

6

Metodologia de Pesquisa

De acordo com Frankfort-Nachmias e Nachmias (1996), metodologia é um sistema de normas e procedimentos explícitos a partir dos quais uma pesquisa é desenvolvida, e contra a qual quaisquer possíveis questionamentos devem ser apresentados. Neste capítulo tem-se por objetivos: apresentar os procedimentos metodológicos utilizados para o desenvolvimento da pesquisa; identificar as questões empregadas como referencial para a formulação do instrumento de pesquisa; e descrever os métodos estatísticos utilizados e os motivos que levaram à escolha de tais métodos. Na Seção 6.1 são descritas as características gerais da pesquisa e o desenvolvimento do instrumento de pesquisa de campo. Na Seção 6.2 são apresentadas as técnicas estatísticas.

Esta tese seguiu a taxonomia na qual o processo de pesquisa pode ser formalmente estabelecido em quatro fases. Na primeira fase, é realizada a revisão bibliográfica. Na segunda, constrói-se o modelo teórico. Na terceira testa-se o modelo formulado e, na quarta, reflete-se sobre os resultados obtidos.

Assim, na primeira fase foram feitos levantamentos bibliográficos com a intenção de conhecer o estado da arte relacionado com o gerenciamento da cadeia de suprimentos e com a avaliação de desempenho. Trata-se de uma fase que assume, entre outras metas, o objetivo de evitar a “reinvenção da roda”. Como trabalho complementar, foram identificados grupos de pesquisa, na mesma área de interesse, a fim de estabelecer contatos com outros pesquisadores.

A segunda fase caracterizou-se pela formulação do modelo, sendo que nesse momento o pesquisador buscou empregar um elevado grau de elaboração, através da análise e síntese dos conceitos estudados. Para tanto, serviu-se a pesquisa dos métodos de **indução** (buscando chegar a conclusões gerais a partir de casos específicos) e de **dedução** (implementando aquele processo lógico que parte de formulações gerais para lidar com casos específicos).

Na fase três confrontou-se o modelo teórico com os resultados da pesquisa de campo, fazendo uso das hipóteses complementares e questões exploratórias, atendendo às expectativas da pesquisa. Vale lembrar que o tipo de teste realizado é função das ambições do pesquisador e da formulação teórica realizada.

A fase quatro representou o momento de reflexão e integração a respeito da pesquisa; nela foram respondidas questões tais como: o que foi aprendido sobre o processo de pesquisa? O que poderia ter sido feito de forma diferente? O que o pesquisador aprendeu com a pesquisa? Tais questões são respondidas de forma implícita ou explícita, sendo que, neste momento, o pesquisador inclui suas contribuições, constituídas por sugestões e inferências para a continuidade e aprofundamento da pesquisa atual.

6.1

Caracterização geral e desenvolvimento do instrumento de pesquisa de campo

6.1.1

Tipos de pesquisa

Nesta subseção serão descritas de forma concisa as estratégias de pesquisa mais comumente utilizadas, tais como: experimento; pesquisa histórica; análise de informações de arquivos (documental), estudo de caso e pesquisa amostral/levantamento (*survey*). Cada uma dessas estratégias pode ser usada para os propósitos exploratórios, descritivos ou explanatórios (causais). A estratégia empregada depende, principalmente, do tipo de questão apresentada pela pesquisa, do grau de controle que o investigador tem sobre os eventos e do foco temporal (eventos contemporâneos *versus* fenômenos históricos).

6.1.1.1

Experimentos

Os experimentos são caracterizados pelo controle total sobre as variáveis, podendo ser desenvolvido em laboratório ou em campo. O controle mencionado deve permitir ao pesquisador a manipulação de variáveis a partir de seus efeitos nos resultados desejados. Quanto maior o controle sobre as variáveis, maior a possibilidade do pesquisador estabelecer relações de causalidade entre as variáveis (Frankfort-Nachmias e Nachmias, 1996). No caso desta tese, existem muitas variáveis internas às organizações e relativas às parcerias entre empresas, o que descredenciava essa estratégia de pesquisa, do ponto de vista da adequação ao problema.

6.1.1.2

Pesquisa histórica

As pesquisas históricas em relacionamentos ocorrem quando não é mais possível acessar os atores envolvidos num dado evento. Para esse tipo de pesquisa, utiliza-se a consulta a dados secundários, culturais e evidências encontradas. Ela, entretanto, não se apresentou como uma boa alternativa, pois havia a necessidade de se investigarem os dados mais recentes possíveis.

6.1.1.3

Análise de arquivo (documental)

A pesquisa de arquivo tem como principal característica obter dados secundários advindos de relatórios gerenciais/contábeis e estatísticas governamentais. É mais utilizada quando da ocorrência de fenômenos especiais ou quando se procuram estabelecer relacionamentos entre eles. A estratégia também não fornece dados atualizados, o que era desejado no caso da presente pesquisa.

6.1.1.4

Estudo de Caso

O estudo de caso é preferido quando o tipo de questão eleito pela pesquisa assume a forma “como” e “por que”; quando o controle do investigador sobre os eventos é muito reduzido; ou quando o foco temporal está em fenômenos contemporâneos dentro de um contexto empírico. O estudo de caso explanatório (causal) pode ser complementado por estudos de caso descritivos ou exploratórios. A necessidade de se utilizar essa estratégia de pesquisa deve nascer do desejo de entender um fenômeno social complexo. A essência de um estudo de caso, ou a tendência central de todos os tipos de estudo de caso, é que eles tentam esclarecer um conjunto de decisões, tais como: “por que elas foram tomadas?”; “como elas foram implementadas?”; e “quais os resultados alcançados?” (Yin, 2001). Como o objetivo da pesquisa não visava a investigar tais aspectos, decidiu-se pela não-utilização dessa estratégia.

6.1.1.5

Pesquisas Amostrais (*Survey*)

Na estratégia de pesquisa amostral, as informações são coletadas de uma amostra representativa da população. Os respondentes, normalmente, pertencem a categorias pré-estabelecidas, o que permite que as informações sejam estatisticamente analisadas. Utilizou-se, então, a pesquisa amostral, que apresentava o conjunto de características que melhor atendia às necessidades da tese.

Três tipos de ferramentas podem ser utilizadas para as pesquisas amostrais: 1 - correspondências via correio; 2 - entrevistas diretas; e 3 - pesquisas por telefone. A escolha do método dependerá dos recursos disponíveis: tempo, dinheiro e acesso às fontes de informação. Nas próximas subseções são descritos os detalhes desse procedimento de pesquisa.

6.1.2

O método de coleta de dados

Os fornecedores de frutas e vegetais frescos do Reino Unido apresentam um alto índice de dispersão, no que tange à distribuição geográfica. Assim sendo, as entrevistas foram descartadas, principalmente por seu custo operacional e pelo tempo disponível. A pesquisa por telefone, por outra via, mostrou-se inviável por três motivos principais. O primeiro foi o custo do acesso telefônico, o segundo a dificuldade de encontrar as pessoas, pois normalmente estão ocupadas com outras atividades e o terceiro a dificuldade do pesquisador em entender diferentes traços lingüísticos – como os sotaques – em entrevistas por telefone. A experiência positiva de realização com pesquisa amostral, utilizando como instrumento a correspondência postal, realizada por Duffy (2002), foi também um fator motivador do uso desse método de coleta de dados.

6.1.3

A definição da população

A população para a pesquisa foi definida como sendo composta por

fornecedores de frutas e vegetais, do Reino Unido¹, que trabalham junto às empresas distribuidoras. Os grandes mercados atacadistas não foram incluídos na pesquisa, pois mesmo sendo fornecedores dos distribuidores, eles funcionam essencialmente como empresas distribuidoras, uma vez que também vendem diretamente aos clientes finais. Outro elemento considerado na definição da população relaciona-se com o tamanho das empresas. Como parâmetro de definição utilizou-se o faturamento anual, por ser este o indicador mais comumente utilizado e o de mais fácil acesso no Reino Unido. Decidiu-se, então, que a população pesquisada seria aquela com faturamento anual acima de £1.000.000,00, em reais algo próximo a R\$ 4.000.000,00. Essa definição foi tomada a fim de reduzir a heterogeneidade das empresas participantes, evitando assim a participação daquelas que fornecem produtos temporariamente (aproveitamento da sazonalidade característica do setor), e daquelas que lidam com os pequenos distribuidores locais. A homogeneidade foi definida, tendo como base principal os argumentos apresentados por Heide e John (1990). Esses autores argumentam que a restrição através da homogeneização minimiza a influência de fatores externos que acarretam variação nas análises dos dados. Por outro lado, Joseph *et al.* (1995) e Monczka, Petersen e Handfield (1998) indicam que a pesquisa realizada em um único setor industrial tem limitações referentes à generalização dos resultados para outros setores.

