

ESCOLA POLITÉCNICA DA UNIVERSIDADE
DE SÃO PAULO

DEPARTAMENTO DE ENGENHARIA DE
PRODUÇÃO



TRABALHO DE FORMATURA

USO DE REGRESSÃO LOGÍSTICA COMO
FERRAMENTA DE DECISÃO

ALUNO: GUILHERME ROCHLITZ QUINTÃO

ORIENTADORA: PROF. LINDA LEE HO

São Paulo
1998

1998
246m

Aos meus pais, por tudo.

AGRADECIMENTOS:

Prof. Linda Lee Ho, pela paciência e incentivo.

Endoclínica de São Paulo, tema de tantos trabalhos.

Marcos Rochlitz, por toda a ajuda nestes anos de Faculdade.

Paulo Godoy e Engrecon, pela colaboração.

Todas as pessoas que contribuíram para este trabalho.

USO DE REGRESSÃO LOGÍSTICA COMO FERRAMENTA DE DECISÃO

ÍNDICE

RESUMO	8
1. ANÁLISE DE DECISÃO	10
1.1 INTRODUÇÃO	10
1.2 SITUAÇÕES DE DECISÃO: PROGRAMADAS E NÃO-PROGRAMADAS	12
1.2.1 ABORDAGEM DE SITUAÇÕES PROGRAMADAS	16
1.2.2 ABORDAGEM DE SITUAÇÕES NÃO-PROGRAMADAS	17
1.3 CRITÉRIOS DE DECISÃO	23
1.3.1 CRITÉRIO DO MAXIMIN	26
1.3.2 CRITÉRIO DO MENOR ARREPENDIMENTO	28
1.3.3 CRITÉRIO DO VALOR MONETÁRIO ESPERADO	30
1.4 FUNÇÃO UTILIDADE	32
1.5 ÁRVORES DE DECISÃO	36
1.6 MODELOS ESTATÍSTICOS	39
2. CONCESSÃO DE CRÉDITO	50

2.1 CLASSIFICAÇÃO DE CRÉDITO	51
2.2 SISTEMAS DE CLASSIFICAÇÃO DE CRÉDITO	52
2.2.1 FUNCIONAMENTO	53
2.2.2 DESEMPENHO	56
2.3 EVOLUÇÃO DOS MÉTODOS DE ESCORAGEM DE CRÉDITO	58
3. MODELOS ESTATÍSTICOS	61
3.1 VARIÁVEIS DE PREDIÇÃO	61
3.2 VARIÁVEIS "DUMMY"	63
3.3 ANÁLISE DISCRIMINANTE	64
3.4 REGRESSÃO LOGÍSTICA	69
3.5 MÉTODOS DE PROGRAMAÇÃO MATEMÁTICA	72
3.6 COMPARAÇÃO ENTRE OS MÉTODOS	75
3.6.1 MÉTODOS DE PROGRAMAÇÃO MATEMÁTICA	76
3.6.2 ANÁLISE DISCRIMINANTE X REGRESSÃO LOGÍSTICA	76
3.6.3 ESCOLHA DO MÉTODO MAIS APROPRIADO	81
4. REGRESSÃO LOGÍSTICA	83
4.1 REGRESSÃO LOGÍSTICA COM UMA VARIÁVEL INDEPENDENTE	83
4.1.1 TESTE DE SIGNIFICÂNCIA DOS COEFICIENTES	87
4.2 REGRESSÃO LOGÍSTICA COM DIVERSAS VARIÁVEIS INDEPENDENTES	88
5. APLICAÇÃO DO MODELO	90
5.1 CONJUNTO DE DADOS	92
5.2 ANÁLISE DESCRITIVA DOS DADOS	93
6. ESTRATÉGIAS PARA OBTENÇÃO DO MODELO	107

6.1 MÉTODO "STEPWISE" _____	108
6.2 INTERPRETAÇÃO DOS DADOS _____	117
7. CONCLUSÃO _____	119
8. BIBLIOGRAFIA _____	121
9. ANEXO A _____	124
9.1 CONJUNTO DE DADOS _____	124
9.2 CONFEÇÃO DOS GRÁFICOS _____	125
9.3 EXEMPLAR DE FICHA CADASTRAL _____	126
10. ANEXO B _____	127
10.1 ANÁLISE ESTATÍSTICA DO CONJUNTO DE DADOS _____	127
10.2 PARTE DO CONJUNTO DE DADOS JÁ CODIFICADA _____	128
10.3 REGRESSÃO LOGÍSTICA: PASSO 1 _____	129
10.4 REGRESSÃO LOGÍSTICA: PASSO 2 _____	130
10.5 REGRESSÃO LOGÍSTICA: PASSO 3 _____	131
10.6 REGRESSÃO LOGÍSTICA: MODELO FINAL _____	132
10.7 CONJUNTO DE VALIDAÇÃO _____	133

LISTA DE TABELAS

Tabela 1.1: Abrangências das situações de decisão em uma organização

Tabela 1.2: Estimativas do número de exposições em campanha publicitária

Tabela 1.3: Critério do Maximin: exposições em campanha publicitária

Tabela 1.4: Critério do Menor Arrependimento: exposições em campanha publicitária

Tabela 1.5: Probabilidades de ocorrência de condições do tempo

Tabela 1.6: Resultados obtidos por ocorrência ou não de incêndio

Tabela 1.7: Resultados de pesquisa para avaliação de um novo consumidor

Tabela 3.1: Codificação através de variáveis “dummy”

Tabela 3.2: Dados de renda anual e tamanho do lote

Tabela 3.3: Variáveis “dummy” no comportamento da variável raça

Tabela 3.4: Classificações por Regressão Logística e Análise Discriminante

Tabela 3.5: Classificação de Regressão Logística e de Análise Discriminante

Tabela 6.1: Codificação da variável sexo em variável "dummy"

Tabela 6.2: Codificação da variável localização em variável "dummy"

Tabela 6.3: Codificação da variável profissão em variável "dummy"

Tabela 6.4: Codificação da variável renda em variável "dummy"

Tabela 6.5: Codificação da variável idade em variável "dummy"

Tabela 6.6: Primeiro passo do método "Stepwise"

Tabela 6.8: Segundo passo do método "Stepwise"

Tabela 6.8: Terceiro passo do método "Stepwise"

Tabela 6.9: Dados de renda e faixa etária do conjunto de dados

LISTA DE FIGURAS

Figura 1.1: Abordagem de situações de decisão não-programadas

Figura 1.2: Valor Monetário Esperado no caso Pennzoil-Getty

Figura 1.3: Classificação de elementos em dois grupos distintos

Figura 1.4: Análise de dez potenciais compradores de um novo produto

Figura 1.5: Função discriminante de elementos em dois grupos distintos

Figura 2.1: Modelo de decisão

Figura 2.2: Processo de concessão de crédito em loja de departamentos

Figura 3.1: Renda anual e tamanho do terreno

Figura 5.1: Árvore de decisão do problema da empresa X

Figura 5.2: Porcentagem de compra dos elementos do conjunto de dados

Figura 5.3: Porcentagem dos elementos do conjunto entre os sexos

Figura 5.4: Porcentagem de compra no sexo masculino

Figura 5.5: Porcentagem de compra no sexo feminino

Figura 5.6: Histograma de Renda dos indivíduos do conjunto de dados

Figura 5.7: Porcentagem de compra dos indivíduos por faixa de renda

Figura 5.8: Histograma de renda dos indivíduos com renda até R\$10.000,00

Figura 5.9: Porcentagem de compra dos indivíduos com renda até R\$10.000,00

Figura 5.10: Porcentagem dos elementos por qualificação da profissão

Figura 5.11: Histograma de idade dos indivíduos

Figura 5.12: Porcentagem de compra por faixa etária

Figura 5.13: Diagrama de localização regional dos indivíduos

Figura 5.14: Porcentagem de compra por localização regional

RESUMO

O presente trabalho discute a aplicação de um modelo estatístico - a Regressão Logística - no auxílio da tomada de decisão.

Inicialmente, são discutidos os três tipos de situação de decisão existentes: programadas, parcialmente programadas e não-programadas. Este último tipo corresponde à situações das quais se tem informação limitada. Para o auxílio da tomada de decisão nestas situações, são apresentados os principais critérios de decisão utilizados (Critérios do Maximin, Minimax e Valor Monetário Esperado) e sugerida a utilização de diagramas de decisão e modelos estatísticos.

Posteriormente, é feita uma análise generalizada dos modelos estatísticos: seus conceitos, formas e exemplos de utilização. Para a adequada compreensão de seu funcionamento, é feito um resumo sobre a análise de crédito que é um campo onde se observou um grande desenvolvimento da aplicação destes modelos em situações de decisão.

Segue-se a introdução dos principais modelos utilizados em problemas na área de crédito e semelhantes: Análise Discriminante, Regressão Logística e Métodos de Programação Matemática. Após a comparação dos mesmos em termos de princípios de funcionamento e adequação à resolução de problemas de tomada de decisão, escolhe-se o modelo de Regressão Logística como tema central do trabalho.

Após uma detalhada observação de suas características, o modelo é aplicado a um problema real: a determinação de potenciais compradores do seguro de vida fornecido por uma empresa administradora de cartões de crédito. É feita uma análise descritiva do conjunto de dados fornecido pela empresa e, com base nesta análise e em estratégias específicas de construção do modelo, chega-se a um modelo final.

Por último, os resultados obtidos com o modelo são interpretados à luz da análise de decisão e é discutida a relevância do tema na Engenharia de Produção da atualidade.

1. ANÁLISE DE DECISÃO

1.1 INTRODUÇÃO

Diariamente, as organizações modernas se deparam com uma variedade de situações onde necessita-se de uma tomada de decisão. Lidar com estas decisões consome grande parte do trabalho dos indivíduos em qualquer nível de uma organização. Algumas situações de decisão encontradas ocorrem com grande frequência e exatamente da mesma forma em cada ocasião. Estas decisões podem ser tomadas eficientemente, estabelecendo-se previamente regras e padrões de comportamento para solucioná-las baseados principalmente em experiências passadas.

Entretanto, um problema pode apresentar uma característica em particular que pode torná-lo diferente em algum ponto importante. Dependendo da situação, outros problemas podem ser novos ou diferentes em todos os aspectos. Neste caso, a intuição e o julgamento das pessoas da organização que já tiveram experiências passadas com problemas similares é o recurso mais eficaz para abordá-los.

Segundo Radford (1981), a essência de uma tomada de decisão está na formulação de alternativas de ação para atender à situação considerada e na escolha de uma alternativa após uma avaliação de sua eficiência no alcance dos objetivos. Um dos componentes mais importantes em um processo decisório é a coleta de informações, de onde pode ser feita uma avaliação da situação de decisão. Se existe informação suficiente para se obter especificações completas de todas as alternativas possíveis e sua eficiência na situação considerada, o processo decisório é relativamente simples.

Por outro lado, muitas vezes a obtenção de toda a informação necessária torna-se impossível na prática, devido às limitações de tempo ou recursos para tal. Além disso, muitas vezes pode ser impossível se determinar o que constitui toda a informação relevante ao problema, mesmo com tempo e recursos disponíveis para coletá-la. Em consequência, o elemento da incerteza fará parte da tomada de decisão - devido à falta de informação completa - levando o indivíduo a ter dúvidas em relação a qual alternativa escolher e à eficiência da mesma.

Neste trabalho serão apresentados alguns modelos estatísticos empregados no auxílio da tomada de decisão, dando maior ênfase ao modelo de Regressão Logística.

Neste capítulo, discutiremos os diversos tipos de situação de decisão existentes: situações programadas, parcialmente programadas e não-programadas. Atenção especial é dada à última, onde não se dispõe de rotinas para sua solução. Serão vistos os diversos critérios de decisão utilizados neste tipo de situação, entre eles as árvores de decisão e os modelos estatísticos.

Uma das áreas onde se desenvolveram modelos estatísticos em situações de decisão é a análise de crédito. No Capítulo 2, será feita uma revisão sobre mesma. Isto será necessário para que se compreenda os métodos utilizados ao longo do trabalho. Veremos como é feita a concessão de crédito a pessoas físicas, a evolução dos métodos de classificação de candidato a empréstimo em alto e baixo risco e os métodos mais utilizados.

No Capítulo 3, estão descritos os modelos estatísticos existentes em situações de decisão e suas principais características. Serão feitas comparações de desempenho e adequação às situações práticas, justificando a escolha do modelo de Regressão Logística.

No Capítulo 4, veremos o modelo de Regressão Logística em detalhes: seus conceitos, evolução, como é aplicado e sua formulação. Primeiramente, veremos o

modelo com uma única variável e, posteriormente, com diversas variáveis. Veremos também os critérios de escolha de variáveis para a construção do modelo.

No Capítulo 5, está apresentado um exemplo de aplicação de Regressão Logística: a classificação dos clientes de uma empresa administradora de cartões de crédito em potenciais compradores ou não compradores de seu Seguro de Vida. Faremos uma análise descritiva do conjunto de dados fornecido pela empresa, visando uma maior familiaridade com o conjunto.

No Capítulo 6, procederemos à construção do modelo final empregando Regressão Logística, descrevendo as estratégias utilizadas. O modelo será testado para a verificação de sua eficiência na resolução do problema proposto.

Por último, no Capítulo 7, estão as conclusões sobre o modelo e sua utilidade em situações de tomada de decisão.

1.2 SITUAÇÕES DE DECISÃO: PROGRAMADAS E NÃO-PROGRAMADAS

Radford (1981) propõe uma distinção entre dois tipos distintos de situações de decisão: as situações programadas e as não-programadas.

As decisões programadas são aquelas que são tomadas freqüentemente e se tornam rotineiras. Estas decisões são "programáveis" porque é possível se estabelecer um procedimento específico para abordá-las. Elas geralmente ocorrem em atividades operacionais e administrativas, principalmente nos níveis mais baixos de uma organização. Nos últimos anos, desenvolveram-se rotinas computadorizadas nestes níveis capazes de tomar decisões quanto, por exemplo, ao uso de recursos, à distribuição de produtos, ao controle de estoques, entre outras situações.

Em contrapartida, as decisões não-programadas não são repetitivas, apresentando sempre um aspecto novo e diferente. Deste modo, não se consegue estabelecer uma rotina de abordagem. Em comparação com as decisões programadas, a informação disponível é normalmente incompleta ou indefinida. Este tipo de situação é encontrada principalmente nos níveis mais altos de uma organização e geralmente são relacionados às atividades de planejamento estratégico.

Um exemplo de situação de decisão não-programada: Du Pont e clorofluorcarbonos

Clemen (1990) ilustra uma situação real de decisão não-programada com um exemplo ocorrido na empresa fabricante de produtos químicos Du Pont Chemical Corporation: Clorofluorcarbonos (CFCs) são substâncias químicas utilizadas como gases refrigerantes em aparelhos de ar condicionado, gases propelentes em "sprays" aerossóis e inúmeros outros produtos. A partir da década de 70, evidências científicas têm apontado que o CFC lançado na atmosfera pode destruir moléculas de ozônio na camada de ozônio que se encontra a cerca de 20 km da superfície terrestre cuja função é proteger a terra da radiação ultravioleta do sol. Encontrou-se um grande buraco na camada de ozônio sobre a Antártida e um estudo realizado em 1988 concluiu que a camada de ozônio sobre o Hemisfério Norte diminuiu 3% desde 1969 - cujas causas foram atribuídas aos CFCs. Além disso, provou-se que a camada de ozônio, uma vez destruída, não se recompõe e a continuação da destruição poderá levar a problemas na agricultura, afetar a ecologia marinha e possivelmente mudar radicalmente os padrões ambientais do globo. Os ecologistas estimam que aproximadamente 30% do CFC lançado na atmosfera provém de aerossóis. Em 1978, o governo norte-americano baniu seu uso como propelente, mas muitos países ainda toleram sua utilização. O mercado mundial de CFC é de cerca de U\$ 25 bilhões ao ano, sendo a Du Pont Chemical Corporation responsável por 25% do total. No início de 1988, a empresa anunciou que iria cessar gradativamente sua produção de CFC e que iria investir no desenvolvimento de produtos substitutivos.

Atualmente, afirma que já tem um substituto para aparelhos de ar condicionado para automóveis, embora a nova substância seja muito cara.

A direção da Du Pont encarregada de tomar a decisão sobre a continuação da produção de CFC enfrentou as seguintes dúvidas:

- Que fatores levar em conta na decisão?
 - Quais são as principais fontes de incertezas que serão encontradas?
 - Quais os objetivos corporativos a serem considerados?
 - É possível que os objetivos corporativos da empresa e a maneira pela qual ela vê o problema tenham evoluído desde a metade da década de 70 quando os CFCs começaram a se tornar uma preocupação?
-

Apesar da distinção entre os dois tipos de situação de decisão, na realidade estes são apenas os extremos de um espectro contínuo de situações, onde existem situações parcialmente programadas. Nestas situações, apenas parte pode ser bem estruturada ou completamente especificada. Normalmente, uma grande parte da informação ou um grande número de decisões complementares podem ser abordadas por uma rotina bem programada. Entretanto, a escolha final entre as alternativas não é bem estruturada, sendo tomada, geralmente, por algum indivíduo encarregado para a tarefa. Isto pode ser ilustrado no seguinte exemplo: em processos seletivos de admissão de novos funcionários em organizações, as decisões de escolha de candidatos para futuras entrevistas geralmente são feitas por funcionários do Departamento de Recursos Humanos. Estas decisões são baseadas nas especificações do cargo disponível e nas características desejadas do candidato. Os candidatos são selecionados por outro funcionário de um nível mais alto na organização, geralmente um administrador, que

fica livre da tarefa rotineira de verificação de todos os currículos recebidos mas que fará a decisão final sobre a escolha do candidato.

O espectro das situações de decisão e suas características está resumido na Tabela 1.1:

Tabela 1.1: Abrangências das situações de decisão em uma organização. Fonte: Radford (1981)

DECISÕES		
PROGRAMADAS	PARCIALMENTE PROGRAMADAS	NÃO-PROGRAMADAS
1. Especificações do procedimento de decisão são decididas antes da tomada de decisão	1. Apenas parte do processo decisório pode ser completamente especificada e programada	1. O processo decisório não pode ser completamente estruturado antes da tomada de decisão
2. Pequena intervenção administrativa durante a decisão	2. O administrador toma a decisão final se baseando em decisões já tomadas e em seu julgamento pessoal	2. Cada indivíduo resolve a situação de acordo com sua experiência e julgamento pessoal
3. Resoluções repetidas com a mesma informação produzem os mesmos resultados	3. Dois administradores podem concordar em alguns pontos importantes, mas discordar da solução final	3. Dois administradores diferentes podem chegar a diferentes conclusões

Radford (1981) e Clemen (1990) apresentam procedimentos para a abordagem das diferentes situações. Nas seções 1.2 e 1.3 estão descritos os procedimentos referentes a situações programadas e não-programadas respectivamente.

1.2.1 ABORDAGEM DE SITUAÇÕES PROGRAMADAS

Segundo Radford (1981), as etapas a serem adotadas na abordagem de uma situação de decisão "programável" seriam:

Inteligência: procura das condições dos ambientes externo e interno ao processo decisório e coleta das informações relativas a estas condições:

- a) *Percepção e formulação da situação:* identificação da situação de decisão e a definição de suas características principais;
- b) *Construção de um modelo da situação:* o modelo é um instrumento para se estimar os possíveis resultados de uma situação de decisão sobre uma variedade de condições possíveis;
- c) *Determinação de tantas medidas de custos e benefícios forem necessárias para a situação considerada:* sistemas uniformes de medida facilitam comparações entre as diferentes alternativas de resolução do problema;

Modelagem: as alternativas de ação disponíveis são determinadas e analisadas como soluções possíveis para o problema decisório detectado:

- a) *Especificação de alternativas disponíveis:* identificação e formulação clara dos possíveis caminhos a serem seguidos;

Escolha: uma das alternativas de ação é escolhida para ser implementada com base na avaliação de sua eficiência relativa ao alcance dos objetivos:

- a) *Avaliação dos custos e benefícios da alternativa de ação disponível:* medição do efeito da implementação de cada alternativa de ação em termos de custos e benefícios;
- b) *Determinação de um critério de escolha entre as diversas alternativas de ação:* especificação de regras cujos resultados são relativos à escolha desejada;
- c) *Resolução da situação de decisão:* a escolha é feita entre as diversas alternativas de ação disponíveis com base em um critério aceitável.

Estes passos não precisam ser seguidos na ordem discutida. Além disso, na prática, o procedimento poderá consistir de uma série de repetições de todos ou alguns destes passos até que a situação seja resolvida.

É possível se criar um roteiro de resolução que será aplicável a todas as situações possíveis do problema, uma vez que estas são conhecidas. Deste modo, o modelo estará livre de incertezas, não necessitando de intervenção.

1.2.2 ABORDAGEM DE SITUAÇÕES NÃO-PROGRAMADAS

Segundo Radford (1981), as situações “não-programáveis” são aquelas onde se encontra pelo menos uma das seguintes características:

Falta de informações completas sobre a situação de decisão e seu ambiente

A falta de informações completas leva à incerteza entre os indivíduos envolvidos na sua resolução quanto às suas características.

Neste caso, é impossível construir um modelo compreensível da situação que inclua todos os fatores relevantes e as relações entre eles com informação incompleta. Com um modelo incompleto da situação, pode-se crer que alguns dos aspectos mais importantes da situação estejam acima da compreensão imediata. Neste caso, necessita-se de esforço e tempo razoáveis para se identificar a natureza da situação considerada.

Quando se tem informação limitada, os diversos indivíduos envolvidos na resolução do problema podem formar opiniões pessoais divergentes quanto às características do mesmo. A percepção de cada indivíduo sobre a situação é baseada na informação disponível vista sob a ótica de experiências passadas e aspectos relacionados à competência, motivação e julgamento destes.

E a total falta de informação leva à incerteza sobre as condições exatas do ambiente da situação de decisão. Neste caso, geralmente existe um período de tempo relativamente longo entre a escolha e a implantação de uma alternativa de ação, possibilitando que as condições do ambiente da decisão sofram alterações significativas tornando-o diferente do que previsto anteriormente. Portanto, quanto maior o tempo de implantação, maior é o risco de que o resultado de uma determinada alternativa de ação seja diferente do que o estimado durante o processo decisório.

Falta de medidas quantitativas para descrever claramente os custos e benefícios das alternativas disponíveis

A escolha entre diversas alternativas de ação é consideravelmente mais fácil quando a comparação entre elas pode ser feita em termos de custos e benefícios bem compreendidos. Esta condição permite que sejam comparadas em termos compatíveis entre si. Entretanto, em muitas situações é muito difícil a criação de medidas puramente quantitativas para se medir os custos e benefícios das diversas alternativas. Por exemplo, existem problemas para se criar uma medida da imagem de uma empresa

no mercado através das diversas maneiras pelas quais ela participa em programas sociais.

Apesar disso, segundo Radford (1981), nestes casos, freqüentemente os indivíduos se sentem tentados a criar medidas quantitativas dos custos e benefícios, mesmo que estes não reflitam claramente a relação entre os fatores envolvidos e os objetivos da situação de decisão. Isto pode causar distorções significativas na situação de decisão, uma vez que as conclusões tiradas de uma análise distorcida da situação muitas vezes não correspondem à realidade. Infelizmente, estas conclusões em geral têm um aspecto de validade, já que são produtos de um pensamento aparentemente lógico.

Quando diversos indivíduos têm diferentes opiniões sobre uma situação de decisão e suas características, eles também podem ter diferentes medidas pessoais sobre custos e benefícios de uma determinada alternativa de ação dificilmente comparáveis.

Existência de objetivos múltiplos como parte da organização, grupo ou indivíduo em questão

Muitas organizações, grupos ou indivíduos têm mais do que um objetivo. Nestas situações, a escolha de uma determinada alternativa de ação pode satisfazer plenamente um objetivo mas não ser tão eficaz no alcance de outros objetivos. Muitas vezes é impossível a escolha de uma alternativa que seja a melhor para todos os objetivos. Nestas situações, a escolha entre alternativas é facilitada pela adoção de uma única e bem compreendida medida de custo e benefício (como a unidade monetária) de modo que a prioridade entre os objetivos possa ser expressa em termos de unidades numéricas.

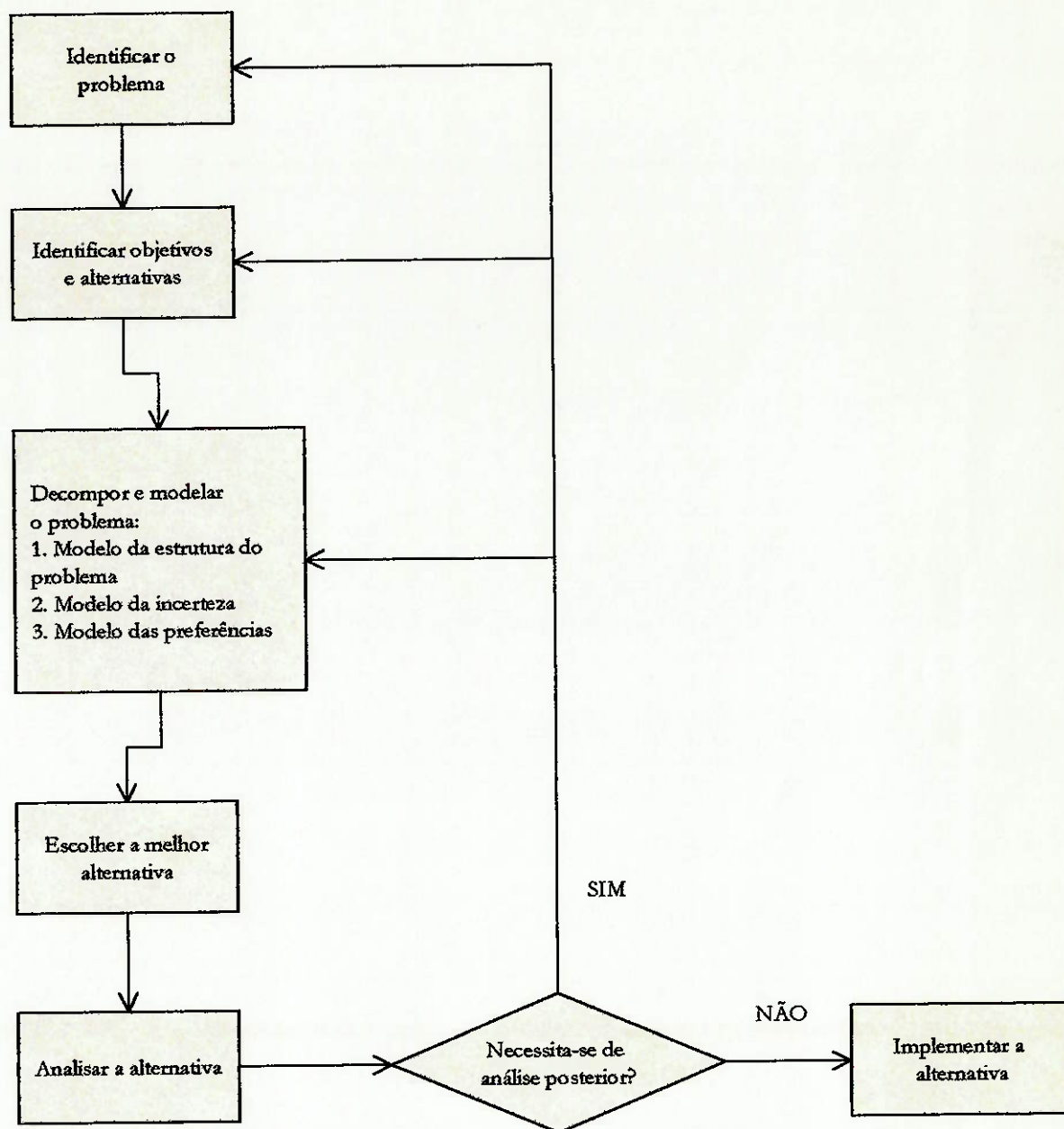
Existência de um ou mais participantes na tomada de decisão, todos com poder de influência no resultado

Em muitas situações de decisão não-programadas o poder de decisão não se encontra nas mãos de um único indivíduo. Muitas vezes, dois ou mais indivíduos são encarregados da tarefa e têm interesses e intenções conflitantes, embora nenhum tenha autoridade para tomar a decisão individualmente. Neste caso, uma solução satisfatória só poderá ser encontrada após um acordo entre os mesmos, resultado de quantas negociações forem necessárias.

Dentre estes fatores, Radford (1981) considera a falta de informações completas como o mais importante e o mais difícil de ser contornado. Em uma situação não-programada, é pouco provável que a escolha de uma alternativa satisfaça todos os indivíduos e objetivos de uma organização. Nestas circunstâncias, a tomada de decisão sempre terá um caráter individual.

Apesar dos problemas citados, Clemen (1990) propõe que se siga sempre uma seqüência lógica de passos para uma decisão plausível, conforme a Figura 1.1:

Figura 1.1: Proposta de abordagem de situações de decisão não-programadas. Fonte: Clemen (1990)



O procedimento é relativamente parecido ao proposto por Radford (1981) para abordagem de situações de decisão programadas, exceto o terceiro passo que é a decomposição e modelagem do problema.

Segundo Clemen (1990), este passo é o mais importante da tomada de decisão em situações deste tipo.

A primeira parte da estruturação de uma abordagem de decisão não-programada seria a decomposição do problema em subproblemas menores e que são mais fáceis de manusear. O próximo passo seria a modelagem do problema, que, segundo Clemen (1990), pode ser dividido em:

1. **Modelagem da estrutura do problema:** propõe a utilização de *diagramas ou árvores de decisão* para representar ou modelar o problema;
2. **Modelagem a incerteza inerente ao problema:** propõe o uso da probabilidade na construção de *modelos probabilísticos (ou estatísticos)*, o que permitiria a observação de fatores que não são aparentes no problema e
3. **Modelagem das preferências relacionadas ao problema:** acredita que a representação matemática do problema poderá facilitar a escolha de uma alternativa preferida.

Deste modo, visto que não se pode construir rotinas de decisão, o que se propõe é a utilização de determinadas ferramentas que possam auxiliar a tomada de decisão - não se esquecendo do julgamento pessoal dos indivíduos envolvidos que, segundo Clemen (1990), sempre será indispensável neste tipo de situação.

A seguir, serão discutidos os critérios de decisão referentes a decisões não-programadas.

1.3 CRITÉRIOS DE DECISÃO

Segundo Gregory (1988), em situações de decisão não-programadas existem diversas maneiras pelas quais um problema pode ser solucionado. Embora estas maneiras sejam diferentes em suas considerações e nas ações resultantes, existem algumas propriedades - baseadas em níveis de plausibilidade aceitáveis - que deverão possuir. Gregory (1988) propõe três propriedades aparentemente simples e triviais, mas que poderão fornecer uma lógica muito útil em determinados momentos do processo decisório:

- a) *“Qualquer alternativa que seja dominada deverá ser desconsiderada pelo tomador de decisão”*: a idéia de domínio é utilizada no sentido de que o indivíduo deverá escolher a melhor alternativa independentemente da ação que seja tomada pela natureza. Gregory (1988) apresenta o seguinte exemplo fictício: suponhamos que uma empresa esteja planejando sua campanha publicitária para o próximo semestre. Ela poderá escolher seu meio de divulgação entre televisão, jornais e “outdoors”, tendo recursos suficientes para apenas um dos três meios. O retorno do investimento da campanha é medido pela estimativa da exposição da campanha, ou seja, o número de clientes em potencial que teriam a oportunidade de ter contato com um dos três meios. Além disso, verificou-se que o número real de exposições depende das condições do tempo, onde um tempo ruim favorece a divulgação por jornais e a televisão e um tempo melhor favorece a divulgação

por “out-doors”. Estimou-se o número de exposições (em milhares de clientes potenciais), cujos valores são apresentados na Tabela 1.2:

Tabela 1.2: Estimativas do número de exposições em campanha publicitária (em milhares de clientes potenciais). Fonte: Gregory (1988)

Meio	CONDIÇÃO DO TEMPO			
	Ruim	Moderado	Bom	Excelente
Televisão	200	190	170	130
Jornais	180	160	150	130
Out-doors	110	140	140	190

Baseando-se nestes dados, observamos que os números de exposições da televisão é sempre maior ou igual ao do jornal, independentemente da condição do tempo, de modo que este último não deverá ser levado em consideração.

Entretanto, suponhamos que a empresa adote a estratégia de maximizar o menor retorno. Neste caso, o menor retorno da campanha em televisão seria 130 (tempo excelente), igual ao da campanha de jornais, e o menor retorno dos “out-doors” seria 110 (tempo ruim). Segundo este critério, seria indiferente a escolha da campanha através de jornais ou televisão, já que, na pior das hipóteses (tempo excelente) o retorno seria o mesmo (130). Porém, a propriedade da alternativa dominada diz que, independentemente de qualquer critério de decisão adotado, a divulgação por jornais nunca será a melhor alternativa neste caso específico. Ou seja, a ação de escolha de divulgação por jornais é uma ação dominada e portanto deverá ser desconsiderada.

- b) “Se, sob qualquer critério de decisão, prefere-se a alternativa *A* do que a alternativa *B* e prefere-se a alternativa *B* do que a alternativa *C*, então deverá se preferir a alternativa *A* do que a alternativa *C*”: esta propriedade não necessita de maiores explicações. Um procedimento decisório cujos resultados são “Prefere-se *A* do que *B*, prefere-se *B* do que *C* e, no entanto, prefere-se *C* do que *A*” é um procedimento incongruente.
- c) “A preferência entre duas alternativas quaisquer não deverá depender da presença ou ausência de uma terceira alternativa”: o seguinte exemplo mostra a violação desta propriedade: um diretor de uma empresa fabricante de automóveis vai até ao fornecedor de autopeças e é informado de que este no momento só tem eixos e espelhos retrovisores em seu estoque. Sendo assim, encomenda apenas eixos. Momentos mais tarde, o fornecedor o informa de que encontrou também escapamentos em seu estoque. O diretor diz: “Neste caso, encomendarei espelhos retrovisores”.

Segundo Gregory (1988), estas considerações embora simples são básicas para qualquer tomada de decisão e em certas situações podem ser muito úteis. Deve-se sempre estar atento para que não sejam violadas.

Sob estas considerações, existem diversos critérios que geralmente são utilizados na tomada de decisão. Os mais conhecidos são:

- ✓ Critério do Maximin
- ✓ Critério do Menor Arrependimento
- ✓ Critério do Maximax

- ✓ Critério de Hurwicz
- ✓ Critério do Valor Monetário Esperado (VME)
- ✓ Critério de Laplace
- ✓ Critério do Custo de Oportunidade Esperado

Estes critérios, como veremos adiante, são apenas parte do processo de decisão que envolve também fatores subjetivos e dificilmente quantificáveis. Segundo Gregory (1988), apenas os critérios do Maximin, Menor Arrependimento e do Valor Monetário Esperado são utilizados com maior frequência. Estes serão descritos nos próximos itens, omitindo a descrição dos demais.

1.3.1 CRITÉRIO DO MAXIMIN

Suponhamos que um indivíduo tenha que escolher entre as alternativas possíveis $A_1, A_2, \dots, A_m$; que a natureza apresente condições possíveis $B_1, B_2, \dots, B_n$ e que o retorno de uma ação A_i tomada sob B_j seja R_{ij} . Quando uma alternativa A_i é escolhida, pode-se ter os valores de retorno $R_{i1}, R_{i2}, \dots, R_{in}$. Considera-se que:

$$R_{i0} = \min (R_{i1}, R_{i2}, \dots, R_{in}) \quad (1.1)$$

e

$$R_{r0} = \max_{1 \leq i \leq m} [R_{i0}, R_{20}, \dots, R_{m0}] \quad (1.2)$$

O indivíduo deverá escolher a ação A_z tal que proporcione sempre um retorno R_{z0} que é maior ou igual ao valor R_{i0} . Pode-se, então, ter a vantagem de saber qual é o retorno mínimo garantido.

Retomando o exemplo da seção anterior onde uma empresa necessita decidir sobre como irá planejar sua campanha publicitária, suponhamos que a empresa tenha reavaliado seu retorno segundo a Tabela 1.3:

Tabela 1.3: Critério do Maximin: valores mínimos de retorno R_{i0} no número de exposições em campanha publicitária (em milhares de clientes potenciais). Fonte: Gregory (1988)

Meio	CONDIÇÃO DO TEMPO			
	Ruim	Moderado	Bom	Excelente
Televisão	220	210	190	140
Jornais	180	170	160	150
Out-doors	130	160	160	160

Neste caso, teríamos os valores mínimos de retorno para cada alternativa como 140, 150 e 130 para a televisão, jornais e "out-doors" respectivamente. Segundo o critério do Maximin, a alternativa a se escolher deverá ser a divulgação por jornais.

Este critério geralmente resulta de um pensamento muito pessimista ou conservador por parte do indivíduo que tomará a decisão. Este pensamento assume que a natureza terá maior probabilidade de tomar a ação mais indesejada possível e que portanto o indivíduo deverá se precaver ao máximo.

1.3.2 CRITÉRIO DO MENOR ARREPENDIMENTO

O segundo critério é similar ao apresentado anteriormente. Propõe que uma alternativa de ação deve ser medida pela quantia em que o retorno de um indivíduo seria aumentado caso ele soubesse qual seria a ação tomada pela natureza.

Suponhamos que um indivíduo tenha que escolher entre as alternativas possíveis $A_1, A_2, \dots, A_m$; a natureza apresente condições possíveis $B_1, B_2, \dots, B_n$ e o retorno de uma ação A_i tomada sob B_j seja R_{ij} . Define-se o *arrependimento* ρ_{ij} da alternativa i sob a ação da natureza j como:

$$\rho_{ij} = R_{kj} - R_{ij} \quad (1.3)$$

onde $R_{kj} = \max_{1 \leq k \leq m} [R_{1j}, R_{2j}, \dots, R_{mj}]$.

Seja $\rho_{kz} = \max_{1 \leq k \leq m} [\rho_{i1}, \rho_{i2}, \dots, \rho_{in}]$, o critério do Menor Arrependimento requer que o indivíduo tome uma determinada ação A_z tal que $A_z = \min_{1 \leq k \leq n} \rho_{kz}$.

O critério visa a segurança sob a pior ação possível tomada pela natureza, de modo a escolher o menor arrependimento de cada ação que este tomar. Há, portanto, uma matriz de termos de *arrependimento*, onde cada coluna contém pelo menos um valor nulo. No exemplo anterior da campanha publicitária, a matriz de *arrependimento* está na Tabela 1.4:

Tabela 1.4: Critério do Menor Arrependimento: valores de arrependimento p_{ij} no número de exposições em campanha publicitária (em milhares de clientes potenciais). Fonte: Gregory (1988)

Meio	CONDIÇÃO DO TEMPO			
	Ruim	Moderado	Bom	Excelente
Televisão	0	0	0	20
Jornais	40	40	30	10
Out-doors	90	50	30	0

Os arrependimentos máximos são 20, 40 e 90 para os meios televisão, jornais e “out-doors” respectivamente e o menor arrependimento é o meio televisão ($i = 1$).

