico e acústico às quais estão submetidos os professores de escolas públicas municipais de João Pessoa.** Joao Pessoa, 2009. Dissertação (Mestrado em engenharia) Programa de Pós Graduação em Engenharia de Produção, Centro de Tecnologia, Universidade Federal da Paraíba.

PEREIRA, Glaucia Guimarães **Avaliação da CAPES: Abordagem quantitativa multivariada dos programas de administração.** Dissertaçao (Mestrado em administração) Departamento de economia, administração e contabilidade da USP, 2005.

PREARO, Leandro Campi **Uso de técnicas estatísticas multivariadas em dissertações e teses sobre o comportamento do consumidor:** um estudo exploratório. São Paulo, 2008 100 p. Dissertação (Mestrado) Universidade de São Paulo.

R Development Core Team. **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org, 2009.

ROYSTON, Patrick; SAUERBREI, Willi **Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for continuous variables.** Ed John Wiley and Sons, 2008.

RYAN, Thomas P. **Modern Regression Methods.** 2ª Ed. John Wiley & Sons, 2009.

SHAN, Gongbing; BOHN, Christiane **Anthropometrical data and coefficients of regression related to gender and race** Applied Ergonomics vol 34, 2003 p.327–337

SHUVAL, Keren; DONCHIN, Milka **Prevalence of upper extremity musculoskeletal symptoms and ergonomic risk factors at a Hi-Tech company in Israel** International Journal of Industrial Ergonomics 35 (2005) 569–581

SILVA, Luiz Bueno da . **Análise da relação entre produtividade e conforto térmico: o caso dos digitadores do centro de processamento de dados e cobrança da Caixa Econômica Federal do Estado de Pernambuco.** Florianópolis, 2001. 84 p. Tese (Doutorado) Programa de Pós-graduação em Engenharia de Produção, Universidade de Santa Catarina.

_____. **Ergonomia.** Apostila da disciplina de Ergonomia. Universidade Federal da Paraíba, 2005.

_____; et al. **Application Of Nonlinear Models To Studies In The Ergonomics Area.** Brazilian Journal of Operations and Production Management, v. 4. Pp. 39-60, 2007

SLACK, N.; CHAMBERS, S.; JOHNSTON, R. **Administração da produção.** Tradução: Maria Teresa Corrêa de Oliveira, Fábio Alher; Revisão técnica Henrique Luiz Corrêa. – 2. Ed. – 7 reimp.-São Paulo, Atlas, 2007.

SOARES, M. M. **21 anos da ABERGO: A Ergonomia brasileira atinge a sua maioridade.** Anais do ABERGO 2004. XIII Congresso Brasileiro de Ergonomia, II Fórum Brasileiro de Ergonomia e I Congresso de Iniciação Científica em Ergonomia. Fortaleza, 29 de agosto a 2 de setembro de 2004.

SOUZA, E. C. **Análise de influência local no modelo de regressão logística.** Piracicaba, 2006. Dissertação (Mestrado em Agronomia) – Universidade de São Paulo, Escola Superior de Agricultura.

SUBRAMANIAN, A.; SILVA, L.B.; COUTINHO, A.S. **Aplicação de método e técnica multivariados para previsão de variáveis termoambientais e perceptivas**. Revista Produção v.17 n.1 p. 052-070, Jan./Abr. 2007

TATCHER, Andrew; GREYLING, Mike **Mental models of the Internet** International Journal of Industrial Ergonomics (1998), v 22 p. 299 – 305

TUOMI, K.; ILMARINE, J.; JAHKOLA, A.; KATAJARINNE, L.; TULKKI, A**. Índice de Capacidade para o Trabalho.** Tradução: Frida Marina Fischer. São Carlos: UFSCar, 2005.

VASCONCELOS, Ernesto dos Santos **Proposta metodológica para o monitoramento da qualidade do combustível gasolina comum tipo C do Estado do Ceará.** João Pessoa, 2006 180 f Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal da Paraíba, Centro de Tecnologia.