6.1.4

A amostra

A maioria dos pesquisadores defende que uma amostra aleatória deve ser extraída da população de interesse. Entretanto, devido à racionalização da base de fornecedores realizada no Reino Unido, tanto por parte dos distribuidores quanto por parte das empresas de serviços de alimentação, a população total, após a definição dos critérios estabelecidos no item anterior, era de 308 empresas. Assim sendo, considerou-se que tanto do ponto de vista financeiro quanto do operacional era possível enviar os questionários para o total da população e assim foi feito.

¹ No Reino Unido a caracterização é feita utilizando o termo vegetais que engloba verduras, legumes e tubérculos

Os dados cadastrais das empresas foram obtidos a partir de publicações relacionadas ao setor, editadas no Reino Unido. Os principais documentos utilizados foram o **Plimsoll Report** (1999), o **Food Trades Directory of the UK and Europe** (2000) e algumas bases de dados eletrônicas www.tradeuk.com, www.tapin.co.uk e www.yell.com. Como havia vários dados conflitantes e sobrepostos, foi necessário compará-los com a base de dados utilizada por Duffy (2002). Além disso, os endereços foram confirmados através do site do correio britânico www.royalmail.co.uk. Após o envio dos questionários², oito retornaram com endereços não confirmados. Destes, cinco indicavam empresas que não estavam instaladas naquele endereço (não sendo possível localizar o novo endereço). Os outros três foram enviados novamente, não havendo nenhum problema de correspondência, restando 303 questionários ao final do processo.

6.1.5

O projeto do questionário

A pesquisa amostral realizada através da aplicação de questionário enfrenta problemas, pois o percentual de respondentes pode comprometer a possibilidade de análise. Além disso, uma vez enviado às empresas, a possibilidade de esclarecimentos de eventuais dúvidas são bastante pequenas. A fim de facilitar eventuais esclarecimentos, na carta de apresentação da pesquisa, presente na primeira página do questionário, foram disponibilizados tanto telefones como o *e-mail* de contato do pesquisador.

O questionário foi desenvolvido ao longo de três meses (início de novembro de 2001 a final de janeiro de 2002). Para o pré-teste do questionário, a versão preliminar do questionário foi enviada para pesquisadores brasileiros e britânicos, e para representantes de entidades ligadas ao setor no Reino Unido. As considerações feitas foram debatidas e a versão final do questionário foi estruturada de forma a conter sete questões.

² Uma cópia do questionário é apresentada no Anexo desta tese.

6.1.5.1

Os objetivos de cada questão

Nesta seção tem-se por objetivo apresentar as justificativas para o desenvolvimento das questões relacionadas ao questionário, considerando-se a importância desse instrumento.

Questão 1: esta questão teve por objetivo identificar o conceito de comprador-intermediário-chave para cada uma das empresas. Seu desenvolvimento demandou várias horas de debate entre o orientando e co-orientador, pois durante a revisão da literatura observou-se que os autores falam do relacionamento entre as empresas, mas não do relacionamento com seu cliente-chave. Entretanto, como o objeto principal do questionário era investigar o relacionamento de parcerias dentro da cadeia de suprimentos, identificou-se a necessidade de investigar não apenas o relacionamento fornecedor e comprador-intermediário na cadeia britânica de suprimentos, mas também o relacionamento do ponto de vista do fornecedor com seu cliente-chave. Nesse momento, surgiu um novo problema: **Qual o conceito de comprador intermediário chave?** Inicialmente, tentou-se desenvolver um conceito que abrangesse todos os conceitos identificados na literatura. Além disso, foi solicitado aos pesquisadores do *Centre for Food Chain Research* que conceituassem cliente-chave. Observou-se então que ou o conceito seria simples e não abrangeria todas as possibilidades, ou ficaria extenso demais e, conseqüentemente, difícil de entender. Surgiu então a possibilidade de se ofertarem diferentes conceitos, a partir dos quais cada empresa responderia o seu questionário. Considera-se esse procedimento uma contribuição metodológica, pois constituiu um elemento diferencial da coleta de dados desta pesquisa, não observada em nenhuma pesquisa realizada e publicada nos periódicos internacionais de maior destaque. Antes de se chegar à versão final, na qual apenas uma das opções deveria ser escolhida, pensou-se em inserir uma questão segundo a qual os respondentes criariam um *ranking* com os conceitos. Entretanto, tal questão foi abandonada, pois a quase totalidade dos pesquisadores e empresários que avaliaram a versão preliminar do questionário aconselhou sua exclusão, devido à dificuldade de resposta e ao tempo demandado.

Questão 2: essa questão possui cinco pontos a serem investigados. Os dois primeiros dizem respeito ao número de empresas de clientes que cada fornecedor

possui e ao número de clientes-chaves. As três últimas cobrem a participação financeira e o tempo de dedicação aos fornecedores. Do ponto de vista conceitual, não houve grandes dificuldades para a estruturação da questão. Sabia-se, através de levantamentos feitos em pesquisas anteriores no *CFCR* que os principais clientes dos fornecedores eram os distribuidores, notadamente as grandes redes de supermercados inglesas citadas no Capítulo 5, além das empresas de serviços de alimentação. Diante disso, o objetivo era identificar quantos clientes em cada uma das áreas eram considerados compradores intermediários chave. O terceiro ponto tinha por objetivo identificar o grau de dependência do fornecedor com relação ao(s) seu(s) cliente(s)-chave(s), considerando-se o grau de dependência da participação do(s) cliente(s)-chave(s) no faturamento do fornecedor. Nos quarto e quinto pontos buscava-se identificar quão dedicados ao(s) seu(s) cliente(s)-chave(s) os fornecedores estão.

Questão 3: essa questão foi desenvolvida com o intuito de se captarem as percepções dos respondentes com relação aos processos que representariam a base do desenvolvimento de parcerias de sucesso com seus compradores intermediários chave. Nesse caso, procurava-se uma visão mais geral da percepção manifestada por eles. Entretanto, no projeto do questionário a questão foi estruturada em quatro eixos principais. Porém, quando a análise fatorial foi realizada, foram identificados sete fatores, de forma que os itens agrupados não correspondiam à teoria. Assim sendo, toda a questão foi descartada da análise.

Questão 4: essa questão tem por objetivo identificar os benefícios alcançados pelos fornecedores no seu relacionamento com seus compradores intermediários chave. Assim sendo, foram apresentadas quinze variáveis que, teoricamente, estavam agrupadas em quatro grandes grupos, conforme pode ser visto no Anexo. Todos eles correspondem ao modelo conceitual para pesquisa de campo apresentado no Capítulo 4 desta tese. A questão não apresentou maiores problemas conceituais para o desenvolvimento, pois os benefícios observados na literatura, com relação às parcerias nas cadeias de suprimentos, apresentam uma convergência. Foi necessária, então, a adaptação de alguns dos itens, como por exemplo o *mix de produtos*, aos critérios importantes para a cadeia de alimentos frescos.

Questão 5: essa questão teve como objetivo principal identificar os processos internos a serem desenvolvidos pelos fornecedores, visando conseguir os benefícios advindos de suas parcerias. Cabe observar que a pergunta foi feita especificamente para a captura dos benefícios resultantes das parcerias. Vale ressaltar que os benefícios alcançados por parte dos fornecedores é um fato importante, considerando-se a assimetria de poder na cadeia de alimentos britânica, e que os fornecedores são o elo mais fraco.