O critério do Menor Arrependimento tem algumas desvantagens em relação ao critério do Maximin. Se, por exemplo, uma ação extrema da natureza é uniformemente ruim para todas as ações do indivíduo, então este não será um fator de influência em sua decisão. Para ambos os critérios, apenas os valores relativos de cada retorno deve ser considerado. Entretanto, para o critério do Menor Arrependimento, a relatividade entre os valores ocorre apenas dentro de uma mesma coluna, de modo que tornando irrelevante o quão ruim estes valores geralmente são quando comparados com outras ações da natureza.

1.3.3 CRITÉRIO DO VALOR MONETÁRIO ESPERADO

Os dois critérios anteriores são expressos em termos de retorno, sem nenhuma consideração sobre a probabilidade de ocorrência das ações da natureza. Algumas decisões são baseadas em resultados que, muitas vezes, são muito pouco prováveis. Em vista deste problema, utiliza-se o critério do Valor Monetário Esperado (VME).

Sejam as probabilidades $q_1, q_2, \dots, q_n$ associadas a cada ação da natureza, tal que $\sum_{i=1}^n q_i = 1$. Estas probabilidades podem ser subjetivas ou baseadas em algum fundamento (como pesquisa de mercado ou dados históricos). O retorno médio da alternativa i é obtido como sendo:

$$A_i = q_1 R_{i1} + q_2 R_{i2} + \dots + q_n R_{in} \quad (1.4)$$

O critério do VME requer que a ação A_w seja tomada tal que $A_w = \max_i A_i$.

Para o exemplo anterior, suponhamos que as probabilidades de ocorrência de cada condição do tempo sejam conhecidas, conforme a Tabela 1.5:

Tabela 1.5: Probabilidades de ocorrência de condições do tempo. Fonte: Gregory (1988)

CONDIÇÃO DO TEMPO	PROBABILIDADE DE OCORRÊNCIA
Ruim	20%
Moderado	40%
Bom	30%
Excelente	10%

Neste caso, os Valores Médios Esperados para cada alternativa são:

$$\text{Televisão: } (0.2)220 + (0.4)210 + (0.3)190 + (0.1)140 = 199$$

$$\text{Jornais: } (0.2)180 + (0.4)170 + (0.3)160 + (0.1)150 = 167$$

$$\text{"Out-doors": } (0.2)130 + (0.4)160 + (0.3)160 + (0.1)150 = 154$$

Assim, segundo o critério do VME, o meio televisão seria o meio recomendado, dado que apresenta o retorno máximo.

Observa-se que, se a mesma decisão é tomada em um grande número de vezes, o critério do VME irá determinar a alternativa que, quando utilizada repetidamente, irá resultar em um maior retorno. Porém, muitos problemas decisórios são problemas que ocorrem uma única vez ou que têm um número limitado de repetições. Nestes casos, embora uma decisão ocorra uma única vez, é possível que outras decisões de mesma importância em termos de retorno necessitem de serem tomadas regularmente por outros indivíduos na organização.

Além disso, segundo Gregory (1988), "um indivíduo tomador de decisão sabe que nem sempre ele estará certo, porém sabe que, se analisar suas opções e suas

consequências a fundo, seus esforços mais cedo ou mais tarde se refletirão nos resultados da organização”.

O critério do Valor Monetário Esperado ainda é o critério mais utilizado em problemas decisórios. Entretanto, como citado anteriormente, os critérios de decisão apenas servem de base para a escolha de uma determinada alternativa na tomada de decisão. Outros fatores mais subjetivos também podem influenciar o processo. Para termos uma idéia de como isto ocorre, veremos a seguir o que é a função utilidade.

1.4 FUNÇÃO UTILIDADE

No exemplo anterior da campanha publicitária, utilizou-se o critério do Valor Monetário esperado para a escolha de uma alternativa de ação. Para se aplicar este critério, precisamos conhecer não apenas as probabilidades dos possíveis resultados mas também os retornos correspondentes. Observa-se que estes retornos são medidos exclusivamente em termos financeiros. Entretanto, em muitas situações, as consequências de um determinado resultado não podem ser medidas exclusivamente através do retorno financeiro. Por exemplo, suponhamos que uma empresa decida investir em programas sociais para seus funcionários. A motivação para uma decisão deste tipo pode ser o desejo por parte dos administradores da empresa de ter funcionários motivados, o que se presume que irá retornar sob a forma de maior eficiência no trabalho. É possível que o retorno financeiro real não corresponda com o que foi investido, mas isto não significa necessariamente que o investimento não deva ser feito. De fato, existem outros benefícios que dificilmente são quantificáveis, como a possibilidade de se trabalhar em um ambiente mais agradável, maior lealdade dos funcionários com a empresa, melhoria da qualidade de vida dos mesmos, entre outros.

O segundo exemplo refere-se ao seguro de instalações de uma fábrica contra incêndio descrito em Gregory (1988): Suponhamos que:

- 1) O ativo total da fábrica seja um total de A unidades monetárias, sendo que a maior parte consiste de suas instalações;
- 2) Para que a fábrica faça o seguro de suas instalações durante um ano, ela deverá pagar uma taxa anual de i unidades monetárias;
- 3) Caso haja algum incêndio ou outro dano nas instalações e a fábrica tenha feito o seguro, ela será inteiramente recompensada pelos prejuízos;
- 4) Caso não faça o seguro e ocorra algum dano, ela tenha uma redução em seu ativo de B unidades monetárias.

Existem duas alternativas possíveis: fazer ou não o seguro. A natureza também tem duas ações possíveis: ocorrência ou não de fogo nas instalações. Em termos de resultados obtidos durante o ano, teríamos:

Tabela 1.6: Resultados obtidos por ocorrência ou não de incêndio. Fonte: Gregory (1988)

AÇÃO	RESULTADO	
	Com incêndio	Sem incêndio
Fazer o seguro	$A - i$	$A - i$
Não fazer o seguro	$A - B$	A

Caso ocorra ou não um incêndio e o seguro tenha sido feito, o ativo total da empresa permanecerá $A - i$. Caso o seguro não tenha sido feito e ocorra um incêndio,

o resultado será o ativo menos a redução do mesmo, ou seja, $A - B$; caso o incêndio não ocorra, o ativo permanecerá o mesmo.

Suponhamos agora que se tenha determinado que a probabilidade real de ocorrência de um incêndio seja α . Se utilizarmos o critério do Valor Monetário Esperado para as duas ações e os valores da Tabela 1.6, deveríamos fazer o seguro apenas se:

$$A - i \geq (1 - \alpha)A + \alpha(A - B) = A - \alpha B \quad (1.5)$$

ou seja,

$$\alpha > i / B \quad (1.6)$$

Segundo Gregory (1988), geralmente isto não é o que ocorre na prática: a probabilidade real de risco α normalmente é muito menor do que i / B . Entretanto, existe um fator que leva as empresas a fazerem o seguro: a *função utilidade* $u_i(x)$. A função $u_i(x)$ expressa o valor que a utilidade de um determinado elemento x tem para o indivíduo i .

Aplicando esta função em nosso exemplo, teríamos que os resultados de se perder grande parte do ativo total da empresa, que se criou durante um longo período, superam a perda monetária do pagamento da tarifa anual do seguro, logo:

$$u(A - i) \geq (1 - \alpha)u(A) + \alpha u(A - B) \quad (1.7)$$

Ou seja, a função utilidade mostra que, para a empresa, a utilidade de seu ativo A subtraída a taxa anual i é maior do que a média ponderada entre a utilidade de seu ativo A e a utilidade do mesmo ativo subtraída da perda B de um eventual incêndio. Observa-se que a função utilidade é um critério subjetivo, porém está presente em grande parte das tomadas de decisão do cotidiano, já que muitas decisões têm um caráter subjetivo e não apenas monetário. Portanto, os critérios citados anteriormente, bem como a utilização de modelos matemáticos e árvores de decisão sugerida por Clemen (1990), servem apenas de base para uma análise completa da situação de decisão. Apesar disso, quando não se tem nenhuma informação sobre uma situação de decisão nova - por exemplo, quando se desconhece as probabilidades de ocorrência de determinadas alternativas ou mesmo o risco envolvido em cada uma - a utilização de modelos matemáticos e árvores de decisão poderá ser muito útil.

Esta abordagem matemática de um problema decisório chama-se análise de decisão. Segundo Gregory (1988), qualquer modelo matemático é uma aproximação da realidade e, portanto, o mesmo ocorre com a análise de decisão. O resultado do modelo matemático será apenas uma evidência utilizada pelo indivíduo para formar sua opinião, e o peso que ele dará para esta evidência dependerá de seu conhecimento sobre a confiabilidade do modelo que criou.

Veremos agora as árvores de decisão e os modelos matemáticos em maiores detalhes.

1.5 ÁRVORES DE DECISÃO

Árvores de decisão são representações gráficas utilizadas para se estimar o Valor Monetário Esperado das diversas alternativas de risco disponíveis em um problema. As possíveis alternativas de uma decisão, representada por um ponto (ou nó), são representadas por traços (ou ramos) de modo que a representação se assemelha a uma árvore. A cada ramo associa-se um custo de decisão e uma probabilidade de ocorrência. O objetivo da representação é ponderar os custos de cada nó para se escolher o(s) ramo(s) associado(s) ao maior Valor Monetário Esperado.

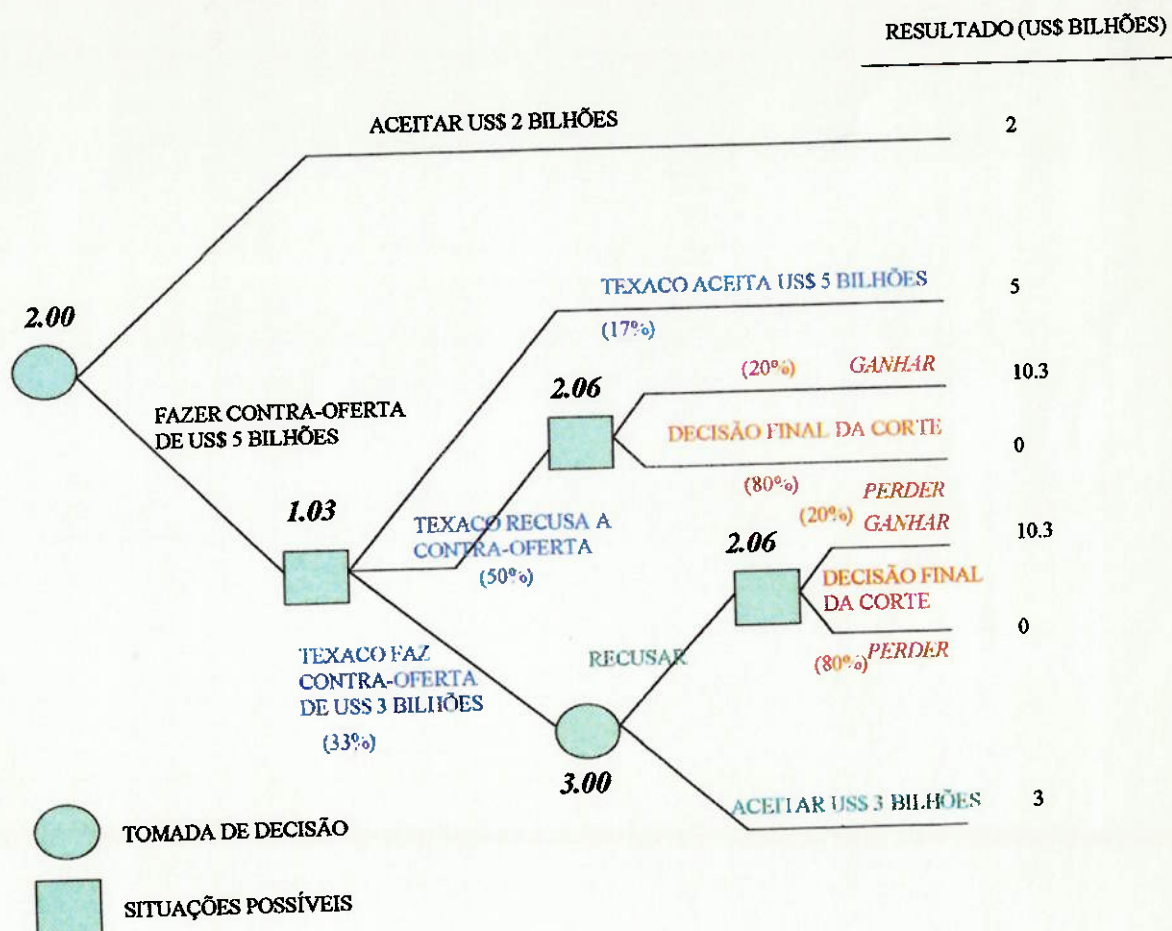
Para melhor compreensão, consideremos o seguinte exemplo extraído de Clemen (1990):

Texaco versus Pennzoil

No início de 1984, as empresas norte-americanas fabricantes de derivados de petróleo Pennzoil e Getty Oil decidiram se fundir. Porém, antes que se assinassem os documentos formalizando a transação, outra empresa do ramo, a Texaco, propôs a compra da Getty Oil por um preço melhor. Pennzoil julgou que a Texaco havia agido de má fé e processou a mesma por ter interferido ilegalmente na transação da Pennzoil-Getty. Ganhou o processo e a corte determinou que a Texaco deveria pagar uma indenização de 11 bilhões de dólares, a maior da história norte-americana até o momento. Texaco apelou para a corte e conseguiu uma redução para 2 bilhões de dólares, porém os juros e outras tarifas elevaram o total de novo para 10.3 bilhões. O presidente da Texaco, James Kinnear, alegou que iria batalhar na Justiça até o final e que o acordo Pennzoil-Getty desobedecia a algumas normas legais. Entretanto, ofereceu à Pennzoil uma indenização de 2 bilhões de dólares para encerrar o caso.

Hugh Liedtke, presidente da Pennzoil, foi aconselhado por seus assessores de que apenas um acordo entre 3 e 5 bilhões de dólares seria justo. Seus assessores também estimaram as probabilidades de vitória caso fosse a julgamento. A Figura 1.2 demonstra árvore de decisão de Hugh Liedtke:

Figura 1.2: Valor Monetário Esperado das possíveis alternativas de decisão de Hugh Liedtke no caso Pennzoil-Getty. Fonte: Clemen (1990)



Segundo método proposto por Clemen (1990) para se calcular o maior Valor Monetário Esperado das possíveis alternativas, começa-se pelos ramos da árvore e caminha-se em direção à origem, da seguinte forma:

- 1) Calculando os valores esperados (definidos momentaneamente) de cada ramo da árvore quando se encontra um nó e
- 2) Escolhendo o ramo com maior valor esperado quando se encontrar um nó da árvore.

Observemos a figura: caso Liedtke faça uma contra-oferta para a Texaco de US\$ 5 bilhões, existem três situações possíveis (letras azuis): 1) Texaco aceitar esta quantia (probabilidade de ocorrência de 17%), 2) Texaco recusá-la (probabilidade de 50%) e 3) Texaco fazer contra-oferta de US\$ 3 bilhões (probabilidade de 33%).

Caso ocorra a situação 3), Liedtke tem duas opções (letras verdes): 3.1) recusar a contra-oferta, situação que irá a julgamento, com probabilidade de êxito de apenas 20% e 3.2) aceitá-la.

Observemos o cálculo do Valor Monetário Esperado (VME) para a opção 3.1):

a) Caso ganhe, $VME = 0.2 * 10.3 = 2.06$

b) Caso perca, $VME = 0.8 * 0 = 0$

Como este valor é menor do que o obtido pela opção 3.2) - US\$ 3 bilhões - , o nó 3) terá $VME = 3$ bilhões.

Seguindo o mesmo raciocínio para a situação 2), o VME deste nó será de US\$ 2.06 bilhões. Multiplicando os valores obtidos nos nós 2), 3) e a situação 1) pelas suas respectivas probabilidades, temos:

$$VME = 0.17 * 5 = 0.85 \text{ bilhão}$$

$$VME = 0.5 * 2.06 = 1.03 \text{ bilhão}$$

$$VME = 0.33 * 3 = 0.99 \text{ bilhão}$$

Comparando os *VMEs* dos nós 1), 2) e 3), temos que o nó terá $VME = 1.03$.

Finalmente, procede-se desta maneira até o início da árvore obtendo-se o *VME* do igual a US\$ 2 bilhões. Em seguida, observa-se as alternativas referentes a este valor encontrado e considera-se esta alternativa como a melhor.

Deste modo, segundo este critério a melhor alternativa para a Pennzoil seria de aceitar a proposta inicial de US\$ 2 bilhões de dólares da Texaco para encerrar o caso.

É, portanto, uma ferramenta simples e útil. Entretanto, carece de mecanismos mais apurados na determinação das probabilidades de ocorrência de cada alternativa, o que pode ser solucionado com o uso de modelos estatísticos. Na Seção 1.6 estão descritas as principais características destes modelos.

1.6 MODELOS ESTATÍSTICOS

A atividade de modelagem de problemas cotidianos com modelos matemáticos já existe desde o início do século, quando eram utilizados na área de Imunologia para a determinação de fatores de risco (Hair, 1979). Porém, foi a partir da década de 40 que se impulsionou sua aplicação, quando a indústria de crédito passou a utilizar os modelos como base para a decisão de se conceder ou não um empréstimo.

Segundo Hand e Henley (1997), os métodos decisórios tradicionais de concessão de crédito para pessoas físicas utilizavam o julgamento subjetivo do cedente do empréstimo, baseado em experiências anteriores e nos dados do candidato ao empréstimo. Entretanto, à medida em que a indústria de crédito expandiu-se, os métodos tradicionais já não eram mais eficazes quando aplicados ao fornecimento de crédito em massa, o que implicava em lentidão e um grande número de omissões de pagamento. A obtenção de disparidades nos resultados de análise feitos pelos analistas de crédito de maneira subjetiva era freqüente. O mesmo acontecia com a análise para crédito a grandes empresas, que necessitava de um número de informações da empresa candidata suficientemente grande.

Como solução para este problema, desenvolveram-se modelos estatísticos que agilizam o processo, tornando-o mais eficiente e principalmente servindo de base para decisão, substituindo o julgamento subjetivo. Deste modo, diversos modelos foram utilizados e se desenvolveram em função desta atividade. Atualmente, existem diversos modelos disponíveis e seu uso é muito abrangente, embora a área de crédito ainda seja onde são mais difundidos

Problemas decisórios semelhantes ao da concessão de crédito são encontrados no cotidiano. O julgamento subjetivo é, em muitos casos, lento e pouco confiável, o que pode ser altamente custoso em determinadas situações. Por este motivo, os modelos têm sido utilizados nas mais diversas áreas, como Medicina, Economia, Marketing, Produção, Transportes, entre outras.

Alguns exemplos de utilização são:

- Estimação da chance de sucesso de um novo produto no mercado
- Classificação de fornecedores de uma indústria por ordem de desempenho
- Determinação da probabilidade de omissão de pagamento de um empréstimo
- Identificação dos consumidores potenciais de um produto
- Identificação das alterações demográficas em uma população

- Estimação da probabilidade de ocorrência de infarto em pacientes
- Elaboração de rotas de ônibus no sistema de transportes urbano
- Identificação das causas de falência de um grupo de empresas

Uma das maneiras de se solucionar problemas como os expostos acima é classificando elementos de um conjunto de dados em grupos ou categorias. Por exemplo, para se determinar as causas de falência de um grupo de empresas, pode-se tomar um conjunto de dados de empresas que faliram no passado e averiguar possíveis causas em comum. Depois, pode-se gerar um modelo estatístico que possa ser aplicado a um conjunto de dados de empresas atuais, para se prever a probabilidade de falência. Neste caso, de acordo com a probabilidade estimada, pode-se classificar uma empresa em um grupo determinado - como por exemplo Alto, Médio ou Baixo Risco de Falência.

Em análise de decisão, os modelos estatísticos funcionam de modo similar a este exemplo. Primeiramente, estima-se a probabilidade de um elemento pertencer a um determinado grupo (através da formulação matemática do problema) e depois compara-se o valor obtido com valores de probabilidade pré-estabelecidos para cada grupo. Com esta comparação, classifica-se o elemento em um grupo determinado de acordo com uma probabilidade e esta classificação servirá de base para a tomada de decisão.

Basicamente, segundo Mizutani (1982), o processo de modelagem matemática de um problema decisório resume-se a:

- 1) Utilizar uma amostra de dados para gerar um modelo matemático, comumente conhecido como função discriminante;
- 2) Utilizar a equação em um conjunto de dados de elementos reais para tentar prever o comportamento destes elementos;
- 3) A partir dos resultados da equação, tomar uma decisão.

A idéia é tomar uma amostra de um conjunto de dados para produzir um modelo matemático que possa ser aplicado ao conjunto de dados inteiro. Isto é feito em situações onde é impossível avaliar todos os dados existentes - geralmente quando se tem conjuntos de dados muito extensos. Portanto, espera-se que o modelo consiga avaliar satisfatoriamente o comportamento do conjunto.

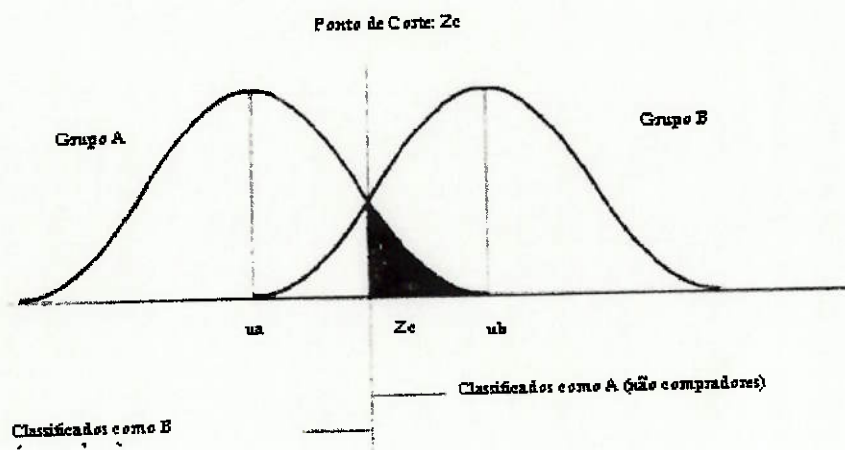
Na prática, o item 2) anteriormente é dividido em duas etapas:

- a) O conjunto de dados é dividido em duas partes em proporções que variam de 50%-50% a 70%-30%, dependendo da situação. A primeira parte é utilizada para a formulação do modelo e é normalmente chamada de "conjunto de treinamento" ("training set");
- b) A segunda parte é utilizada para o teste ou validação do modelo, normalmente chamada de "conjunto de validação" ("validation set"), e tem o objetivo de verificar a eficiência do modelo criado.

No caso dos modelos estatísticos utilizados em análise de decisão, que visam a classificação de elementos em grupos, esta eficiência é medida pela porcentagem de acertos/erros ou previsões corretas/incorretas dos elementos em um determinado grupo.

Graficamente, podemos representar de modo simplificado o problema de classificação de elementos em dois grupos distintos através da Figura 1.3 :

Figura 1.3: Classificação de elementos em dois grupos distintos. Fonte: Hair et alli (1979)



No eixo da Figura 1.3, tem-se a probabilidade de ocorrência da característica em questão no problema e a distribuição dos elementos em função da mesma, que é determinada por um ponto de corte Z_c (ou critério de corte). Os elementos à esquerda do ponto de corte são classificados como pertencentes ao Grupo A e os elementos à direita, como pertencentes ao Grupo B. Observa-se a concentração dos elementos da amostra em dois grupos claramente distintos, cujas curvas de distribuição têm médias claramente distintas μ_a , μ_b . Entretanto, observa-se também que há uma área de interseção das curvas. Esta área corresponde aos elementos que foram classificados de modo incorreto, conhecidos como erros do tipo 1 e 2:

Erro tipo 1: elementos que na realidade pertencem ao Grupo A mas que foram classificados como pertencentes ao Grupo B;

Erro tipo 2: elementos que na realidade pertencem ao Grupo B mas que foram classificados como pertencentes ao Grupo A.

A seguir, temos um exemplo de situação onde pode-se utilizar um modelo estatístico para auxílio no processo decisório, citado em Hair, Anderson, Tatham (1979).

Exemplo: Suponhamos que uma empresa tenha decidido avaliar se seu novo produto terá ou não sucesso no mercado. Após ter definido como isto seria feito, a empresa decidiu identificar, se possível, quais consumidores comprariam o novo produto e quais não comprariam. Para auxiliar o processo de identificação de compradores potenciais, a empresa desenvolveu três escalas de avaliação a serem utilizadas pelos consumidores na avaliação do produto. Os fatores escolhidos foram durabilidade (X1), desempenho (X2) e aparência (X3), e os resultados possíveis variavam em uma escala de valor mínimo 0 (muito ruim) e máximo 10 (excelente). Para a empresa, uma combinação ponderada dos três fatores seria mais precisa na avaliação de um potencial comprador do que cada fator considerado isoladamente.

A Tabela 1.7 mostra os resultados obtidos para os três fatores considerados na avaliação de dez donas-de-casa entrevistadas que seriam potenciais compradoras. Após a avaliação do produto sob os três fatores perguntou-se a cada dona-de-casa se compraria ou não o produto. Cinco donas-de-casa responderam que comprariam e cinco responderam que não comprariam o produto.

Tabela 1.7: Resultados de pesquisa para avaliação de um novo consumidor. Fonte: Hair et alli (1979)

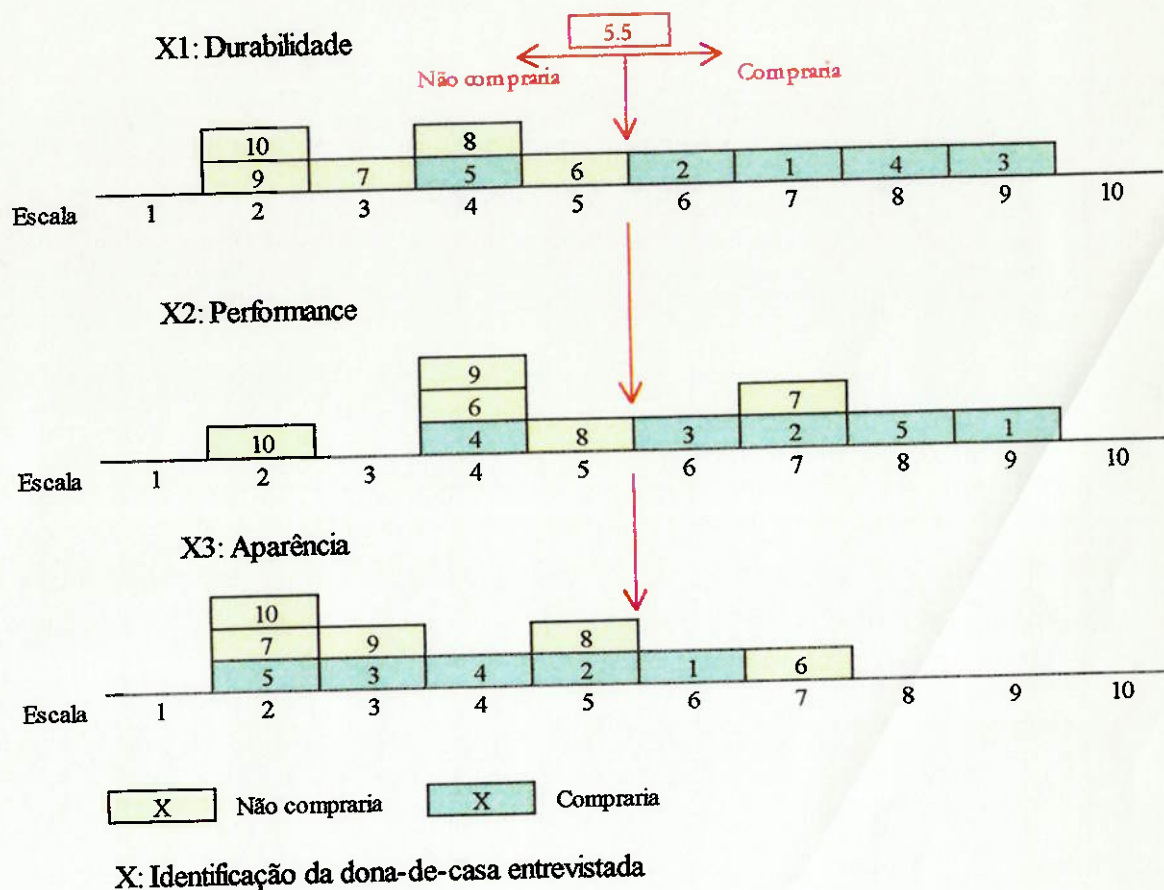
Intenção de Compra	Número do Entrevistado	Avaliação do Novo Produto		
		X1 Durabilidade	X2 Performance	X3 Aparência
Grupo 1 Compraria o produto	1	8	9	6
	2	6	7	5
	3	10	6	3
	4	9	4	4
	5	4	8	2
	Média do Grupo	7,4	6,8	4,0
Grupo 2 Não compraria o produto	6	5	4	7
	7	3	7	2
	8	4	5	5
	9	2	4	3
	10	2	2	2
	Média do Grupo	3,2	4,4	3,8
Diferença entre as Médias		4,2	2,4	0,2

Observando-se a Tabela 1.7, temos que no fator durabilidade (X1) há uma diferença de 4,2 entre as médias dos compradores (Grupo 1) e não compradores (Grupo 2). Isto pode ser um sinal de que a durabilidade pode ser um fator discriminante entre os grupos, sendo uma característica importante para os compradores (Grupo 1). As diferenças entre as médias no fator aparência (X3) é de

0,2, indicando que este fator possa ser menos discriminante em termos de identificar os compradores.

Apesar de a quantidade de informação ser relativamente pequena, a título de ilustração vamos observar o comportamento dos padrões de resposta dos entrevistados na Figura 1.4:

Figura 1.4: Análise de dez potenciais compradores de um novo produto. Fonte: Hair et alli (1970)



Observa-se uma maior concentração de célula de mesma cor no fator durabilidade (X1), enquanto que no fator desempenho (X2) a concentração não é tão grande e no fator aparência (X3) as células coloridas se misturam ao longo da escala.

Considerando-se apenas o fator durabilidade (X1), observa-se uma discriminação quase perfeita nos dois grupos, entre os valores 5 e 6 da escala. Se utilizássemos um valor de corte entre estes valores, por exemplo 5.5, para separarmos os entrevistados em dois grupos distintos (potenciais compradores ou não) para o fator considerado, conforme indicado na figura, observaremos que praticamente todos os elementos estariam classificados no grupo correto, com exceção do entrevistado número 5, que seria erroneamente classificado no grupo de não compradores.

Observando-se o fator desempenho (X2), nota-se que existe uma diferenciação entre os grupos, embora não seja tão grande. Se fosse utilizado o mesmo valor de corte - 5.5 - o entrevistado número 5 estaria corretamente classificado no grupo de potenciais compradores, enquanto os entrevistados número 4 e 7 estariam erroneamente classificados. Entretanto, os entrevistados que seriam classificados no grupo errado utilizando-se X2 estariam corretamente separados no fator X1.

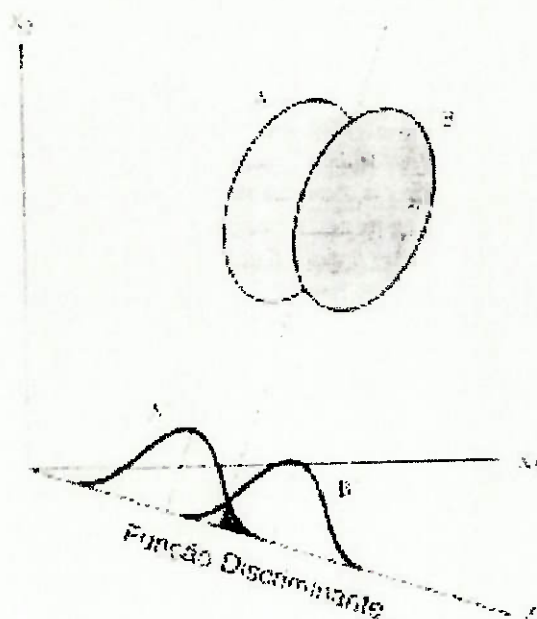
E, finalmente, observando-se o fator aparência (X3), nota-se que há pouca diferenciação entre os grupos. Portanto, X1, X2 e X3 podem ser utilizados *em combinação* para se prever o grupo em que um entrevistado se enquadra, criando-se uma função discriminante, atribuindo X1 e X2 em maior peso e X3 em menor peso, tal que maximize a separação dos grupos em uma determinada amostra.

Os modelos estatísticos procedem de modo similar a este exemplo hipotético, identificando as variáveis com maior poder de discriminação entre os grupos e atribuindo pesos para cada variável que reflita estas diferenças, para em seguida utilizá-los na formação de uma função matemática que separe os entrevistados em cada grupo. Deste modo, a empresa poderá utilizar a função para prever se, em um determinado mercado consumidor, seu produto terá sucesso ou não.

A Figura 1.5 representa o problema anteriormente discutido:

A Figura 1.5 representa o problema anteriormente discutido:

Figura 1.5: Função discriminante de elementos em dois grupos distintos. Fonte: Hair et alii (1979)



Para que possamos entender a necessidade dos modelos estatísticos na tomada de decisão, veremos a seguir o que é a análise de crédito e como os modelos estatísticos são utilizados na concessão de crédito. Entretanto, o motivo principal é conhecer o sistema de "escoragem" de crédito utilizado para a modelagem de problemas de análise de crédito cujos procedimentos são semelhantes à maioria dos sistemas utilizados na modelagem estatística de problemas cotidianos relacionados à análise de decisão.

2. CONCESSÃO DE CRÉDITO

O termo crédito utilizado neste trabalho refere-se à quantia que é emprestada a um devedor, geralmente por uma instituição financeira. Esta quantidade deve ser reposta, com juros, sob a forma de parcelas, normalmente em intervalos regulares.

A necessidade por parte do candidato ao empréstimo em fazê-lo demonstra sua necessidade de recursos dos quais não dispõe no momento, mas que espera dispor em um futuro próximo. Entretanto, devido à impossibilidade de se prever a situação futura do devedor e de que maneira esta irá influir na sua capacidade de cumprir com as obrigações financeiras que assumiu junto ao credor, existe um risco para o credor na concessão do empréstimo.

Frente a este risco, o credor deverá estimar a futura situação de seu cliente para avaliar as possibilidades deste saldar seus compromissos com a instituição financeira. A probabilidade de reposição do empréstimo pelo credor (ou a probabilidade de omissão do pagamento) é estimada através de informações fornecidas pelo credor à instituição financeira. Esta estimativa servirá de base para uma decisão da instituição de fornecer ou não o empréstimo. Uma estimativa precisa será vantajosa para a mesma, já que diminuirá o risco do empréstimo. Este processo chama-se análise de crédito.

O processo tem como objetivo a classificação de um credor em duas classes distintas: aqueles com probabilidade satisfatória de reposição do empréstimo feito e que portanto receberão o empréstimo e aqueles com probabilidade abaixo da desejada e que portanto não receberão o empréstimo. Por este motivo também é chamado de classificação de crédito ("credit scoring").

O processo de classificação de crédito apresenta as seguintes dificuldades:

- Grande quantidade de informações envolvida - no caso de um grande número de candidatos a empréstimo;

- Diversidade de aspectos a serem considerados nas várias informações;
- Importância relativa a ser dada a cada informação;
- Multiplicidade de enfoques envolvida na decisão de concessão;
- Necessidade de um processo rápido e confiável.

Existe, portanto, a necessidade de um sistema que auxilie o processo de decisão em análise de crédito, procurando diminuir o risco envolvido.

Neste caso, os modelos estatísticos têm como objetivo a classificação de um candidato a empréstimo em duas categorias - alto e baixo risco de empréstimo. Esta classificação servirá de base para o credor decidir pela concessão de empréstimo ou não.

Sua utilização traz os seguintes benefícios:

- Aumento da confiabilidade dos resultados;
- Redução do tempo de avaliação;
- Concentração das informações do cliente, tornando a avaliação mais global;
- Possibilidade de utilização de um sistema de informações, que possibilita uma série de outros recursos como confecção de gráficos, históricos e simulação de dados dos clientes.

2.1 CLASSIFICAÇÃO DE CRÉDITO

A classificação de crédito é uma atividade altamente difundida na sociedade brasileira, visto que é necessária para a concessão de empréstimo para inúmeras atividades, como a compra de automóveis, imóveis ou mesmo para empreendimentos

de pessoas físicas a empresas de grande porte. Nos últimos anos, tem sido muito utilizada para a concessão de crédito para empresas, principalmente as microempresas, que geralmente iniciam suas atividades após a realização de um empréstimo.

Segundo a empresa norte-americana de concessão de crédito Freddie Mac (URL: www.freddiemac.com/function/fm-homby/creditsc.htm), a classificação de crédito bem feita traz benefícios tanto para quem concede, quanto para quem requisita o empréstimo.

Para quem concede (no caso, as instituições financeiras privadas ou governamentais), isto significa uma diminuição na porcentagem de empréstimos que não são pagos. Geralmente, estas despesas com empréstimos que não são pagos são compensadas com um aumento nos juros aplicados às parcelas de pagamento do empréstimo. Com a diminuição nesta porcentagem, os juros cobrados são menores, já que o risco de empréstimo é menor. Isto, a médio e longo prazo, significa um aumento de lucro gerado por dois fatores: 1) diminuição dos custos diretos: os empréstimos que não são pagos e 2) aumento na demanda pela diminuição dos juros cobrados. Consequentemente, quem requisita o empréstimo pode obtê-lo a juros mais baixos, o que no caso de uma empresa poderá significar uma economia considerável.

2.2 SISTEMAS DE CLASSIFICAÇÃO DE CRÉDITO

Com o aumento da competição entre as empresas, aumentou-se também a demanda de crédito. Isto levou à utilização de diversos modelos estatísticos para auxiliarem o processo decisório de fornecimento do crédito, o que foi facilitado pelo desenvolvimento da indústria eletrônica como um todo. Estes modelos são chamados

de sistemas estatísticos de classificação de crédito. Discutiremos a seguir seu funcionamento e desempenho.