VENABLES, W.N.; et al. **An Introduction to R.** 1992

WEISBERG, Sanford. **Applied linear regression.** 2005 3ª Ed John Wiley and Sons, Inc ISBN 0-471-66379-4

APÊNDICES

A.1 - Rotina no software R utilizada na análise

```
#carregar as variáveis coletadas no estudo
ICT<-scan("ICTcat.dat")
SEXO<-scan("sexo.dat")
IDADE<-scan("idade.dat")
IMCc<-scan("imccat.dat")
IMC<-scan("imc.dat")
E_CIVIL<-scan("e_civil.dat")
ESC<-scan("esc.dat")
TSERV<-scan("tserv.dat")
TER<-scan("ter.dat")
CAR<-scan("car.dat")
INS<-scan("ins.dat")
NOT<-scan("not.dat")
ATV<-scan("atv.dat")
I0<-scan("i0.dat")
ACE<-scan("ace.dat")
PER<-scan("per.dat")
PRE<-scan("pre.dat")
AVA<-scan("ava.dat")
TG<-scan("tg.dat")A
TBS<-scan("tbs.dat")
TBU<-scan("tbu.dat")
RUI1<-scan("rui1.dat")
RUI2<-scan("rui2.dat")
ILM1<-scan("ilm1.dat")
ILM2<-scan("ilm2.dat")
#Carregar os pacotes utilizados na análise
library (Design)
library (MASS)
#Transformar a VI em um vetor contendo fatores
ICTf<- as.factor(ICT)
#
#modelo de regressão multinomial ordinal
#Passo 0: Todas as variáveis exceto as variáveis excluídas por correlação
Modelo<-polr
(ICTf~SEXO+E_CIVIL+ESC+TSERV+TER+CAR+INS+NOT+ATV+I0+ACE+PER+PRE+
AVA+TG+RUI1+ILM1)
Modelo1<-
lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+TER+CAR+INS+NOT+ATV+I0+ACE+PER+P
RE+AVA+TG+RUI1+ILM1)
Modelo
Modelo1
#Passo 1 - retirado TER
Modelo<-polr
(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+INS+NOT+ATV+I0+ACE+PER+PRE+AVA
+TG+RUI1+ILM1)
```

```
Modelo1<-
lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+INS+NOT+ATV+I0+ACE+PER+PRE+A
VA+TG+RUI1+ILM1)
Modelo
Modelo1
#Passo 2 - retirado INS
Modelo<-polr
(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+I0+ACE+PER+PRE+AVA+TG+
RUI1+ILM1)
Modelo1<-
lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+I0+ACE+PER+PRE+AVA+
TG+RUI1+ILM1)
Modelo
Modelo1
#Passo 3: Retirado I-0
Modelo<-polr
(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+ACE+PER+PRE+AVA+TG+RU
I1+ILM1)
Modelo1<-
lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+ACE+PER+PRE+AVA+TG+
RUI1+ILM1)
Modelo
Modelo1
#Passo 4: Retirado PRE
Modelo<-polr
(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+ACE+PER+AVA+TG+RUI1+IL
M1)
Modelo1<-
lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+ACE+PER+AVA+TG+RUI1
+ILM1)
Modelo
Modelo1
#Passo 5: Retirado ACE
Modelo<-polr
(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+PER+AVA+TG+RUI1+ILM1)
Modelo1<-
lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+ATV+PER+AVA+TG+RUI1+ILM
1)
Modelo
Modelo1
#Passo 6: Retirar ATV
Modelo<-polr
(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+PER+AVA+TG+RUI1+ILM1)
Modelo1<-
lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+PER+AVA+TG+RUI1+ILM1)
Modelo
Modelo1
#Passo 7: Retirar AVA
Modelo<-polr (ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+PER+TG+RUI1+ILM1)
Modelo1<-lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+PER+TG+RUI1+ILM1)
```

```
Modelo
Modelo1
#Passo 8: Retirar ILM1
Modelo<-polr (ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+PER+TBS+RUI1)
Modelo1<-lrm(ICTf~SEXO+E_CIVIL+ESC+TSERV+CAR+NOT+PER+TBS+RUI1)
#Passo 9: Retirar ESC
Modelo<-polr (ICTf~SEXO+E_CIVIL+TSERV+CAR+NOT+PER+TBS+RUI1)
Modelo1<-lrm(ICTf~SEXO+E_CIVIL+TSERV+CAR+NOT+PER+TBS+RUI1)
```