Questão 6: essa questão tem como objetivo principal identificar os processos externos ou processos de desenvolvimento conjunto entre fornecedores e compradores. Aqui, o interesse foi identificar quais processos os fornecedores consideravam importantes para serem desenvolvidos em parcerias com os compradores, a fim de manter e aprimorar as parcerias desenvolvidas. Tanto na Questão 5 quanto nesta procurava-se, inicialmente, confrontar os processos identificados por Fearne e Hughes (1998), quando da pesquisa realizada junto aos compradores.

6.1.5.2

Posicionamento do questionário no conceito de relacionamento

A aplicação do questionário deu-se de forma a obter informações das percepções e realidades do ponto de vista dos fornecedores. Considerando que numa cadeia de suprimentos existe relacionamento entre as empresas, torna-se possível perceber que a pesquisa de campo possui uma limitação, pois investiga apenas um elo de ligação dentro da cadeia. Além disso, alguns pesquisadores defendem que, numa pesquisa sobre relacionamento, deve-se investigar as mesmas questões em ambos os lados. Entretanto, do ponto de vista operacional, tal procedimento é extremamente complexo, pois haveria a necessidade de identificação dos compradores intermediários chave e dos fornecedores de cada uma das empresas. De acordo com experiências anteriores do professor co-orientador, as empresas britânicas não se apresentam dispostas a fornecer esse tipo de informação. Ainda assim, o grau de distorção dessa visão do relacionamento não apresentava grandes riscos às análises, pois o trabalho desenvolvido por Fearne e Hughes (1998) havia identificado alguns elementos importantes para o relacionamento por parte dos compradores. Outro fator atenuante é a continuidade do projeto de pesquisa dentro do *CFCR*. Em trabalhos desenvolvidos por outros

autores, podem-se observar as dificuldades típicas do desenvolvimento de pesquisas que investigam os dois lados da parceria (Provan e Gassenheimer, 1994; Heide e John, 1990; Kumar, Scheer e Seenkamp, 1995).

6.1.5.3

A escolha da escala para as questões 4, 5 e 6

Decidiu-se pela utilização da escala Likert, com cinco pontos, de acordo com a recomendação da Sociedade Britânica de Pesquisa em Marketing, conforme citação por pesquisadores como Rodeghier (1996), que afirma que as escalas de cinco pontos ou mais são suficientemente confiáveis e consistentes para a obtenção das informações desejadas. Decidiu-se, por outro lado, não utilizar a escala com sete pontos, pois alguns autores consideram que, para pesquisas de relacionamento, esta segunda escala apresenta maior grau de dificuldade de resposta, aumentando o tempo de aplicação. Além disso, a experiência do orientador do projeto indicou que a escala eleita apresentaria melhores resultados no mercado britânico. Considera-se que não há necessidade de se justificar a escolha das outras escalas utilizadas nas demais questões.

6.1.5.4

A composição dos fatores

O desenvolvimento do questionário teve como referencial o modelo teórico apresentado na Figura 4.2. Cada uma das questões descritas anteriormente **possuía um número de variáveis maior que o número de fatores** que compõem o modelo. Assim sendo, cada um dos fatores era composto por mais de uma variável. Tal procedimento foi empregado para reduzir o índice de erro de medição, que ocorre quando se utiliza apenas uma variável para medir um fator. Existem diferentes tipos de erros que podem estar envolvidos quando da representação única de um fator por apenas uma variável (Churchill, 1979; Bryman e Cramer, 1997). Um dos motivos para se utilizar mais de um fator é a dificuldade de se representar o conceito em sua totalidade numa única questão. Assim, o pesquisador está sujeito a cometer o erro de criar uma questão complexa, do ponto de vista do entendimento conceitual. Outro ponto positivo relativo à utilização de escalas múltiplas de medição está no fato de dispersar o erro de interpretação pelos diferentes respondentes. Como complementação a esse

argumento, pode-se destacar Hair *et al.* (1998), que observaram a redução do erro de medição relativa a escala simples a partir do uso da média dos valores obtidos por escala múltipla .

6.1.6

O uso de questões reversas

O uso ou não de questão reversa foi analisado durante o desenvolvimento do questionário. As questões reversas são utilizadas a fim de reduzir a possibilidade de complacência por parte dos respondentes, com respostas iguais para a maioria das questões. Por outro, a formulação das questões reversas faz com que o questionário fique mais explicativo e, portanto, mais lento e difícil de ser respondido.

Em discussão realizada com o professor titular da cadeira de Métodos Quantitativos do Imperial College at Wye, estudou-se a possibilidade de formular um questionário com questões afirmativas e outro com questões negativas (reversas). Os diferentes questionários seriam enviados para empresas diferentes, configurando-se então dois grupos de respondentes, um para os questionários com questões afirmativas, e o outro com as reversas. Esse método, inicialmente, pareceu interessante, pois havia a possibilidade de comparar os resultados utilizando-se dois instrumentos de pesquisa nos quais apenas um fator se diferenciava. Entretanto, considerando-se o total da população, que continha 303 empresas, e o número mínimo necessário de respondentes, para atender à análise fatorial, em torno de 88 empresas, decidiu-se não recorrer a tal expediente, que apresentava o risco de fazer perder a oportunidade realizar a análise fatorial e, conseqüentemente, as análises posteriores.

O questionário foi enviado contendo somente questões afirmativas. Pelo total de 99 empresas respondentes, verificou-se que a escolha foi acertada, pois não seria possível realizar a análise fatorial com apenas metade dos questionários. Além disso, o acerto da escolha foi confirmado durante a tabulação dos dados, quando se observou que os casos de complacência parcial por parte dos respondentes poderiam ser desconsiderados. Dessa forma, o procedimento adotado não proporcionou distorções relevantes aos resultados. Pode-se destacar, ainda, outros dois fatores que contribuíram para que os respondentes não fossem

complacentes. O primeiro reside no fato de não haver obrigatoriedade de resposta, pois não havia promessa alguma de benefício a ser recebido caso o questionário fosse respondido. Isso fez imaginar que os empresários que não desejassem preencher o documento, simplesmente os descartariam. O segundo fator relaciona-se com a não-obrigatoriedade de identificação por parte das empresas. No questionário, foi enviado um número que nos dava a possibilidade de identificar as empresas e, do total de 99 questionários respondidos, nenhum respondente negou-se a fornecer sua identificação, o que fez concluir que os participantes confiaram no sigilo das informações individuais prometido a eles.

6.1.7

A taxa de resposta obtida

A primeira remessa dos questionários foi enviada em meados de fevereiro de 2002. De acordo com a experiência do coordenador do projeto junto ao setor, de recomendações encontradas na literatura e da expectativa de tempo de entrega do correio britânico, decidiu-se que a segunda remessa seria enviada três semanas após, ou seja, no início de março 2002. Assim sendo, no início de março, aproximadamente 70 questionários preenchidos haviam retornado. A segunda remessa foi preparada e enviada para as empresas cujos questionários não haviam sido recebidos devidamente preenchidos até o dia 08 de março de 2002. A única diferença existente entre os questionários da primeira remessa e da segunda foi a carta de apresentação. Em meados de abril estavam à disposição 99 questionários para as análises, encontrando-se a maioria já em adiantado estágio de tabulação. Considerando-se então o total de 303 empresas que receberam os questionários, obteve-se uma taxa de retorno de 32,67%. De acordo com a revisão de literatura, é comum trabalhar-se com uma taxa de retorno entre 20% e 30% (Tan *et al*, 1999; Heide e John, 1990; Kumar, Scheer e Seenkamp, 1995). Deve-se ressaltar que os 32,67% são da população total e não da amostra, o que assegurava ainda mais tranquilidade quanto ao número de respostas obtidas. Pode-se creditar esse bom índice de resposta primeiramente ao conhecimento do co-orientador a respeito do ambiente empresarial pesquisado e, em segundo lugar, ao tempo demandado, de menos de 15 minutos para cada questionário. O período de resposta foi estimado, pois as respostas não demandavam consultas a relatórios gerenciais ou contábeis, além de conterem apenas questões objetivas.

6.1.8

Aleatoriedade dos respondentes.