2.2.1 FUNCIONAMENTO

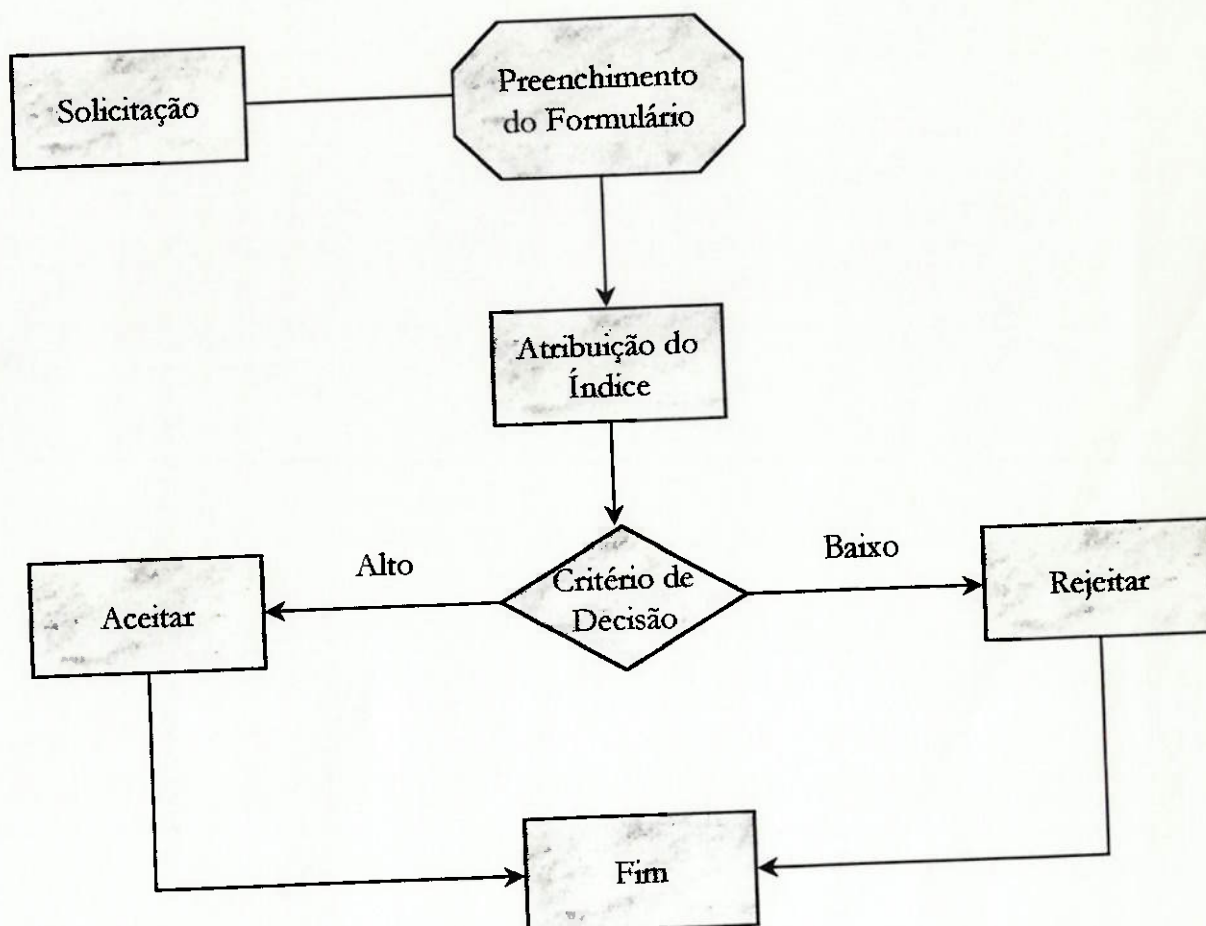
Os sistemas de classificação de crédito utilizam dados pessoais do candidato a empréstimo para o auxílio na decisão de se conceder ou não o empréstimo. Geralmente, o resultado obtido é um índice que reflete a probabilidade de o candidato repor o empréstimo.

Segundo Mizutani (1982), estes sistemas são compostos pelos seguintes elementos:

- 1) Expressão para o cálculo da escoragem de crédito, uma combinação de:
 - a) Conjunto de fatores considerados: composto pelos dados do candidato, por exemplo, o número de dependentes ou posse de casa própria (sim ou não);
 - b) Pesos relativos dos fatores: a cada variável é associada uma constante que denota o peso relativo de cada variável considerada, ou seja, coeficientes de ponderação;
- 2) Pontos de corte, associados a uma probabilidade (ou critério de corte): candidatos com “escoragem” maior que o critério de corte são considerados “aceitos”, caso contrário, é recusada a proposta;
- 3) Rotina operacional: sistema computadorizado, capaz de realizar a rotina de atribuição do índice.

A Figura 2.1 representa o modelo de decisão de crédito:

Figura 2.1: Modelo de decisão. Fonte: Mizutani (1982)



Quanto ao tipo de informação utilizada, os diversos sistemas de classificação de crédito disponíveis no mercado podem ser divididos em dois grupos (Hand e Henley, 1997):

- 1) Sistemas genéricos: são baseados exclusivamente na declaração de renda do candidato;

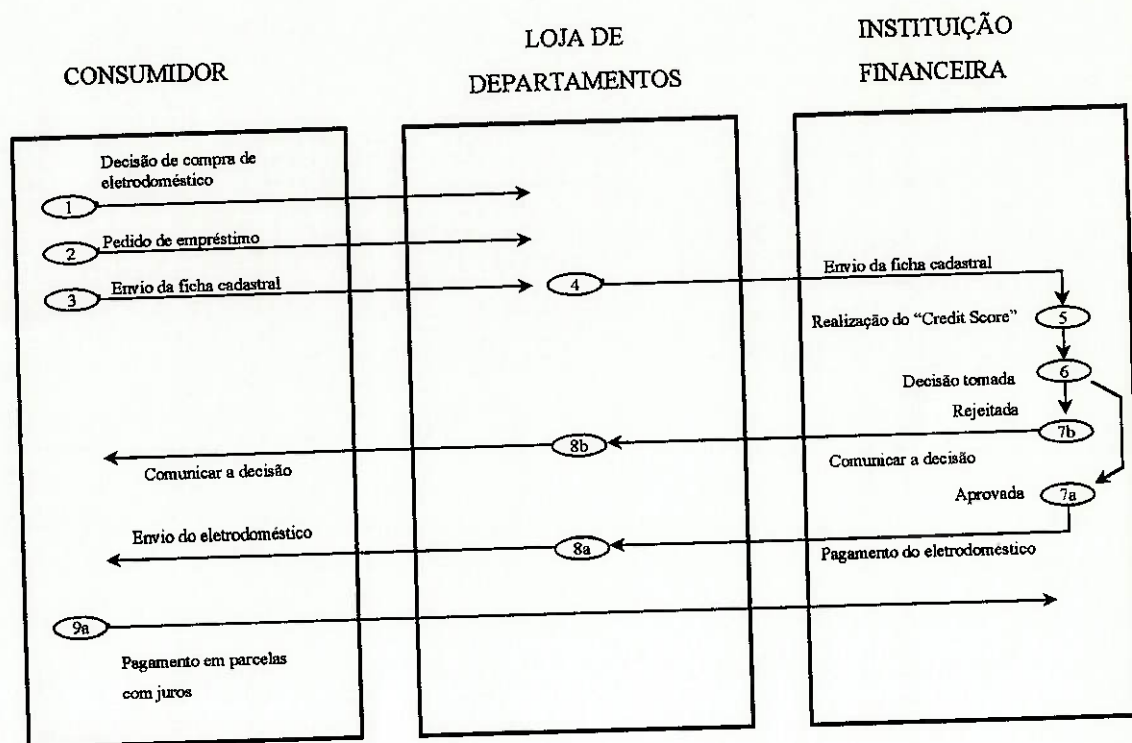
2) Sistemas personalizados: são baseados, além da declaração de renda do candidato, no histórico de empréstimos do mesmo.

Na maioria dos sistemas, o principal dado observado no candidato é o seu histórico de pagamento de dívidas. Outros fatores também são importantes: segundo Mizutani (1982), é prática comum a utilização de dados como idade de contas bancárias, rendimentos, profissão, tempo de serviço e valor dos bens, entre outros.

Uma aplicação muito comum dos sistemas de escoragem de crédito está no financiamento de crédito ao consumidor na compra, por exemplo, de um eletrodoméstico, como ilustra o fluxograma da Figura 2.2:

Figura 2.2: Processo de concessão de crédito para o consumidor em uma loja de departamentos.

Elaborado pelo autor



2.2.2 DESEMPENHO

Com relação ao desempenho destes sistemas, Rosenberg e Gleit (1994) apud Hand e Henley (1997) demonstraram que os modelos estatísticos produzem estimativas mais precisas do que as baseadas no julgamento subjetivo. Também demonstraram que apenas os modelos estatísticos conseguem lidar com um grande número de transações, devido à quantidade de informações envolvida.

Entretanto, os sistemas de análise de crédito podem apresentar problemas. Treinamento inadequado de pessoal e sistemas informatizados deficientes podem impedir o acompanhamento e monitoramento do desempenho dos modelos, prejudicando seu funcionamento.

Além disso, embora se utilize um sistema com desempenho satisfatório, o julgamento subjetivo ainda é necessário em certas circunstâncias. Por exemplo, algumas exceções podem ser feitas para um candidato a um empréstimo que, embora tenha um histórico de pagamentos favorável, não foi capaz de arcar com suas obrigações financeiras devido a problemas de saúde ou perda do emprego, ou mesmo para um candidato que não tenha um histórico de pagamento de dívidas ou que tenha um histórico recente e que não forneça informações suficientes.

Outra situação possível é a de um candidato que tenha no passado ido à falência e, mesmo assim, tenha uma pontuação alta, enquanto outro, sem nenhuma omissão anterior de dívida, tenha uma pontuação mais baixa. Conforme dito anteriormente, estes seriam erros do tipo 1 e 2:

Erro tipo 1: candidato de baixo risco que tem seu pedido de empréstimo negado

Erro tipo 2: candidato de alto risco que tem seu pedido aceito

Isto decorre do fato de que as pontuações refletem uma combinação de fatores de risco, explicitadas nas informações fornecidas pelo candidato. Estes fatores de risco são numerosos porém sutis, de modo que nenhum fator atua individualmente. Coletivamente, são capazes de demonstrar um risco alto de empréstimo, resultado de uma pontuação baixa.

Uma outra necessidade de auxílio do julgamento subjetivo pode surgir quando se obtém pontuações muito próximas do limite mínimo estabelecido para a decisão, o que

demandava uma revisão de todos os dados do candidato - o que não pode ser feito pelo sistema.

Mesmo assim, o julgamento subjetivo nunca é tão rápido e confiável quanto os sistemas de classificação de crédito e, para instituições de porte que fornecem empréstimos em massa, também é uma opção mais econômica.

2.3 EVOLUÇÃO DOS MÉTODOS DE ESCORAGEM DE CRÉDITO

O primeiro estudo publicado sobre um sistema de classificação de crédito foi realizado por Durand (1941). Centenas de empréstimos "bons" e "maus" (ou seja, saldados e não saldados), retirados de arquivos de bancos comerciais, empresas de financiamento à pessoa física, empresas de financiamento para a compra de automóveis e eletrodomésticos e empresas de financiamento à indústria, foram analisados utilizando-se a Análise Discriminante. Os resultados mostraram uma boa predição de pagamento dos empréstimos feitos em 20 bancos de varejo e em 9 empresas de financiamento à indústria.

Um posterior estudo feito por Wolbers (1949) em uma rede de lojas de departamentos internacional. Desenvolveu um modelo de análise formado por um conjunto de fatores multiplicados por seus respectivos coeficientes de ponderação e aplicou-o a um determinado conjunto de dados. Posteriormente, aplicou o mesmo modelo em outro conjunto de dados e checkou os resultados, conseguindo demonstrar que o número de perdas com empréstimos não saldados poderia ser reduzido em 7%, ou melhor, seu método conseguiu classificar negócios saldados e não saldados com

uma precisão de 7% a mais do que a obtida com o julgamento subjetivo utilizado na empresa até então.

Na década seguinte, Myers e Cordner (1957) realizaram um estudo em uma empresa de empréstimos a pessoas físicas em Los Angeles (EUA), demonstrando uma possível redução de perdas de 24% com empréstimos não saldados. Outro estudo, realizado por McGrath (1960) apontou uma possível redução de 20% nas perdas.

Entretanto, Myers (1963) apontou algumas dificuldades de aceitação destes modelos no mercado da época:

- 1) Uma relutância natural por parte dos analistas de crédito mais experientes a abandonar o julgamento pessoal por testes relativamente novos;
- 2) A incapacidade inicial dos computadores de executarem os modelos com uma precisão de acertos suficientemente grande que resultasse em uma redução substancial nas perdas;

Nos anos posteriores estas dificuldades foram superadas, principalmente pelo desenvolvimento dos modelos estatísticos vinculado ao advento de computadores de maior desempenho (nota do autor).

Atualmente, seu uso é muito difundido. O número de variáveis consideradas é muito maior e o tamanho das amostras utilizadas chegam a 100.000 unidades. Segundo Eisenbeis (1996), existem empresas norte-americanas especializadas no desenvolvimento de modelos para produtos específicos e também na venda de “escoragens” de crédito padronizadas para o mercado. Além de servirem para uma decisão de fornecer ou não o empréstimo, os modelos vêm sendo utilizados para o estabelecimento do valor e juros do empréstimo de acordo com seu risco e para se prever o faturamento proveniente da utilização de cartões de crédito.

A difusão da utilização de escoragens de crédito também é resultado de um esforço de automação do processo decisório de concessão de crédito, impulsionado por diversos fatores, entre eles:

- 1) A diminuição das margens de lucro tem forçado as instituições a diminuir seus custos e aumentar sua eficiência, resultando em um tratamento mais sistemático do processo;
- 2) A existência de grandes bancos de dados tem possibilitado os pesquisadores investigar padrões de omissão e não omissão de pagamento, aumentando a eficiência dos modelos;
- 3) Uma tendência de se estabelecer programas de crédito para pequenas empresas, o “microcrédito”. Segundo Eisenbeis (1996), esta é a tendência atual do mercado. Uma particularidade do desenvolvimento de escoragens de crédito para esta finalidade é uma maior precisão obtida com dados dos altos executivos destas pequenas empresas, principalmente dados referentes à utilização de seus cartões de crédito.

A seguir, veremos quais são os modelos estatísticos mais utilizados na análise de crédito.

3. MODELOS ESTATÍSTICOS

Como visto anteriormente, os modelos estatísticos aplicados à análise de crédito utilizam variáveis de predição provenientes de questionários preenchidos pelos candidatos ao empréstimo para estimar as probabilidades de omissão do pagamento.

No Anexo A, encontra-se um típico exemplar de ficha cadastral utilizada neste processo. A decisão de concessão ou não do empréstimo é tomada através da comparação da probabilidade estimada com uma probabilidade limite preestabelecida.

Dentre os diversos modelos existentes, apenas os seguintes serão descritos aqui: Análise Discriminante, Regressão Logística e Métodos de Programação Matemática, visto que são os mais utilizados. Faremos uma descrição superficial dos modelos, de modo a possibilitar a comparação entre as características básicas de cada um, o que será feito na Seção 3.6. A descrição detalhada do modelo de Regressão Logística será feita no Capítulo 4.

Antes que se prossiga, é conveniente que se faça uma descrição dos termos que serão utilizados tanto na Análise Discriminante como na Regressão Logística, em especial as variáveis de predição e “dummy”.

3.1 VARIÁVEIS DE PREDIÇÃO

Dada uma função da seguinte forma:

$$Y = f(X_1, X_2, X_3, \dots, X_n) \quad (3.1)$$

A variável Y é chamada de variável resposta e é função das variáveis $X_1, X_2, X_3, \dots, X_n$. Estas são chamadas de variáveis de predição porque irão determinar o valor da variável resposta Y . Considera-se que o resultado de uma variável seja independente do resultado de outra, de modo são chamadas de “variáveis independentes”.

De um modo geral, as variáveis independentes são divididas em dois grupos distintos:

- 1) Variáveis métricas: mensuráveis de algum modo. Podem ser de dois tipos:
 - a) Contínuas: obedecem a valores contínuos, sem interrupções. Exemplos: pressão sanguínea, idade, altura de um indivíduo, área de seu domicílio;
 - b) Discretas: obedecem a valores pontuais, geralmente resultados de contagens. Exemplos: número de cômodos e número de filhos de uma família;
- 2) Variáveis não-métricas: não podem ser quantificadas, portanto, não são mensuráveis. Podem ser de dois tipos:
 - a) Ordinais: embora não mensuráveis, é possível colocar os possíveis resultados em uma ordem determinada, crescente ou decrescente. Entretanto, a distância entre dois resultados possíveis é subjetiva. Exemplos: grau de satisfação do cliente (muito satisfeito, satisfeito, pouco satisfeito, insatisfeito), grau de escolaridade (alto, médio, baixo);
 - b) Nominiais ou categóricas: não é possível se estabelecer uma ordem entre os possíveis resultados. Exemplos: estado civil (casado, solteiro, outro), cor dos olhos (azul, verde, castanho, negro).

3.2 VARIÁVEIS "DUMMY"

As variáveis não-métricas podem assumir $k \geq 2$ resultados possíveis. É interessante transformar a variável não-métrica com $k > 2$ resultados possíveis em $k - 1$ variáveis "dummy", porém todas elas com apenas 2 resultados possíveis de modo a facilitar a interpretação dos coeficientes (ou pesos) na Análise Discriminante e Regressão Logística. Geralmente, utiliza-se resultados possíveis 0 ou 1.

Considere o seguinte exemplo onde uma variável nominal assume 4 resultados possíveis: A, B, C ou D. A transformação procederá do seguinte modo:

Tabela 3.1: Codificação através de variáveis "dummy". Fonte: Hosmer e Lemeshow (1989). Adaptado pelo autor

CATEGORIAS ($k = 4$)	VARIÁVEIS "DUMMY" ($K - 1 = 3$)		
	D1	D2	D3
A	0	0	0
B	1	0	0
C	0	1	0
D	0	0	1

Observa-se que a categoria A equivale às variáveis D1, D2 e D3 iguais a zero; a categoria B equivale à variável D1 = 1 e as demais iguais a zero e a categoria C equivale à variável D2 = 1 e as demais iguais a zero. Deste modo, o número mínimo de variáveis "dummy" para k categorias é sempre $k - 1$.

Procederemos agora à descrição do modelo de Análise Discriminante.

3.3 ANÁLISE DISCRIMINANTE

O valor da função discriminante, Y , é estimado através das informações de um conjunto de variáveis de predição $X_1, X_2, X_3, \dots, X_n$ segundo a expressão (3.1). A função $f(X_1, X_2, X_3, \dots, X_n)$ pode ser uma função linear, isto é:

$$f(X_1, X_2, X_3, \dots, X_n) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (3.2)$$

onde $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ são constantes ou pesos.

Ou seja, a variável resposta Y é função das variáveis do problema multiplicadas por seus respectivos coeficientes de ponderação. Poderá assumir apenas dois valores distintos: 0, caso a característica de interesse no problema esteja ausente ou 1, caso tal característica esteja presente.

Segundo Johnson e Wichern (1982), existem vários modelos de funções discriminantes. A mais comum é a função discriminante de Fischer (ou Regra de Fischer):

Dadas duas populações π_1 e π_2 , respectivamente representadas pelos vetores $\bar{x}_1 = [\bar{x}_{11} \ \bar{x}_{12} \ \dots \ \bar{x}_{1n}]'$ e $\bar{x}_2 = [\bar{x}_{21} \ \bar{x}_{22} \ \dots \ \bar{x}_{2n}]'$ onde $\bar{x}_{ij}$ é média da j -ésima característica da população i (exemplo: idade, altura, renda). Consideremos também que as matrizes de variância-covariância amostral S_1^2 e S_2^2 sejam as estimativas das variâncias e covariâncias das características dos desvios-padrão de cada população.

Sob a suposição de que possuem a mesma estrutura de variância-covariância, será construída uma matriz combinada S^2 , a *matriz variância-covariância ponderada* (vide Johnson e Wichern, 1982).

Segundo a Regra de Fischer, deve-se alocar um determinado elemento $x_0 = [x_{01} \ x_{02} \ \dots \ x_{0n}]'$ em π_1 se:

$$y_0 = (\bar{x}_1 - \bar{x}_2)' S^{-1} x_0 \geq \hat{m} = \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 + \bar{x}_2) \quad (3.3)$$

ou

$$y_0 - \hat{m} \geq 0$$

Deve-se alocar elemento $x_0 = [x_{01} \ x_{02} \ \dots \ x_{0n}]'$ para π_2 se:

$$y_0 < \hat{m} \quad (3.4)$$

ou

$$y_0 - \hat{m} < 0$$

Além deste modelo, existem outras que envolvem os custos de classificação (Johnson e Wichern, 1982) que não serão descritas aqui.

Devido à relação de dependência da variável resposta Y com as demais variáveis de predição, esta variável é chamada de "variável dependente".

Pelo fato de que assume apenas dois valores possíveis, esta variável é uma variável categórica binária ou dicotômica. Em consequência, a Análise Discriminante é amplamente utilizada como ferramenta de decisão - que em muitos casos resume-se a

escolher uma alternativa dentre duas alternativas possíveis. Também é frequentemente aplicada a situações onde o objetivo principal é identificar um grupo ao qual um objeto (uma pessoa, um produto ou uma empresa) pertence. Pode ser utilizada, por exemplo, para prever se uma empresa terá sucesso ou não em determinado empreendimento, classificar funcionários de uma empresa em áreas de interesse profissional ou até mesmo prever o sucesso ou fracasso de um produto novo a ser lançado no mercado. Seu uso é altamente difundido e está presente nos principais *softwares* estatísticos disponíveis no mercado.

A seguir, veremos um exemplo citado em Johnson e Wichern, 1982:

Consideremos duas populações em uma determinada cidade — População 1: habitantes que possuem cortador de grama, e População 2: habitantes que não possuem cortador de grama. Um fabricante de cortadores de grama deseja classificar as famílias da cidade em dois grupos distintos: potenciais compradores e não compradores de cortador de grama, de modo a direcionar sua campanha publicitária. Utilizou-se o seguinte conjunto aleatório de dados de renda anual (X_1) e tamanho do terreno da residência dos habitantes (X_2):

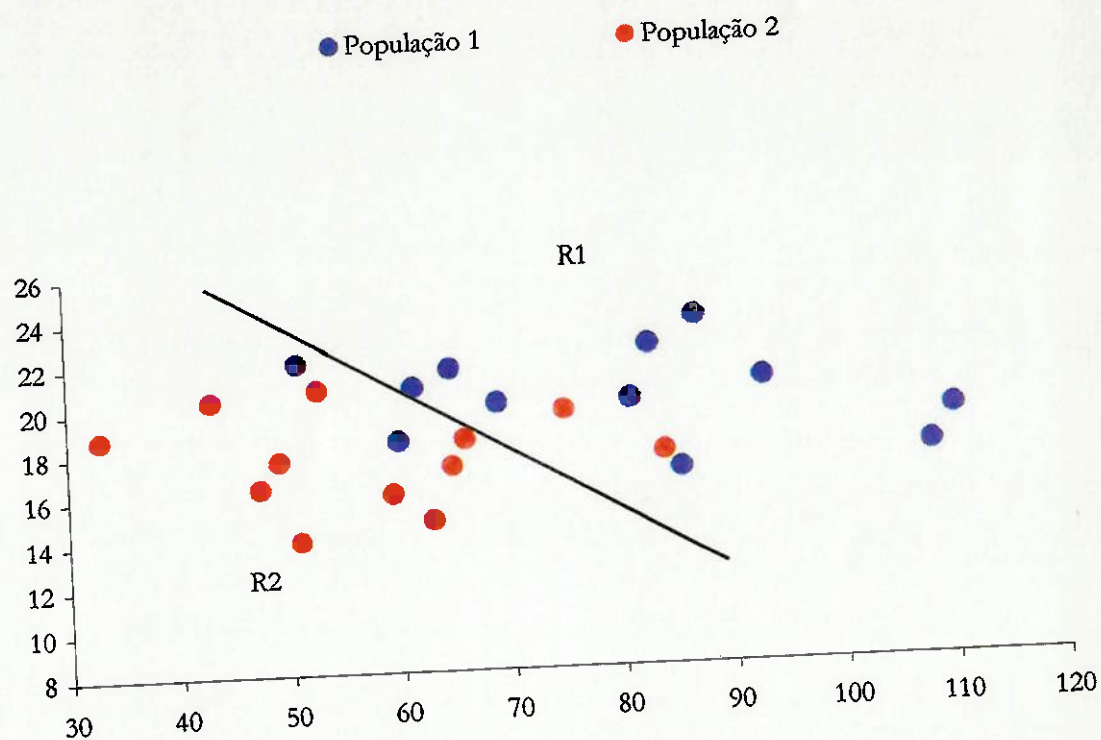
Tabela 3.2: Dados de renda anual e tamanho do lote dos habitantes de uma determinada cidade.

Fonte: Johnson e Wichern (1982)

População 1: habitantes que possuem o produto	
X1 (renda anual em milhares de dólares)	X2 (tamanho do terreno em milhares de pés quadrados)
60,0	18,4
85,5	16,8
64,8	21,6
61,5	20,8
87,0	23,6
110,1	19,2
108,0	17,6
82,8	22,4
69,0	20,0
93,0	20,8
51,0	22,0
81,0	20,0
População 2: habitantes que não possuem o produto	
X1 (renda anual em milhares de dólares)	X2 (tamanho do terreno em milhares de pés quadrados)
75,0	19,6
52,8	20,8
64,8	17,2
43,2	20,4
84,0	17,6
49,2	17,6
59,4	16,0
66,0	18,4
47,4	16,4
33,0	18,8
51,0	14,0
63,0	14,8

Estes valores estão ilustrados na Figura 3.1. Observa-se que compradores do produto tendem a ter renda anual e tamanho do terreno maiores do que não compradores.

Figura 3.1: Renda anual e tamanho do terreno de compradores e não compradores de cortador de grama. Fonte: Johnson e Wichern (1982)



Através da Regra de Fischer, é possível se estabelecer limites que, no caso de duas dimensões (renda e tamanho do terreno), é uma reta. De acordo com a Figura 3.1, a reta separa os elementos da amostra em não compradores e compradores do produto

(R_1 e R_2 , respectivamente). Deste modo, a classificação de elementos pode ser feita, além da Regra de Fischer, pela localização gráfica dos mesmos nas regiões R_1 e R_2 de acordo com as rendas anuais e tamanhos dos terrenos.

3.4 REGRESSÃO LOGÍSTICA

Na Regressão Logística, considera-se que a probabilidade π de ocorrência de um certo evento pode ser escrita como função das variáveis de predição $X_1, X_2, X_3, \dots, X_n$ dada por:

$$\pi(x_1, x_2, \dots, x_n) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}} \quad (3.5)$$

Para maior facilidade no manuseio de $\pi(x_1, x_2, \dots, x_n)$, utiliza-se a transformação logarítmica da função $\pi(x_1, x_2, \dots, x_n)$:

$$g(x_1, x_2, x_3, \dots, x_n) = \ln \left[\frac{\pi(x_1, x_2, x_3, \dots, x_n)}{1 - \pi(x_1, x_2, x_3, \dots, x_n)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3.6)$$

Para esta função, associa-se uma variável resposta Y , dependente e categórica binária, determinada do seguinte modo:

$$Y = 1, \text{ se } \pi(X_1, X_2, X_3, \dots, X_n) \geq \alpha \quad (3.7)$$

$$Y = 0, \text{ se } \pi(X_1, X_2, X_3, \dots, X_n) \leq \alpha$$

$$\text{dado } 0 \leq \alpha \leq 1$$

Atualmente é muito utilizada em modelos econométricos, previsões de crescimento de economias e modelos de desempenho de uma empresa no mercado de acordo com diversas estratégias possíveis de serem adotadas. Está disponível nos principais *softwares* estatísticos disponíveis no mercado.

A seguir, veremos um exemplo de sua aplicação na área da Medicina. Apesar de a área desta aplicação não ser diretamente relacionada à Engenharia de Produção, o exemplo a seguir mostra o tratamento estatístico de variáveis de predição, o que é feito de modo semelhante em aplicações do método na Engenharia em geral, além de ilustrar a ampla gama de aplicações do método.

Exemplo: (Hosmer e Lemeshow, 1989) Realizou-se um estudo para se identificar fatores de risco associados à gestação de bebês de baixo peso (pesando menos do que 2.500 gramas). O estudo foi realizado em 189 mulheres, das quais 59 deram luz à bebês de baixo peso e 130 deram luz à bebês normais. Considerou-se quatro variáveis que se poderiam ser significativas: idade (variável IDA), peso da mãe após seu último período menstrual (variável PESO), raça (variáveis “dummy” D1 e D2) e números de visitas ao ginecologista durante o primeiro trimestre de gravidez (variável VPT). O número de elementos na amostra é significativamente grande e, em consequência, o conjunto de dados não será colocado aqui.

Como visto anteriormente, as variáveis “dummy” criadas para a variável RAÇA são as seguintes:

Tabela 3.3: Variáveis "dummy" utilizadas para descrever o comportamento da variável RAÇA.

Fonte: Hosmer e Lemeshow (1989)

RAÇA	Variável "Dummy"	
	D1	D2
Branca	0	0
Negra	1	0
Outra	0	1

A Regressão Logística utiliza métodos que medem o nível de significância de cada variável no modelo formulado. No caso específico, isto teria o objetivo de determinar quais dos fatores efetivamente são determinantes para a gestação de bebês de baixo peso.

O procedimento de estimação dos parâmetros (coeficientes) da Regressão Logística é feito através de um método chamado de Máxima Verossimilhança¹, que permite que se chegue à seguinte equação para o presente modelo (no momento, a descrição do processo será omitida, visando a simplificação do exemplo):

$$g(IDA, PESO, D1, D2, VPT) = 1.295 - 0.024 \times IDA - 0.014 \times PESO + 1.004 \times D_1 + 0.433 \times D_2 - 0.049 \times VPT \quad (3.8)$$

¹ Estimadores de máxima verossimilhança são estimadores que geram valores para parâmetros desconhecidos da formulação da Regressão Logística e que maximizam a probabilidade de se obter valores semelhantes ao da amostra.

A função probabilidade em função da idade, peso, raça e número de visitas se torna é dada por:

$$\pi(IDA, PESO, D1, D2, VPT) = \frac{e^{1.295 - 0.024 \times IDA - 0.014 \times PESO + 1.004 \times D_1 + 0.433 \times D_2 - 0.049 \times VPT}}{1 + e^{1.295 - 0.024 \times IDA - 0.014 \times PESO + 1.004 \times D_1 + 0.433 \times D_2 - 0.049 \times VPT}} \quad (3.9)$$

Observa-se que todas as variáveis examinadas no problema têm influência na gestação de bebês de baixo peso, em maior ou menor grau. Quando aplicadas à equação, fornecem um determinado resultado que demonstra a probabilidade de que o bebê vá nascer com baixo peso. Segundo a mesma, quanto maiores forem a idade e o peso antes do último período menstrual das mulheres, menor é a chance de que estas dêem luz a bebês considerados de baixo peso e mulheres da raça negra e outras têm mais chances de darem luz à bebês de baixo peso do que mulheres da raça branca.

3.5 MÉTODOS DE PROGRAMAÇÃO MATEMÁTICA

Métodos de programação matemática podem ser utilizados quando se tem uma função objetivo a se maximizar ou minimizar.

No caso da análise de crédito, deseja-se minimizar o risco de empréstimo. Segundo modelo proposto por Kolesar e Showers (1985), consideram-se dois fatores que representam a capacidade de pagamento do candidato no futuro:

1) Risco θ associado à concessão de crédito ao candidato ao empréstimo:

$$\theta = \begin{Bmatrix} b \\ a \end{Bmatrix}, \text{ onde} \quad \begin{array}{l} b : \text{baixo risco} \\ a : \text{alto risco} \end{array} \quad (3.10)$$

2) Decisão d a ser tomada em relação a cada candidato:

$$d = \begin{Bmatrix} c \\ r \end{Bmatrix}, \text{ onde} \quad \begin{array}{l} c : \text{concessão} \\ r : \text{rejeição} \end{array} \quad (3.11)$$

Considere a *função perda*, $L(\theta, d)$, que representa a consequência da tomada de decisão d quando o verdadeiro estado do candidato é θ :

$$L(\theta, d) = \begin{cases} \gamma > 0 & \text{se } \theta = b \text{ e } d = r \\ \delta > 0 & \text{se } \theta = a \text{ e } d = c \\ 0, & \text{caso contrário} \end{cases} \quad (3.12)$$

onde γ, δ são respectivamente as perdas de se rejeitar crédito a um candidato de baixo risco e se conceder crédito a um candidato de alto risco.

Considera-se agora uma *regra de decisão* $d(x)$ a ser tomada quando o perfil x do candidato é observado. O desempenho desta regra é quantificado por uma *função risco* $R(\theta, d)$ que reflete o custo esperado da tomada de decisão $d(x)$ quando o verdadeiro estado do candidato é θ .

Considere π_b a probabilidade de um candidato de baixo risco de empréstimo.

Se denotarmos por $M_b(d)$ e $M_a(d)$ as probabilidades de classificação incorreta de um candidato sob a regra de decisão $d(x)$, e utilizando a função perda (3.12), o risco esperado é dado por:

$$R(\theta, d) = \begin{cases} \gamma \cdot M_b(d), & \text{para } d = r \\ \delta \cdot M_a(d), & \text{para } d = c \end{cases} \quad (3.13)$$

Portanto, a primeira função objetivo a ser minimizada é $\min f(d)$:

$$f(d) = \gamma \cdot \pi_b \cdot M_b(d) + \delta \cdot \pi_a \cdot M_a(d) \quad (3.14)$$

onde $\pi_a = 1 - \pi_b$

As probabilidades $M_b(d) \leq \alpha$ e $M_a(d) \leq \beta$, tal que $\alpha, \beta \in [0,1]$, geram a segunda e a terceira função objetivo:

$$\min M_a(d), \text{ sujeito a } M_b \leq \alpha \quad (3.15)$$

$$\min M_a(d), \text{ sujeito a } \pi_b \cdot M_b(d) + \pi_a \cdot [1 - M_a(d)] \leq \beta \quad (3.16)$$

A seguir, veremos um exemplo de sua aplicação, proposto por Kolesar e Showers (1985).

Exemplo: Showers e Chacrin (1981) apud Kolesar e Showers (1985) realizaram um extenso estudo sobre a empresa norte-americana AT&T, considerando a concessão de linha telefônica a consumidores, utilizando três diferentes métodos (Análise Discriminante e mais outros dois métodos desenvolvidos especificamente para o problema; omitiremos a descrição dos mesmos pois fogem do escopo deste trabalho).

Neste estudo, analisou-se uma amostra de 87.000 consumidores que requisitaram à empresa uma linha telefônica. As variáveis do problema foram fornecidas pelos candidatos (exemplo: posse ou não de cartão de crédito, renda anual, número de anos na residência atual).

Posteriormente, Kolesar e Showers (1985), utilizando a mesma amostra deste estudo, realizaram outro estudo utilizando método de programação matemática. Utilizando as funções objetivo (3.14), (3.15) e (3.16), gerou-se uma escoragem de crédito apropriada. Algumas variáveis da amostra que haviam sido consideradas no estudo anterior não foram consideradas; outras variáveis que não haviam sido consideradas anteriormente foram incluídas. Apesar desta modificação, o resultado obtido - em termos de porcentagem de candidatos classificados de modo incorreto - foi muito parecido com o obtido no estudo anterior.

Agora, iremos discutir e comparar os três modelos ora apresentados.

3.6 COMPARAÇÃO ENTRE OS MÉTODOS

Segundo Hand e Henley (1997), em geral, não existe um método ideal. O melhor método para um determinado problema depende de seus detalhes, como características utilizadas, estrutura da informações disponíveis ou até que ponto se consegue formar categorias de classificação utilizando as informações disponíveis. Além disso, a precisão da classificação não é o único aspecto relevante na escolha de um método a ser utilizado. Outros fatores também são importantes, como a velocidade de classificação, velocidade pela qual o processo pode ser revisto, a facilidade de se compreender o método e porque, através dele, se chegou a determinada conclusão.

3.6.1 MÉTODOS DE PROGRAMAÇÃO MATEMÁTICA

A aplicação dos métodos de Programação Matemática em problemas de tomada de decisão é relativamente recente. Alguns trabalhos, como o de Hand (1981), descrevem como utilizar estes métodos para maximizar um fator crítico - no caso, uma função linear da soma dos fatores correspondentes aos candidatos classificados corretamente. Entretanto, segundo Kolesar e Showers (1985), devido ao fato de este métodos gerarem soluções determinísticas, geralmente chega-se a sistemas de equações de difícil resolução. Outro problema é o grande volume de dados, o que pode tornar o sistema de equações inviável. Este modelo não está disponível nos *softwares* estatísticos existentes no mercado.

Por outro lado, a Regressão Logística e a Análise Discriminante são muito utilizados pela relativa facilidade de estruturação do modelo e, conseqüentemente, os mais desenvolvidos. Compararemos a seguir estes dois métodos.

3.6.2 ANÁLISE DISCRIMINANTE X REGRESSÃO LOGÍSTICA

Os primeiros trabalhos sobre Análise Discriminante datam da década de 40, quando Durand (1941) demonstrou que este método produz boas previsões de pagamento de empréstimos. Entretanto, Eisenbeis (1996) apresentou a primeira crítica à aplicação econômica da mesma: "Uma das principais suposições na Análise

Discriminante é que as variáveis independentes que descrevem os elementos dos grupos que estão sendo medidos são *multivariadas e normalmente distribuídas*".

Na realidade, considera-se que as variáveis de predição tenham uma distribuição normal (como, por exemplo, uma variável do tipo "expectativa média de vida de um determinado equipamento (anos)"). Porém, muitas vezes isto não ocorre (por exemplo, a distribuição de renda da população da Grande São Paulo). Observa-se também que a utilização de variáveis não-métricas - que evidentemente não obedecem a uma distribuição normal - viola esta condição. Este é um problema muito comum, uma vez que muitos fenômenos presentes na natureza e que são utilizados em modelos estatísticos são categóricos e não obedecem a uma distribuição normal, como no caso de um produto ter sido comprado ou não, uma pessoa ter boa saúde ou não, uma empresa ter falido ou não.

Quando ocorre esta utilização indevida, os estimadores da Análise Discriminante tornam-se inconsistentes, perdendo-se a eficiência de predição, até mesmo em grandes amostras. Logo, apesar de amplamente utilizada e estar presente nos principais *softwares* estatísticos disponíveis no mercado, existem objeções quanto a sua aplicação.

O modelo de Regressão Logística tem sido utilizado durante anos; porém apenas Truett, Cornfield e Kannel (1967) apud Press e Wilson (1978) conseguiram demonstrar uma aplicação de todo seu potencial, em um estudo sobre doenças cardíacas. A partir desta publicação - considerada um marco na utilização da Regressão Logística - a literatura sobre o método cresceu rapidamente, paralelamente à diversidade de sua aplicação.

Na Regressão Logística, não há restrições quanto à distribuição das variáveis, inexistindo impecilhos quanto ao uso de variáveis não-métricas através de variáveis de substituição métricas. Considera-se que: 1) as variáveis independentes sejam dicotômicas ou 2) que algumas sejam dicotômicas e outras obedeçam a uma distribuição normal. Portanto, uma de suas vantagens é que o modelo é relativamente flexível: considerações diferentes geram uma mesma formulação logística.

Press e Wilson (1978), em estudo realizado em problemas classificatórios de dois grupos envolvendo variáveis mistas (variáveis métricas binárias e contínuas) demonstrou empiricamente que os estimadores de máxima verossimilhança utilizados na Regressão Logística são mais eficientes do que os estimadores da Análise Discriminante, ou seja, a Regressão Logística produz uma classificação mais precisa do que a Análise Discriminante.

O estudo mostra ainda que os tempos de compilação e execução computadorizada dos modelos de Regressão Logística são entre 1 e 2 vezes maiores do que os modelos de Análise Discriminante. Entretanto, o estudo foi realizado na década de 70 quando não existiam os computadores de alto desempenho atuais. Portanto, é possível que este argumento não seja mais aplicável aos dias atuais.

A seguir, veremos outros dois estudos comparativos entre os métodos, também apresentados por Press e Wilson (1978). O primeiro exemplo é uma aplicação dos métodos na área da Medicina, e o segundo, na área Social. Apesar das aplicações em áreas distintas da Engenharia, os exemplos são muito valiosos ao comparar diretamente os dois métodos, além de também ilustrarem a ampla gama de aplicações dos modelos estatísticos.

Estudo 1: Realizou-se um estudo no British Columbia Cancer Institute, Canadá, entre 1955 e 1963, com o objetivo de classificar pacientes com câncer de mama a partir de evidências clínicas e históricas dos mesmos. A classificação tinha o objetivo de medir a extensão de metástases nodais na formação do câncer de mama. As variáveis do estudo eram mistas, ambas contínuas ou discretas, sendo que algumas eram binárias. Utilizou-se a Regressão Logística e a Análise Discriminante para a modelagem do problema.

Observaram-se 173 pacientes com câncer de mama - cujas metástases nodais foram determinadas cirurgicamente - e com um histórico clínico completo. Os pacientes

foram divididos aleatoriamente em dois grupos: o primeiro, com 115 pacientes foi utilizado para os procedimentos de classificação (conjunto de treinamento²), e o segundo, com 58 pacientes foi utilizado para validar as funções de estimação do primeiro grupo (conjunto de validação).

A variável dependente de cada modelo foi definida como 0, caso os nódulos linfáticos não estivessem associados com a metástase, e 1, caso estivessem. As variáveis independentes utilizadas foram: número de filhos, histórico de histerectomia (0 ou 1), histórico de doenças de mama benignas durante o período de lactação (0 ou 1) e duração dos sintomas em meses. Portanto, três variáveis independentes eram binárias, uma categórica e uma contínua. Os resultados obtidos foram:

Tabela 3.4: Resumo das classificações de pacientes com câncer de mama por Regressão Logística e Análise Discriminante. Fonte: Press e Wilson (1978)

Dados	Grupo Real	Análise Discriminante			Regressão Logística		
		0	1	% de Classificação	1	% de Classificação	
Conjunto de Treinamento	0	71	5	67	65	11	71
	1	33	6		22	17	
	Total	104	11		87	28	
Conjunto de Validação	0	31	0	59	25	6	62
	1	24	3		16	11	
	Total	55	3		41	17	

² Ver Seção 1.7.1

A Regressão Logística classificou corretamente 82 pacientes (65+17) dos 115 pacientes do conjunto de treinamento o que equivale a uma porcentagem de classificação de 71.3%. De modo análogo, a Análise Discriminante teve uma porcentagem de classificação no conjunto de treinamento de 62.07%, inferior à logística, o que também ocorreu no conjunto de validação.

O estudo também forneceu dados que permitiram a conclusão de que a ausência de metástases nodais estava associada com um grande número de filhos e também com o histórico de histerectomia. Uma duração dos sintomas mais longa estava levemente associada à ausência de metástases. A presença de metástases nodais estava associada com um histórico de doenças de mama benignas durante a lactação e mudanças no mamilo como primeiro sintoma da doença.

Estudo 2: Um estudo demográfico realizado nos Estados Unidos na década de 70 coletou informações procurando explicações sobre mudanças populacionais. As informações foram obtidas de um censo de 1973. A mudança percentual da população entre o censo de 1960 e o de 1973 para cada estado foi codificada como 0 ou 1, se a mudança foi abaixo ou acima da mudança média para todos os estados. As variáveis independentes utilizadas para o ano de 1973 foram: renda per capita (em milhares de dólares americanos), taxa de natalidade (percentual), taxa de mortalidade (percentual), urbanização da população (0 ou 1 se a população é menor ou maior do que 70% urbanizada) e ausência ou presença de linha costeira (0 ou 1). As variáveis independentes eram, portanto, contínuas (três primeiras) e binárias (duas últimas, respectivamente). O conjunto de dados não será mostrado aqui, devido à grande quantidade de elementos.

Dividiram-se os 50 estados em 5 grupos de 10 estados cada. Dividiu-se os grupos em dois grupos distintos: o primeiro, com 4 grupos, para ser o conjunto de treinamento; o segundo, com o grupo restante, para ser o conjunto de validação. Os resultados obtidos foram:

*Tabela 3.5: Porcentagem de classificação correta de Regressão Logística e de Análise Discriminante.**Fonte: Press e Wilson (1978)*

Método	%de classificação correta	
	Conjunto de Treinamento	Conjunto de Validação
Regressão Logística	80	70
Análise Discriminante	60	68

Demonstrou-se que a os estados com população em crescimento tinham taxas de natalidade mais baixas e taxas de mortalidade mais baixas do que os estados com população em decrescimento. Também demonstrou-se que o crescimento populacional estava associado a uma maior renda per capita e à presença de linha costeira, o que levou os pesquisadores à conclusão que os norte-americanos, na época, estavam procurando um bom nível de vida - maiores salários e um ambiente agradável.

3.6.3 ESCOLHA DO MÉTODO MAIS APROPRIADO

Tendo em vista as seguintes características do modelos apresentadas nos itens anteriores quanto à

- 1) eficiência;
- 2) violação de preceitos básicos;
- 3) possibilidade de utilização do modelo no problema;

- 4) características do problema - problema decisório do tipo "sim ou não" (ser ou não comprador potencial do produto), com variáveis mistas,

considerou-se neste trabalho a Regressão Logística um modelo mais apropriado e adequado para problemas envolvendo análise de decisão.

Veremos agora o modelo de Regressão Logística em maiores detalhes do que visto até o momento, explicando seu funcionamento e, em seguida, aplicando-o em um problema real.

4. REGRESSÃO LOGÍSTICA

Métodos de Regressão tem sido parte de praticamente todos os trabalhos publicados que descrevem a relação entre uma variável de resposta e uma ou mais variáveis explicativas. O objetivo do método é encontrar o modelo mais apropriado para se descrever a relação entre uma variável de resposta (variável dependente) e um conjunto de variáveis explicativas (variáveis independentes).

Veremos com mais detalhes, inicialmente, a Regressão Logística para um modelo com apenas uma variável independente e, na Seção 4.3, para um modelo com várias variáveis independentes.

4.1 REGRESSÃO LOGÍSTICA COM UMA VARIÁVEL INDEPENDENTE

Conforme dito anteriormente, na Regressão Logística considera-se que a probabilidade $\pi(x)$ de ocorrência de uma variável independente x em relação a um determinado vetor de coeficientes $\beta = (\beta_0, \beta_1)'$ seja expressa por:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (4.1)$$

Além disso, para maior facilidade no manuseio de $\pi(x)$, utiliza-se a função logaritmo de modo a tornar a função linear em função de x :

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x \quad (4.2)$$

A importância desta transformação é que $g(x)$ tem propriedades lineares desejáveis, que podem ser contínuas e variar de variar de $-\infty$ até $+\infty$, dependendo do valor de x .

Suponhamos agora que, no exemplo da empresa AT&T da Seção 3.5, se tenha o banco de dados utilizado na elaboração do modelo para o problema. Este banco de dados conteria os diversos elementos da amostra (no caso, os candidatos à linha telefônica) e que chamaremos de x_i , e também outro dado - que chamaremos de y_i - associado a cada variável x_i e que indica se o indivíduo adquiriu ou não a linha telefônica.

Portanto, a variável y_i é uma variável binária e consideramos que ela assume apenas os valores 0 (caso o candidato x_i tenha adquirido a linha telefônica) e 1 (caso o candidato x_i não tenha adquirido).

Seja $\pi(x_i)$ a probabilidade de o candidato i adquirir a linha telefônica. Segundo uma distribuição de Bernoulli, pode-se enunciar o problema do seguinte modo:

- a) Se y_i é igual a 1, a probabilidade de o candidato x_i adquirir o produto será $\pi(x_i)$;
- b) Se y_i é igual a 0, a probabilidade de o candidato x_i não adquirir o produto será $1 - \pi(x_i)$.

Uma maneira conveniente de se expressar estas probabilidades para um determinado par (x_i, y_i) é através do termo $\zeta(x_i)$ abaixo, chamado de *função de verossimilhança*.

$$\zeta(x_i) = \pi(x_i)^{y_i} \cdot [1 - \pi(x_i)]^{1-y_i} \quad (4.3)$$

Assim, em um determinado conjunto de dados, têm-se uma determinada probabilidade de ocorrência dos n diversos pares (x_i, y_i) , que denotaremos como $P(x_1, y_1; x_2, y_2; \dots; x_n, y_n)$. Como cada par (x_i, y_i) é um evento probabilístico independente, temos que:

$$P(x_1, y_1; x_2, y_2; \dots; x_n, y_n) = P(x_1, y_1) \cdot P(x_2, y_2) \cdot \dots \cdot P(x_n, y_n) \quad (4.4)$$

A expressão (4.4) pode ser rescrita do seguinte modo:

$$l(\beta) = \prod_{i=1}^n \zeta(x_i) \quad (4.5)$$

A expressão (4.5) é conhecida como *função de verossimilhança*. E o *estimador de verossimilhança* é um determinado vetor de coeficientes $\beta = (\beta_0, \beta_1)^t$ tal que maximiza (4.5). Maximizar (4.5) equivale a maximizar $L(\beta) = \ln[l(\beta)]$, como:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (4.6)$$

Esta expressão chama-se *função de log-verossimilhança* ("log-likelihood"). Para se determinar o vetor $\beta = (\beta_0, \beta_1)'$ que a maximiza, diferencia-se $L(\beta)$ em relação a β_0 e β_1 e iguala-se a 0, resultando nas seguintes equações:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (4.7)$$

e

$$\sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0 \quad (4.8)$$

Estas equações são chamadas de *equações de verossimilhança* e as soluções das mesmas é o estimador de máxima verossimilhança. Na Regressão Logística, estas equações não são lineares, de modo que as equações (4.7) e (4.8) necessitam de métodos especiais para sua resolução. Geralmente, estes métodos são iterativos e, portanto, necessitam de *softwares* para sua resolução - que comumente utilizam o método de Newton-Raphson, cujos detalhes de funcionamento serão omitidos.

Retomando o vetor $\beta = (\beta_0, \beta_1)'$ que da equação (4.6), denota-se por $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ o vetor que a maximiza e, conseqüentemente, o símbolo $\hat{}$ é utilizado para denotar o estimador de máxima verossimilhança.

4.1.1 TESTE DE SIGNIFICÂNCIA DOS COEFICIENTES

Após a obtenção da estimativa dos coeficientes associados às variáveis independentes, segundo Hosmer e Lemeshow (1989), é interessante testar a sua significância no modelo. Isto pode ser feito de duas maneiras. A primeira é comparar os valores da função de log-verossimilhança do modelo (Equação 4.6) com e sem a presença da variável independente. A segunda é através do Teste de Wald, que consiste em testar $H_0: \beta_1 = 0$. Como os estimadores de máxima verossimilhança são assintoticamente normais, a estatística do teste é dada por:

$$Z = \frac{\hat{\beta}_1}{\sigma(\hat{\beta}_1)} \quad (4.9)$$

Rejeita-se $H_0: \beta_1 = 0$ se módulo de Z for menor que Z ao nível de $\alpha/2$.

Neste presente trabalho, o Teste de Wald será utilizado para identificar a significância das variáveis independentes.

4.2 REGRESSÃO LOGÍSTICA COM DIVERSAS VARIÁVEIS INDEPENDENTES

Generalizando a Regressão Logística com uma única variável independente para diversas variáveis independentes, considera-se que a probabilidade $\pi(x_1, x_2, \dots, x_n)$ de ocorrência de um conjunto de variáveis $x' = (x_1, x_2, \dots, x_n)$ em relação a um determinado vetor de coeficientes $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$ seja expressa por:

$$\pi(x_1, x_2, \dots, x_n) = \frac{e^{g(x_1, x_2, \dots, x_n)}}{1 + e^{g(x_1, x_2, \dots, x_n)}} \quad (4.10)$$

onde

$$g(x_1, x_2, \dots, x_n) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p = \ln \left[\frac{\pi(x_1, x_2, \dots, x_n)}{1 - \pi(x_1, x_2, \dots, x_n)} \right] \quad (4.11)$$

Caso se utilize $k_j - 1$ variáveis "dummy"³ para a j -ésima variável independente, denotadas por D_{ju} , a equação (4.11) se torna:

³ Ver Seção 3.2

$$g(x_1, x_2, \dots, x_n) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \sum_{u=1}^{k_j-1} \beta_{ju} D_{ju} + \beta_p x_p \quad (4.12)$$

onde $\beta_{ju}, u = 1, 2, \dots, k_j - 1$ são os coeficientes das variáveis “dummy” D_{ju} .

A estimação dos parâmetros também é feita através da máxima verossimilhança, assim como o teste de significância dos coeficientes também é feito pelo Teste de Wald.

Segundo Hosmer e Lemeshow (1989), uma atenção especial deve ser dada ao caso de regressão com várias variáveis independentes. Semelhantemente à Regressão Linear, deve-se traçar estratégias para escolha do modelo ótimo segundo algum critério.

No Capítulo 5, iremos apresentar um exemplo prático do modelo de Regressão Logística e no Capítulo 6 está detalhada a estratégia utilizada para escolher o modelo ótimo.

5. APLICAÇÃO DO MODELO

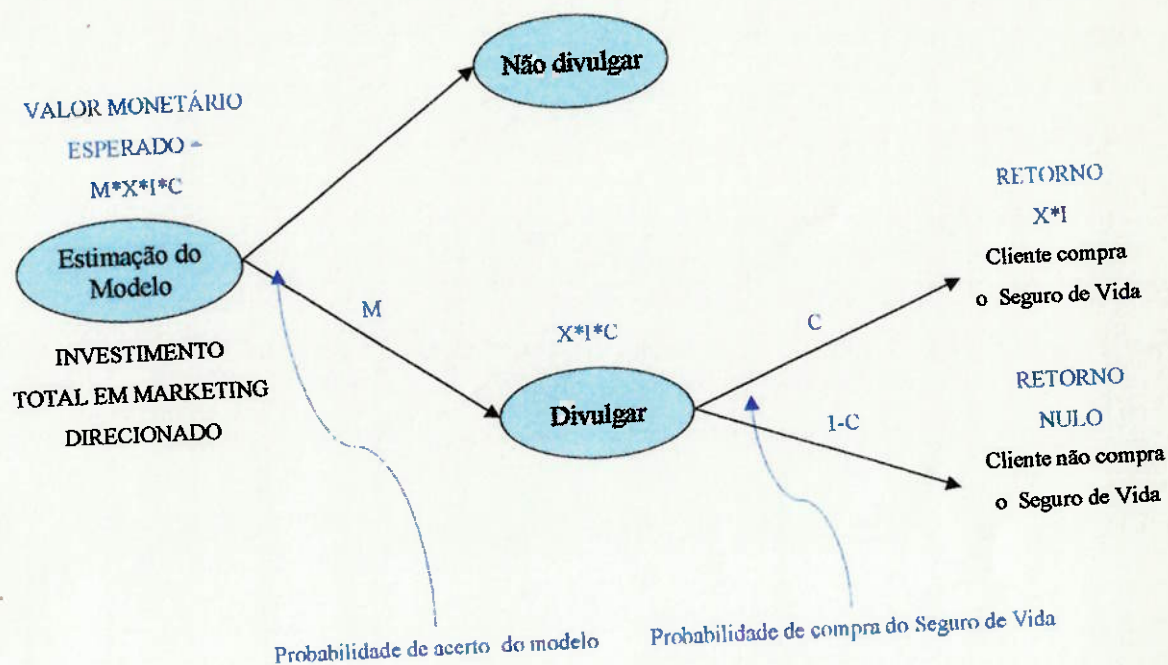
O presente trabalho apresenta a aplicação do modelo de Regressão Logística em um problema encontrado em uma empresa X de cartões de crédito. Ela é uma administradora de cartões de crédito multinacional e de grande porte, com mais de um milhão de clientes em todo o Brasil. Dentre os produtos que oferece, destaca-se o Seguro de Vida, que já existe há alguns anos e é divulgado a todos os clientes que utilizam o cartão de crédito. Como a empresa tem um número muito grande de clientes, a divulgação do produto para todos eles é muito custosa e lenta. Assim, necessita-se saber quais clientes são potenciais compradores do Seguro de Vida para que então seja feito um programa de Marketing direcionado, como por exemplo o envio de mala-direta.

A empresa possui os dados pessoais de cada cliente que adquiriu o Seguro de Vida, provenientes das fichas cadastrais preenchidas por cada um ao adquirir seu cartão de crédito. Iremos utilizar a Regressão Logística para criar um modelo que possa ser utilizado no conjunto total de clientes da empresa no Brasil procurando identificar potenciais compradores do Seguro de Vida.

O modelo matemático irá estimar com uma determinada probabilidade quais são os potenciais compradores do produto, direcionando o investimento de Marketing para estes indivíduos. Apesar de não dispormos de dados a respeito do total a ser investido, faremos a seguir uma *ilustração* da situação através de um diagrama de árvore de decisão.

Após a divulgação do Seguro de Vida, existe a possibilidade de o indivíduo comprar o produto ou não. Deste modo, segundo o critério do Valor Monetário Esperado da Seção 1.4.2, teremos o seguinte modelo:

Figura 5.1: Árvore de decisão do problema da empresa X. Elaborado pelo autor



onde

M : Probabilidade de acerto do modelo, ou seja, a confiabilidade

C : Probabilidade de que os indivíduos comprem o produto

I : Investimento total do Marketing direcionado

Sendo $X \cdot I$ a estimativa de retorno do investimento total (ou seja, o quanto a empresa estima que terá de retorno investindo uma determinada quantia I), onde X é um valor maior do que 1 para que o investimento valha a pena, e C a probabilidade de que um indivíduo tendo recebido a divulgação do produto vá comprá-lo, o retorno esperado será $C \cdot X \cdot I$. Entretanto, existe a probabilidade de acertos M do modelo, e portanto espera-se que o Valor Monetário Esperado do investimento I seja

$M \cdot C \cdot X \cdot I$, como mostra a Figura 5.1. Observa-se que quanto maior é M , maior será o Valor Monetário Esperado para a empresa.

O problema da Empresa X, portanto, resume-se a classificar clientes em compradores potenciais ou não - o que é semelhante a um problema de crédito já que seu objetivo é classificar clientes alto ou baixo risco de empréstimo. Ou seja, o objetivo dos dois problemas é o de classificar elementos em dois grupos distintos, para servir de base para uma decisão de duas alternativas, mutuamente exclusivas (no caso do Seguro de Vida, divulgar ou não o produto).

5.1 CONJUNTO DE DADOS

A empresa X forneceu um conjunto de dados reais provenientes dos formulários que são preenchidos quando os clientes adquirem o cartão de crédito. O conjunto de dados foi coletado aleatoriamente entre os clientes da empresa, contendo assim indivíduos que compraram e indivíduos que não compraram o Seguro de Vida. Os dados encontram-se no Anexo A.

O conjunto de dados tem **2.134 elementos** dos quais coletou-se as seguintes informações:

- 1) Sexo
- 2) Renda
- 3) Idade
- 4) Profissão
- 5) Código de Endereçamento Postal (CEP)
- 6) Compra do produto

Veremos cada item anterior em detalhes na Seção 5.2. As tabelas utilizadas na confecção dos gráficos, bem como a análise estatística completa de cada item, encontram-se nos Anexos A e B respectivamente.

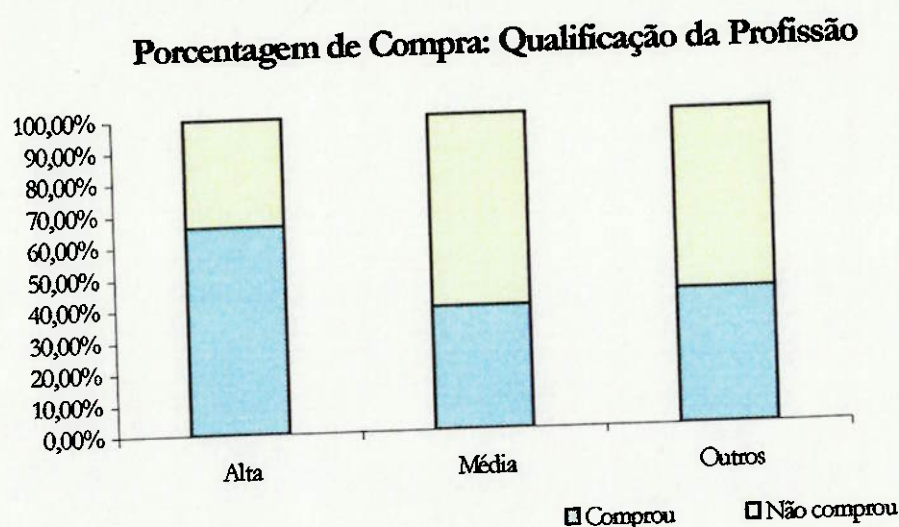
5.2 ANÁLISE DESCRITIVA DOS DADOS

Antes de utilizarmos o modelo de Regressão Logística propriamente dito, é indispensável que observemos o comportamento dos dados do conjunto, através de gráficos e tabelas (como, por exemplo, uma correlação entre idade e compra ou não do produto). Isto permitirá uma maior familiaridade com o conjunto.

Posteriormente, após realizado o modelo de Regressão Logística, deve-se verificar se os dados obtidos no modelo são compatíveis com os dados observados graficamente neste estudo prévio.

Primeiramente, o dado de compra informa se o indivíduo comprou ou não o Seguro de Vida: 1 – comprou e 0 – não comprou. Observa-se na Figura 5.2 que 43,35% dos elementos do conjunto de dados comprou o produto:

Figura.5.2: Porcentagem de compra dos elementos do conjunto de dados. Elaborado pelo autor



Observa-se que a porcentagem de elementos do conjunto de dados que compraram o Seguro de Vida é menor do que a porcentagem de elementos que não o compraram. Entretanto, a diferença entre as porcentagens não é muito grande.

Quando ao sexo, observa-se na Figura 5.3 que 71,32% dos elementos da amostra pertence ao sexo masculino. Além disso, observa-se nas Figuras 5.4 e 5.5 que 59,66% dos elementos do sexo masculino comprou o produto, enquanto que entre as mulheres apenas 50,65% comprou.

Figura 5.3: Porcentagem dos elementos do conjunto entre os sexos. Elaborado pelo autor

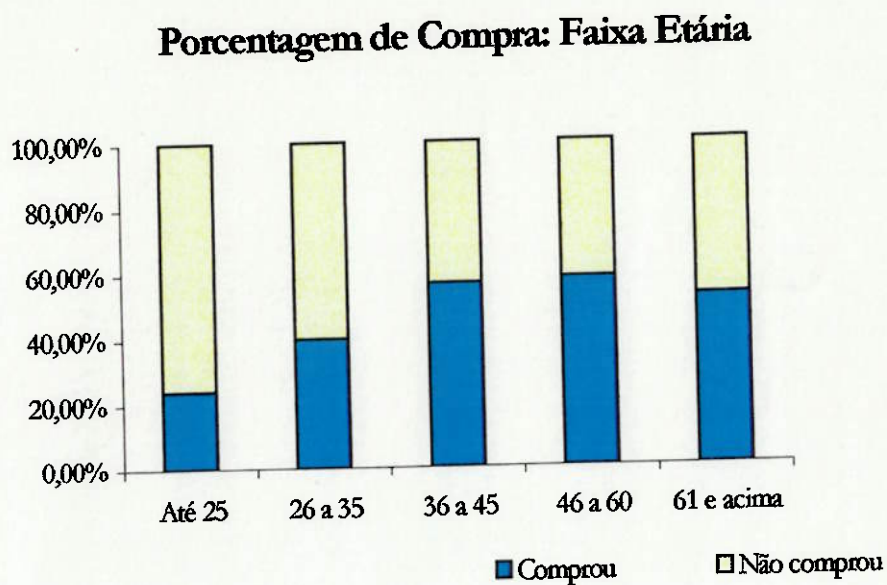


Figura 5.4: Porcentagem de compra dos elementos do conjunto de dados de sexo masculino. Elaborado pelo autor

Divisão Percentual da Amostra entre Regiões

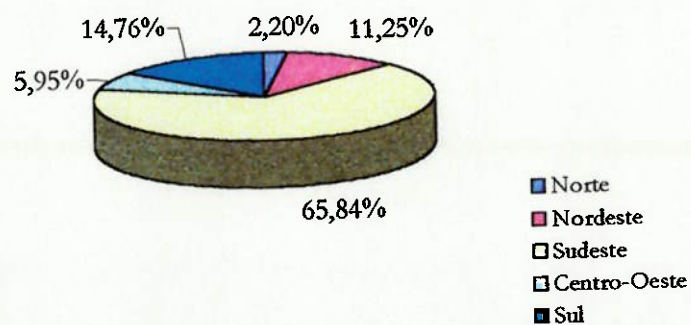
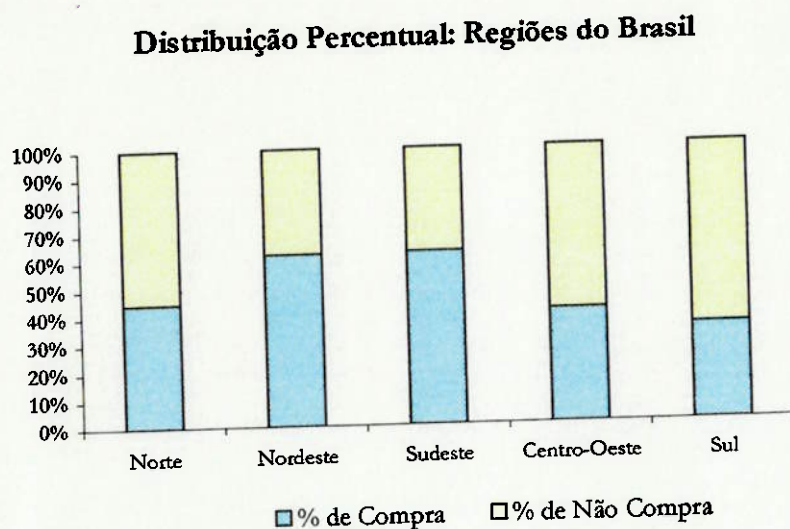
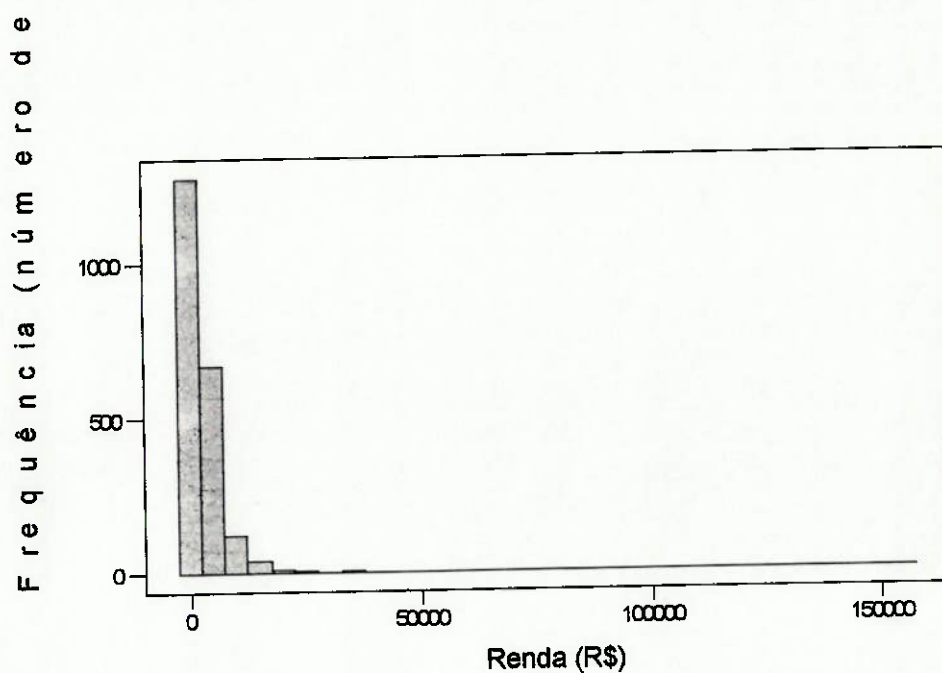


Figura 5.5: Porcentagem de compra dos elementos do conjunto de dados do sexo feminino. Elaborado pelo autor



Quanto à renda, pode-se observar pela Figura 5.6 que grande parte dos elementos do conjunto de dados está situada na faixa de até R\$10.000,00:

Figura 5.6: Histograma de Renda dos indivíduos do conjunto de dados. Elaborado pelo autor

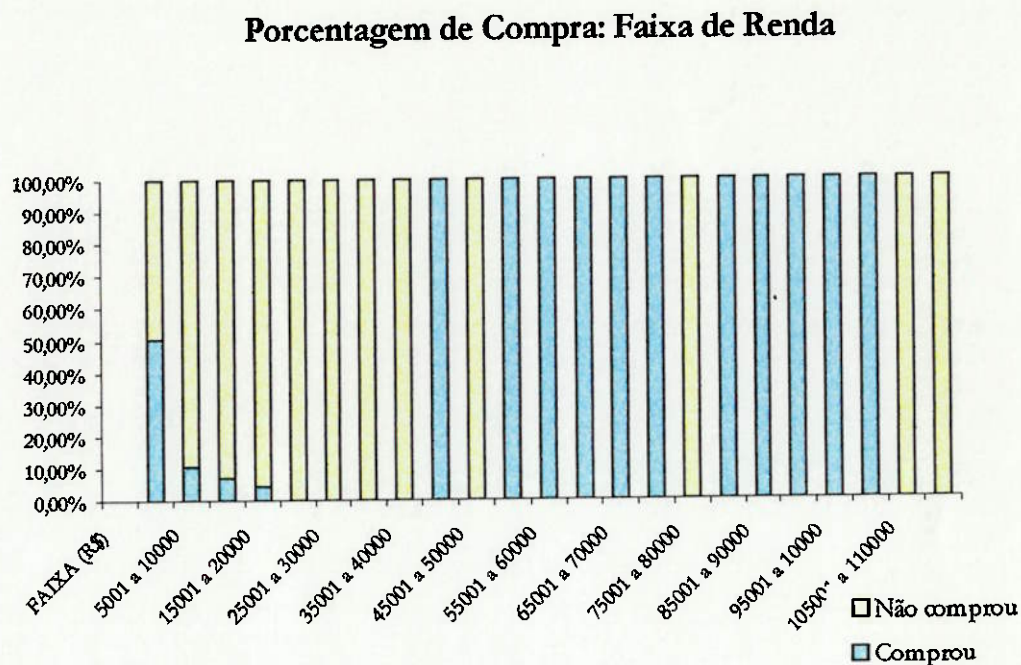


No Anexo A.2 (“Confecção dos Gráficos”), pode-se observar que aproximadamente 95% da amostra está situada na faixa de até R\$ 10.000,00.

Observemos agora o comportamento do conjunto quanto à compra do produto por faixas de renda:

Figura 5.7: Porcentagem de compra dos indivíduos do conjunto de dados por faixa de renda.

Elaborado pelo autor

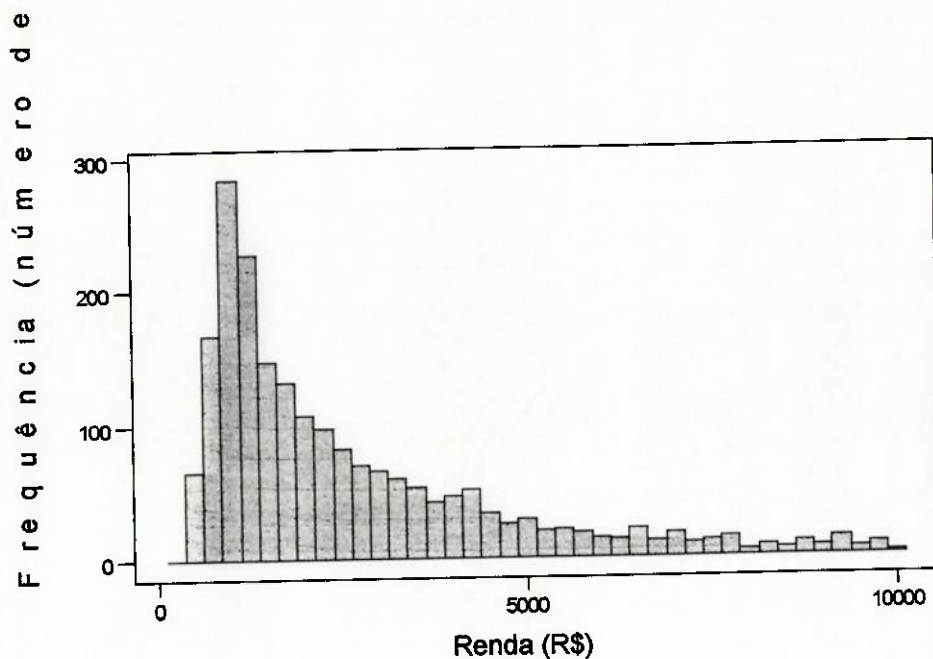


Observa-se pela Figura 5.7 que a porcentagem de elementos que comprou o produto é maior a partir da faixa de renda de R\$ 4.000,00.

Vejamos agora o comportamento do conjunto de dados dos elementos com renda até R\$10.000,00 para que possamos observar apenas os elementos mais representativos do conjunto:

Figura 5.8: Histograma de renda dos indivíduos do conjunto de dados com renda até R\$10.000,00.

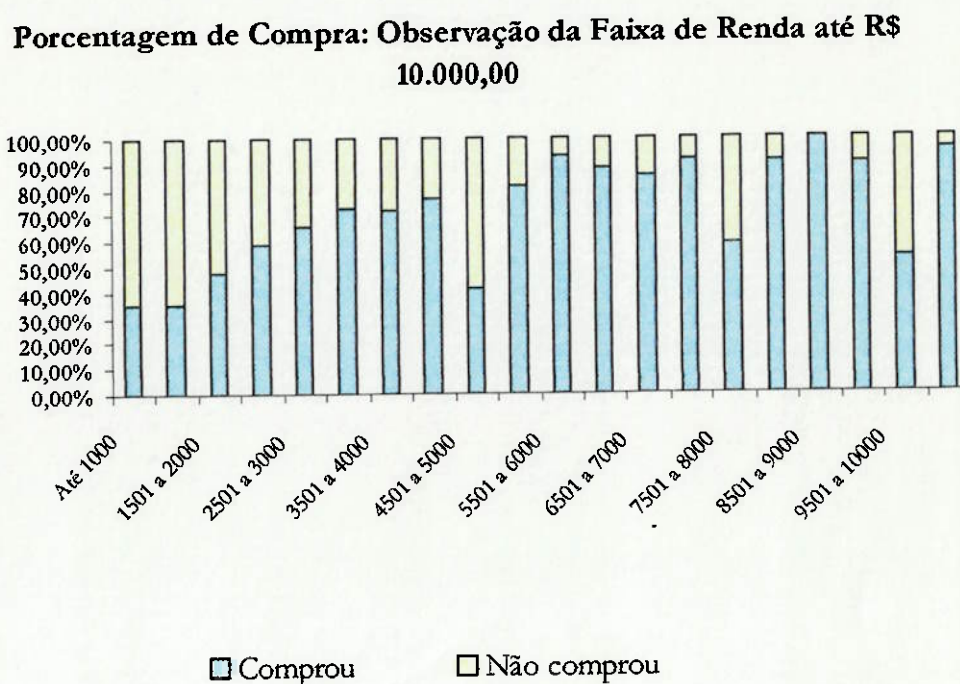
Elaborado pelo autor



Observa-se pela Figura 5.8 que na faixa de renda até R\$ 10.000,00 há uma concentração em torno de R\$ 2.000,00. Pela Figura 5.9, pode-se observar também que nesta faixa de renda há uma tendência de crescimento na porcentagem de compra:

Figura 5.9: Porcentagem de compra dos indivíduos do conjunto de com renda até R\$10.000,00.

Elaborado pelo autor



A profissão dos indivíduos foi codificada segundo critério utilizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). No presente trabalho, as profissões foram agrupadas segundo o critério:

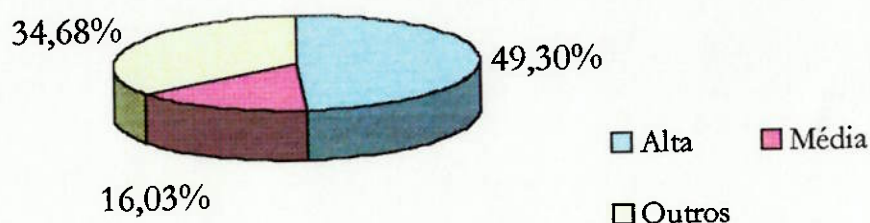
- Alta qualificação*: profissões que exigem um alto nível de especialização para serem exercidas. Exemplos: arquiteto, médico, jornalista, psicólogo, piloto de aeronaves;
- Média qualificação*: profissões que exigem um nível menor de especialização para serem exercidas. Exemplos: trabalhador florestal, motorista, jornaleiro, barbeiro;

- c) *Outros*: nesta categoria estão incluídos os indivíduos de baixa qualificação (exemplos: manobrista, trabalhador de construção civil, auxiliar de enfermagem, doméstica), inativos (exemplos: pensionista, militar reformado, profissional do lar) e diversos (exemplos: empresário, proprietário de estabelecimento de serviços, bolsista/estagiário, capitalista recebendo rendimento de aplicações financeiras, sacerdote ou membro de ordem ou seita religiosa, autônomo).