Apesar de se ter conseguido um bom número de respostas, existe a necessidade de realizar testes a fim de verificar se os respondentes representavam satisfatoriamente as 303 empresas pesquisadas. O teste verifica se existe aleatoriedade dos respondentes com relação aos critérios de controles estabelecidos. Para tanto, os valores percentuais do total de respondentes foram comparados com os percentuais do total da população utilizando dois critérios, quais sejam: faturamento anual e posicionamento geográfico. Com relação ao primeiro fator, pode-se observar a distribuição no gráfico apresentado na Figura 6.1. Com o objetivo de testar a consonância entre o faturamento dos respondentes e da população, a hipótese nula foi estabelecida *Ho: a frequência relativa da distribuição por faturamento dos respondentes é igual a da população*. Com o auxílio do software Excel calculou-se, então, o valor *qui-quadrado* (χ^2) para verificar se havia diferença entre o valor observado e o valor esperado. O valor de $\chi^2 = 4,12$ foi encontrado e comparado com o valor tabelado de $\chi^2_{0,05} = 9,48$, com 4 graus de liberdade. Assim sendo, torna-se difícil rejeitar a hipótese nula *Ho*, ou seja, não há evidências suficientes para indicar diferença entre as médias da população e dos respondentes.

$$\chi^2 = \sum_j \frac{(O - E)^2}{E}$$

O > valor observado

E > valor esperado

j > total de pares

Equação 6.1

Com relação ao posicionamento geográfico, a população de empresas apresentava-se dispersa por todo o Reino Unido. Do total de 86 *counties*³, havia empresas em 46. Observou-se também que 15 deles possuíam no mínimo uma das empresas pesquisadas e que três detinham mais de 25 empresas da população. Em seguida, foram organizados dez grupos (denominados reg1, reg2,..., reg10),

³ O Reino Unido divide-se em unidades administrativas chamadas *counties*. Não se tratam de estados, por não possuírem autonomia administrativas e financeira. Por outro lado, também não representam simplesmente uma divisão geográfica, pois possuem autonomia sobre algumas atividades. Assim sendo, julgou-se melhor não arriscar uma tradução.

com número aproximado de participantes em torno de 30 empresas. No gráfico apresentado na Figura 6.2 pode-se observar o percentual de empresas por região a partir do total da população e o percentual de respondentes. Com o objetivo de testar a igualdade entre as médias, foi estabelecida a hipótese nula ***H₀***: *a frequência relativa da distribuição geográfica dos respondentes é igual à da população*. Com o auxílio do Excel calculou-se então o valor *qui-quadrado* (χ^2), para verificar se existe diferença entre o valor observado e o valor esperado. O valor de $\chi^2 = 5,62$ foi encontrado e comparado com o valor tabulado de $\chi^2_{0,05} = 16,91$ com 9 graus de liberdade. Novamente, seria difícil rejeitar a hipótese nula ***H₀***, uma vez que não há evidências suficientes para indicar diferença entre as médias da população e dos respondentes.

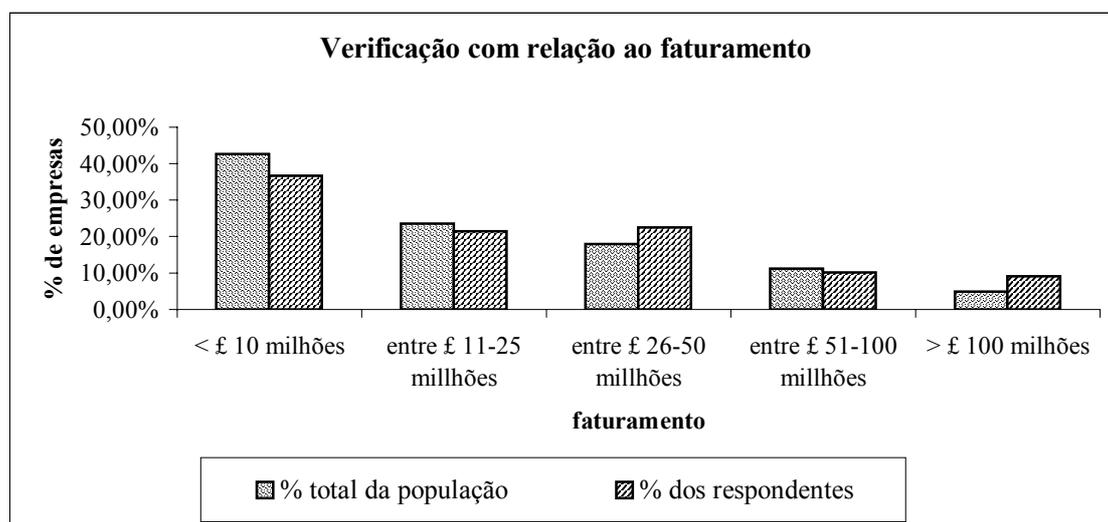


Figura 6.1: Faturamento anual da população e dos respondentes.

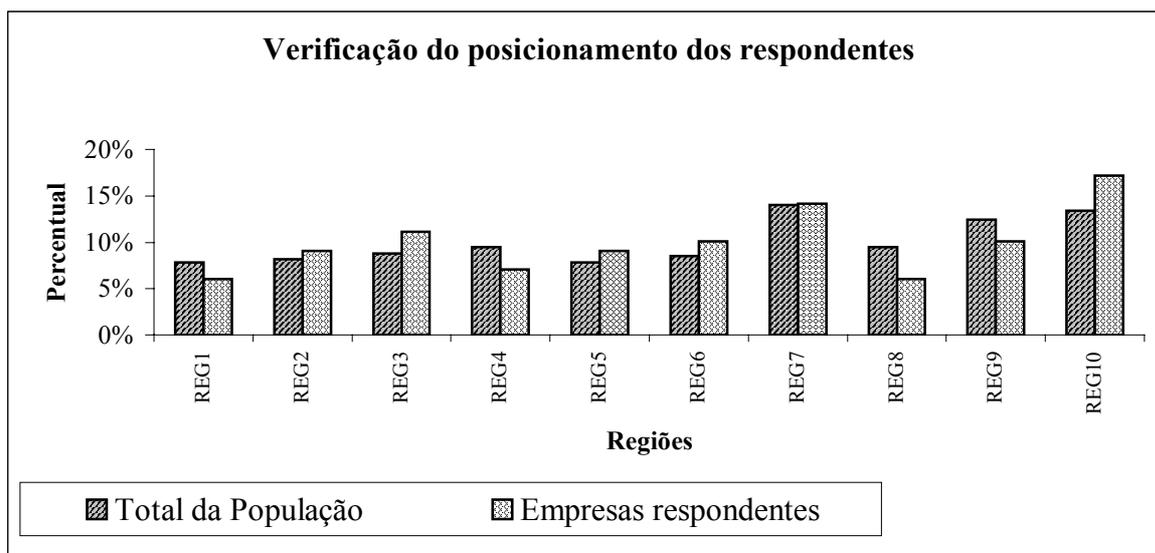


Figura 6.2: Posicionamento geográfico da população e dos respondentes.

6.1.9

Codificação dos dados para tabulação

O software SPSS (*Statistic Package for Social Study*) foi escolhido como ferramenta de análise dos questionários devido a quatro motivos principais: 1 - o conhecimento prévio por parte do pesquisador e do orientador do projeto das principais rotinas necessárias às análises; 2 - a grande difusão de uso do referido software entre os pesquisadores, em escala mundial; 3 - a disponibilidade de licença de seu uso no *CFCR*, e 4 - o atendimento às necessidades demandadas pelas investigações.

Para a entrada de dados foi estabelecida uma codificação dos questionários. Na Quadro 6.1 utilizou-se o sistema binário, sendo que para a resposta escolhida inseria-se 1 (um); em caso contrário, assinalava-se 0 (zero). Na questão dois, para os dois primeiros itens foram utilizados os totais de empresas citadas e para as três últimas questões utilizou-se uma codificação categórica convencional de 1 (um) a 5 (cinco), da esquerda para a direita no questionário. A questão três seguiu também essa codificação categórica, assumindo o código 1 (um) para *not all important* e código 4 (quatro) para *critically important*. Para as questões 4, 5 e 6 foram utilizadas as mesmas codificações apresentadas no quadro a seguir:

Resposta possível	<i>Discordo Completamente</i>	<i>Discordo</i>	<i>Neutro</i>	<i>Concordo</i>	<i>Concordo Totalmente</i>
Codificação	1	2	3	4	5

Quadro 6.1: Codificação das questões 4,5,6.

Após a codificação das variáveis, foi realizado o procedimento de entrada de dados. Após essa etapa, todas as transformações de variáveis e composições de novas variáveis foram realizadas utilizando-se os recursos disponíveis no SPSS. Finalmente, a fim de reduzir a margem de erros, dez dias após a entrada de todos os dados realizou-se um processo de verificação dos valores imputados.

6.2

Técnicas estatísticas

6.2.1

Confiabilidade da escala

Um dos testes realizados após a tabulação de dados é o teste de confiabilidade da escala, chamado de *Cronbach's Alfa* (α). De acordo com Churchill (1979), o cálculo do coeficiente α deve ser o primeiro procedimento para verificar a qualidade da escala utilizada, após a coleta e tabulação dos dados. O teste calcula a correlação entre as variáveis utilizadas no questionário. Se os coeficientes de correlação entre os fatores forem baixos, as variáveis utilizadas no sistema de múltiplos itens para cada fator serão consideradas *pobres*.

Vários trabalhos publicados na área de relacionamento indicam que os testes de confiabilidade baseados no Cronbach's α mostraram-se suficientes na maioria dos casos em que foram utilizados (Monczka, Petersen e Handfield, 1998; Heide e John, 1992; Noordewier *et al*, 1990; Kumar, Scheer e Seenkamp, 1995).

6.2.2

Análise Fatorial

6.2.2.1

Visão geral

Na revisão da literatura identificaram-se alguns autores que utilizaram análise fatorial como ferramenta de *purificação* ou *depuração* de escala, dentre os

quais podemos citar: Frazier (1983); Reve e Stern (1985), Kumar, Stern e Achrol (1992).

A análise fatorial é uma técnica estatística multivariada que tem por objetivo agrupar variáveis que estão altamente correlacionadas em conjuntos de variáveis chamados **fatores**. Essa técnica estatística permite a visualização de fenômenos envolvendo diversas variáveis que seriam difíceis ou impossíveis de se visualizar sem o agrupamento. É importante ressaltar que, além do agrupamento, a análise fatorial pode ser utilizada para redução de dados.

No agrupamento de variáveis, os fatores refletem os relacionamentos que existem entre as variáveis. Cabe ao pesquisador a interpretação dos fatores (ou novas variáveis) formados a partir do questionário inicial, tendo como base a teoria utilizada no desenvolvimento do questionário (Tabachnick e Fidell, 1989; Everitt e Dunn, 2001). A redução de dados é obtida a partir do resultado constituído pelo agrupamento das variáveis. Assim sendo, calculam-se os valores dos fatores a partir dos valores das variáveis envolvidas com cada um dos fatores, e tem-se, então, um novo conjunto fatores (variáveis) com seus respectivos novos valores.

Segundo Souza-Neto, Backer e Souza (2002), o método de análise fatorial é explicado, principalmente, em termos de álgebra das matrizes, que é uma maneira de lidar com grupos de indivíduos ao invés de indivíduos um a um. A análise fundamenta-se em duas propriedades estatísticas gerais:

1 - as variáveis aleatórias observadas são expressas por variáveis hipotéticas independentes, chamadas de "fatores";

2 - o número de variáveis independentes ou explicativas é inferior ao número de variáveis observáveis.

A análise fatorial representa cada variável observada como uma combinação linear de um grupo reduzido de fatores comuns mais um fator único para cada variável resposta. Assim:

$$\begin{aligned}x_1 &= a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + e_1 \\x_2 &= a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + e_2 \\&\vdots \\&\vdots \\&\vdots \\x_p &= a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + e_p\end{aligned}$$

Equação 6.1

onde as seguintes pressuposições são feitas:

1. m é o número de fatores comuns (este número é sempre menor que p)
2. F_1, F_2, \dots, F_m são fatores comuns, com x distribuído normalmente segundo uma média zero e variância unitária.
3. a_{ij} é o coeficiente de saturação de F_j na combinação linear de x_i .
4. $e_1, e_2, e_3, \dots, e_p$ são os fatores únicos, cada um relacionado a uma variável original.

De acordo com Sharma (1996), em duas dimensões a análise fatorial pode ser geometricamente representada por:

$$x_1 = a_{11}F_1 + a_{12}F_2$$

$$x_2 = a_{21}F_1 + a_{22}F_2$$

Equação 6.2

Conforme apresentado na Figura 6.1, os vetores X'_1 e X'_2 de n observações podem ser representados em uma dimensão n -dimensional. Entretanto, os dois vetores estão inseridos em um subespaço tetrádico definido pelos eixos ortogonais F_1, F_2, e_1 e e_2 . Mais especificamente, X'_1 é representado em três dimensões pelos vetores F_1, F_2, e_1 . O vetor X'_2 é definido pelo subespaço F_1, F_2, e_2 . O objetivo da análise fatorial é definir este subespaço tetrádico.

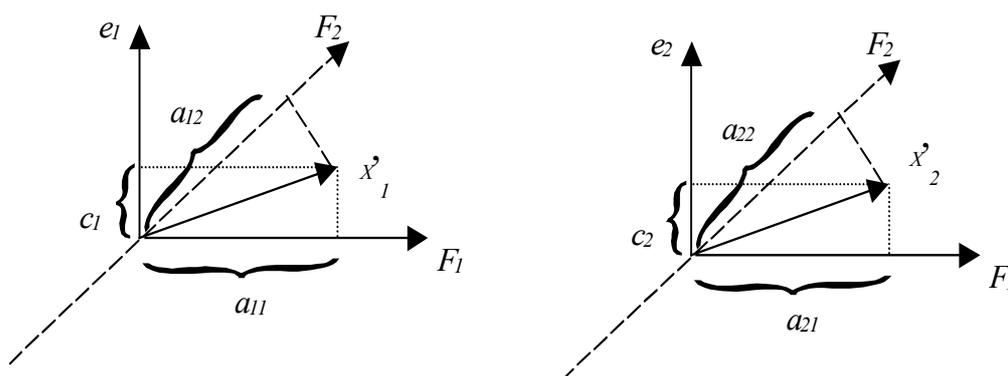


Figura 6.1: Representação dos vetores X'_1 e X'_2
 Fonte: adaptado de Sharma (1996, p.98)

Conforme pode ser observado na Figura 6.1, os coeficientes a_{11}, a_{12} e c_1 são as projeções de X'_1 nos eixos de F_1, F_2, e_1 e, respectivamente, os coeficientes $a_{21},$

a_{22} e c_2 são as projeções de X_2 em F_1, F_2, e_2 . Assim sendo, do teorema de Pitágoras pode-se estabelecer as relações matemáticas apresentadas na seqüência.

$$\|x_1'\|^2 = a_{11}^2 + a_{12}^2 + c_1^2$$

$$\|x_2'\|^2 = a_{21}^2 + a_{22}^2 + c_2^2 \quad \text{Equação 6.3}$$

Esta equação $a_{21}^2 + a_{22}^2$ dá a comunalidade da variável X_1 ; e $a_{21}^2 + a_{22}^2$ dá o valor da comunalidade da variável X_2 . Os aspectos numéricos da análise fatorial dizem respeito às estimativas dos coeficientes de saturação (a_{ij}) e à comunalidade.

6.2.2.2

Pré-condições para análise

Com relação ao tamanho da amostra para a realização da análise fatorial, Hair *et al.* (1998) recomendam que não se deve fazer análise fatorial com amostras menores do que cinqüenta respondentes, mas de preferência o número de respondentes deve ser maior ou igual a cem. Como regra geral, é preferível haver pelo menos cinco observações para cada variável que compõe as questões. Quando possível, deve-se experimentar uma razão de dez para um. Se houver necessidade de trabalhar amostras pequenas e/ou poucos casos por variável, o pesquisador deve estar atento ao valor parcial das conclusões, pois os resultados podem estar distorcidos. No caso da pesquisa desenvolvida, foram utilizados 99 respondentes. As questões que passaram por análise fatorial foram as seguintes: questão quatro, com 15 variáveis (razão 6,6); questão cinco, com 13 variáveis (razão 7,6); e questão seis, com 21 variáveis (razão 4,7). Considerando o total de respondentes, não se verificaram problemas relativos ao não-cumprimento das condições iniciais.