Pela Figura 5.10, observa-se que 49,30% dos elementos do conjunto de dados pertence ao grupo de profissões de alta qualificação:

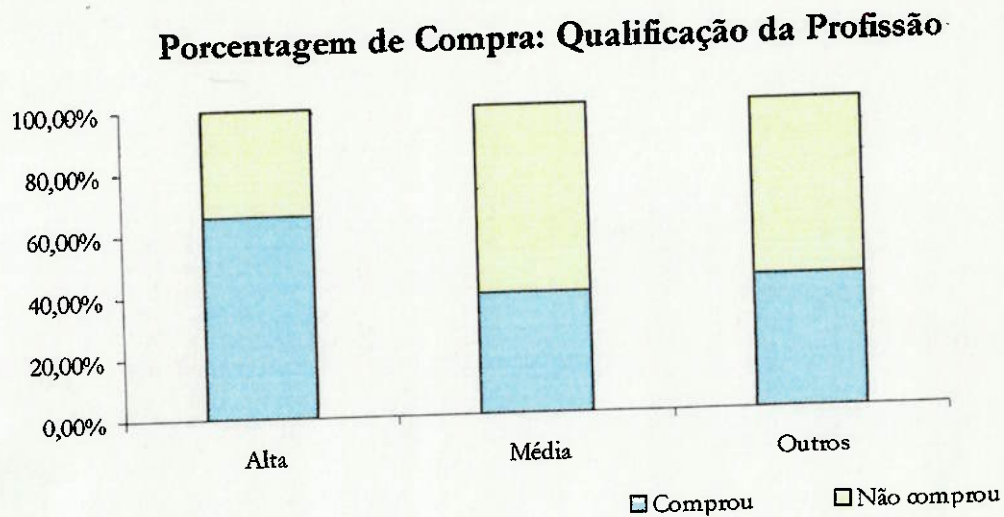
Figura 5.10: Porcentagem dos elementos do conjunto por qualificação da profissão. Elaborado pelo autor

Porcentagem de Qualificação da Profissão



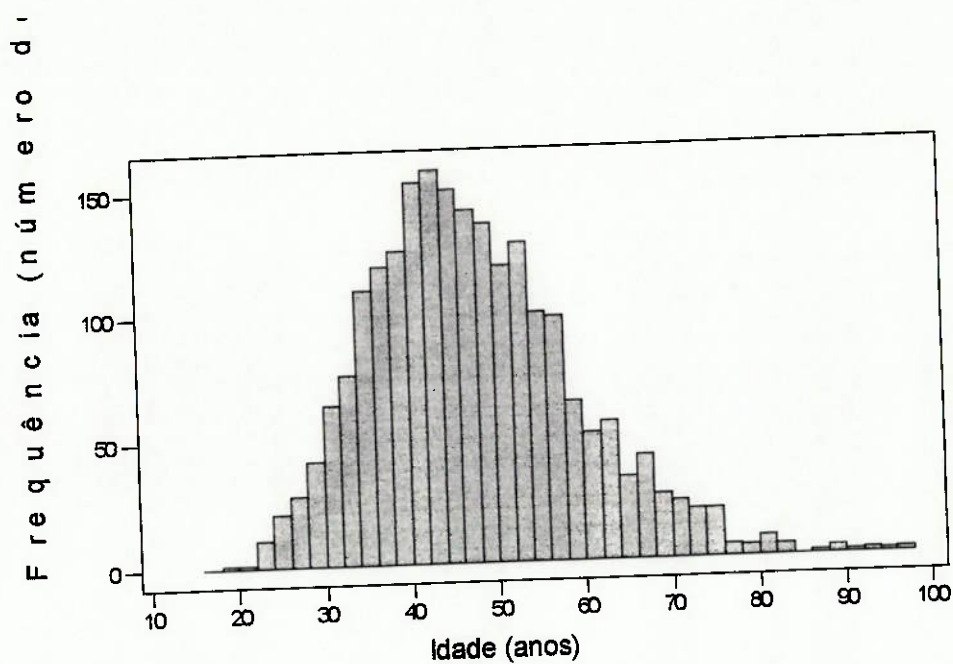
Dentre os grupos de profissões, a porcentagem de compra é maior entre os elementos de alta qualificação profissional, conforme a Figura 5.11:

Figura 5.11: Porcentagem de compra dos elementos do conjunto por qualificação da profissão.
Elaborado pelo autor



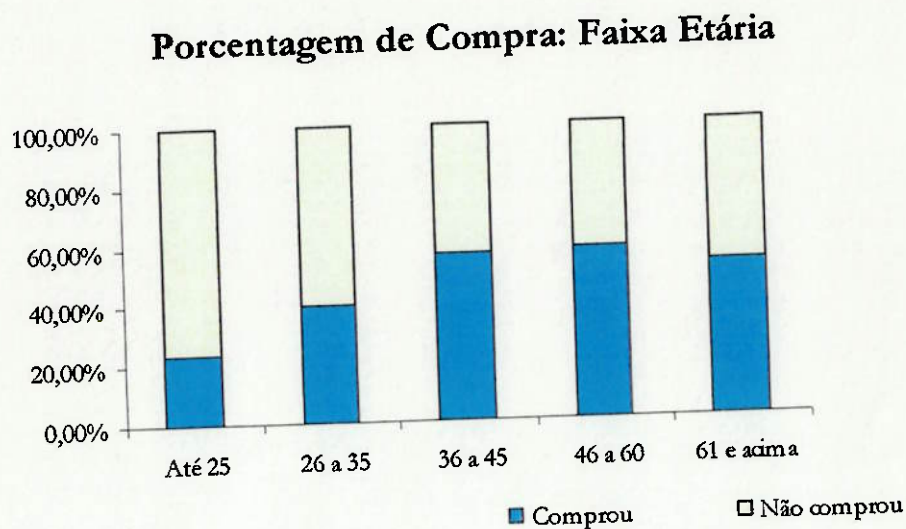
A idade do indivíduo se refere à sua idade na época em que adquiriu o cartão de crédito, expressa em anos. Podemos observar pela Figura 5.11 que a média de idade dos elementos é em torno de 40 anos:

Figura 5.11: Histograma de idade dos indivíduos do conjunto de dados. Elaborado pelo autor



Também observamos pela Figura 5.12 que a porcentagem de compra é crescente até os 60 anos, sendo maior na faixa dos 36 a 60 anos e decaindo na faixa etária seguinte:

Figura 5.12: Porcentagem de compra dos elementos do conjunto por faixa etária. Elaborado pelo autor

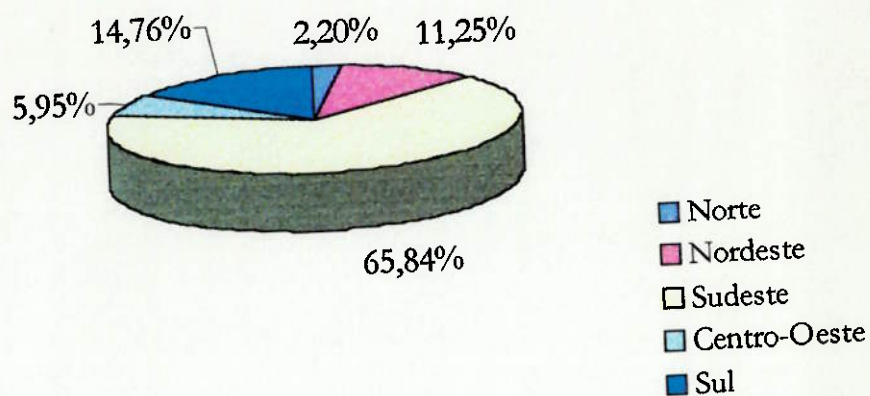


A localização da residência de um elemento do conjunto de dados é expressa no conjunto de dados pelos dois primeiros dígitos de seu Código de Endereçamento Postal (CEP). No presente trabalho, os dígitos foram agrupados segundo as seis regiões do país, conforme tabela no Anexo A ("Tabelas utilizadas na confecção dos gráficos").

Pela Figura 5.13, observa-se que 65,84% dos elementos da amostra pertencem à região Sudeste e, da parte restante, a maioria pertence às regiões Sul e Nordeste:

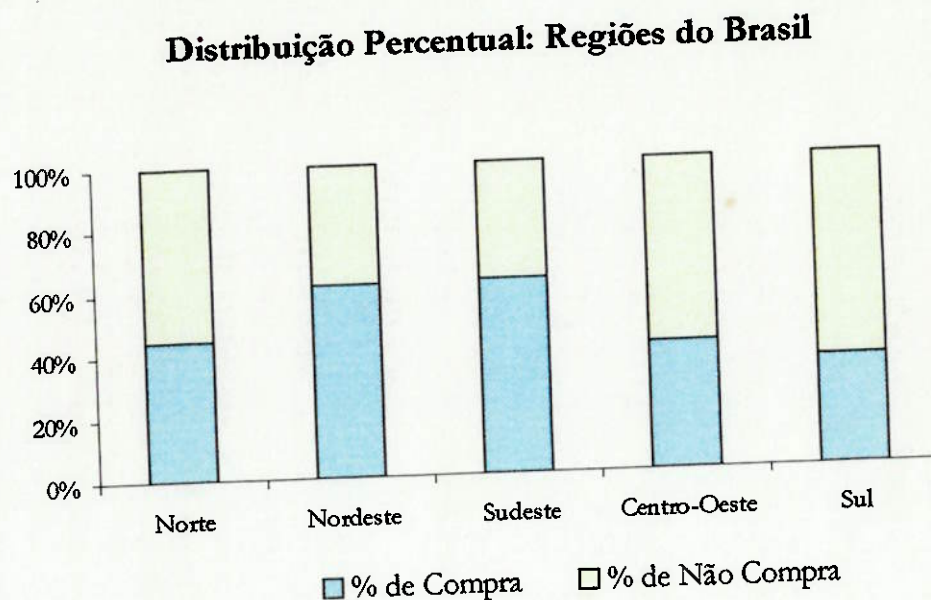
Figura 5.13: Diagrama de localização regional dos indivíduos do conjunto de dados. Elaborado pelo autor

Divisão Percentual da Amostra entre Regiões



Além disso, as regiões Nordeste e Sudeste apresentam as maiores porcentagens de compra do produto, conforme ilustra a Figura 5.14:

Figura 5.14: Porcentagem de compra dos elementos do conjunto por localização regional. Elaborado pelo autor



Em posse do conjunto de dados e conhecendo suas principais características, iremos retomar a discussão sobre as estratégias para obtenção do modelo.

6. ESTRATÉGIAS PARA OBTENÇÃO DO MODELO

Conforme exposto anteriormente, iremos traçar estratégias para a obtenção do modelo ótimo de Regressão Logística.

Segundo Hosmer e Lemeshow (1989), estas estratégias devem possuir: 1) plano para a seleção das variáveis do modelo e 2) método para averiguação do modelo em termos das variáveis independentes individuais e das variáveis em conjunto. Devem também procurar obter o modelo com o menor número de variáveis possível, desde que não percam a eficiência. Isto fará com que o modelo fique numericamente mais estável e que possa ser mais generalizado. Quanto mais variáveis são incluídas em um modelo, mais dependente se torna o modelo do conjunto de dados.

Hosmer e Lemeshow (1989) propõem as seguintes etapas para a seleção de variáveis independentes para o modelo de Regressão Logística:

- 1) Análise minuciosa de cada variável independente,
- 2) Seleção das variáveis independentes para uma análise conjunta e
- 3) Verificação da importância de cada variável incluída no modelo.

Quanto ao método de averiguação do modelo, propõem a utilização de um dos seguintes métodos: Método "Best Subsets" ou Método "Stepwise". Segundo estes autores, o segundo método é o mais utilizado, de modo que será o escolhido para a construção do modelo deste trabalho. Veremos este método a seguir em maiores detalhes, conforme proposto pelos autores.

6.1 MÉTODO "STEPWISE"

A disponibilidade da metodologia para este método na Regressão Logística é recente (Hosmer, Wang, Lin e Lemeshow, 1978 apud Hosmer e Lemeshow, 1989). O procedimento fornece uma maneira rápida e eficiente de se verificar um grande número de variáveis independentes e simultaneamente resolver uma grande quantidade de equações da Regressão Logística.

O método é baseado em um algoritmo estatístico que checa a significância estatística de cada variável independente no modelo. Uma das formas de se proceder neste método é a inclusão gradativa (em passos) de variáveis independentes no modelo, seguida da verificação do comportamento em conjunto de todas as variáveis independentes já incluídas, do seguinte modo:

A inclusão ou exclusão de uma variável independente no modelo é feita com base em um critério de decisão, que é o teste de seu valor p (" p -value"). Este valor denota a probabilidade de o valor da variável ser maior do que o valor observado da variável no modelo. Deste modo, quanto menor o valor p , mais estatisticamente significativa é a variável no modelo. O critério de decisão utilizado neste trabalho para o valor p é 5%, acima do qual rejeita-se H_0 : *variável é significativa* e exclui-se a variável do modelo.

Posterior à análise do valor p , checa-se o valor da função de log-verossimilhança (" $\log$ -likelihood") do modelo. Quanto menor é o módulo deste valor (ou quanto mais próximo de zero ele estiver), melhor é o modelo. Além disso, checa-se o módulo da diferença entre o valores de log-verossimilhança e graus de liberdade entre dois modelos (sem e com a variável independente que se deseja incluir) com base nos valores tabelados da distribuição quadrada para o nível de significância do critério de decisão (no presente trabalho, 5%). Caso o módulo da diferença seja maior do que o

valor tabelado da distribuição qui-quadrada para o módulo da diferença nos valores de graus de liberdade, isto significa que a inclusão da variável não produziu um modelo melhor. Deste modo, tem-se o seguinte algoritmo:

Passo 1: verificação de todos os modelos contendo apenas uma variável independente. Escolha do modelo com menor valor de log-verossimilhança e averiguação do valor p da variável sob um determinado nível de significância. Caso a variável tenha valor p abaixo do limite desejado, procede-se para o passo seguinte;

Passo 2: verificação de todos os modelos possíveis combinando a variável já escolhida e as restantes. Escolha do modelo com menor valor de log-verossimilhança e averiguação do valor p de cada variável sob o mesmo nível de significância. Comparação do módulo da variação do valor de log-verossimilhança com o valor tabelado da distribuição qui-quadrada para o nível de significância adotado e o módulo da variação do valor de graus de liberdade. Caso ambas as variáveis tenham valor p abaixo do limite desejado e o módulo da variação do valor de log-verossimilhança seja abaixo do tabelado, procede-se para o passo seguinte e assim sucessivamente, parando quando uma das duas condições forem violadas.

No presente trabalho, temos as seguintes variáveis independentes: sexo, renda, idade, profissão e localização. A variável resposta é compra. As variáveis independentes sexo, profissão e localização, necessitam de codificação em variáveis "dummy" para serem incluídas no modelo. Procedeu-se da seguinte maneira:

- 1) **Sexo:** esta variável foi codificada em variáveis "dummy" da seguinte maneira:

Tabela 6.1: Codificação da variável sexo em variável "dummy". Elaborado pelo autor

QUALIFICAÇÃO DA PROFISSÃO	D1
Masculino	0
Feminino	1

A variável D1 será chamada de DMSEXO.

- 2) **Localização:** utilizou-se a seguinte codificação segundo a localização regional:

Tabela 6.2: Codificação da variável localização em variáveis "dummy". Elaborado pelo autor

LOCALIZAÇÃO REGIONAL	DÍGITOS DO CEP	D1	D2	D3	D4
Norte	66 a 69	0	0	0	0
Nordeste	40 a 65	1	0	0	0
Sudeste	1 a 39	0	1	0	0
Centro-Oeste	70 a 79	0	0	1	0
Sul	80 a 99	0	0	0	1

As variáveis D1, D2, D3 e D4 serão chamadas de DMLOC1, DMLOC2, DMLOC3 e DMLOC4, respectivamente.

3) **Profissão:** utilizou-se a seguinte codificação segundo agrupamento proposto:

Tabela 6.3: Codificação da variável profissão em variáveis "dummy". Elaborado pelo autor

QUALIFICAÇÃO DA PROFISSÃO	CÓDIGO	D1	D2
Alta	101 a 388	0	0
Média	391 a 599	1	0
Outros	601 a 999	0	1

As variáveis D1 e D2 serão chamadas de DMPRO1 e DMPRO2, respectivamente.

Quando às variáveis renda e idade, apesar de não necessitarem de codificação obrigatória em variáveis "dummy", decidiu-se verificar se elas produzem modelos melhores quando expressas em categorias. Deste modo, tem-se:

4) **Renda:** dividiu-se o espectro de renda em cinco categorias. Isto poderá ser útil visto que a distribuição é muito concentrada na faixa de R\$ 2.000,00, embora existam elementos no conjunto de dados com renda até R\$ 150.000,00.

Tabela 6.4: Codificação da variável renda em variáveis "dummy". Elaborado pelo autor

RENDA (R\$)	D1	D2	D3	D4
Até 1.000	0	0	0	0
1.001 a 2.500	1	0	0	0
2.501 a 5.000	0	1	0	0
5.001 a 10.000	0	0	1	0
10.001 a 150.000	0	0	0	1

As variáveis D1, D2, D3 e D4 serão chamadas de DM1REN, DM2REN, DM3REN e DM4REN, respectivamente. A variável renda contínua será chamada de REN.

5) **Idade:** apesar de a distribuição ser aparentemente normal, codificou-se da seguinte maneira:

Tabela 6.5: Codificação da variável idade em variáveis "dummy". Elaborado pelo autor

IDADE (Anos)	D1	D2	D3	D4
Até 25	0	0	0	0
26 a 35	1	0	0	0
36 a 45	0	1	0	0
46 a 60	0	0	1	0
61 em diante	0	0	0	1

As variáveis D1, D2, D3 e D4 serão chamadas de DM1IDA, DM2IDA, DM3IDA e DM4IDA, respectivamente. A variável idade contínua será chamada de IDA. No Anexo B, está apresentada uma parte do conjunto de dados codificado.

Em posse das variáveis devidamente codificadas, procederemos ao primeiro passo do método "Stepwise". Utilizou-se o pacote estatístico Minitab para se obter o modelo de Regressão Logística. As saídas do Minitab estão no Anexo B.

O conjunto de dados foi dividido aleatoriamente em duas partes, cada qual com 1067 elementos, formando o conjunto de treinamento e de validação respectivamente. Utilizando o Minitab no conjunto de treinamento, procedeu-se ao primeiro passo do método:

PASSO 1: avaliou-se o resultado dos sete modelos possíveis contendo apenas uma variável independente. O modelo com a melhor log-verossimilhança foi obtido com a variável renda, conforme a Tabela 6.6:

Tabela 6.6: Primeiro passo do método "Stepwise". Elaborado pelo autor

VALORES DE LOG-VEROSSIMILHANÇA

MODELO	"LOG-LIKELIHOOD"	GRAUS DE LIBERDADE
SEXO	-720,655	1
LOCALIZAÇÃO	-698,549	4
RENDA DUMMY	-631,325	4
IDADE	-723,280	1
IDADE DUMMY	-679,181	4
PROFISSÃO DUMMY	-703,495	2
RENDA	-627,544	1

COEFICIENTES DO MELHOR MODELO

	COEFICIENTE	DESVIO-PADRÃO	P-VALUE	
CONSTANTE	-0,853	0,1168	0,000	OK
RENDA	0,000	0,000465	0,000	OK

Observa-se que os valores p obtidos são adequados.

PASSO 2: avaliou-se o resultado de todos os modelos possíveis combinando a variável renda. Por motivos óbvios, se excluiu a combinação com variáveis referentes à renda categorizada (DM1REN, DM2REN, DM3REN e DM4REN). O melhor módulo de log-verossimilhança obtido foi no modelo com a variável idade categorizada, segundo a Tabela 6.7:

Tabela 6.7: Segundo passo do método "Stepwise". Elaborado pelo autor

VALORES DE LOG-VEROSSIMILHANÇA				
MODELO	"LOG-LIKELIHOOD"	GRAUS DE LIBERDADE		
RENTA X SEXO	-627,235	2		
RENTA X IDADE DUMMY	-590,563	5		
RENTA X LOCALIZAÇÃO	-604,231	5		
RENTA X PROFISSÃO	-625,852	3		
RENTA X IDADE	-626,622	2		

COEFICIENTES DO MELHOR MODELO				
	COEFICIENTE	DESVIO-PADRAO	P-VALUE	
CONSTANTE	-0,330	0,134	0,011	OK
RENTA	0,00046456	0,00004602	0,000	OK
DM1IDA	-1,1665	0,2593	0,000	OK
DM2IDA	-0,6844	0,2970	0,021	OK
DM3IDA	-0,3312	0,2930	0,258	NÃO OK
DM4IDA	-1,1311	0,1708	0,000	OK

Nota-se que houve uma variação de log-verossimilhança de 36,981 indicando ser significativa a inclusão da idade no modelo. Entretanto, observa-se que a variável DM3IDA não é significativa.

PASSO 3: Este modelo se inicia com a exclusão da variável DM3IDA, por ter sido rejeitada no teste do valor p . Além disso, desconsiderou-se a possível combinação com a variável idade contínua (IDA), por termos considerado a escolha da variável idade categorizada. Prosseguindo com a análise, temos que o melhor modelo com três variáveis é dado por:

Tabela 6.8: Terceiro passo do método "Stepwise". Elaborado pelo autor

VALORES DE LOG-VEROSSIMILHANÇA			
MODELO	"LOG-LIKELIHOOD"	GRAUS DE LIBERDADE	
REDA X IDADE DUMMY X SEXO	-590,900	5	
REDA X IDADE DUMMY X PROFISSÃO	-589,186	6	
REDA X IDADE DUMMY X LOCALIZAÇÃO	-590,364	8	

COEFICIENTES DO MELHOR MODELO			
	COEFICIENTE	DESVIO-PADRAO	P-VALUE
CONSTANTE	-0,1566	0,1711	0,360
REDA	0,00043582	0,00004731	0,000
DM1IDA	-1,1221	0,2580	0,000
DM2IDA	0,6539	0,2967	0,028
DM4IDA	-1,3118	0,1686	0,000
DM1PRO	-0,1510	0,2047	0,461
DM2PRO	-0,3240	0,1616	0,045

Note que a variação da log-verossimilhança foi de apenas 1,377 indicando não ser significativa a inclusão da variável no modelo (valor Qui-quadrado a 5% e 1 grau de liberdade é 3,841). Portanto, o modelo escolhido é dado por:

$$g(REN, DM1IDA, DM2IDA, DM3IDA) = -0,3739 + 0,0005 * REN - 1,1319 * DM1IDA - 0,6498 * DM2IDA - 1,2964 * DM4IDA \quad (6.1)$$

6.2 INTERPRETAÇÃO DOS DADOS

O modelo obtido considera a renda do indivíduo (variável REN), bem como se este está situado na faixa de idade de 26 a 35 anos (variável DM1IDA), na faixa de 36 a 45 (variável DM2IDA) ou na faixa acima de 61 anos (variável DM4IDA). Deste modo, não considera aqueles indivíduos com idade abaixo de 25 anos ou na faixa de 46 a 60 anos.

Segundo os coeficientes da equação obtida, a renda contribui positivamente para a chance de compra do seguro de vida e a idade, negativamente. A faixa de idade que menos diminui a chance está na faixa de 36 a 45 anos.

Para os indivíduos com idades até 25 anos ou na faixa dos 46 aos 60 anos, observa-se que o fator determinante é a renda. Observemos a Tabela 6.9:

Tabela 6.9: Dados de renda e faixa etária do conjunto de dados. Elaborado pelo autor

ELEMENTOS	FAIXA ETÁRIA	Até 25	26 a 35	36 a 45	46 a 60	61 e acima
QUE COMPRARAM		1848	3012	4131	5316	6007
QUE NÃO COMPRARAM	RENDAMÉDIA	1080	1402	1802	1946	2185

Os valores de renda média de cada grupo de elementos (aqueles que compraram e aqueles que não compraram) para as faixas etárias de até 25 e de 46 a 60 anos, quando aplicados à equação (6.1), fornecem chances de compra muito próximas às expostas na Figura 5.12.

Outra observação importante é o fato de que as médias de renda dos elementos que compraram o produto são mais altas do que as dos elementos que não compraram, mais uma vez comprovando determinação da renda nas chances de compra.

Entretanto, as evidências do modelo de que a faixa etária de 36 a 45 anos diminuem menos as chances de compra não foram comprovadas nem na análise descritiva, nem na Tabela 6.9. Seria necessária uma análise mais profunda do modelo para que fosse estabelecido o motivo. Uma sugestão é a divisão do conjunto de dados em faixas etárias mais estreitas.

Segundo cálculo feito pelo Minitab (cuja saída se encontra no Anexo B), utilizando-se o modelo obtido no conjunto de treinamento, seria alcançado um índice de acertos de 73,8%. Ou seja, o modelo conseguiu prever com esta precisão quais são os potenciais compradores do Seguro de Vida.

Para verificarmos a eficiência do modelo, o aplicamos no conjunto de validação, obtendo um índice de acertos de 64,4%. Um valor mais baixo já era esperado, uma vez que o modelo não dispõe de nenhuma informação sobre o conjunto.

7. CONCLUSÃO

Segundo o método de Regressão Logística desenvolvido, obteve-se um modelo que utiliza dados de renda e idade do indivíduo para se prever a probabilidade que este tem de adquirir o Seguro de Vida da empresa X. Segundo este modelo, será possível prever com uma exatidão de 64,4% se um indivíduo é um potencial comprador ou não do produto. É uma porcentagem razoável e certamente será muito útil no direcionamento dos investimentos de Marketing da empresa. Entretanto, em situações do cotidiano geralmente os índices de acerto obtidos com a Regressão Logística são maiores. Diversos fatores podem ter influído no resultado:

- 1) Na prática, utiliza-se conjunto de dados muito maiores, da ordem de 100.000 elementos - o que aumenta o poder de predição do modelo.
- 2) Arranjaram-se os elementos em três conjuntos de acordo com a qualificação da profissão: alta, média e outros. Observa-se que a terceira categoria é muito diversificada, contendo profissões tão diversas quanto sacerdotes e empresários. Na prática, utiliza-se uma divisão mais apurada, com categorias mais específicas. Para tal, necessita-se de um conjunto de dados maior.
- 3) A divisão dos elementos entre cinco regiões do país, apesar de útil, também não é ótima uma vez que dentro de um mesmo bloco regional existem grandes disparidades sociais. Na prática, utiliza-se uma divisão também mais apurada.

Além disso, os fatores determinantes na compra do produto (idade e renda), segundo o modelo, são conceitualmente coerentes com a situação de compra: é razoável que a renda do indivíduo determine seu comportamento de compra, o mesmo acontecendo com sua idade. Ambos fatos podem ser observados na análise descritiva.

Uma outra consideração é a falta de dados a respeito do montante do investimento da empresa direcionado ao Marketing, impossibilitando a estimação de economias provenientes da utilização do método.

O trabalho demonstrou como um modelo matemático pode ser utilizado no auxílio de situações de decisão, o que é uma ferramenta muito útil nas mais diversas áreas. A utilização do modelo é cada vez mais explorada, bem como a diversidade de sua aplicação. Entretanto, deve-se sempre procurar interpretar o modelo obtido e estabelecer ligações com observações gráficas ou estatísticas do conjunto de dados. Um bom entendimento das evidências fornecidas pelo modelo é fundamental no processo de tomada de decisão. Mesmo assim, é importante lembrar que o processo decisório tem um caráter subjetivo, de modo que os modelos não devem substituir o julgamento pessoal e sim auxiliá-lo.

O modelo de Regressão Logística, como demonstrado ao longo deste trabalho, é um instrumento eficaz e com uma gama infinita de aplicações. Os benefícios de sua utilização na Engenharia de Produção são inúmeros, dada as diversas situações de decisão encontradas no cotidiano - muitas delas de grande importância. Trata-se, portanto, de um assunto de extrema relevância para o futuro da área.

8. BIBLIOGRAFIA

- CLEMEN, R. T. **Making decisions: an introduction to decision analysis.** Duxbury Press, Belmont, California, 1990.
- DURAND, D. Risk elements in consumer installment financing (Study #8). **National Bureau of Economic Research**, New York, 1941.
- EISENBEIS, R. A. Recent developments in the application of credit-scoring techniques to the evaluation of commercial loans. **IMA Journal of Mathematics Applied in Business & Industry**, 7, 271-90, 1996.
- GREGORY, G. **Decision analysis.** Pitman Publishing, 1988.
- HAIR, JOSEPH F.; ANDERSON, RALPH E.; TATHAM, RONALD L.; GRABLOWSKY, BERNIE J. **Multivariate data analysis.** Tulsa, OK: Petroleum Publishing Company, 1979.
- HAND, D. J.; HENLEY, W. E. Statistical classification methods in consumer credit scoring: a review. **Journal of Royal Statistical Society**, 523-39, 1997.
- HAND, D. J. **Discrimination and Classification.** Chichester: Wiley, 1981.
- HENLEY, W. E. **Statistical aspects of credit scoring.** PhD Thesis. The Open University, Milton Keynes, 1995.
- HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression.** Massachusetts, Wiley, 1989.
- JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis.** Prentice Hall, Englewood Cliffs, New Jersey, 522-5, 1982.
- KOLESAR, P.; SHOWERS, J. L. A robust credit screening model using categorical data. **Management Science**, 31, 123-33, 1985.

KOTZ, S.; JOHNSON, N. L. **Encyclopedia of statistical sciences**. Vol.5, 301-7, John Wiley and Sons, 1982.

KOTZ, S.; JOHNSON, N. L. **Encyclopedia of statistical sciences**. Vol.5, 128-32. John Wiley and Sons, 1982.

KOTZ, S.; JOHNSON, N. L. **Encyclopedia of statistical sciences**. Vol.4, 647-50. John Wiley and Sons, 1982.

MCGRATH, J. J. Improving credit evaluation with a weighted application bank. **Journal of Applied Psychology**, 44, 325-28, 1960.

MIZUTANI, P. I. **Trabalho de formatura: um modelo matemático para análise de crédito para pessoas físicas**. Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Produção, 1982.

MYERS, J. H.; FORGY, E. W. The development of numerical credit evaluation systems. **Journal of American Statistical Association**, 58, 799-806, 1963.

MYERS, J. H. E CORDNER, W. Increase credit operation profits. **The Credit World**, February, 12-3, 1957.

PRESS, J.; WILSON, S. Choosing between logistic regression and discriminant analysis. **Journal of American Statistical Association**, 73, 699-705, 1978.

RADFORD, K. J. **Modern managerial decision making**. Reston Publishing Company, Inc, Reston, Virginia, 1981.

WALKER, S. H.; DUNCAN, D. B. **Biometrika**, 54, 167-79.

WOLBERS, H. L. **The use of the biographical data bank in predicting good and potentially poor credit risks**. M.A. Thesis, University of Southern California, 1949.

Internet:

[URL: banktechnews.faulknergray.com/buyers/buy30b.htm]

[URL: ncua.gov/ref/opinion_letters/96-0431.html]

[URL: www.bob-taylor.com/fico3.htm]

[URL: www.freddiemac.com/function/fm-homby/creditsc.htm]

[URL: www.getinnovative.com/mortgage/creditsc.htm]

[URL: www.gt.com/gtonline/finance/finadv/071897d.html]

[URL: www.ibge.gov.br/]

9. ANEXO A

9.1 CONJUNTO DE DADOS

Idade	Sexo	Renda	Código da Profissão	Dígitos do CEP	Compra
66	M	5471	922	22	1
61	M	2196	301	13	1
46	M	900	904	4	1
34	M	2488	905	70	1
60	M	1480	131	11	1
70	M	2558	923	22	1
45	M	1427	904	13	1
51	M	3440	101	11	1
67	M	7952	117	70	1
53	M	14677	904	11	1
55	M	532	101	4	1
55	M	1817	902	1	1
48	M	1082	172	41	1
46	M	3809	101	59	1
46	M	3142	131	22	1
40	M	5091	101	21	1
68	M	2045	111	25	1
53	M	1544	923	26	1
57	M	2763	125	5	1
43	M	1959	125	5	1
46	M	1997	999	5	1
51	F	1072	131	58	1
50	F	2605	111	24	1
29	M	11231	902	22	1
44	F	1214	131	4	1
68	F	3103	922	20	1
52	M	6812	101	5	1
54	F	4030	102	3	1
49	M	12226	115	23	1
40	F	4392	111	22	1
45	M	4030	395	13	1
56	M	1818	295	21	1
36	M	1308	117	13	1
50	M	12776	301	88	1
39	M	3930	101	82	1
41	M	5121	101	1	1
50	M	2681	124	57	1
49	M	4323	111	30	1
43	M	3069	902	9	1
42	M	3717	101	5	1
37	M	5144	296	79	1
45	M	1330	125	88	1
36	F	1400	395	7	1
41	M	4236	101	4	1
68	F	983	143	22	1
31	M	2098	902	4	1
39	M	891	902	30	1
58	M	5175	131	1	1
52	M	3350	142	4	1
59	F	1682	143	40	1
38	M	6834	101	4	1
39	M	4320	903	4	1
41	F	3146	902	2	1

35	M	4267
50	M	3849
41	M	10798
42	M	1120
37	M	3387
51	M	4236
45	M	3071
41	M	4239
57	M	2965
48	M	2932
54	M	6686
32	F	5774
34	M	1676
29	M	883
42	M	2260
44	M	4000
52	F	2882
74	M	1526
35	M	2363
43	F	2229
45	F	877
34	F	1753
71	M	2635
42	M	3366
49	F	3479
25	M	3977
33	M	1040
29	M	1370
34	F	1722
52	F	1121
54	M	1468
51	F	1250
47	M	2908
37	M	1857
42	F	2321
41	M	4811
62	M	1167
44	M	975
46	F	2841
70	M	7270
26	F	818
40	M	5060
54	F	1687
47	M	1901
34	M	4962
51	M	2365
40	M	2552
45	F	1474
55	M	8634
54	M	1935
56	M	960
39	M	808
56	M	1399
44	M	6058

101
297
101
532
125
101
191
125
395
902
922
401
402
902
124
301
296
922
125
102
904
131
101
395
132
181
905
101
902
906
391
131
164
151
906
532
541
905
136
409
115
189
393
902
213
702
402
113
101
301
395
999
189
108

58
90
41
2
88
41
74
28
59
37
55
31
86
9
58
41
60
52
1
74
80
88
30
38
53
1
22
84
8
11
1
2
30
45
22
92
13
17
2
11
95
20
5
4
86
12
12
5
2
4
8
37
21
12

[illegible]

44	F	3683
57	F	3363
53	M	616
36	F	1915
46	F	1644
44	M	750
46	M	4954
33	F	3443
56	M	3183
35	F	1050
45	M	2021
41	M	1026
55	M	1204
75	F	12840
30	M	2816
31	M	1096
49	F	1937
65	M	1505
31	M	1042
58	M	5723
64	F	9637
43	M	539
71	M	1260
57	M	2386
81	F	1452
56	M	1083
43	M	1088
38	F	1011
52	F	1164
43	F	927
52	F	936
42	M	1034
70	M	947
25	M	691
51	M	1537
38	F	1286
61	M	9337
35	M	953
33	M	6079
78	M	4717
50	M	3331
52	M	1017
33	M	1661
50	M	989
68	F	985
91	M	780
40	M	1492
59	M	926
37	F	986
39	F	1215
31	M	729
39	M	2586
63	M	675
77	M	1862

125
999
709
113
143
403
124
182
143
154
903
112
702
143
704
103
395
923
101
155
924
905
999
142
922
181
115
902
905
101
905
901
901
904
902
395
299
532
395
923
107
905
903
904
297
923
296
531
297
393
905
402
923
921

22	1
4	1
15	1
4	1
86	1
3	1
4	1
24	1
38	1
13	1
96	1
13	1
13	1
1	1
9	1
96	1
20	1
20	1
13	1
3	1
20	1
15	0
90	0
67	0
59	0
68	0
38	0
4	0
28	0
15	0
5	0
13	0
95	0
22	0
88	0
4	0
97	0
13	0
70	0
3	0
74	0
35	0
13	0
6	0
20	0
4	0
66	0
21	0
78	0
4	0
14	0
66	0
36	0
1	0

47	M	2782
54	M	8532
56	M	3099
66	M	12026
57	M	7444
48	F	1920
48	M	4131
42	M	2972
51	F	2596
65	F	1186
44	M	18079
53	M	8313
49	M	12568
45	M	1317
57	M	1875
54	M	6052
45	M	4734
46	M	13602
54	M	1671
54	M	1897
31	M	936
41	F	1712
37	M	2671
44	M	541
44	M	713
39	M	2204
46	M	11616
43	F	4430
59	F	768
55	M	6493
47	M	1077
44	M	1511
35	M	5075
49	M	5283
37	F	6372
44	M	2342
52	F	2821
73	M	2454
43	M	9871
58	F	2619
36	F	2499
63	M	4658
43	M	1196
36	M	5488
49	M	9043
63	M	3313
57	F	7644
35	F	3218
47	F	2831
46	M	3989
37	M	3786
57	M	2142
62	M	16646
39	M	3576

22	1
50	1
49	1
18	1
4	1
7	1
14	1
5	1
12	1
1	1
66	1
20	1
13	1
13	1
4	1
22	1
82	1
1	1
4	1
4	1
22	1
20	1
69	1
21	1
50	1
8	1
6	1
69	1
1	1
30	1
31	1
30	1
13	1
20	1
4	1
8	1
22	1
22	1
5	1
21	1
78	1
51	1
37	1
3	1
23	1
90	1
30	1
3	1
25	1
60	1
30	1
3	1
22	1
16	1

57	F	5474
30	F	713
32	F	1363
33	M	800
38	F	1065
54	M	808
32	F	1306
50	F	917
53	F	3257
26	F	958
34	M	950
28	F	2907
34	M	800
49	F	1264
53	F	1933
45	M	1686
41	M	2998
46	F	2914
37	M	1659
41	M	1062
53	F	762
47	M	873
67	F	600
29	M	2960
72	F	4920
51	F	9585
45	F	726
46	F	6944
26	M	800
50	F	1247
40	M	916
31	F	598
27	F	1551
49	F	713
39	M	2325
42	M	1015
29	M	2287
51	F	1482
44	F	1287
42	M	1385
43	M	1074
32	F	799
40	M	1202
36	F	2557
45	M	2409
47	F	2746
58	F	3559
67	M	4314
65	M	6832
45	M	8191
40	M	1183
56	M	3438
44	M	3552
57	M	19395

902
409
114
905
159
601
395
511
297
411
293
126
905
922
143
299
297
143
903
159
905
143
902
101
906
294
125
291
301
143
905
902
115
599
707
716
609
134
132
159
172
904
395
296
711
143
923
111
301
171
396
181
904
101

91	0
78	0
52	0
96	0
36	0
14	0
72	0
99	0
70	0
72	0
39	0
89	0
98	0
88	0
2	0
85	0
21	0
9	0
30	0
13	0
37	0
3	0
91	0
5	0
22	0
38	0
4	0
13	0
88	0
15	0
17	0
85	0
80	0
89	0
12	0
13	0
68	0
2	0
13	0
8	0
9	0
22	0
45	0
66	0
95	0
23	1
70	1
24	1
4	1
21	1
40	1
5	1
4	1
24	1

57	M	1185
62	M	1236
66	M	1093
56	M	3553
33	M	579
39	F	1165
50	M	759
44	F	854
64	M	1013
38	M	1164
71	F	829
57	F	1451
50	F	1518
44	M	2740
44	M	1155
37	M	1554
42	F	1635
33	M	900
48	F	2786
48	M	747
45	F	3283
52	M	1217
29	F	1112
60	M	3519
43	F	1982
38	M	1249
56	M	540
57	M	1083
45	M	1977
72	M	2903
51	F	909
62	M	826
25	M	1994
48	M	666
47	M	1716
47	M	961
45	M	1109
57	M	1646
58	F	903
49	M	905
25	F	1573
68	F	2379
46	M	508
33	M	1193
54	F	989
66	F	1850
58	M	833
33	M	1344
39	M	2760
39	M	2486
41	F	867
38	M	668
38	F	1163
41	M	777

902
296
999
707
904
143
396
143
923
719
901
293
143
142
296
716
103
905
296
903
395
295
297
293
296
707
923
411
119
901
902
902
702
392
293
903
403
923
903
168
115
924
411
601
297
922
531
902
902
395
902
902
395
905

12	0
70	0
20	0
3	0
19	0
6	0
40	0
19	0
9	0
15	0
15	0
57	0
18	0
59	0
79	0
13	0
60	0
4	0
20	0
89	0
89	0
21	0
17	0
29	0
71	0
68	0
33	0
95	0
72	0
14	0
18	0
83	0
13	0
26	0
28	0
87	0
86	0
14	0
93	0
85	0
13	0
2	0
56	0
73	0
35	0
20	0
16	0
46	0
89	0
87	0
80	0
86	0
90	0
86	0

48	M	3959	125	20	1
54	F	1583	131	50	1
34	M	4657	101	2	1
39	M	7406	301	60	1
40	M	8776	303	5	1
55	M	3658	181	4	1
46	M	5430	213	57	1
45	M	1149	101	4	1
31	M	1922	172	22	1
38	F	1471	395	2	1
64	M	5778	923	22	1
58	M	609	923	21	1
69	F	3725	922	59	1
38	M	1734	529	2	1
42	M	3151	401	21	1
52	M	24095	904	1	1
49	F	1729	391	20	1
60	M	9386	301	95	1
38	M	900	999	20	1
37	M	4114	104	91	1
42	F	7971	111	22	1
60	M	5645	902	28	1
48	M	7031	111	97	1
41	M	3184	216	79	1
44	M	4345	101	4	1
44	M	6453	395	5	1
35	F	2380	906	74	1
44	F	1500	903	13	1
56	F	2086	922	18	1
45	M	17934	125	82	1
45	M	3917	216	20	1
35	F	2875	393	70	1
52	M	1495	923	51	1
55	M	4215	391	31	1
43	F	1912	902	3	1
46	F	1659	132	22	1
66	M	1611	923	66	1
60	M	2788	922	58	1
51	M	1681	904	31	1
62	M	1054	923	90	1
42	F	3244	395	15	1
37	M	5725	395	4	1
42	F	2492	395	60	1
54	M	7501	131	59	1
38	M	1815	922	6	1
35	M	1428	395	14	1
31	M	5422	125	52	1
55	F	2625	297	91	1
42	M	924	904	4	1
36	M	1444	902	2	1
32	F	2998	182	22	1
55	M	10385	303	4	1
31	M	4313	126	4	1
53	M	2736	707	13	1

52	F	2533	144	15	0
40	M	1339	394	20	0
58	M	1834	923	8	0
45	M	2240	703	13	0
42	F	2696	113	1	0
41	F	960	143	12	0
27	F	870	902	70	0
33	F	1247	296	20	0
51	M	1325	716	9	0
94	F	5254	906	1	0
42	M	1648	707	13	0
34	M	1002	539	26	0
25	F	797	904	85	0
47	F	1017	303	90	0
43	M	1419	303	2	0
45	M	2661	111	20	0
61	M	6986	906	91	1
38	M	2334	111	40	1
68	M	725		22	1
46	M	110453	301	1	1
53	M	26714	301	6	1
40	M	2828	303	1	1
46	M	4081	101	28	1
75	M	11307	101	40	1
67	M	3967	216	40	1
68	M	4138	107	41	1
54	M	5345	922	22	1
45	M	8026	115	80	1
59	M	806	923	5	1
42	M	8435	301	2	1
51	M	1257	999	22	1
51	M	10760	403	24	1
48	M	8803	903	7	1
57	M	16887	301	32	1
39	M	1319	156	3	1
64	M	1841	131	1	1
40	M	5653	395	20	1
52	M	14663	101	4	1
37	M	1060	125	80	1
64	F	1184	904	4	1
50	M	1083	903	9	1
49	M	1968	103	6	1
58	M	5640	115	1	1
52	M	3110	101	3	1
47	M	4099	101	20	1
49	M	1166	902	4	1
70	M	4299	902	24	1
48	M	12671	111	4	1
58	M	5508	216	22	1
29	M	572	301	12	1
41	F	3630	395	1	1
48	M	30197	301	3	1
54	M	1758	101	21	1
51	F	2046	132	22	1

62	M	558
44	F	786
47	F	1417
32	M	904
64	M	892
47	M	3235
41	F	1409
44	M	1075
34	F	1112
45	M	838
47	F	3041
31	F	1135
43	F	979
68	M	2088
56	M	958
26	F	1435
42	M	2054
46	M	776
43	M	1534
46	M	2527
44	M	1048
62	M	1196
82	M	1391
58	M	1007
45	M	1045
51	M	533
22	M	1253
41	M	1031
47	F	2496
41	M	1506
42	F	2792
52	F	1060
36	M	3047
39	F	1344
57	M	1252
23	M	1495
53	F	2952
38	M	1004
50	M	3463
35	M	520
70	F	1242
56	M	2575
42	M	2434
58	M	845
46	F	892
61	F	875
46	M	537
50	F	1843
33	M	600
49	F	802
33	F	1485
32	F	634
45	M	890
32	M	1038

905
902
143
905
901
542
394
902
142
902
143
298
292
294
901
409
701
303
901
901
402
921
923
902
395
905
168
156
393
999
393
296
395
393
701
716
143
101
901
902
906
297
296
923
902
902
905
292
131
296
171
902
709
542

36	0
20	0
14	0
13	0
19	0
69	0
58	0
68	0
5	0
60	0
70	0
18	0
51	0
85	0
35	0
3	0
75	0
93	0
87	0
39	0
30	0
80	0
13	0
95	0
99	0
95	0
13	0
92	0
91	0
95	0
4	0
21	0
51	0
35	0
25	0
13	0
38	0
36	0
14	0
44	0
22	0
78	0
78	0
8	0
24	0
78	0
15	0
38	0
80	0
29	0
30	0
35	0
8	0
82	0

35	F	1935	143	8	1
38	M	1558	159	13	1
42	M	4200	395	90	1
60	M	6056	111	13	1
44	M	5190	301	5	1
72	M	576	923	3	1
60	M	2558	309	90	1
48	M	1015	923	1	1
37	M	2114	391	1	1
74	F	2822	924	22	1
53	M	2208	904	28	1
35	M	3130	402	41	1
42	M	655	101	13	1
51	M	4786	155	2	1
53	M	3148	903	2	1
25	M	1272	155	9	1
49	M	1750	903	20	1
46	F	2713	142	30	1
48	M	2166	392	9	1
45	M	9962	303	12	1
47	M	1714	901	17	1
28	F	1432	939	30	1
53	M	3116	707	38	1
44	M	4914	903	4	1
30	F	5251	111	79	1
59	M	699	902	4	1
70	M	4628	922	24	1
40	M	1857	401	57	1
44	M	1770	142	12	0
24	F	933	905	36	0
52	M	6680	301	88	0
66	M	879	901	86	0
59	M	1707	902	78	0
39	M	1635	501	9	0
66	M	2601	216	59	0
60	M	1039	924	96	0
41	M	4322	108	18	0
36	M	2929	395	37	0
57	M	2398	922	70	0
37	F	1218	296	4	0
53	M	6584	213	24	0
66	M	4325	297	3	0
53	M	2536	115	74	0
36	M	574	905	13	0
62	M	3405	906	4	0
53	M	789	923	30	0
55	M	1324	539	72	0
74	M	6315	303	89	0
49	M	1758	904	40	0
36	F	2557	393	13	0
58	M	799	922	53	0
53	M	594	903	95	0
66	F	729	999	5	0
28	M	507	905	15	0

41	M	6079	395	21	1
41	M	6616	395	70	1
45	M	4956	101	22	1
52	M	1875	121	24	1
67	M	4539	111	1	1
42	F	9600	593	11	1
44	F	1643	297	74	1
42	M	807	905	17	1
47	M	5979	401	80	1
43	M	3009	395	69	1
49	M	3208	125	7	1
35	M	2234	145	1	1
48	F	800	125	8	1
48	F	987	143	13	1
40	F	756	902	35	1
42	M	4347	101	21	1
36	M	2151	395	13	1
52	M	1800	902	13	1
32	M	2304	111	5	1
40	F	578	303	50	1
70	M	15995	906	90	1
47	M	7306	101	90	1
48	M	3898	121	40	1
36	M	1460	111	1	1
30	M	3358	303	78	1
44	M	1458	902	89	1
41	M	3316	101	42	1
28	M	2021	126	9	1
36	M	996	101	4	1
56	M	1130	409	84	1
43	M	2267	401	49	1
48	M	7473	159	12	1
50	M	2392	143	12	1
48	M	9303	101	13	1
54	F	1570	298	11	1
45	M	744	402	15	1
30	M	2992	125	22	1
32	F	882	122	1	1
47	F	2574	182	22	1
73	M	1243	902	22	1
27	M	2486	303	52	1
34	F	779	904	5	1
40	M	4263	111	35	1
83	M	3133	921	20	0
41	M	1520	541	15	0
59	M	3634	921	78	0
66	M	1311	906	40	0
39	M	2613	539	8	0
56	M	1225	923	85	0
31	F	596	143	29	0
77	M	1378	923	8	0
66	M	1153	922	21	0
61	M	846	923	58	0
38	M	641	902	13	0

44	M	2250	902	54	1
38	M	1433	902	89	1
42	M	933	905	13	1
54	M	12067	101	5	1
56	F	15786	922	91	1
50	M	106908	131	22	1
49	F	1595	125	40	1
38	F	2029	395	14	1
45	F	1758	904	22	1
40	M	3510	391	4	1
67	M	6970	303	4	1
52	F	4398	142	60	1
36	M	3279	126	31	1
34	M	1994	395	21	1
37	M	1746	142	17	1
48	M	9274	101	45	1
54	M	7814	111	5	1
39	F	1132	393	3	1
44	M	753	145	2	1
42	M	1175	903	81	1
48	M	8792	125	3	1
40	F	1130	126	5	1
51	M	2849	115	14	1
57	M	6870	293	24	1
44	F	5051	291	21	1
30	M	1609	155	81	1
44	M	3103	109	81	1
40	M	6377	401	13	1
32	M	1751	126	30	1
40	M	5025	159	91	1
47	M	5806	101	30	1
39	M	3067	111	90	1
41	M	2056	166	3	1
56	M	9192	131	39	1
44	M	1778	703	4	1
52	M	34694	301	3	1
56	F	4750	296	22	1
42	F	525	905	1	1
40	F	2659	303	1	1
49	M	9125	101	41	1
53	F	900	125	4	1
58	M	3159	922	59	1
34	M	940	902	9	1
52	M	2736	395	12	1
36	M	2992	902	51	1
43	M	4571	101	21	1
63	M	12249	922	31	1
29	M	954	301	30	1
49	F	3056	135	5	1
49	M	5966	303	5	1
55	M	1212	902	22	1
38	F	3316	102	4	1
48	M	6831	303	5	1
57	M	5614	213	26	1

53	F	644
42	M	3407
60	M	1189
58	F	5255
35	M	800
40	F	3202
70	M	13637
58	M	3000
41	M	23624
51	M	3709
54	M	1386
48	M	6083
46	M	6479
61	M	24550
51	M	9045
44	M	1192
64	M	4588
63	M	956
37	M	1125
41	F	11144
43	F	1048
54	M	7932
46	M	3973
59	M	895
52	M	3321
37	F	1845
35	M	5183
70	M	2473
47	F	1400
52	M	2100
45	M	1369
47	M	1000
51	F	657
61	M	2056
50	M	9692
44	M	5340
57	M	9810
68	M	7263
33	M	1017
37	F	1518
46	M	6927
40	M	1711
43	M	2359
61	M	1342
45	F	3764
46	M	2320
52	M	11444
61	M	1214
62	M	4245
49	M	1350
49	M	4987
39	M	1333
38	F	2963
57	M	2936

902
707
923
923
904
395
211
301
125
125
111
301
902
125
103
902
111
213
905
172
904
142
923
923
301
143
111
922
132
403
297
902
923
171
904
902
156
902
402
125
111
902
395
923
131
904
131
402
301
101
395
902
125
999

28	0
3	0
57	0
34	0
79	0
2	1
50	1
3	1
20	1
74	1
4	1
22	1
1	1
5	1
13	1
7	1
25	1
35	1
3	1
1	1
4	1
71	1
60	1
90	1
18	1
24	1
52	1
1	1
22	1
78	1
24	1
3	1
5	1
58	1
4	1
1	1
4	1
20	1
6	1
6	1
90	1
9	1
90	1
90	1
90	1
96	1
4	1
4	1
3	1
13	1
38	1
12	1
90	1
11	1

45	M	1491	292	72	0
43	M	1024	151	29	0
36	M	636	301	52	0
39	M	792	539	38	0
41	M	1038	902	88	0
37	M	1835	156	38	0
59	M	1888	923	14	0
37	F	785	143	86	0
38	M	2052	156	88	0
39	M	1577	707	35	0
39	M	1922	701	81	0
45	M	1644	129	65	0
37	M	854	902	83	0
36	F	1149	924	70	0
40	M	765	999	29	0
40	M	1685	549	28	0
43	M	3337	111	26	0
36	M	1404	707	35	0
44	M	1554	295	40	0
53	M	2865	902	18	0
56	M	1208	142	15	0
24	M	880	902	78	0
46	M	1200	702	13	0
38	F	1832	298	80	0
32	F	729	999	91	0
45	F	652	905	95	0
27	F	1884	395	70	0
33	M	1914	124	89	0
32	M	1125	905	96	0
46	M	4573	214	17	0
46	M	958	902	15	0
81	F	1531	924	35	0
55	F	1685	143	96	0
40	M	1594	707	35	0
29	M	1434	189	91	0
49	M	2391	298	9	0
56	F	1743	143	30	0
45	M	2082	171	15	0
31	M	1193	303	55	0
24	M	1843	216	70	0
68	M	2870	906	4	0
42	M	1000	903	13	0
45	M	2347	597	29	0
46	M	1220	904	2	0
37	M	1570	999	27	0
45	M	3842	153	15	0
48	F	1052	144	2	0
45	M	1391	703	6	0
39	M	1592	111	13	0
23	M	808	531	35	0
63	M	583	902	12	0
40	F	2375	124	11	0
30	F	2799	115	20	0
31	M	1154	411	21	0

52	F	2466
45	F	3995
49	M	1882
36	F	4048
57	F	2001
43	F	1249
49	M	2109
61	M	2147
31	M	1592
55	F	1328
47	M	3275
50	F	622
58	M	1031
77	M	830
46	F	1079
50	M	1810
49	M	1268
41	M	1354
35	M	1909
73	F	1164
55	M	728
28	M	871
57	M	695
26	F	892
59	M	1130
41	M	1769
58	M	1417
44	M	663
58	F	2042
46	M	1142
44	M	733
48	M	1695
62	F	1041
46	M	2170
60	F	1121
48	M	1125
51	M	14691
46	M	2612
50	F	2658
50	M	900
33	M	1210
40	F	900
63	F	1431
65	M	1158
43	M	1908
38	M	3414
36	M	1255
49	M	2083
40	M	3095
69	F	837
54	M	661
57	M	911
64	F	4656
45	F	1400

297
395
703
101
906
296
155
929
159
298
111
297
929
922
142
295
124
531
702
924
532
905
709
399
901
296
906
922
903
296
905
999
922
299
922
402
395
103
924
532
112
903
142
922
394
395
411
129
126
924
909
531
111
296

60	0
62	0
26	0
4	0
91	0
75	0
68	0
36	0
11	0
13	0
85	0
91	0
91	0
2	0
13	0
70	0
45	0
71	0
68	0
1	0
19	0
17	0
86	0
3	0
19	0
59	0
18	0
88	0
9	0
53	0
96	0
99	0
95	0
98	0
17	0
13	0
4	0
99	0
23	0
35	0
66	0
75	0
82	0
24	0
31	0
95	0
74	0
13	0
30	0
12	0
18	0
14	0
20	0
78	0

40	F	5593
56	M	6927
44	M	1500
39	M	972
38	M	2971
53	M	2064
48	F	4056
70	M	5399
38	M	1161
48	M	7375
54	F	711
66	F	1287
65	F	4194
25	M	822
22	M	988
68	M	1402
88	M	3612
81	M	12702
57	M	695
47	M	858
39	M	1025
68	M	1301
27	M	1709
54	M	1231
74	M	3985
31	F	523
38	F	855
37	F	1135
37	F	1664
63	M	1736
54	M	3386
61	F	1300
46	F	1000
66	F	1628
58	M	3454
65	M	1364
39	M	939
56	M	790
37	M	1873
53	M	3609
35	F	1243
66	M	1101
63	M	5152
53	M	571
50	F	1100
53	M	3599
53	M	1861
63	M	3303
57	F	1114
52	M	1565
55	M	1215
52	M	2359
47	M	1661
52	F	559

142
156
295
904
111
403
297
101
301
303
923
922
906
126
402
922
297
922
923
901
902
922
115
115
922
902
297
143
395
922
111
922
145
923
922
115
905
923
125
902
158
923
121
923
999
155
117
216
114
143
923
902
296
143

96	1
13	1
22	1
5	1
5	1
27	1
88	1
22	1
21	1
4	1
9	1
91	1
4	1
85	1
35	0
50	0
58	0
59	0
40	0
15	0
14	0
40	0
28	0
35	0
30	0
26	0
19	0
3	0
4	0
70	0
60	0
22	0
4	0
23	0
96	0
13	0
14	0
9	0
1	0
30	0
5	0
21	0
3	0
13	0
4	0
21	0
37	0
79	0
72	0
1	0
36	0
99	0
62	0
65	0

39	M	11167
42	M	2947
56	F	2972
45	M	8342
62	M	9572
46	M	1919
43	M	6540
49	M	1464
48	M	3892
32	M	800
60	F	5634
29	M	3165
25	M	2410
31	M	1926
42	F	2248
44	F	9188
37	M	900
36	M	572
34	F	3947
58	M	875
53	M	7101
62	F	1743
31	M	900
59	F	2006
29	F	1917
57	M	7635
45	F	5400
50	M	1233
42	M	6944
52	M	2034
46	M	8722
58	M	5933
50	F	529
42	M	2573
33	M	6481
36	M	2461
55	M	1359
52	M	507
48	F	1104
56	M	1917
53	F	5072
47	F	2217
53	M	4340
49	M	683
50	M	2540
50	M	990
47	M	7308
49	F	4411
50	M	892
51	M	797
51	F	2186
41	M	667
42	M	2312
48	M	574

297
301
297
171
216
903
303
902
303
902
906
125
126
156
999
131
904
905
151
902
395
115
904
132
115
171
144
902
301
115
129
903
909
910
126
126
923
923
904
703
924
142
142
905
172
902
402
172
295
905
905
903
719
719

24	1
14	1
66	1
70	1
71	1
93	1
6	1
22	1
9	1
12	1
1	1
1	1
13	1
13	1
51	1
24	1
90	1
36	1
9	1
45	1
5	1
22	1
5	1
1	1
2	1
90	1
4	1
3	1
80	1
3	1
32	1
2	1
13	1
17	1
3	1
5	1
7	1
9	1
3	1
12	1
23	1
1	1
20	1
63	1
6	1
5	1
5	1
4	1
3	1
19	1
5	1
2	1
3	1
7	1

52	M	16137
41	M	767
53	M	2151
46	M	7257
30	M	2854
56	M	4514
42	F	3846
45	M	15182
49	F	2080
37	M	3740
49	M	4386
38	M	935
72	F	3477
50	F	9659
58	M	7807
34	M	1567
44	M	6470
40	M	2509
43	M	6187
35	M	1200
70	M	8565
36	F	1265
32	M	3184
53	F	2643
52	F	8947
36	F	2278
43	F	875
45	F	1974
38	F	1024
57	M	5221
35	F	2310
71	F	1782
54	M	14776
37	M	2469
58	M	2131
38	M	1167
46	F	9343
35	M	2379
35	M	4417
60	F	3384
51	M	2995
33	F	958
54	M	3738
37	M	3747
42	F	2785
35	M	1504
36	F	1853
45	M	1375
27	F	2550
59	M	2241
44	F	1245
53	M	1163
34	M	4736
37	M	3928

303
101
902
101
131
923
411
301
297
101
301
904
922
211
301
126
125
905
101
902
213
124
402
999
216
297
145
924
902
156
111
292
183
395
906
905
296
126
121
215
298
419
902
126
395
395
303
707
171
923
902
131
499
395

12	1
13	1
21	1
13	1
5	1
22	1
22	1
23	1
20	1
20	1
31	1
88	1
41	1
4	1
2	1
4	1
22	1
20	1
24	1
1	1
5	1
83	1
12	1
4	1
2	1
13	1
45	1
5	1
13	1
40	1
60	1
88	1
22	1
82	1
1	1
37	1
4	1
91	1
22	1
70	1
2	1
30	1
5	1
2	1
80	1
29	1
23	1
9	1
1	1
96	1
57	1
79	1
15	1
71	1

60	M	3250
52	M	5017
54	M	7111
37	M	3873
47	M	3442
49	M	800
54	M	1122
60	F	1147
67	F	955
53	M	1263
50	M	1084
54	M	1634
39	M	1679
40	M	8776
43	F	7943
58	M	2578
46	M	7646
55	F	2154
49	F	4124
47	M	5196
37	M	7198
42	M	1255
42	M	4219
44	M	4316
62	M	2625
58	M	1233
35	M	10094
49	M	800
41	M	1500
48	M	960
42	M	10394
46	F	1457
45	M	2009
46	F	3761
54	M	13996
40	F	3260
45	M	148703
36	M	3071
42	M	1941
50	F	6133
53	M	1172
41	F	1942
49	M	2148
37	M	1761
52	F	3464
38	M	1672
44	M	3973
41	M	2944
63	F	3770
46	M	1695
43	M	1658
34	M	3196
40	F	2281
48	M	1051

902
142
131
131
111
902
902
301
901
101
101
902
902
131
296
902
301
922
126
303
131
101
125
124
297
904
395
402
126
101
301
902
401
297
111
395
301
297
296
297
904
999
115
297
125
394
395
125
999
124
903
101
159
303

5	1
13	1
80	1
40	1
36	1
5	1
28	1
60	1
90	1
1	1
5	1
25	1
12	1
58	1
70	1
12	1
80	1
1	1
51	1
29	1
8	1
49	1
5	1
4	1
78	1
2	1
5	1
20	1
22	1
24	1
22	1
18	1
13	1
1	1
5	1
65	1
1	1
2	1
60	1
21	1
14	1
4	1
23	1
2	1
29	1
68	1
95	1
13	1
29	1
11	1
5	1
4	1
58	1
4	1

34	M	992
33	M	610
37	M	713
52	M	1171
49	F	873
42	F	1140
36	M	1634
34	F	569
32	F	1652
59	M	1046
57	M	1167
48	M	1000
30	M	663
27	M	2397
52	F	1074
51	F	2347
35	M	1100
42	M	1302
38	M	4058
49	F	1319
34	M	1029
34	M	821
47	F	2369
50	M	803
45	F	1937
54	F	758
28	M	1767
47	M	1262
42	M	2166
59	F	1998
24	M	1089
39	M	6572
40	M	3095
40	F	1312
41	M	1756
57	F	861
37	M	2008
48	F	3938
79	M	1800
46	M	4028
61	M	3472
38	M	4602
35	M	12542
50	M	8134
50	M	3265
72	M	16241
67	M	12162
48	M	27436
55	M	79755
75	M	7840
55	M	1003
40	M	6523
52	M	7750
60	M	2406

902
999
905
409
411
999
707
905
131
922
999
903
532
216
924
922
103
121
394
142
124
402
143
923
394
904
124
156
395
399
159
111
395
297
394
143
599
116
111
125
216
395
125
111
131
301
906
301
101
211
904
101
409
923

91	0
53	0
85	0
5	0
96	0
21	0
12	0
89	0
90	0
13	0
19	0
81	0
96	0
72	0
64	0
13	0
84	0
29	0
65	0
14	0
80	0
82	0
35	0
34	0
9	0
22	0
45	0
4	0
2	0
4	0
38	0
23	0
19	0
1	0
13	0
7	0
4	0
70	0
4	0
13	0
21	1
3	1
1	1
29	1
40	1
2	1
22	1
5	1
12	1
70	1
13	1
80	1
9	1
1	1

51	M	4184	101	21	1
45	F	3538	924	24	1
78	M	4720	906	8	0
42	M	1560	101	81	0
39	M	1723	143	17	0
42	M	854	531	2	0
40	M	1866	702	58	0
69	M	1184	923	6	0
50	M	947	539	8	0
57	M	1000	102	69	0
38	M	4364	999	4	0
49	M	3582	121	80	0
73	F	696	923	21	0
20	M	750	301	52	0
64	F	743	297	66	0
57	M	1473	999	41	0
35	F	2139	296	6	0
63	M	1197	923	13	0
38	F	942	131	15	0
35	M	1128	999	13	0
55	F	524	922	95	0
31	F	1222	529	9	0
66	M	1079	923	35	0
81	M	907	901	35	0
42	M	1249	112	60	0
34	M	1079	298	13	0
43	M	1320	901	62	0
26	M	1087	711	9	0
50	F	2392	139	30	0
53	M	1500	902	4	0
37	M	3059	395	78	0
43	F	980	143	38	0
75	M	950	923	38	0
38	M	1148	131	18	0
42	F	700	905	68	0
45	F	1663	394	92	0
45	M	1732	532	95	0
92	F	3773	301	24	0
59	F	1043	905	37	0
59	M	2248	296	20	0
41	M	1348	541	89	0
40	M	6262	395	96	0
50	F	1064	143	90	0
36	F	3313	182	4	0
33	F	1742	142	13	0
36	F	3658	297	90	0
43	M	3361	395	87	0
41	M	1250	535	84	0
40	M	1213	532	35	0
38	M	4139	395	79	0
48	M	2201	101	20	0
49	F	10456	211	4	0
38	M	2260	701	89	0
38	F	866	112	96	0

56	M	2547
39	F	611
38	F	5211
34	M	772
49	M	1524
46	M	1212
52	M	1047
36	F	2157
34	M	4571
30	F	1862
61	F	904
61	M	1336
53	M	810
56	F	4200
42	M	1415
41	F	971
50	M	1035
41	F	1449
27	F	773
42	F	1767
48	M	3107
28	M	615
46	F	1319
44	M	7311
30	F	2298
40	M	607
42	M	2293
47	M	1122
49	F	728
42	M	1588
29	M	824
51	F	4735
44	M	3047
44	F	1000
43	F	896
30	M	693
40	M	1793
23	M	542
54	F	1567
43	M	580
69	M	1646
45	M	4320
38	M	2422
41	M	2613
32	F	1953
88	M	1079
28	M	1269
54	M	3521
65	M	13094
70	M	32814
55	M	9218
55	M	2688
35	M	8734
47	F	1000

531
143
115
902
719
395
901
113
901
296
905
924
902
922
906
143
902
702
499
298
124
905
301
303
293
902
719
298
905
703
903
213
125
905
904
905
531
905
906
902
923
395
702
301
399
923
299
923
922
101
922
902
303
904

5	0
35	0
30	0
97	0
14	0
23	0
16	0
82	0
97	0
72	0
86	0
27	0
13	0
74	0
31	0
79	0
2	0
27	0
95	0
29	0
3	0
88	0
2	0
3	0
78	0
21	0
11	0
8	0
86	0
18	0
88	0
2	0
23	0
85	0
87	0
39	0
9	0
88	0
30	0
3	0
95	0
93	0
7	0
60	1
2	1
30	1
72	1
31	1
22	1
22	1
4	1
20	1
13	1
22	1

47	M	958	902	4	1
44	M	26605	111	11	1
66	M	7852	923	24	1
62	M	9597	922	14	1
54	M	4228	923	1	1
75	M	7637	921	21	1
96	M	863	904	24	1
52	M	2649	303	85	1
43	M	3422	903	40	1
57	M	1545	403	9	1
28	M	3329	395	53	1
70	F	2579	924	22	1
43	M	7624	101	20	1
43	M	6509	303	60	1
57	M	1521	923	6	1
57	M	14527	906	22	1
54	M	1557	903	24	1
31	F	3005	395	60	1
65	F	1087	923	20	1
52	M	1809	301	3	1
45	M	1644	999	22	1
64	M	1866	902	20	1
63	M	3634	906	22	1
53	M	1518	125	11	1
65	M	1480	301	90	1
52	M	3659	111	22	1
40	M	1119	902	25	1
42	F	4417	171	22	1
40	M	4025	144	21	1
47	M	840	902	9	1
35	M	1300	903	30	1
33	M	5082	126	21	1
37	M	2360	395	15	1
41	F	2925	309	8	1
36	M	1336	125	29	1
33	M	2161	902	13	1
33	M	2472	189	7	1
56	F	2616	395	36	1
52	M	7465	923	4	1
44	F	2890	132	21	1
48	F	2000	301	52	1
53	F	1669	924	4	1
57	F	2836	142	90	1
35	F	3960	125	13	1
37	M	4699	125	3	1
55	F	1083	172	5	1
41	M	8868	903	4	1
41	M	10322	395	74	1
31	M	5502	403	4	1
55	M	14204	101	13	1
68	M	9347	102	4	1
54	M	3975	395	24	1
31	M	3051	303	42	1
47	M	1369	301	24	1

45	M	1319	296	60	0
29	F	2312	392	6	0
72	M	2247	906	18	0
30	F	1332	394	7	0
46	M	1594	707	35	0
38	M	2320	395	69	0
46	M	3199	101	91	0
46	M	744	905	13	0
72	F	5792	143	13	0
63	M	1282	923	24	0
49	M	570	905	21	0
35	M	974	296	65	0
41	M	4487	107	90	0
35	F	1368	129	46	0
51	F	550	906	95	0
64	M	976	923	13	0
37	F	836	902	91	0
35	M	900	301	64	0
50	M	1169	904	7	0
35	F	1579	143	86	0
33	M	780	901	79	0
39	M	1333	903	91	0
45	M	825	301	95	0
30	M	658	901	76	0
40	M	1108	905	28	0
51	F	1985	143	13	0
49	M	1161	905	4	0
24	M	1039	543	95	0
31	M	1198	295	98	0
33	M	1100	124	37	0
27	M	1973	125	41	0
45	M	5505	707	11	0
42	F	1654	143	82	0
30	F	900	999	36	0
72	F	1080	999	22	0
54	M	1598	707	8	0
28	M	722	101	81	0
26	M	1198	159	31	0
50	F	548	905	89	0
40	F	4218	395	6	0
41	F	5412	303	3	0
34	F	617	905	27	0
29	M	623	112	85	0
42	M	1688	401	12	0
45	F	832	905	8	0
38	M	2334	703	78	0
60	M	1125	707	9	0
60	M	951	923	5	1
66	M	3976	131	30	1
49	F	4667	103	14	1
52	M	1190	104	1	1
50	M	15493	131	22	1
46	M	7684	101	27	1
70	M	18226	301	21	1

51	M	2265	158	30	1
32	M	2382	298	4	1
34	F	2375	395	4	1
38	F	2206	126	41	1
40	M	3251	101	29	1
39	M	842	401	38	1
38	F	5466	519	8	1
31	F	1943	143	4	1
46	F	1580	394	45	1
55	M	2171	902	12	1
38	M	1178	905	36	1
44	M	2083	905	3	1
40	M	2374	116	11	1
42	M	2648	172	5	1
46	M	683	902	92	1
46	F	3410	904	1	1
52	F	1117	151	4	1
47	F	7527	215	66	1
60	M	5608	101	22	1
40	M	926	903	12	1
39	M	1139	905	89	1
59	M	4334	929	66	0
56	M	2416	402	66	0
23	M	633	902	85	0
57	M	4926	139	87	0
39	F	850	409	82	0
59	M	1110	391	31	0
63	M	3424	923	93	0
26	M	2062	216	21	0
49	M	2225	124	65	0
55	M	1653	923	35	0
38	M	2883	597	11	0
74	M	679	922	22	0
34	F	900	592	4	0
64	M	1182	155	37	0
53	M	588	501	89	0
68	M	1185	711	20	0
58	M	1000	902	58	0
54	M	4067	923	4	0
56	F	1012	922	16	0
45	M	1685	303	64	0
33	F	1153	172	4	0
51	M	3019	151	64	0
43	M	900	905	5	0
71	M	1904	902	78	0
31	F	1434	293	18	0
78	M	1431	906	1	0
32	M	1101	509	5	0
45	M	896	904	60	0
40	M	2822	395	79	0
33	F	1127	395	13	0
31	F	2279	114	16	0
76	F	593	297	40	0
66	M	1018	906	3	0

42	F	1440	599	4	1
45	M	7416	125	51	1
41	M	6833	101	22	1
50	M	5301	922	25	1
53	M	1000	131	91	1
35	M	10913	303	4	1
46	M	4222	411	13	1
40	F	1373	395	46	1
38	M	9166	111	4	1
39	M	3715	707	13	1
31	M	1636	126	5	1
33	M	4011	126	70	1
35	M	2634	124	12	1
40	F	3950	144	37	1
32	M	3064	303	31	1
46	F	2967	142	73	1
53	M	3213	125	80	1
37	M	15941	395	40	1
25	F	1370	394	5	1
50	M	9333	922	20	1
59	M	9713	921	21	1
49	M	5766	125	71	1
55	M	6755	126	52	1
73	M	8862	395	41	1
53	F	4550	213	71	1
44	F	1059	393	3	1
47	M	3237	124	4	1
51	M	6060	101	22	1
32	M	5763	111	4	1
55	M	2280	102	30	1
72	M	12983	216	80	1
67	M	10613	213	23	1
42	M	3116	303	16	1
33	M	1208	151	32	1
28	F	1325	131	9	1
34	M	11057	125	3	1
34	F	1485	115	18	1
53	M	1167	902	5	1
45	M	5264	101	52	1
68	M	2999	922	90	1
38	M	1116	401	3	1
50	M	3055	216	53	1
63	M	4598	216	13	1
29	M	2224	402	30	1
31	M	527	905	4	1
31	M	2370	125	4	1
23	F	853	902	88	1
47	M	7501	125	22	1
43	M	7552	402	67	1
49	F	2692	297	74	1
56	M	9157	111	64	1
47	M	4738	142	78	1
62	M	2863	906	2	1
55	F	3379	922	40	1

37	F	1279	171	80	0
54	M	1250	999	79	0
48	M	1209	902	29	0
26	M	1487	159	48	0
32	M	1986	156	34	0
31	M	1137	999	83	0
48	M	1150	999	6	0
37	F	1600	143	4	0
45	M	1076	708	13	0
42	M	1987	707	13	0
39	F	1928	393	21	0
49	M	1603	531	8	0
25	M	1219	901	97	0
34	M	3754	101	27	0
36	M	1494	395	21	0
47	F	2366	299	32	0
51	M	729	903	8	0
28	M	958	902	64	0
49	F	4087	142	88	0
33	F	1640	395	89	0
47	M	4073	101	5	0
43	F	1359	395	16	0
74	F	1951	924	24	0
43	F	1008	193	5	0
42	M	1405	154	12	0
35	M	1050	707	12	0
56	F	550	905	36	0
35	M	4220	301	82	0
24	M	660	905	86	0
42	F	1229	511	21	0
30	M	3888	189	20	0
39	M	1069	904	13	0
75	F	638	902	83	0
62	M	920	906	88	0
45	F	973	923	89	0
39	F	544	905	93	0
52	F	2122	145	36	0
52	F	1084	143	28	0
41	M	1407	391	19	0
47	F	1092	409	71	1
43	M	3603	111	29	1
49	F	6322	297	41	1
50	M	4144	142	18	1
51	M	6569	101	1	1
56	M	4578	922	70	1
63	M	6141	923	70	1
65	M	2166	906	1	1
40	M	15411	301	96	1
62	M	950	923	4	1
54	M	17583	301	4	1
24	M	984	113	21	1
63	M	9945	921	22	1
68	M	24743	121	13	1
48	M	6275	111	18	1

47	F	1415	511	7	0
59	M	916	419	37	0
49	M	1100	904	95	0
57	M	861	301	35	0
42	M	752	297	40	0
48	M	842	532	13	0
38	M	1451	295	3	0
46	M	1724	111	22	0
42	F	1148	392	82	0
47	M	538	902	35	0
69	F	1175	297	4	0
51	F	1815	299	15	0
61	M	1073	922	25	0
40	F	1206	395	38	0
42	F	2172	399	88	0
50	M	3740	111	4	0
35	M	1379	999	40	0
47	F	2048	143	30	0
49	F	1668	393	21	0
47	F	2253	104	22	0
42	F	2525	395	91	0
51	F	2230	395	21	0
43	M	2288	395	85	0
53	M	4637	143	57	0
44	M	928	119	88	0
50	M	1215	902	22	0
72	M	1200	923	6	0
68	M	3533	142	14	0
42	M	1030	905	95	0
39	M	2418	395	80	0
48	M	6205	297	29	0
40	M	1011	295	88	0
36	F	4489	902	78	0
36	M	1132	702	96	0
37	F	1047	999	77	0
72	F	3836	924	80	0
46	M	1068	905	12	0
41	M	1098	296	39	0
55	F	2854	309	18	0
43	M	4193	115	29	0
43	M	2766	101	35	0
42	M	972	905	89	0
51	F	875	143	24	0
65	F	5238	924	91	0
36	F	1100	905	14	0
30	F	1387	296	77	0
34	M	1029	299	49	0
33	M	1054	129	95	0
65	F	1091	922	37	0
36	M	1083	905	81	0
51	M	3833	923	37	0
34	M	1798	155	89	0
30	M	1047	904	58	0
34	M	1421	295	18	0

52	M	1763	296	68	0
40	F	2063	395	17	0
26	M	1135	295	54	0
53	M	2320	922	58	0
64	M	551	923	87	0
63	F	1159	135	1	0
55	M	2123	901	16	0
47	M	2466	298	77	0
45	F	1772	929	69	0
49	M	4461	101	50	0
60	M	2484	923	13	0
38	M	1269	923	99	0
52	F	1368	923	13	0
35	M	1518	299	2	0
35	M	1187	391	15	0
33	M	1392	395	16	0
72	M	687	905	72	0
61	M	1400	124	5	0
35	M	2624	299	14	0
44	F	2694	103	66	0
56	F	1229	922	19	0
59	F	1262	143	14	0
51	F	1048	143	63	0
51	M	1978	292	57	0
37	M	2787	714	7	0
44	F	840	923	11	0
54	M	1578	922	98	0
57	M	784	905	44	0
72	M	3244	922	20	0
51	M	5031	298	11	0
67	M	1306	296	79	0
35	M	3735	111	4	0
43	M	1756	297	72	0
47	M	1021	905	39	0
62	M	1200	901	96	0
28	M	1789	125	6	0
47	F	1014	117	11	0
30	M	1124	297	13	0
27	M	881	532	14	0
30	F	1700	296	75	0
63	M	2580	295	9	0
80	M	4739	924	6	0
67	M	558	901	86	0
48	M	1538	395	78	0
57	M	2559	295	9	0
62	M	2410	921	80	0
52	M	805	901	87	0
42	F	1867	124	74	0
47	M	2000	902	92	0
52	F	3021	924	31	0
50	F	891	297	89	0
40	M	2784	905	68	0
61	M	1543	298	37	0
43	M	667	999	72	0