Outra pré-condição para a análise fatorial é a multicolinearidade, atributo segundo o qual o comportamento de uma variável pode ser explicado pela outra, ou seja, permite a verificação da influência de uma variável sobre outra. De acordo com Hair *et al.* (1998), um pouco de colinearidade é desejado entre as

variáveis para que se realize a análise fatorial. No SPSS, o inter-relacionamento entre as variáveis pode ser medido utilizando a rotina *Kaiser-Mayer-Olkin* (KMO), que é uma medida de adequação da amostra *MSA*. Conforme classificação realizada por Hair *et al.* (1998), as variáveis que apresentarem coeficientes maiores ou iguais a 0,8 são **meritosos**; maiores ou iguais a 0,7, e menores que 0,8, **medianos**; entre 0,6 e 0,7, **medíocres**; entre 0,5 e 0,6, **miseráveis**; e, finalmente, abaixo de 0,5, **inaceitáveis**. Tabachnick e Fidell (1989) sugerem que o índice de adequação seja pelo menos 0,6, para justificar a análise fatorial.

6.2.2.3

Extração dos Fatores

Os dois procedimentos mais comuns para extração dos fatores são: a análise dos componentes principais (*Principal Components Analysis – PCA*) e a análise comum dos fatores (*Common Factor Analysis – FA*). A principal diferença entre os dois métodos é a variância utilizada para extrair os fatores. Na análise fatorial, três tipos de variância total existem: 1 - variância comum; 2 - variância específica, também conhecida como “variância única”; e 3 - variância do erro. A variância comum pode ser definida como a variância da variável que é repartida por todas as outras variáveis envolvidas. A variância específica é aquela relativa a apenas uma variável. A variância do erro está relacionada à medida do erro e aos componentes aleatórios relativos ao fenômeno da medição (Hair *et al.*, 1998; Tabachnick e Fidell, 1989).

A escolha do método a ser utilizado depende, basicamente, dos objetivos da pesquisa. Geralmente, a análise dos componentes principais-PCA é recomendada nos casos em que se está procurando determinar um número mínimo de fatores capazes de promover o registro do maior número de variância na análise, ou seja, trata-se de fatores que explicam a variação, o que permite assumi-los como instrumentos eficientes para se fazer previsões em análises subsequentes. Por outro lado, a análise comum dos fatores-FA é recomendada quando o objetivo principal é identificar dimensões pré-estabelecidas, o que faz com que haja menor possibilidade de contaminação pela variância específica e pela variância do erro, posto que utiliza apenas a variância comum na extração (Hair *et al.*, 1998; Malhotra, 1999).

O objetivo da análise fatorial é a identificação dos fatores-chaves de sucesso no relacionamento dos fornecedores e compradores dentro da cadeia de suprimentos do Reino Unido. A identificação desses fatores responderá à primeira questão central de investigação formulada. Após a análise, foram identificados, num primeiro momento, os relacionamentos entre esses fatores e, em seguida, o relacionamento entre os fatores identificados e as características das empresas.

As justificativas apresentadas anteriormente conduziram à escolha da análise dos componentes principais (*Principal Components Analysis - PCA*), por considerar-se esse método mais apropriado ao caso. Foram identificados vários trabalhos, durante a revisão de literatura, que utilizam o mesmo método com objetivos conceitualmente similares aos desta tese, entre os quais Monczka, Petersen e Handfield (1998); Tan *et al* (1999); Gundlach, Achroc e Mentzer (1995).

6.2.2.4

Comunalidade

Comunalidades são as estimativas de partilhamento das variâncias entre as variáveis. A dimensão da comunalidade é um indicador útil para se ter acesso ao quanto uma variável é considerada na solução apresentada. Comunalidades com valores elevados indicam que a variância da variável foi extraída quando da solução do problema. Por outro lado comunalidades baixas mostram que uma porção relevante da variância daquela variável não foi considerada na solução (Hair et al. 1998). Assim, se uma variável apresenta comunalidade de 0,30 na solução, significa que esta variável é menos comum com as outras variáveis que compõem o fator, do que as variáveis que apresentarem comunalidades superiores a 0,30 e vice-versa.

Hair et al. (1998) recomenda que cada variável deve ter comunalidade de pelo menos 0,45, ou seja, deve partilhar pelo menos 45% de sua variância com as outras variáveis. Nos casos que as comunalidades forem inferiores a este valor, esses autores recomendam que a variável deve ser retirada da solução. Gassenheimer, Sterling e Robicheaux (1989) utilizam na análise dos componentes principais itens com valores superiores a 0,40.

6.2.2.5

Número de fatores a serem extraídos

Conforme descrito na subseção anterior, o método utilizado fará a extração das variáveis de forma a obter o máximo da variância possível. Uma vez que os fatores tendem a captar o máximo da variância, é necessário que se estabeleça um critério de corte ou critério de parada para cada fator. Segundo Hair *et al.* (1998), o critério de parada mais utilizado é o *latent root*, de acordo com o qual um fator deve reter a variância de pelo menos uma das variáveis, o que permitirá justificá-lo como fator. Considerando-se que cada variável contribui com pelo menos uma unidade para a variância total, chamada *autovalor*, o critério mínimo aceitável indica que é necessário que cada fator tenha pelo menos uma unidade em seu *autovalor* (Hair *et al.*, 1998; Tabachnick e Fidell, 1989).

Um segundo método utilizado é o teste *scree*, a partir do qual se traça um gráfico dos *eigenvalores* contra o número de fatores, sendo que o formato final da curva é utilizado para definir o critério de parada. Entretanto, Tabachnick e Fidell (1989) acreditam que esse método não é exato, e que depende muito de critérios estabelecidos pelo pesquisador para decidir o número de fatores a serem utilizados. A terceira abordagem utilizada relaciona-se com o total da variância acumulada explicada pelo total de fatores. Hair *et al.* (1998) sugerem que em ciências sociais não deveriam ser aceitos valores inferiores a 60% da variância explicada. Defendem também que o método é útil para testar a teoria existente ou para replicar pesquisas. Levando em consideração tal conjunto de fatores, esta tese utiliza-se do método *latent root* com *autovalor* de, no mínimo, 1 para cada fator e com a variância total superior a 60%.

6.2.2.6

Métodos de rotação

O resultado da análise fatorial é apresentado através de uma matriz inicial na qual se explicita o relacionamento, denominado *loadings* no SPSS, entre os cinco fatores extraídos e as variáveis individuais, conforme visto na Tabela 6.1, a seguir. Entretanto, a interpretação dessa matriz inicial, chamada matriz não-rotacionada, é um pouco difícil, pois a extração é feita a partir da ordem de importância de cada fator com o primeiro fator representando a maior variância; a partir do segundo

fator, a extração é feita com a variância residual. É muito comum o primeiro fator tender a ser um fator geral no qual todas as variáveis apresentam suas maiores significâncias (Hair *et al.*, 1998).

Variáveis	Fator (Component)				
	1	2	3	4	5
Var.D	.701	.292	1.575E-02	-.122	.211
Var.I	.663	-.364	-9.649E-02	-9.797E-03	-.236
Var.E	.630	-.151	.107	-.132	-.508
Var.H	.604	-.339	-.411	-.235	1.924E-02
Var.F	.575	-.460	-.402	-2.800E-02	4.304E-03
Var.C	.571	-.113	.539	.178	.297
Var.B	.560	-2.191E-02	.440	-.131	-.148
Var.J	.548	-.165	-5.923E-02	.518	-.296
Var.L	.519	.227	-.183	.334	.420
Var.M	.481	.455	-6.059E-02	-.249	-.355
Var.O	.513	.687	-.178	-.135	3.413E-02
Var.N	.493	.687	-.153	-.165	.110
Var.G	.454	-.518	-.235	-.198	.441
Var.A	.491	-.196	.622	-.319	.176
Var.K	.513	.141	6.806E-02	.711	-2.951E-02

Extraction Method: Principal Component Analysis.