39	M	1113	903	29	1
35	F	2020	111	4	1
52	M	3801	142	24	1
42	F	6658	126	5	1
36	M	1665	904	9	1
37	M	5198	213	29	1
92	M	1841	902	98	1
47	M	3894	159	30	1
66	F	721	922	91	1
66	F	5975	301	91	1
58	M	629	101	3	1
55	F	2935	922	90	1
43	M	3553	124	50	1
74	F	5715	924	90	1
54	F	3236	906	50	1
40	M	1959	402	30	1
50	M	2025	124	63	1
30	F	987	111	21	1
36	F	1876	114	13	1
57	F	10965	301	5	1
51	M	3519	402	13	1
50	M	5741	131	1	1
44	M	1249	549	8	1
31	F	1681	392	4	1
30	M	1662	115	5	1
35	F	1011	904	5	1
38	M	4126	303	14	1
48	M	1371	409	21	1
56	M	552	923	2	1
56	F	2911	199	3	1
35	M	4405	101	4	1
54	F	2514	143	1	1
45	M	1044	905	13	1
30	M	1000	902	21	1
48	F	1676	296	21	1
40	F	880	905	4	1
70	F	1166	145	22	1
49	M	1309	923	44	1
46	F	4871	296	59	1
66	F	6994	906	21	1
48	M	2671	101	13	1
40	F	750	902	13	1
40	F	1160	101	58	1
53	M	1261	295	65	0
40	M	2371	101	84	0
63	M	7115	301	95	0
61	M	1473	922	59	0
41	F	1447	143	3	0
75	M	17082	923	2	0
54	M	9412	111	60	0
25	M	933	542	24	0
82	M	2096	923	95	0
28	M	1212	125	1	0
62	M	1406	923	80	0

40	M	3303
48	M	2593
49	F	2201
59	F	2449
38	F	4229
65	M	12961
49	M	4116
50	M	14891
52	M	9830
49	F	13154
42	F	2082
68	F	4187
40	M	4150
37	M	3810
37	F	4040
45	F	2808
42	F	14441
50	M	7775
60	M	6495
32	M	2844
39	M	3410
36	M	1126
42	M	2730
33	M	2125
37	M	3753
29	M	1856
70	M	3280
42	M	1181
36	M	1989
37	M	1799
59	M	2821
36	F	2063
35	M	2580
51	M	4591
52	M	4547
47	M	3431
56	M	8796
57	M	2355
40	M	3953
42	M	4717
37	F	1593
46	M	11707
60	M	21035
56	M	863
47	M	3837
55	F	1052
36	M	1061
55	M	1860
35	F	2481
29	M	3700
52	M	2443
37	M	2637
33	M	3205
53	F	6352

101
904
904
292
297
142
395
111
301
171
135
922
126
395
101
145
131
902
213
126
395
101
401
101
392
296
216
125
296
101
131
393
126
101
101
709
137
296
106
395
395
303
121
151
129
924
121
923
142
309
702
126
101
124

91	1
8	1
5	1
60	1
82	1
24	1
14	1
70	1
30	1
52	1
22	1
1	1
80	1
23	1
20	1
1	1
90	1
11	1
88	1
3	1
90	1
31	1
9	1
32	1
3	1
4	1
71	1
1	1
52	1
7	1
20	1
3	1
4	1
60	1
49	1
2	1
20	1
78	1
12	1
13	1
31	1
4	1
4	1
78	1
4	1
9	1
14	1
4	1
91	1
39	1
66	1
5	1
13	1
75	1

57	M	8323	296	62	0
44	F	1508	902	18	0
74	F	1289	922	21	0
55	M	1873	906	14	0
40	M	2805	101	9	0
56	M	2137	296	57	0
53	M	1166	401	60	0
59	M	908	922	47	0
57	F	1156	143	12	0
62	M	2617	922	24	0
36	M	636	903	8	0
80	F	950	924	11	0
33	M	3836	707	9	0
61	M	897	143	96	0
55	M	3410	395	96	0
41	F	971	395	60	0
64	M	1997	921	27	0
33	M	1206	704	9	0
58	M	3043	303	5	0
38	F	2695	142	4	0
33	F	1103	707	13	0
38	M	899	702	69	0
52	M	1417	923	27	0
49	M	618	601	89	0
48	M	990	154	93	0
66	M	1441	922	36	0
30	M	582	905	18	0
55	M	937	904	20	0
74	M	2203	922	85	0
66	M	4027	296	13	0
49	M	1904	549	92	0
29	M	1210	710	13	0
28	F	2396	392	1	0
59	F	1258	906	5	0
26	F	700	931	13	0
38	M	817	296	74	0
44	M	721	905	13	0
42	F	974	296	31	0
37	F	1620	296	54	0
38	M	1865	902	50	0
38	F	728	902	66	0
47	M	1136	903	13	0
38	F	1122	296	40	0
39	M	2232	999	91	0
43	M	2361	391	89	0
49	M	1433	905	36	0
38	F	832	539	96	0
43	F	6507	111	40	0
53	M	1060	409	9	0
53	F	1498	922	97	0
55	M	900	903	88	0
63	F	1246	922	90	0
36	F	603	509	60	0
35	M	529	142	50	0

33	F	712	902	76	0
32	M	613	904	75	0
41	M	1117	155	22	0
61	M	1686	544	13	0
37	F	2200	144	39	0
64	F	2092	922	82	0
34	F	1154	905	78	0
26	M	1164	158	98	0
45	M	1364	394	16	0
45	F	2003	906	30	0
43	F	1960	922	38	0
30	M	1080	902	85	0
41	M	2258	293	29	0
37	F	1045	296	26	0
30	M	1243	905	38	0
37	F	1922	113	20	0
30	F	1388	391	59	0
59	F	806	923	31	0
52	F	1825	143	18	0
46	M	1933	103	13	0
40	M	725	901	19	0
33	M	1000	541	22	0
44	M	1068	509	1	0
41	M	4529	391	4	0
41	M	1489	710	13	0
47	F	1765	292	9	0
47	M	1675	403	9	0
44	F	957	115	5	0

9.2 CONFECÇÃO DOS GRÁFICOS

*TABELA UTILIZADA NA CONFECÇÃO DO
GRÁFICO DE COMPRA TOTAL*

	Número de Elementos	Porcentagem dos Elementos
Comprou	925	43,35%
Não comprou	1209	56,65%
TOTAL	2134	100,00%

*TABELA UTILIZADA NA CONFECÇÃO
DOS GRÁFICOS DE SEXO*

	Número de Elementos	Porcentagem dos Elementos
Sexo Feminino	612	28,68%
Sexo Masculino	1522	71,32%
TOTAL	2134	100,00%
Sexo Feminino		
Comprou	310	50,65%
Não comprou	302	49,35%
TOTAL	612	100,00%
Sexo Masculino		
Comprou	908	59,66%
Não comprou	614	40,34%
TOTAL	1522	100,00%

TABELA UTILIZADA NA CONFECCÃO DOS GRÁFICOS DE RENDA

FAIXA (R\$)	Número de Elementos	Porcentagem dos Elementos	Indivíduos que Não Compraram	Indivíduos que Compraram	% de Compra	% de Não Compra
0 a 5000	1774	83,13%	892	882	49,72%	50,28%
5001 a 10000	258	12,09%	27	231	89,53%	10,47%
10001 a 15000	58	2,72%	4	54	93,10%	6,90%
15001 a 20000	24	1,12%	1	23	95,83%	4,17%
20001 a 25000	7	0,33%	0	7	100,00%	0,00%
25001 a 30000	3	0,14%	0	3	100,00%	0,00%
30001 a 35000	4	0,19%	0	4	100,00%	0,00%
35001 a 40000	1	0,05%	0	1	100,00%	0,00%
40001 a 45000	0	0,00%	0	0	0,00%	100,00%
45001 a 50000	1	0,05%	0	1	100,00%	0,00%
50001 a 55000	0	0,00%	0	0	0,00%	100,00%
55001 a 60000	0	0,00%	0	0	0,00%	100,00%
60001 a 65000	0	0,00%	0	0	0,00%	100,00%
65001 a 70000	0	0,00%	0	0	0,00%	100,00%
70001 a 75000	0	0,00%	0	0	0,00%	100,00%
75001 a 80000	1	0,05%	0	1	100,00%	0,00%
80001 a 85000	0	0,00%	0	0	0,00%	100,00%
85001 a 90000	0	0,00%	0	0	0,00%	100,00%
90001 a 95000	0	0,00%	0	0	0,00%	100,00%
95001 a 10000	0	0,00%	0	0	0,00%	100,00%
100001 a 105000	0	0,00%	0	0	0,00%	100,00%
105001 a 110000	1	0,05%	0	1	100,00%	0,00%
110001 a 115000	2	0,09%	0	2	100,00%	0,00%
TOTAL	2134	100,00%				

TABELA UTILIZADA NA CONFECCÃO DOS GRÁFICOS DE LOCALIZAÇÃO

REGIÃO	Número de Elementos	Porcentagem dos Elementos	Indivíduos que Compraram	% de Compra	Indivíduos que Não Compraram	% de Não Compra
Norte		2,20%	21	44,68%	26	55,32%
Nordeste	47	11,25%	149	62,08%	91	37,92%
Sudeste	240	65,84%	880	62,63%	525	37,37%
Centro-Oeste	1405	5,95%	52	40,94%	75	59,06%
Sul	127	14,76%	110	34,92%	205	65,08%
TOTAL	315	100,00%	1212		922	

TABELAS UTILIZADAS NA CONFEÇÃO DOS GRÁFICOS DE LOCALIZAÇÃO REGIONAL

REGIÃO	ESTADOS COBERTOS PELA EMPRESA	1o. E 2o. DÍGITOS DO CEP	FAIXA
Norte	Pará	66 a 68	66 a 69
	Amazonas	69	
Nordeste	Maranhão	65	40 a 65
	Bahia	40 a 48	
	Ceará	60 a 63	
	Rio Grande do Norte	59	
	Paraíba	58	
	Pernambuco	50 a 56	
	Alagoas	57	
	Sergipe	49	
	Piauí	64	
Sudeste	Minas Gerais	30 a 39	1 a 39
	São Paulo	1 a 19	
	Rio de Janeiro	20 a 28	
	Espírito Santo	29	
Centro-Oeste	Mato Grosso	78	70 a 79
	Goiás	74 a 77	
	Mato Grosso do Sul	79	
	Distrito Federal	70 a 73	
Sul	Paraná	80 a 87	80 a 99
	Santa Catarina	88 a 89	
	Rio Grande do Sul	90 a 99	

QUALIFICAÇÃO DAS PROFISSÕES

CÓDIGO	PROFISSÃO	GRUPO	QUALIFICAÇÃO
101	Engenheiro	10	ALTA QUALIFICAÇÃO
102	Arquiteto	10	ALTA QUALIFICAÇÃO
103	Agrônomo	10	ALTA QUALIFICAÇÃO
104	Químico	10	ALTA QUALIFICAÇÃO
105	Desenhista Industrial	10	ALTA QUALIFICAÇÃO
106	Físico	10	ALTA QUALIFICAÇÃO
107	Geólogo	10	ALTA QUALIFICAÇÃO
108	Tecnólogo	10	ALTA QUALIFICAÇÃO
109	Outros (Engenheiros/Físicos/Químicos/Assemelhados)	11	ALTA QUALIFICAÇÃO
111	Médico	11	ALTA QUALIFICAÇÃO
112	Veterinário/Zootecnista	11	ALTA QUALIFICAÇÃO
113	Enfermeiro/Nutricionista	11	ALTA QUALIFICAÇÃO
114	Fisioterapeuta/Terapeuta Ocupacional	11	ALTA QUALIFICAÇÃO
115	Odontólogo	11	ALTA QUALIFICAÇÃO
116	Biólogo/Biomédico	11	ALTA QUALIFICAÇÃO
117	Farmacêutico	11	ALTA QUALIFICAÇÃO
118	Fonoaudiólogo	11	ALTA QUALIFICAÇÃO
119	Outros (Médicos/Odontólogos/Enfermeiros/Assemelhados)	12	ALTA QUALIFICAÇÃO
121	Economista	12	ALTA QUALIFICAÇÃO
122	Estatístico	12	ALTA QUALIFICAÇÃO
123	Atuário e Matemático	12	ALTA QUALIFICAÇÃO
124	Contador	12	ALTA QUALIFICAÇÃO
125	Administrador	12	ALTA QUALIFICAÇÃO
126	Analista de Sistema	12	ALTA QUALIFICAÇÃO
127	Geógrafo	12	ALTA QUALIFICAÇÃO
128	Astrônomo/Meteorologista	12	ALTA QUALIFICAÇÃO
129	Outros (Economista/Estatístico/Contador/Assemelhados)	13	ALTA QUALIFICAÇÃO
131	Advogado	13	ALTA QUALIFICAÇÃO
132	Psicólogo	13	ALTA QUALIFICAÇÃO
133	Sociólogo	13	ALTA QUALIFICAÇÃO
134	Assistente Social	13	ALTA QUALIFICAÇÃO
135	Bibliotecário, Arquivista, Museólogo, Arqueólogo	13	ALTA QUALIFICAÇÃO
136	Comunicólogo	13	ALTA QUALIFICAÇÃO
137	Relações Públicas	13	ALTA QUALIFICAÇÃO
138	Profissionais de Letras e Artes	13	ALTA QUALIFICAÇÃO
139	Outros (Advogado/Psicólogo/Bibliotecário/Assemelhados)	14	ALTA QUALIFICAÇÃO
142	Professor de Ensino Superior	14	ALTA QUALIFICAÇÃO
143	Professor de Primeiro e Segundo Grau	14	ALTA QUALIFICAÇÃO
144	Diretor de Estabelecimento de Ensino	14	ALTA QUALIFICAÇÃO
145	Outros Trabalhos de Nível Superior ligados ao Ensino	15	ALTA QUALIFICAÇÃO
151	Técnico de Contabilidade e Estatística	15	ALTA QUALIFICAÇÃO
152	Técnico de Biologia	15	ALTA QUALIFICAÇÃO
153	Técnico de Agronomia e Agrimensura	15	ALTA QUALIFICAÇÃO
154	Técnico de Química	15	ALTA QUALIFICAÇÃO
155	Técnico de Mecânica	15	ALTA QUALIFICAÇÃO
156	Técnico de Eletricidade, Eletrônica, Telecomunicação	15	ALTA QUALIFICAÇÃO
157	Técnico de Laboratório e Raio X	15	ALTA QUALIFICAÇÃO
158	Desenhista Técnico	15	ALTA QUALIFICAÇÃO
159	Outros Técnicos	16	ALTA QUALIFICAÇÃO
161	Empresário/Produtor de Espetáculos Públicos	16	ALTA QUALIFICAÇÃO
162	Ator/Diretor de Espetáculos	16	ALTA QUALIFICAÇÃO
163	Cantor/Compositor	16	ALTA QUALIFICAÇÃO
164	Músico	16	ALTA QUALIFICAÇÃO
165	Coreógrafo/Bailarino	16	ALTA QUALIFICAÇÃO
166	Locutor/Comentarista	16	ALTA QUALIFICAÇÃO
167	Operador de Câmara de TV e Cinema	16	ALTA QUALIFICAÇÃO
168	Atleta Profissional e Técnico em Desporto	16	ALTA QUALIFICAÇÃO
169	Outros (Espetáculos Públicos/Prática Desportiva)	17	ALTA QUALIFICAÇÃO
171	Jornalista	17	ALTA QUALIFICAÇÃO
172	Publicitário	17	ALTA QUALIFICAÇÃO
173	Relações Públicas	17	ALTA QUALIFICAÇÃO
179	Outros (Jornalista/Assembleado)	18	ALTA QUALIFICAÇÃO
181	Piloto de Aeronaves	18	ALTA QUALIFICAÇÃO
182	Comissários de Bordo		

		18	ALTA QUALIFICAÇÃO
		18	ALTA QUALIFICAÇÃO
183	Comandante de Embarcações	19	ALTA QUALIFICAÇÃO
189	Outros (Navegação Aérea/ Marítima e Interior)	19	ALTA QUALIFICAÇÃO
191	Escultor/ Pintor/ Assemelhados	19	ALTA QUALIFICAÇÃO
192	Desenhista Comercial	19	ALTA QUALIFICAÇÃO
193	Decorador	20	ALTA QUALIFICAÇÃO
199	Outros (Profissão Científica/ Técnica/ Artística)	20	ALTA QUALIFICAÇÃO
201	Membros do Poder Legislativo	20	ALTA QUALIFICAÇÃO
203	Membros do Poder Executivo	20	ALTA QUALIFICAÇÃO
205	Membros do Poder Judiciário	29	ALTA QUALIFICAÇÃO
209	Outros	29	ALTA QUALIFICAÇÃO
291	Ocupante Cargo de Direção/ Assessoramento Intermediário	29	ALTA QUALIFICAÇÃO
292	Agente Administrativo	29	ALTA QUALIFICAÇÃO
293	Serventuário de Justiça	29	ALTA QUALIFICAÇÃO
294	Tabelião	29	ALTA QUALIFICAÇÃO
295	Militar em Geral	29	ALTA QUALIFICAÇÃO
296	Servidor Público Federal	29	ALTA QUALIFICAÇÃO
297	Servidor Público Estadual	29	ALTA QUALIFICAÇÃO
298	Servidor Público Municipal	30	ALTA QUALIFICAÇÃO
299	Outros (Servidor Civil/ Militar/ Agente Adm. Pública)	30	ALTA QUALIFICAÇÃO
301	Diretor de Empresas	30	ALTA QUALIFICAÇÃO
303	Gerente	30	ALTA QUALIFICAÇÃO
309	Outros (Assemelhados a Diretores e Gerentes)	38	ALTA QUALIFICAÇÃO
309	Assessor	38	ALTA QUALIFICAÇÃO
380	Analista de Pesquisa/ Financeiro/ Técnico de Planejamento	38	ALTA QUALIFICAÇÃO
381	Analista	38	ALTA QUALIFICAÇÃO
385	Analista de Sistemas	38	ALTA QUALIFICAÇÃO
386	Operador Computador	38	ALTA QUALIFICAÇÃO
387	Digitador	39	MÉDIA QUALIFICAÇÃO
388	Programador	39	MÉDIA QUALIFICAÇÃO
391	Chefe Intermediário	39	MÉDIA QUALIFICAÇÃO
392	Serviços de Contabilidade/ Caixa/ Assemelhados	39	MÉDIA QUALIFICAÇÃO
393	Secretário/ Estenógrafo/ Datilógrafo/ Recepcionista/ Telefonista	39	MÉDIA QUALIFICAÇÃO
394	Auxiliar de Escritório	39	MÉDIA QUALIFICAÇÃO
395	Bancário e Econômiário	39	MÉDIA QUALIFICAÇÃO
396	Securitário	39	MÉDIA QUALIFICAÇÃO
398	Consultor	40	MÉDIA QUALIFICAÇÃO
399	Outros (Administrativos e Assemelhados)	40	MÉDIA QUALIFICAÇÃO
401	Supervisor, Inspetor, Agente de Compras e Vendas	40	MÉDIA QUALIFICAÇÃO
402	Vendedor Praticista, Representante Comercial, Caixeiro Viajante	40	MÉDIA QUALIFICAÇÃO
403	Corretor de Imóveis, Seguros, Títulos e Valores	40	MÉDIA QUALIFICAÇÃO
404	Leiloeiro, Avaliador e Assemelhados	40	MÉDIA QUALIFICAÇÃO
405	Agenciador de Propaganda	41	MÉDIA QUALIFICAÇÃO
409	Outros (Supervisores/ Corretores/ Assemelhados)	41	MÉDIA QUALIFICAÇÃO
410	Vendedora Autônoma	41	MÉDIA QUALIFICAÇÃO
411	Vendedor de Comércio Varejista e Atacadista	41	MÉDIA QUALIFICAÇÃO
412	Jornaleiro	41	MÉDIA QUALIFICAÇÃO
413	Feirante	41	MÉDIA QUALIFICAÇÃO
414	Agente de Aeroporto/ Navegação/ Transportes	49	MÉDIA QUALIFICAÇÃO
419	Outros (Vendedores e Assemelhados)	49	MÉDIA QUALIFICAÇÃO
491	Demonstrador	49	MÉDIA QUALIFICAÇÃO
492	Modelo de Modas	50	MÉDIA QUALIFICAÇÃO
499	Outros (Trabalhadores do Comércio e Assemelhados)	50	MÉDIA QUALIFICAÇÃO
501	Porteiro, Ascensorista, Garagista, Faxineiro de Edifício	51	MÉDIA QUALIFICAÇÃO
509	Outros (Serviços Administrativos/ Conservação/ Limpeza de Edifício)	51	MÉDIA QUALIFICAÇÃO
511	Barbeiro, Cabeleireiro, Manicura, Pedicura, Maquiador, Esteticista, Massagista	52	MÉDIA QUALIFICAÇÃO
519	Outros (Serviços de Beleza)	52	MÉDIA QUALIFICAÇÃO
521	Camareiro/ Porteiro/ Garçom/ Cozinheiro de Hotel	53	MÉDIA QUALIFICAÇÃO
529	Outros (Serviços de Hospedagem e Restaurante)	53	MÉDIA QUALIFICAÇÃO
531	Motorista (Transporte de Passageiros)	53	MÉDIA QUALIFICAÇÃO
532	Motorista (Transporte de Carga)	53	MÉDIA QUALIFICAÇÃO
533	Contramestre de Embarcações	53	MÉDIA QUALIFICAÇÃO
534	Marinheiro e Assemelhados	53	MÉDIA QUALIFICAÇÃO
535	Maquinista/ Foguista de Embarcações e Locomotivas	53	MÉDIA QUALIFICAÇÃO
536	Motorista	54	MÉDIA QUALIFICAÇÃO
539	Outros (Trabalhadores de Serviços de Transporte)		
541	Mecânico de Manutenção de Veículos e Máquinas		

		54	MEDIA QUALIFICAÇÃO
		54	MEDIA QUALIFICAÇÃO
		54	MEDIA QUALIFICAÇÃO
542	Eletricista de Manutenção	54	MEDIA QUALIFICAÇÃO
543	Pintor/ Lanterneiro de Veículos	54	MEDIA QUALIFICAÇÃO
544	Bombeiro, Instalador de Água, Luz, Esgoto e Assemelhados	59	MEDIA QUALIFICAÇÃO
549	Outros (Trabalhadores de Serviços de Manutenção)	59	MEDIA QUALIFICAÇÃO
591	Alfaiate	59	MEDIA QUALIFICAÇÃO
592	Protético	59	MEDIA QUALIFICAÇÃO
593	Despachante	59	MEDIA QUALIFICAÇÃO
594	Agente de Viagem/ Guia Turístico	59	MEDIA QUALIFICAÇÃO
595	Agente Funerário	59	MEDIA QUALIFICAÇÃO
596	Auxiliar de Laboratório	59	MEDIA QUALIFICAÇÃO
597	Estivador/ Carregador/ Embalador	59	MEDIA QUALIFICAÇÃO
598	Empregado Doméstico	59	MEDIA QUALIFICAÇÃO
599	Outros (Trabalhadores de Serviços e Assemelhados)	60	OUTROS
601	Trabalhador Agrícola	60	OUTROS
602	Trabalhador da Pecuária	60	OUTROS
603	Trabalhador Florestal	60	OUTROS
604	Trabalhador da Pesca	60	OUTROS
605	Garimpeiro	60	OUTROS
609	Outros (Trabalhador Agrícola/ Pecuária, Florestal/ Pesca/ Extração)	70	OUTROS
701	Mestre e Contramestre	70	OUTROS
702	Mecânico/ Montador/ Preparador/ Operador (Indústria)	70	OUTROS
703	Eletricista/ Assemelhados da Indústria	70	OUTROS
704	Trabalhador de Instalação de Processo Químico	70	OUTROS
705	Trabalhador de Fabricação de Roupas	70	OUTROS
706	Trabalhador de Tratamento e Fabricação de Fumo	70	OUTROS
707	Trabalhador Metalúrgico/ Siderúrgico	70	OUTROS
708	Trabalhador de Usinagem de Metais	70	OUTROS
709	Trabalhador de Construção Civil	71	OUTROS
710	Trabalhador de Fabricação e Preparação de Alimentos e Bebidas	71	OUTROS
711	Trabalhador de Artes Gráficas	71	OUTROS
712	Trabalhador de Produtos Têxteis (Exceto Roupas)	71	OUTROS
713	Trabalhador de Artefatos de Madeira	71	OUTROS
714	Trabalhador de Fabricação de Papel e Papelão	71	OUTROS
715	Trabalhador de Fabricação de Calçados e Artefatos de Couro	71	OUTROS
716	Trabalhador de Produtos de Borracha e Plástico	71	OUTROS
717	Joalheiro e Ourives	71	OUTROS
719	Outros Trabalhadores da Produção Industrial	74	OUTROS
740	Costureira	75	OUTROS
750	Operador Produção	75	OUTROS
759	Impressor	76	OUTROS
760	Inspetor Mecânico	76	OUTROS
761	Instrumentista	76	OUTROS
762	Mecânico Industrial	76	OUTROS
764	Maquinista	77	OUTROS
769	Operador/ Forjador	77	OUTROS
771	Bombeiro	80	OUTROS
774	Marceneiro	80	OUTROS
801	Auxiliar de Escritório	80	OUTROS
802	Auxiliar de Enfermagem	80	OUTROS
803	Fornecido/ Montador/ Assistente de Produção	80	OUTROS
805	Consertador/ Servente	80	OUTROS
806	Manobrista	81	OUTROS
807	Agente de Segurança	90	OUTROS
811	Doméstica	90	OUTROS
901	Proprietário de Estabelecimento Agrícola, Pecuária, Florestal	90	OUTROS
902	Proprietário de Estabelecimento Comercial	90	OUTROS
903	Proprietário de Estabelecimento Industrial	90	OUTROS
904	Proprietário de Estabelecimento de Serviços	90	OUTROS
905	Proprietário de Microempresa	90	OUTROS
906	Proprietário de Imóvel Recebendo Rendimento de Aluguel	90	OUTROS
907	Capitalista Recebendo Rendimento de Aplicações Financeiras	90	OUTROS
908	Empresário	91	OUTROS
909	Outros Proprietários e Capitalistas	91	OUTROS
910	Sacerdote ou Membro de Ordem ou Seita Religiosa	92	OUTROS
919	Outros Trabalhadores Religiosos		
921	Militar Reformado		

		92	OUTROS
		92	OUTROS
922	Funcionário Público Civil Aposentado	92	OUTROS
923	Aposentado (Exceto Funcionário Público)	92	OUTROS
924	Pensionista	92	OUTROS
929	Outros (Aposentados e Pensionistas)	93	OUTROS
931	Bolsista, Estagiário e Assemelhados	93	OUTROS
932	do Lar	93	OUTROS
939	Outros Estudantes	94	OUTROS
949	Espólios	99	OUTROS
998	Autônomo	99	OUTROS
999	Outras Ocupações		

TABELA UTILIZADA NA CONFECCÃO DOS GRÁFICOS DE PROFISSÃO

QUALIFICAÇÃO	NÚMERO DE ELEMENTOS	PORCENTAGEM	% de Compra	% de Não Compra
Alta	1052	49,30%	66,06%	33,94%
Média	342	16,03%	39,09%	60,91%
Outros	740	34,68%	43,11%	56,89%
TOTAL	2134	100,00%		

TABELA UTILIZADA NA CONFECCÃO DOS GRÁFICOS DE IDADE

FAIXA	Número de Elementos	Porcentagem dos Elementos	Indivíduos que Não Compraram	% Não Compra	Indivíduos que Compraram	% Compra
18 a 24	21	0,98%	16	76,19%	5	23,81%
25 a 29	82	3,84%	44	53,66%	38	46,34%
30 a 34	187	8,76%	118	63,10%	69	36,90%
35 a 39	305	14,29%	133	43,61%	172	56,39%
40 a 44	377	17,67%	163	43,24%	214	56,76%
45 a 49	362	16,96%	143	39,50%	219	60,50%
50 a 54	298	13,96%	108	36,24%	190	63,76%
55 a 59	207	9,70%	114	55,07%	93	44,93%
60 a 64	121	5,67%	57	47,11%	64	52,89%
65 a 69	84	3,94%	34	40,48%	50	59,52%
70 a 74	51	2,39%	23	45,10%	28	54,90%
75 a 97	39	1,83%	27	69,23%	12	30,77%
TOTAL	2134	100,00%	980		1154	

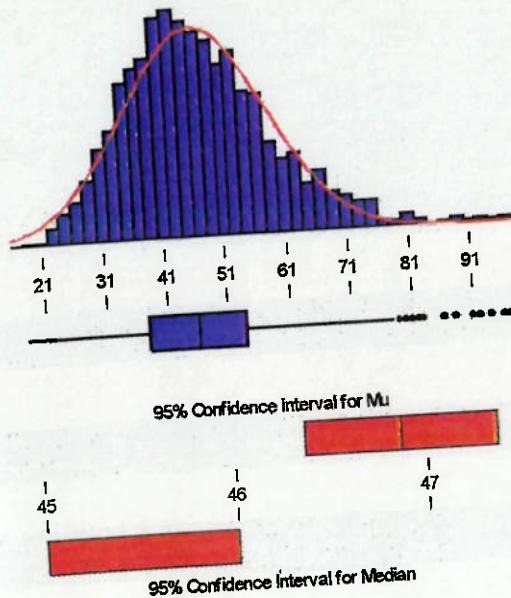
9.3 EXEMPLAR DE FICHA CADASTRAL

10. ANEXO B

10.1 ANÁLISE ESTATÍSTICA DO CONJUNTO DE DADOS

Descriptive Statistics for IDADE

Variable: C3



Anderson-Darling Normality Test

A-Squared: 9.043
P-Value: 0.000

Mean: 46.8673
StDev: 11.8027
Variance: 139.305
Skewness: 0.592428
Kurtosis: 0.509286
N: 2133

Minimum: 18.0000
1st Quartile: 38.0000
Median: 46.0000
3rd Quartile: 54.0000
Maximum: 97.0000

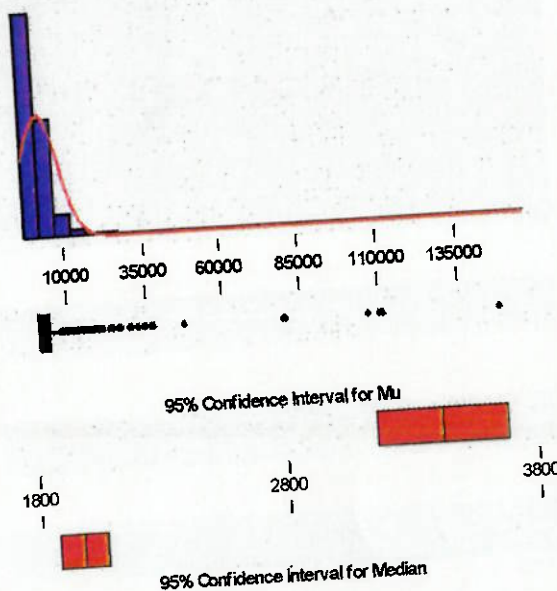
95% Confidence Interval for Mu
46.3662 47.3685

95% Confidence Interval for Sigma
11.4589 12.1680

95% Confidence Interval for Median
45.0000 46.0000

Descriptive Statistics for Renda

Variable: C1



Anderson-Darling Normality Test

A-Squared: 315.945
P-Value: 0.000

Mean: 3420.85
StDev: 6066.09
Variance: 36797482
Skewness: 13.1503
Kurtosis: 252.001
N: 2133

Minimum: 507
1st Quartile: 1132
Median: 1968
3rd Quartile: 3867
Maximum: 148703

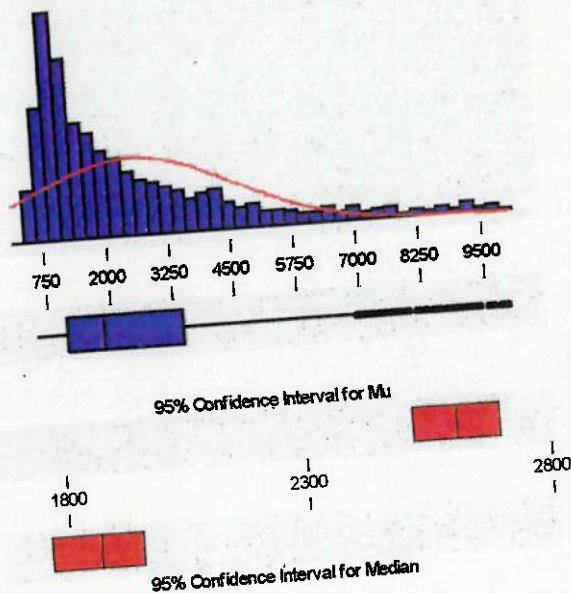
95% Confidence Interval for Mu
3163 3678

95% Confidence Interval for Sigma
5889 6254

95% Confidence Interval for Median
1872 2063

Descriptive Statistics for RENDA ATE R\$ 10.000,00

Variable: C2



Anderson-Darling Normality Test

A-Squared: 112.724
P-Value: 0.000

Mean: 2609.09
StDev: 2047.49
Variance: 4192205
Skewness: 1.50396
Kurtosis: 1.77794
N: 2032

Minimum: 507.00
1st Quartile: 1117.50
Median: 1863.50
3rd Quartile: 3439.50
Maximum: 9962.00

95% Confidence Interval for Mu
2520.01 2698.16

95% Confidence Interval for Sigma
1986.42 2112.46

95% Confidence Interval for Median
1761.67 1948.33

10.2 PARTE DO CONJUNTO DE DADOS JÁ CODIFICADA

Idade	Sexo	Renda	Código da Profissão	Dígitos do CEP	Compu	DMSEXO	DM1PRO	DM2PRO	DM1LOC	DM2LOC	DM3LOC	DM4LOC	DM1REN	DM2REN	DM3REN	DM4REN	DM1IDA	DM2IDA	DM3IDA	DM4IDA
66	M	5471	922	22	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0
61	M	2196	301	13	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
46	M	900	904	4	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1
34	M	2488	905	70	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0
60	M	1480	131	11	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
70	M	2558	923	22	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0
45	M	1427	904	13	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
51	M	3440	101	11	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1
67	M	7952	117	70	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
53	M	14677	904	11	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
55	M	532	101	4	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0
55	M	1817	902	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
48	M	1082	172	41	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
46	M	3809	101	59	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
46	M	3142	131	22	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
40	M	5091	101	21	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
68	M	2045	111	25	1	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0
53	M	1544	923	26	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
57	M	2763	125	5	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0

10.3 REGRESSÃO LOGÍSTICA: PASSO 1

Worksheet size: 100000 cells

Retrieving worksheet from file: C:\GUILHERME\POLI\TF\CONJUNTO DE DADOS\TREINAMENTO.M

TW

Worksheet was saved on 1/ 9/1996

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	0.3621	0.2563	1.41	0.158	1.00	0.99	1.01
Idade	-0.000366	0.005285	-0.07	0.945			

Log-Likelihood = -723.280

Test that all slopes are zero: G = 0.005, DF = 1, P-Value = 0.945

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	83.178	64	0.054
Deviance	93.691	64	0.009
Hosmer-Lemeshow	13.698	8	0.090

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	56	63	76	87	72	66	63	69	64	8	624
Exp	63.0	68.4	69.6	76.1	67.9	66.2	62.7	72.7	64.5	12.9	
0											
Obs	52	54	43	43	44	47	44	55	46	14	442
Exp	45.0	48.6	49.4	53.9	48.1	46.8	44.3	51.3	45.5	9.1	
Total	108	117	119	130	116	113	107	124	110	22	1066

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	71818	26.0%	Somers D	0.01
Discordant	68457	24.8%	Goodman-Kruskal Gamma	0.02
Ties	135533	49.1%	Kendalls Tau-a	0.01
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
Constant	-0.8253	0.1168	-7.07	0.000	1.00	1.00	1.00
Renda	0.00045695	0.00004446	10.28	0.000			

Log-Likelihood = -627.544

Test that all slopes are zero: G = 191.475, DF = 1, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1404.059	922	0.000
Deviance	1059.173	922	0.001
Hosmer-Lemeshow	11.933	8	0.154

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	35	44	36	47	50	67	73	80	91	101	624
Exp	39.8	43.1	44.9	48.3	53.2	58.0	66.5	76.5	90.2	103.6	
0											
Obs	71	63	70	60	58	39	33	26	16	6	442
Exp	66.2	63.9	61.1	58.7	54.8	48.0	39.5	29.5	16.8	3.4	
Total	106	107	106	107	108	106	106	106	107	107	1066

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	203128	73.6%	Somers D	0.48
Discordant	71287	25.8%	Goodman-Kruskal Gamma	0.48
Ties	1393	0.5%	Kendalls Tau-a	0.23
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
						Lower	Upper

Constant	0.43358	0.07363	5.89	0.000	0.73	0.56	0.95
DMSEXO	-0.3174	0.1383	-2.30	0.022			

Log-Likelihood = -720.655

Test that all slopes are zero: G = 5.253, DF = 1, P-Value = 0.022

* NOTE * No goodness of fit tests performed.
 * The model uses all degrees of freedom.

Measures of Association:
 (Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	64722	23.5%	Somers D	0.06
Discordant	47120	17.1%	Goodman-Kruskal Gamma	0.16
Ties	163966	59.4%	Kendalls Tau-a	0.03
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	0.74222	0.09346	7.94	0.000			
DM1PRO	-0.5134	0.1854	-2.77	0.006	0.60	0.42	0.86
DM2PRO	-0.8569	0.1385	-6.19	0.000	0.42	0.32	0.56

Log-Likelihood = -703.495

Test that all slopes are zero: G = 39.575, DF = 2, P-Value = 0.000

* NOTE * No goodness of fit tests performed.
 * The model uses all degrees of freedom.

Measures of Association:
 (Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	114779	41.6%	Somers D	0.21
Discordant	58131	21.1%	Goodman-Kruskal Gamma	0.33
Ties	102898	37.3%	Kendalls Tau-a	0.10
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624

0 442
Total 1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-0.5108	0.4216	-1.21	0.226	2.39	0.95	6.02
DM1LOC	0.8726	0.4704	1.86	0.064	3.12	1.35	7.23
DM2LOC	1.1376	0.4289	2.65	0.008	1.15	0.44	3.02
DM3LOC	0.1431	0.4904	0.29	0.770	1.10	0.45	2.66
DM4LOC	0.0953	0.4505	0.21	0.832			

Log-Likelihood = -698.549

Test that all slopes are zero: G = 49.465, DF = 4, P-Value = 0.000

* NOTE * No goodness of fit tests performed.
* The model uses all degrees of freedom.