Tabela 6.1: Matriz dos fatores componentes (Component Matrix).

Fonte: resultado de uma análise fatorial no SPSS.

Diante do problema de interpretação dos resultados, o método prevê um processo que se chama de rotação. De acordo com Tabachnich e Fidell (1989), quando se rotaciona a matriz componente a variância é distribuída entre todos os fatores, o que conduz normalmente a um melhor agrupamento das variáveis de uma forma mais evidente, em que as correlações mais fortes são maximizadas e as mais fracas minimizadas.

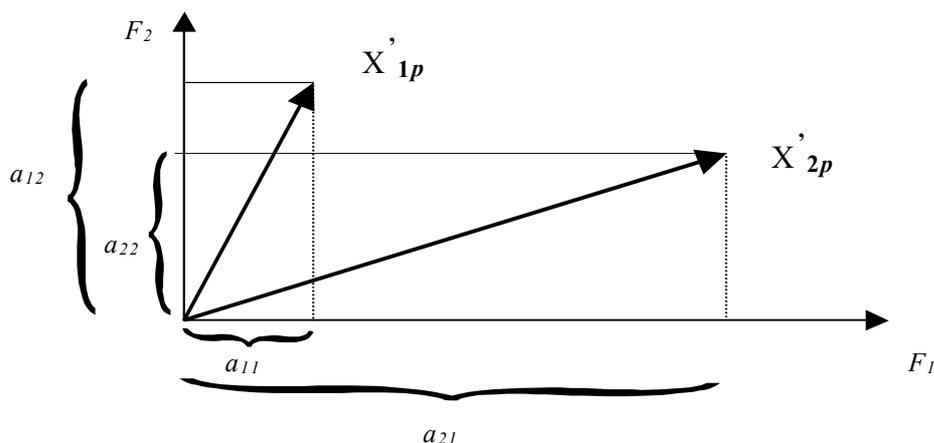


Figura 6.1: Gráfico representativo da rotação de vetores na análise fatorial.

Fonte: Adaptado de Sharna (1996, p.10)

De acordo com Sharma (1996), se na formulação apresentada a seguir, assumir-se que os eixos e_1 e e_2 são identificados e fixos (i.e., as comunalidades foram estimadas), os vetores X'_1 e X'_2 podem ser projetados em um subespaço bidimensional representado por F_1 e F_2 . A figura 5.4, apresenta o resultado das projeções dos vetores X'_{1p} e X'_{2p} . Os vetores de projeção X'_{1p} e X'_{2p} podem ser projetados em um subespaço unidimensional definido pelo vetores F_1 e F_2 . Tais componentes (i.e., a projeção dos vetores) são os coeficientes (*loadings*) e também os *loadings*-padrões para o modelo de rotação ortogonal.

Conforme apresentado na Figura 6.1 os valores a_{11} , a_{12} são os *loadings* de X_1 para F_1 e F_2 , respectivamente; e a_{21} , a_{22} são os *loadings* para X_2 para F_1 e F_2 , também respectivamente. O quadrado dos respectivos *loadings* fornece suas comunalidades de cada variável, que são a soma da comunalidade de cada variável com cada um dos dois fatores

$$\begin{aligned} \|x'_{1p}\|^2 &= a_{11}^2 + a_{12}^2 \\ \|x'_{2p}\|^2 &= a_{21}^2 + a_{22}^2 \end{aligned} \quad \text{Equação 6.1}$$

O comprimento da projeção dos vetores representa a comunalidade das variáveis. Os eixos apresentados na Figura 6.1 podem ser rotacionados sem que a orientação e magnitude dos vetores X'_{1p} e X'_{2p} sejam alterados, ou seja, sem que o total da comunalidade das variáveis seja alterado. Durante a rotação, o total da comunalidade não sofre alteração, porém a comunalidade de cada uma das variáveis se altera. Em outras palavras, a decomposição do total das Equação 6.1, que representa a comunalidade.

Existem, basicamente, duas categorias de rotação, quais sejam: 1 - rotação ortogonal e 2 - rotação oblíqua. Na rotação ortogonal não se permite que haja correlação entre os fatores; na rotação oblíqua essa correlação é permitida. Na realidade os dois métodos têm a tendência de apresentar resultados finais muito similares (Tabachnick e Fidell, 1989). Hair *et al.* (1998) defendem que os *constructors*, ou fatores que compõem os relacionamentos na área de

comportamento humano, apresentam correlações e portanto são necessariamente não-ortogonais. Além disso, Tabachnick e Fidell (1989) ressaltam que a rotação ortogonal deve ser utilizada somente se os pesquisadores tiverem plena convicção da total independência dos fatores. Os autores afirmam que uma forma prática de se decidir pelo uso da rotação oblíqua é observar os valores iniciais dos relacionamentos: se os valores em sua maioria forem superiores a 0,3, então existirá sobreposição de variâncias, o que indicará a não-independência dos fatores, o que conduz ao uso do referido procedimento. Como esta pesquisa envolve fatores de relacionamento entre fornecedores e compradores, decidiu-se pelo uso da rotação oblíqua.

6.2.2.7

Significância dos valores dos relacionamentos (*loadings*)

A interpretação da matriz rotacionada dá-se pela identificação dos maiores valores de relacionamento entre os fatores e as variáveis, considerando que esses *loadings* representam a correlação entre as variáveis e os fatores. Hair *et al.* (1998) defendem que valores inferiores a 0,3 devem ser sumariamente desconsiderados da análise; valores maiores que 0,4 já podem ser considerados significantes; e os valores acima de 0,5 podem ser considerados realmente significantes. Além disso, o tamanho da amostra tem influência em tais valores, sendo que apenas os valores acima de 0,45 devem ser considerados no caso de uma amostra com 150 respondentes, recomendação seguida na execução desta pesquisa.

Entretanto, outros dois pontos devem ser considerados na definição dos fatores. O primeiro diz respeito ao fato de dois fatores possuírem valores idênticos para a mesma variável. Nesse caso, os autores recomendam que a variável seja estabelecida como componente do fator ao qual estiver mais conceitualmente ligada. O segundo ponto, relacionado à seleção, refere-se à presença de apenas uma variável representativa para o fator. Segundo Tabachnick e Fidell (1989), esse fator só deverá ser mantido se o seu *loadings* for bastante representativo e se a variável for importante para as análises subseqüentes. Senão, deve-se evitar que apenas uma variável componha todo um fator.

6.2.3

Normalidade dos fatores (novas variáveis)

Para o desenvolvimento desta subseção utilizar-se-ão como referência Hair *et al.* (1998) e SPSS (1999). Deve-se considerar que, após a realização da análise fatorial, haverá um conjunto de novas variáveis, denominadas fatores. É preciso verificar se eles atendem às condições iniciais de análise para os testes a serem realizados na seqüência, como por exemplo a análise de regressão. Nesse momento, o teste de normalidade deverá ser realizado.

A normalidade é requerida para a realização dos testes estatísticos de ANOVA e de Regressão Linear. O diagnóstico mais simples de normalidade pode ser feito graficamente utilizando-se um histograma que compara os valores observados com a curva normal, conforme demonstrado pela figura 5.5. Esse método, entretanto, é muito simplista e pode apresentar problemas com amostras pequenas, diante das quais é recomendável utilizar o teste de normalidade da distribuição. Nele, compara-se a distribuição acumulada de probabilidade dos pontos observados com a distribuição acumulada, conforme a figura 5.6. Se os valores obtidos seguirem conjuntamente a linha diagonal, então os dados apresentam distribuição normal.

O terceiro e último teste de normalidade proposto baseia em análise estatística e pode ser realizado no SPSS, utilizando a rotina de Kolmogorov-Smirnov. Este método utiliza o teste de significância e tem como hipótese nula *H₀*: a distribuição da população é normal (ou seja, se não for possível rejeitar *H₀*, a distribuição será considerada normal). Na Tabela 6.1 é apresentado um resultado com o objetivo de exemplificação. Pelo exemplo apresentado pode-se verificar que os fatores 1, 2 e 4 rejeitam *H₀* com significância menor que 5%, ou seja, esses três fatores não apresentam distribuição normal. Entretanto, o teste falha em rejeitar o fator 3 com 5% de significância, o que conduz à conclusão de que esse fator apresenta distribuição normal.