Measures of Association: (Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	102657	37.2%	Somers D	0.20
Discordant	47145	17.1%	Goodman-Kruskal Gamma	0.37
Ties	126006	45.7%	Kendalls Tau-a	0.10
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-0.5053	0.1482	-3.41	0.001	1.41	1.00	1.99
DM1REN	0.3421	0.1763	1.94	0.052	4.62	3.08	6.92
DM2REN	1.5306	0.2063	7.42	0.000	15.05	7.92	28.60
DM3REN	2.7111	0.3277	8.27	0.000	25.42	7.63	84.68
DM4REN	3.2354	0.6140	5.27	0.000			

Log-Likelihood = -631.325

Test that all slopes are zero: G = 183.914, DF = 4, P-Value = 0.000

* NOTE * No goodness of fit tests performed.
* The model uses all degrees of freedom.

Measures of Association: (Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	161265	58.5%	Somers D	0.43
Discordant	43377	15.7%	Goodman-Kruskal Gamma	0.58
Ties	71166	25.8%	Kendalls Tau-a	0.21
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
Constant	0.85847	0.08944	9.60	0.000			
DM1IDA	-1.2441	0.2396	-5.19	0.000	0.29	0.18	0.46
DM2IDA	-0.8936	0.2796	-3.20	0.001	0.41	0.24	0.71
DM3IDA	-0.3989	0.2756	-1.45	0.148	0.67	0.39	1.15
DM4IDA	-1.2955	0.1544	-8.39	0.000	0.27	0.20	0.37

Log-Likelihood = -679.181

Test that all slopes are zero: G = 88.202, DF = 4, P-Value = 0.000

* NOTE * No goodness of fit tests performed.
* The model uses all degrees of freedom.

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	131382	47.6%	Somers D	0.30
Discordant	49498	17.9%	Goodman-Kruskal Gamma	0.45
Ties	94928	34.4%	Kendalls Tau-a	0.14
Total	275808	100.0%		

10.4 REGRESSÃO LOGÍSTICA: PASSO 2

Worksheet size: 100000 cells

Retrieving worksheet from file: C:\GUILHERME\POLI\TF\CONJUNTO DE DADOS\TREINAMENTO.M
TW
Worksheet was saved on 1/ 9/1996

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-0.4663	0.2882	-1.62	0.106			
Renda	0.00045924	0.00004442	10.34	0.000	1.00	1.00	1.00
Idade	-0.007815	0.005762	-1.36	0.175	0.99	0.98	1.00

Log-Likelihood = -626.622

Test that all slopes are zero: G = 193.320, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1513.977	1059	0.000
Deviance	1253.244	1059	0.000
Hosmer-Lemeshow	8.376	8	0.398

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	38	35	43	46	52	63	74	79	93	101	624
Exp	38.6	42.9	45.0	48.5	53.2	58.2	67.1	76.7	90.3	103.5	
0											
Obs	68	72	63	61	55	43	33	27	14	6	442
Exp	67.4	64.1	61.0	58.5	53.8	47.8	39.9	29.3	16.7	3.5	
Total	106	107	106	107	107	106	107	106	107	107	1066

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	203349	73.7%	Somers D 0.48
Discordant	71258	25.8%	Goodman-Kruskal Gamma 0.48
Ties	1201	0.4%	Kendalls Tau-a 0.23
Total	275808	100.0%	

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
Constant	-0.7861	0.1268	-6.20	0.000			
Renda	0.00045468	0.00004456	10.20	0.000	1.00	1.00	1.00
DMSEXO	-0.1167	0.1484	-0.79	0.432	0.89	0.67	1.19

Log-Likelihood = -627.235

Test that all slopes are zero: G = 192.094, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1439.205	966	0.000
Deviance	1106.145	966	0.001
Hosmer-Lemeshow	12.173	8	0.144

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	33	44	38	49	51	62	79	76	91	101	624
Exp	39.5	43.0	45.1	48.3	52.7	58.0	67.2	76.5	90.2	103.6	
0											
Obs	73	63	68	58	56	44	28	30	16	6	442
Exp	66.5	64.0	60.9	58.7	54.3	48.0	39.8	29.5	16.8	3.4	
Total	106	107	106	107	107	106	107	106	107	107	1066

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	203649	73.8%	Somers D	0.48
Discordant	70845	25.7%	Goodman-Kruskal Gamma	0.48
Ties	1314	0.5%	Kendalls Tau-a	0.23
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-0.6317	0.1562	-4.04	0.000	1.00	1.00	1.00
Renda	0.00043238	0.00004583	9.43	0.000	0.83	0.56	1.22
DM1PRO	-0.1893	0.1974	-0.96	0.338	0.76	0.56	1.02
DM2PRO	-0.2800	0.1544	-1.81	0.070			

Log-Likelihood = -625.852

Test that all slopes are zero: G = 194.860, DF = 3, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1324.121	995	0.000
Deviance	1165.110	995	0.000
Hosmer-Lemeshow	6.274	8	0.616

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	39	35	42	45	58	60	74	76	93	102	624
Exp	38.0	41.2	45.0	49.4	53.4	58.8	67.7	77.0	90.1	103.4	
0											
Obs	68	71	64	62	49	46	33	30	14	5	442
Exp	69.0	64.8	61.0	57.6	53.6	47.2	39.3	29.0	16.9	3.6	
Total	107	106	106	107	107	106	107	106	107	107	1066

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	203068	73.6%	Somers D	0.48
Discordant	71592	26.0%	Goodman-Kruskal Gamma	0.48
Ties	1148	0.4%	Kendalls Tau-a	0.23
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compr	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-1.8560	0.4862	-3.82	0.000	1.00	1.00	1.00
Renda	0.00047039	0.00004570	10.29	0.000	2.97	1.07	8.26
DM1LOC	1.0899	0.5211	2.09	0.036	3.63	1.42	9.29
DM2LOC	1.2894	0.4793	2.69	0.007	1.00	0.34	2.93
DM3LOC	-0.0028	0.5495	-0.01	0.996	1.35	0.50	3.61
DM4LOC	0.2978	0.5030	0.59	0.554			

Log-Likelihood = -604.231
 Test that all slopes are zero: G = 238.102, DF = 5, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1716.384	973	0.000
Deviance	1096.512	973	0.003
Hosmer-Lemeshow	15.364	8	0.052

Table of Observed and Expected Frequencies:
 (See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	21	40	50	44	50	68	77	81	93	100	624
Exp	24.9	38.7	49.4	51.9	54.9	60.7	68.8	79.5	91.0	104.1	
0											
Obs	85	67	59	63	56	38	29	25	13	7	442
Exp	81.1	68.3	59.6	55.1	51.1	45.3	37.2	26.5	15.0	2.9	
Total	106	107	109	107	106	106	106	106	106	107	1066

Measures of Association:
 (Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	210861	76.5%	Somers D	0.53
Discordant	63932	23.2%	Goodman-Kruskal Gamma	0.53
Ties	1015	0.4%	Kendalls Tau-a	0.26
Total	275808	100.0%		

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-0.3394	0.1340	-2.53	0.011	1.00	1.00	1.00
Renda	0.00046456	0.00004602	10.10	0.000	0.31	0.19	0.52
DM1IDA	-1.1665	0.2593	-4.50	0.000	0.50	0.28	0.90
DM2IDA	-0.6844	0.2970	-2.30	0.021	0.72	0.40	1.28
DM3IDA	-0.3312	0.2930	-1.13	0.258	0.26	0.19	0.37
DM4IDA	-1.3311	0.1708	-7.79	0.000			

Log-Likelihood = -590.563
 Test that all slopes are zero: G = 265.438, DF = 5, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1827.401	987	0.000

Deviance	1085.673	987	0.015
Hosmer-Lemeshow	17.685	8	0.024

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	22	27	46	55	50	66	82	82	94	100	624
Exp	23.9	31.6	43.9	56.0	58.0	63.0	70.2	80.7	92.3	104.4	
0											
Obs	84	80	60	55	56	40	24	24	12	7	442
Exp	82.1	75.4	62.1	54.0	48.0	43.0	35.8	25.3	13.7	2.6	
Total	106	107	106	110	106	106	106	106	106	107	1066

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	215855	78.3%	Somers D	0.57
Discordant	59116	21.4%	Goodman-Kruskal Gamma	0.57
Ties	837	0.3%	Kendalls Tau-a	0.28
Total	275808	100.0%		

Worksheet size: 100000 cells

Retrieving worksheet from file: C:\GUILHERME\POLI\TF\CONJUNTO DE DADOS\TREINAMENTO.M
TW
Worksheet was saved on 1/ 9/1996

Binary Logistic Regression

Step	Log-Likelihood
0	-723.282
1	-651.136
2	-606.035
3	-592.347
4	-591.203
5	-591.193
6	-591.193
7	-591.193

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-0.3739	0.1303	-2.87	0.004	1.00	1.00	1.00
Renda	0.00046450	0.00004593	10.11	0.000	0.32	0.19	0.53
DM1IDA	-1.1319	0.2573	-4.40	0.000	0.52	0.29	0.93
DM2IDA	-0.6498	0.2953	-2.20	0.028	0.27	0.20	0.38
DM4IDA	-1.2964	0.1678	-7.72	0.000			

Log-Likelihood = -591.193

Test that all slopes are zero: G = 264.177, DF = 4, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1852.765	978	0.000
Deviance	1073.071	978	0.018
Hosmer-Lemeshow	22.348	8	0.004
Brown:			
General Alternative	30.320	2	0.000
Symmetric Alternative	27.854	1	0.000

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	22	27	48	54	48	64	85	82	94	100	624
Exp	23.9	31.6	44.3	54.4	58.3	64.2	70.3	80.6	92.1	104.3	
0											
Obs	84	80	58	53	59	44	21	24	12	7	442
Exp	82.1	75.4	61.7	52.6	48.7	43.8	35.7	25.4	13.9	2.7	

Total 106 107 106 107 107 108 106 106 106 107 1066

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	215751	78.2%	Somers D	0.57
Discordant	59201	21.5%	Goodman-Kruskal Gamma	0.57
Ties	856	0.3%	Kendalls Tau-a	0.28
Total	275808	100.0%		

10.5 REGRESSÃO LOGÍSTICA: PASSO 3

Worksheet size: 100000 cells

Retrieving worksheet from file: C:\GUILHERME\POLI\TF\CONJUNTO DE DADOS\TREINAMENTO.M
TW

Worksheet was saved on 1/ 9/1996

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-0.3332	0.1407	-2.37	0.018			
Renda	0.00046212	0.00004606	10.03	0.000	1.00	1.00	1.00
DM1IDA	-1.1341	0.2573	-4.41	0.000	0.32	0.19	0.53
DM2IDA	-0.6638	0.2959	-2.24	0.025	0.51	0.29	0.92
DM4IDA	-1.2954	0.1679	-7.71	0.000	0.27	0.20	0.38
DMSEXO	-0.1188	0.1549	-0.77	0.443	0.89	0.66	1.20

Log-Likelihood = -590.900

Test that all slopes are zero: G = 264.765, DF = 5, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1878.591	1008	0.000
Deviance	1105.526	1008	0.017
Hosmer-Lemeshow	20.320	8	0.009

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	23	27	47	47	50	69	81	84	96	100	624
Exp	23.8	31.7	44.4	54.3	58.3	62.9	70.8	80.5	93.0	104.3	
0											
Obs	83	80	59	60	57	37	26	22	11	7	442
Exp	82.2	75.3	61.6	52.7	48.7	43.1	36.2	25.5	14.0	2.7	
Total	106	107	106	107	107	106	107	106	107	107	1066

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	216151	78.4%	Somers D
Discordant	58804	21.3%	Goodman-Kruskal Gamma
Ties	853	0.3%	Kendalls Tau-a
Total	275808	100.0%	

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-0.1566	0.1711	-0.92	0.360	1.00	1.00	1.00
Renda	0.00043582	0.00004731	9.21	0.000	0.33	0.20	0.54
DM1IDA	-1.1221	0.2580	-4.35	0.000	0.52	0.29	0.93
DM2IDA	-0.6539	0.2967	-2.20	0.028	0.27	0.19	0.37
DM4IDA	-1.3118	0.1686	-7.78	0.000	0.86	0.58	1.28
DM1PRO	-0.1510	0.2047	-0.74	0.461	0.72	0.53	0.99
DM2PRO	-0.3240	0.1616	-2.01	0.045			

Log-Likelihood = -589.186

Test that all slopes are zero: G = 268.193, DF = 6, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1608.168	1019	0.000
Deviance	1124.645	1019	0.011
Hosmer-Lemeshow	12.423	8	0.133

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	1	2	3	4	5	6	7	8	9	10	Total
1											
Obs	21	34	39	51	58	61	81	84	95	100	624
Exp	22.9	31.7	44.4	52.9	58.5	63.8	71.8	81.0	92.9	104.1	
0											
Obs	85	73	67	56	49	45	26	22	12	7	442
Exp	83.1	75.3	61.6	54.1	48.5	42.2	35.2	25.0	14.1	2.9	
Total	106	107	106	107	107	106	107	106	107	107	1066

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	216036	78.3%	Somers D
Discordant	58989	21.4%	Goodman-Kruskal Gamma
Ties	783	0.3%	Kendalls Tau-a
Total	275808	100.0%	

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	0.0331	0.9940	0.03	0.973	1.00	1.00	1.00
Renda	0.00046822	0.00004638	10.10	0.000	0.32	0.19	0.53
DM1IDA	-1.1370	0.2590	-4.39	0.000	0.53	0.29	0.97
DM2IDA	-0.6363	0.3067	-2.07	0.038	0.15	0.03	0.84
DM4IDA	-1.8834	0.8689	-2.17	0.030	0.64	0.10	4.13
DM1LOC	-0.4525	0.9548	-0.47	0.636	0.66	0.10	4.63
DM2LOC	-0.4103	0.9909	-0.41	0.679	1.00	0.34	2.93
DM3LOC	-0.0022	0.5492	-0.00	0.997	1.35	0.50	3.60
DM4LOC	0.2969	0.5027	0.59	0.555			

Log-Likelihood = -590.364
Test that all slopes are zero: G = 265.837, DF = 8, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1920.721	992	0.000
Deviance	1103.636	992	0.007
Hosmer-Lemeshow	23.206	8	0.003

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

	Group										Total
Value	1	2	3	4	5	6	7	8	9	10	
1											
Obs	23	28	48	51	46	67	84	82	95	100	624
Exp	23.9	31.8	43.8	54.4	58.2	63.0	70.8	80.7	93.1	104.4	
0											
Obs	83	79	58	56	61	39	23	24	12	7	442
Exp	82.1	75.2	62.2	52.6	48.8	43.0	36.2	25.3	13.9	2.6	
Total	106	107	106	107	107	106	107	106	107	107	1066

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	216213	78.4%	Somers D	0.57
Discordant	58800	21.3%	Goodman-Kruskal Gamma	0.57
Ties	795	0.3%	Kendalls Tau-a	0.28
Total	275808	100.0%		

10.6 REGRESSÃO LOGÍSTICA: MODELO FINAL

Worksheet size: 100000 cells

Retrieving worksheet from file: C:\GUILHERME\POLI\TF\CONJUNTO DE DADOS\TREINAMENTO.M

TW

Worksheet was saved on 1/ 9/1996

Binary Logistic Regression

Link Function: Logit

Response Information

Variable	Value	Count
Compra	1	624
	0	442
Total		1066

Logistic Regression Table

Predictor	Coef	StDev	Z	P	Odds Ratio	95% CI	
Constant	-0.3739	0.1303	-2.87	0.004	1.00	1.00	1.00
Renda	0.00046450	0.00004593	10.11	0.000	0.32	0.19	0.53
DM1IDA	-1.1319	0.2573	-4.40	0.000	0.52	0.29	0.93
DM2IDA	-0.6498	0.2953	-2.20	0.028	0.27	0.20	0.38
DM4IDA	-1.2964	0.1678	-7.72	0.000			

Log-Likelihood = -591.193

Test that all slopes are zero: G = 264.177, DF = 4, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	1852.765	978	0.000
Deviance	1073.071	978	0.018
Hosmer-Lemeshow	22.348	8	0.004

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	22	27	48	54	48	64	85	82	94	100	624
Exp	23.9	31.6	44.3	54.4	58.3	64.2	70.3	80.6	92.1	104.3	
0											
Obs	84	80	58	53	59	44	21	24	12	7	442
Exp	82.1	75.4	61.7	52.6	48.7	43.8	35.7	25.4	13.9	2.7	
Total	106	107	106	107	107	108	106	106	106	107	1066

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	215751	78.2%	Somers D	0.57
Discordant	59201	21.5%	Goodman-Kruskal Gamma	0.57
Ties	856	0.3%	Kendalls Tau-a	0.28
Total	275808	100.0%		

10.7 CONJUNTO DE VALIDAÇÃO

PI (X)	Compra Esperada	Compra Real	Acertos
0,208268287	0	1	0
0,219353729	0	1	0
0,220800333	0	1	0
0,224417535	0	1	0
0,226352119	0	1	0
0,231612224	0	1	0
0,234923302	0	1	0
0,235269467	0	1	0
0,237689546	0	1	0
0,24159505	0	1	0
0,244402384	0	1	0
0,245949706	0	1	0
0,246122038	0	1	0
0,250469285	0	1	0
0,2510676	0	1	0
0,251692102	0	1	0
0,252918874	0	1	0
0,255207636	0	1	0
0,255560959	0	1	0
0,255724989	0	1	0
0,256800118	0	1	0
0,258844396	0	1	0
0,259736509	0	1	0
0,260899236	0	1	0
0,260899236	0	1	0
0,261885708	0	1	0
0,272328629	0	1	0
0,277340934	0	1	0
0,279955168	0	1	0
0,28455249	0	1	0
0,288649937	0	1	0
0,290944351	0	1	0
0,291040184	0	1	0
0,297794488	0	1	0
0,301009779	0	1	0
0,305425933	0	1	0
0,306609687	0	1	0
0,306792979	0	1	0
0,306792979	0	1	0
0,309268147	0	1	0
0,31435491	0	1	0
0,315958978	0	1	0
0,318989508	0	1	0
0,319292302	0	1	0
0,319685172	0	1	0
0,321720127	0	1	0
0,324361179	0	1	0
0,326298286	0	1	0
0,328021763	0	1	0
0,332969917	0	1	0
0,333474611	0	1	0
0,333485946	0	1	0

0,370423366	0	0	1
0,370748403	0	0	1
0,371073556	0	0	1
0,371181967	0	0	1
0,371724214	0	0	1
0,371941204	0	0	1
0,372701073	0	0	1
0,374331479	0	0	1
0,374549083	0	0	1
0,375420006	0	0	1
0,375528928	0	0	1
0,376400758	0	0	1
0,376946059	0	0	1
0,378037595	0	0	1
0,379786623	0	0	1
0,380552808	0	0	1
0,381429178	0	0	1
0,381538778	0	0	1
0,381648391	0	0	1
0,384169004	0	0	1
0,385712449	0	0	1
0,386152774	0	0	1
0,386935889	0	0	1
0,38824305	0	0	1
0,388688238	0	0	1
0,390025626	0	0	1
0,390897901	0	0	1
0,390910044	0	0	1
0,393112048	0	0	1
0,393678449	0	0	1
0,394220777	0	0	1
0,395108546	0	0	1
0,398666491	0	0	1
0,400226487	0	0	1
0,400895679	0	0	1
0,403029485	0	0	1
0,403128976	0	0	1
0,403687932	0	0	1
0,406038264	0	0	1
0,408280685	0	0	1
0,408505139	0	0	1
0,409852661	0	0	1
0,413002116	0	0	1
0,413452626	0	0	1
0,4177396	0	0	1
0,418304616	0	0	1
0,4198878	0	0	1
0,420223371	0	0	1
0,421246129	0	0	1
0,425101143	0	0	1
0,428750231	0	0	1
0,429648109	0	0	1
0,430331199	0	0	1

0,322125707	0	0	1
0,322216005	0	0	1
0,32241543	0	0	1
0,325572722	0	0	1
0,326082892	0	0	1
0,328241342	0	0	1
0,328742401	0	0	1
0,32884491	0	0	1
0,329867366	0	0	1
0,332130241	0	0	1
0,33254251	0	0	1
0,334301074	0	0	1
0,334714688	0	0	1
0,334818131	0	0	1
0,335025065	0	0	1
0,337420751	0	0	1
0,337513203	0	0	1
0,338356011	0	0	1
0,340114614	0	0	1
0,343664432	0	0	1
0,344296952	0	0	1
0,345136348	0	0	1
0,345871612	0	0	1
0,345871612	0	0	1
0,346292095	0	0	1
0,346712818	0	0	1
0,347765672	0	0	1
0,348187231	0	0	1
0,348289027	0	0	1
0,34955894	0	0	1
0,351991907	0	0	1
0,352627866	0	0	1
0,352946042	0	0	1
0,353052129	0	0	1
0,354751493	0	0	1
0,356028428	0	0	1
0,356785959	0	0	1
0,358172928	0	0	1
0,359122793	0	0	1
0,359443575	0	0	1
0,36062085	0	0	1
0,360727958	0	0	1
0,36083508	0	0	1
0,360942216	0	0	1
0,362332544	0	0	1
0,364593017	0	0	1
0,364915903	0	0	1
0,365346607	0	0	1
0,365666065	0	0	1
0,366328317	0	0	1
0,366963661	0	0	1
0,368583716	0	0	1
0,368799948	0	0	1

0,265027401	0	0	1
0,265130943	0	0	1
0,265402537	0	0	1
0,26557063	0	0	1
0,265764938	0	0	1
0,267841722	0	0	1
0,268128227	0	0	1
0,2687536	0	0	1
0,269955177	0	0	1
0,27004673	0	0	1
0,270229895	0	0	1
0,271055105	0	0	1
0,27168477	0	0	1
0,272157844	0	0	1
0,272433967	0	0	1
0,273171152	0	0	1
0,273355642	0	0	1
0,274834362	0	0	1
0,274926946	0	0	1
0,275853858	0	0	1
0,276132308	0	0	1
0,27799308	0	0	1
0,278166128	0	0	1
0,278552813	0	0	1
0,279019786	0	0	1
0,279019786	0	0	1
0,279767938	0	0	1
0,280142474	0	0	1
0,280329857	0	0	1
0,283244053	0	0	1
0,28362141	0	0	1
0,289222531	0	0	1
0,291697698	0	0	1
0,291711541	0	0	1
0,292864553	0	0	1
0,294695605	0	0	1
0,295178567	0	0	1
0,2962427	0	0	1
0,297018009	0	0	1
0,297115005	0	0	1
0,297503172	0	0	1
0,300326097	0	0	1
0,301596497	0	0	1
0,306609687	0	0	1
0,306807228	0	0	1
0,307682768	0	0	1
0,309878154	0	0	1
0,313051417	0	0	1
0,313165727	0	0	1
0,313265646	0	0	1
0,315970001	0	0	1
0,316260231	0	0	1
0,317765338	0	0	1

0,226363852	0	0	1
0,226526583	0	0	1
0,226689397	0	0	1
0,2277497	0	0	1
0,22807665	0	0	1
0,228473992	0	0	1
0,228567695	0	0	1
0,229469874	0	0	1
0,231446932	0	0	1
0,231517647	0	0	1
0,231931117	0	0	1
0,234935345	0	0	1
0,235938699	0	0	1
0,236022446	0	0	1
0,236513257	0	0	1
0,236777088	0	0	1
0,237100899	0	0	1
0,238026367	0	0	1
0,2381949	0	0	1
0,241680169	0	0	1
0,243032526	0	0	1
0,243288977	0	0	1
0,243301312	0	0	1
0,243386839	0	0	1
0,243386839	0	0	1
0,243386839	0	0	1
0,243802429	0	0	1
0,245175221	0	0	1
0,245777456	0	0	1
0,246035862	0	0	1
0,247330634	0	0	1
0,247330634	0	0	1
0,249424309	0	0	1
0,250382092	0	0	1
0,250456707	0	0	1
0,251329714	0	0	1
0,252042205	0	0	1
0,252042205	0	0	1
0,252042205	0	0	1
0,252042205	0	0	1
0,254489242	0	0	1
0,255106624	0	0	1
0,256445671	0	0	1
0,256610073	0	0	1
0,257053367	0	0	1
0,257154884	0	0	1
0,257154884	0	0	1
0,258577152	0	0	1
0,260899236	0	0	1
0,261078416	0	0	1
0,263041649	0	0	1
0,263505117	0	0	1
0,264317229	0	0	1

0,475068319	0	1	0
0,477260821	0	1	0
0,477501453	0	1	0
0,477861858	0	1	0
0,479115283	0	1	0
0,479368721	0	1	0
0,479578991	0	1	0
0,479703784	0	1	0
0,481571726	0	1	0
0,482834748	0	1	0
0,48364671	0	1	0
0,485266891	0	1	0
0,485842161	0	1	0
0,485851029	0	1	0
0,485967061	0	1	0
0,487836525	0	1	0
0,493860184	0	1	0
0,495709105	0	1	0
0,495730729	0	1	0
0,497240278	0	1	0
0,497937012	0	1	0
0,499425	0	1	0
0,499433875	0	1	0
0,5000145	0	1	0
0,50012175	0	1	0
0,50024675	0	1	0
0,500479	0	1	0
0,50406991	0	1	0
0,504423135	0	1	0
0,505356045	0	1	0
0,505579393	0	1	0
0,506168812	0	1	0
0,506740467	0	1	0
0,50732985	0	1	0
0,195547202	0	0	1
0,196059197	0	0	1
0,196425528	0	0	1
0,198634372	0	0	1
0,199152449	0	0	1
0,201982575	0	0	1
0,205069703	0	0	1
0,205676135	0	0	1
0,211426033	0	0	1
0,216424952	0	0	1
0,216661362	0	0	1
0,218004548	0	0	1
0,220309701	0	0	1
0,220960207	0	0	1
0,222070069	0	0	1
0,223921175	0	0	1
0,22498398	0	0	1
0,225227052	0	0	1
0,226352119	0	0	1

0,385052313	0	1	0
0,385064389	0	1	0
0,390897901	0	1	0
0,392791807	0	1	0
0,398443801	0	1	0
0,398666491	0	1	0
0,400127233	0	1	0
0,400561037	0	1	0
0,400684821	0	1	0
0,401242666	0	1	0
0,401465876	0	1	0
0,404694683	0	1	0
0,411426481	0	1	0
0,411551318	0	1	0
0,412214074	0	1	0
0,412564111	0	1	0
0,412664329	0	1	0
0,413127095	0	1	0
0,414241367	0	1	0
0,414466802	0	1	0
0,41570345	0	1	0
0,415707336	0	1	0
0,41898292	0	1	0
0,42022727	0	1	0
0,423751938	0	1	0
0,424533646	0	1	0
0,429534287	0	1	0
0,429888275	0	1	0
0,431124548	0	1	0
0,435234287	0	1	0
0,435348467	0	1	0
0,437747814	0	1	0
0,443128184	0	1	0
0,445651326	0	1	0
0,447028775	0	1	0
0,448407038	0	1	0
0,448879253	0	1	0
0,450705895	0	1	0
0,45312196	0	1	0
0,453940453	0	1	0
0,454964261	0	1	0
0,457512087	0	1	0
0,461998924	0	1	0
0,463380686	0	1	0
0,466157821	0	1	0
0,467194321	0	1	0
0,470659747	0	1	0
0,470900048	0	1	0
0,471028488	0	1	0
0,471594494	0	1	0
0,471722955	0	1	0
0,472533304	0	1	0
0,474489171	0	1	0

0,336682815	0	1	0
0,337305512	0	1	0
0,337617072	0	1	0
0,337928773	0	1	0
0,339798342	0	1	0
0,340844749	0	1	0
0,342516463	0	1	0
0,343249072	0	1	0
0,344413338	0	1	0
0,346923269	0	1	0
0,350933142	0	1	0
0,351368096	0	1	0
0,351885965	0	1	0
0,353052129	0	1	0
0,353052129	0	1	0
0,353052129	0	1	0
0,35359444	0	1	0
0,355602554	0	1	0
0,355815463	0	1	0
0,356561091	0	1	0
0,358267991	0	1	0
0,35902763	0	1	0
0,361811578	0	1	0
0,361918839	0	1	0
0,362443569	0	1	0
0,363302694	0	1	0
0,363732584	0	1	0
0,364593017	0	1	0
0,36847562	0	1	0
0,370098447	0	1	0
0,37020674	0	1	0
0,370315047	0	1	0
0,370419635	0	1	0
0,373147496	0	1	0
0,373678972	0	1	0
0,375964743	0	1	0
0,376291735	0	1	0
0,376618841	0	1	0
0,377067136	0	1	0
0,377382524	0	1	0
0,377819189	0	1	0
0,377819189	0	1	0
0,378037595	0	1	0
0,378802407	0	1	0
0,379677216	0	1	0
0,379896041	0	1	0
0,381100448	0	1	0
0,382745179	0	1	0
0,383074449	0	1	0
0,383403826	0	1	0
0,383623471	0	1	0
0,383965089	0	1	0
0,384059117	0	1	0

0,565127995	1	1	1
0,565242147	1	1	1
0,565356291	1	1	1
0,568212402	1	1	1
0,569697238	1	1	1
0,570376396	1	1	1
0,570485438	1	1	1
0,57197312	1	1	1
0,572532931	1	1	1
0,573690514	1	1	1
0,579008618	1	1	1
0,579452803	1	1	1
0,580928345	1	1	1
0,582054003	1	1	1
0,582392957	1	1	1
0,582623625	1	1	1
0,582639917	1	1	1
0,582966337	1	1	1
0,585101761	1	1	1
0,586454247	1	1	1
0,58791797	1	1	1
0,588264131	1	1	1
0,588388984	1	1	1
0,589042855	1	1	1
0,589380139	1	1	1
0,590512397	1	1	1
0,590728434	1	1	1
0,592420456	1	1	1
0,592869009	1	1	1
0,597895864	1	1	1
0,59890904	1	1	1
0,599012089	1	1	1
0,599235211	1	1	1
0,599792836	1	1	1
0,600350203	1	1	1
0,602910724	1	1	1
0,602919223	1	1	1
0,604363984	1	1	1
0,605687516	1	1	1
0,607797982	1	1	1
0,608244615	1	1	1
0,608900055	1	1	1
0,611891016	1	1	1
0,613535947	1	1	1
0,614311244	1	1	1
0,614526712	1	1	1
0,618488593	1	1	1
0,619255523	1	1	1
0,620122922	1	1	1
0,620143299	1	1	1
0,62089317	1	1	1
0,621762881	1	1	1
0,621762881	1	1	1

0,621892453	1	1	1
0,622317261	1	1	1
0,624163114	1	1	1
0,624933875	1	1	1
0,626348177	1	1	1
0,626348177	1	1	1
0,626665943	1	1	1
0,631867316	1	1	1
0,631887436	1	1	1
0,635318019	1	1	1
0,636509227	1	1	1
0,637153796	1	1	1
0,637797873	1	1	1
0,638004256	1	1	1
0,640254059	1	1	1
0,644522322	1	1	1
0,647186519	1	1	1
0,650459363	1	1	1
0,651628233	1	1	1
0,651733671	1	1	1
0,653515802	1	1	1
0,654154632	1	1	1
0,656664358	1	1	1
0,656986463	1	1	1
0,658442464	1	1	1
0,65867081	1	1	1
0,659807165	1	1	1
0,663335163	1	1	1
0,664889381	1	1	1
0,668612769	1	1	1
0,671787653	1	1	1
0,672511991	1	1	1
0,672606467	1	1	1
0,675878964	1	1	1
0,676379768	1	1	1
0,678316314	1	1	1
0,68154335	1	1	1
0,68214794	1	1	1
0,68325476	1	1	1
0,688658045	1	1	1
0,689063882	1	1	1
0,690348592	1	1	1
0,692338533	1	1	1
0,692635277	1	1	1
0,694601963	1	1	1
0,696470895	1	1	1
0,696765399	1	1	1
0,698724626	1	1	1
0,699017888	1	1	1
0,699603925	1	1	1
0,699708994	1	1	1
0,699817298	1	1	1
0,700889569	1	1	1

0,701170939

0,701570779

0,701941544

0,704075137

0,706103466

0,708322963

0,709473221

0,71032685

0,71032685

0,711281683

0,711765681

0,712900599

0,714703545

0,71527148

0,718382927

0,718382927

0,719235075

0,719321698

0,719790367

0,721006667

0,72241318

0,726762727

0,727964062

0,728970877

0,729986188

0,730169261

0,731897577

0,733551925

0,733720372

0,736527291

0,737697417

0,739841998

0,739931393

0,740751663

0,741632912

0,743756494

0,744729063

0,749031856

0,750171938

0,751212136

0,753286149

0,753717523

0,754757317

0,755695293

0,755781039

0,761148326

0,761907517

0,762665048

0,763346227

0,764091405

0,765342151

0,767417771

0,767749237

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

0,432610066
0,434321095
0,435805256
0,436490643
0,437061984
0,440035619
0,440150077
0,440379012
0,446450783
0,448177272
0,450016012
0,452891761
0,459225251
0,460729176
0,460960002
0,461203518
0,465571069
0,467313931
0,467326627
0,468007766
0,468924231
0,469039909
0,469271276
0,470196871
0,470428303
0,470428303
0,471710246
0,472743298
0,47367848
0,474257531
0,474827799
0,474836652
0,476690284
0,478303883
0,480042734
0,480047601
0,483173856
0,48364671
0,483869849
0,489323748
0,49118104
0,492695145
0,492922473
0,493270781
0,494664078
0,495012415
0,496298568
0,497567019
0,499076626
0,49955
0,49965725
0,500005625
0,50012175

[illegible][illegible][illegible]

0,770306911
0,772028263
0,772600023
0,773659176
0,776332095
0,779142369
0,782170824
0,782323019
0,787582174
0,789595596
0,790052351
0,790129387
0,799669898
0,8007153
0,800857838
0,803511328
0,805562184
0,807657897
0,807946366
0,812304683
0,813435179
0,814420053
0,818045901
0,819838988
0,820792282
0,821343435
0,822018824
0,823171176
0,824183085
0,824720905
0,825056384
0,827393363
0,828124126
0,828452172
0,828713301
0,828979643
0,829765679
0,830224476
0,830616947
0,832314266
0,836930221
0,841438581
0,850455935
0,850633075
0,852278046
0,853036678
0,857407391
0,858148387
0,861392134
0,862056306
0,865341292
0,865985365
0,867276117

[illegible][illegible][illegible]

0,947566802
0,948366948
0,950223418
0,953457856
0,95490119
0,956221134
0,956627667
0,956706162
0,956781587
0,957316209
0,957316209
0,957806981
0,959536674
0,959949442
0,960657705
0,962090508
0,962224516
0,962877502
0,963010795
0,968795765
0,970329675
0,972384645
0,972435861
0,972927774
0,973122331
0,974173203
0,976106937
0,976150232
0,976225813
0,977486492
0,9786803
0,97946071
0,979757614
0,980077541
0,980464191
0,980807808
0,981059739
0,981316944
0,981762992
0,982352577
0,984292557
0,985111049
0,987258714
0,98738662
0,988455477
0,988836075
0,990283059
0,991157839
0,993450087
0,993640717
0,993719476
0,994681448
0,995781028

[illegible][illegible][illegible]

0,995977333
0,996612077
0,996783158
0,996856837
0,997385929
0,997675138
0,997838328
0,998022333
0,998297395
0,998309256
0,998401458
0,998561866
0,998882649
0,998912297
0,999050568
0,999148666
0,999329384
0,999373037
0,99943046
0,999587723
0,999736986
0,999822272
0,999917022
0,999945803
0,999979969
0,999988045
0,999994066
0,999998014
0,999998823
1,000000
1,000000
0,510107123
0,511035707
0,511383907
0,511616034
0,511964216
0,512080274
0,512892642
0,513008689
0,513481731
0,514061916
0,514865243
0,515106151
0,515329319
0,518113195
0,51823803
0,519397635
0,521359617
0,521475529
0,522634519
0,523802121
0,524024985
0,524256693

[illegible][illegible][illegible]

0,524604237
0,524844772
0,525067591
0,526697834
0,527152138
0,527504332
0,52785164
0,528194061
0,529582908
0,530170303
0,530508553
0,532243567
0,532359208
0,532714945
0,534208972
0,5353646
0,537097328
0,537221637
0,538367427
0,539983196
0,540675413
0,540914932
0,54125214
0,541828758
0,542174674
0,544018868
0,544018868
0,546782882
0,547367157
0,547473442
0,547703588
0,547933714
0,547933714
0,548623968
0,549084033
0,549893772
0,551153278
0,551497982
0,55241695
0,552761472
0,552876301
0,552885076
0,553105942
0,556671187
0,556785817
0,556900441
0,557027643
0,559204207
0,560098667
0,561937736
0,564223266
0,564908397
0,565013837

[illegible][illegible][illegible]

0,566962295
0,56730439
0,568570696
0,569355599
0,570371617
0,570599252
0,571290671
0,572419246
0,574464396
0,574485541
0,577980595
0,578215837
0,579792343
0,582054003
0,584312217
0,588255532
0,588817952
0,590616128
0,591065296
0,591186144
0,591626542
0,596443256
0,597225641
0,598900513
0,599681332
0,604244423
0,605132697
0,609452997
0,609674095
0,611682487
0,615186696
0,616285726
0,617273839
0,618260984
0,621544379
0,622753863
0,624925554
0,628077601
0,628077601
0,629708271
0,63294713
0,633702208
0,634672058
0,635318019
0,637901475
0,640468006
0,648344229
0,648450125
0,649089635
0,649931131
0,659173337
0,659694925
0,661473378

[illegible][illegible][illegible]

0,662089237
0,669016456
0,67373062
0,675667639
0,678815142
0,681950707
0,683560479
0,683857608
0,685662441
0,690547147
0,693121946
0,694700488
0,700091791
0,712995661
0,715082244
0,715845956
0,717724653
0,718202129
0,721473612
0,722406062
0,727039359
0,734988961
0,736610532
0,736790733
0,738767704
0,742788281
0,74296573
0,746497863
0,752508397
0,757491674
0,770306911
0,779142369
0,781848037
0,783900898
0,784843641
0,788197257
0,796300377
0,803217819
0,816104388
0,819833744
0,820245035
0,821201861
0,822499215
0,836801208
0,837431906
0,84530401
0,855695199
0,871493285
0,876857201
0,87895039
0,880671186
0,897410058
0,924723203

[illegible][illegible][illegible]

0,94540071	1	0	0
0,970463113	1	0	0
0,98197797	1	0	0
0,983346358	1	0	0
0,994276108	1	0	0
0,998100549	1	0	0