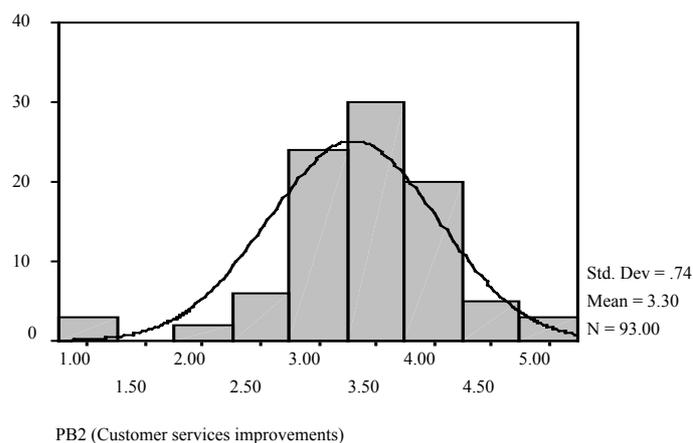


Figura 6.5: Teste de normalidade distribuição acumulada

Fonte: resultado qualquer no SPSS

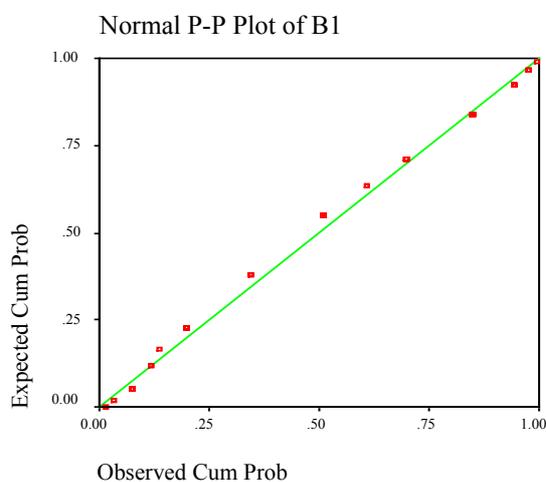


Figura 6.1: Teste de significância dos fatores

Fonte: resultado qualquer no SPSS

Test of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
fator 1	.206	98	.020
fator 2	.210	98	.016
fator 3	.189	98	.080
fator 4	.122	98	.030

a. Lilliefors Significance Correction

Tabela 6.1: Teste de normalidade usando significância 5%

Fonte: resultado qualquer no SPSS

6.2.5

Teste Qui-quadrado

O método *Pearson qui-quadrado* (χ^2) é um teste que investiga o relacionamento entre duas variáveis, utilizando para isso variáveis categóricas. Ele é um dos mais comuns, e baseia-se no cálculo do valor esperado das frequências em cada uma das células, caso não haja relacionamento entre duas variáveis. De acordo com Sincich (1996), o objetivo do teste χ^2 é determinar se existe interdependência entre duas variáveis qualitativas ou categóricas. A única condição inicial requerida para o teste χ^2 é que a frequência esperada para cada célula não pode ser pequena. A razão para isso é dada pelo fato de o teste qui-quadrado verificar as probabilidades em cada uma das células; quando as frequências esperadas apresentam-se baixas, as referidas probabilidades não podem ser estimadas com precisão, fazendo com que o teste não tenha consistência (Everitt, 1977; Hays, 1988; Kendall e Stuart, 1979).

A verificação das hipóteses para teste de independência das variáveis é formulada da seguinte maneira:

Hipótese nula H_0 : *as duas direções (linhas e colunas) das tabelas de contingências são independentes*

Hipótese alternativa H_a : *as duas direções (linhas e colunas) das tabelas de contingências são dependentes.*

a formulação matemática é:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \text{Equação 6.1}$$

onde

r = número de linhas da tabela.

c = número de colunas da tabela

O_{ij} = número de respostas observadas na célula da linha i e coluna j .

E_{ij} = estimativa do número de respostas na célula (ij)

O SPSS MANUAL (1999) utiliza como hipótese nula H_0 : as variáveis são independentes. Na Tabela 6.1 tem-se um exemplo de resultado para o teste *Pearson qui-quadrado* (χ^2). Pode-se observar que a hipótese nula é rejeitada a partir de um coeficiente Pearson de 3%. Nesse caso, as variáveis são dependentes. Além disso, pode-se observar que não há nenhuma falha na condição inicial para teste com todas as células tendo mais que 5% de frequência esperada.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.697 ^b	1	.030		
Continuity Correction ^a	3.547	1	.060		
Likelihood Ratio	5.392	1	.020		
Fisher's Exact Test				.045	.025
Linear-by-Linear Association	4.649	1	.031		
N of Valid Cases	99				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.82.

Tabela 6.1: Teste qui-quadrado para variáveis categóricas de um resultado qualquer no SPSS.

6.2.6

Regressão Logística

O modelo geral de regressão logística pode ser escrito como:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad \text{Equação 6.1}$$

A simplificação do modelo com apenas uma variável independente pode ser escrita como:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 \quad \text{Equação 6.2}$$

sendo então

$$p = \frac{p}{1 + e^{-(\beta_0 + \beta_1 X_1)}} \quad \text{Equação 6.3}$$

As variáveis independentes X_1, X_2, \dots, X_k podem ser a combinação de variáveis categóricas ou contínuas. Como $\frac{p}{1-p}$ é a chance de um evento ocorrer em função das variáveis independentes, então o log de $\frac{p}{1-p}$ é denominado com *logit*, caracterizando um processo comumente denominado *regressão logística*.

A regressão logística binária é um bom método quando a variável dependente tem apenas dois grupos; além disso, esse tipo de regressão não se depara com restrições rígidas. A regressão logística utiliza como mecanismo de predição as probabilidades. Apesar da probabilidade poder variar de 0 a 1, os valores da predição não devem assumir os valores extremos 0 ou 1. A regressão logística assume que o relacionamento entre duas variáveis pode ser representado por uma curva S, conforme apresentado na Figura 6.1. Nesse gráfico, o eixo Y representa a probabilidade de um evento ocorrer, enquanto o eixo X representa o grau de dependência entre as variáveis. Em um nível muito baixo de independência das variáveis, a probabilidade de ocorrer um evento aproxima-se de zero. À medida que o nível de independência cresce, a probabilidade também encontra valores progressivamente positivos.

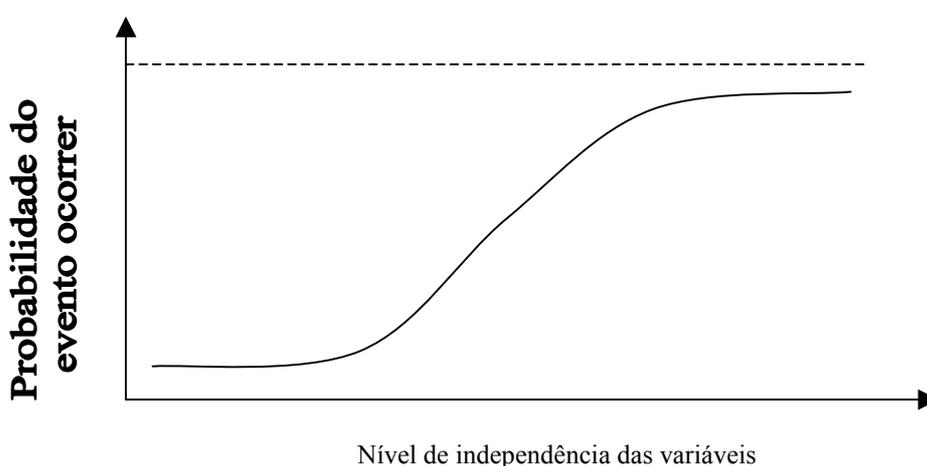


Figura 6.1: Comportamento da curva de regressão logística.

Fonte: Hair *et al.* (1998, p.277).

⁴ Probabilidade de o evento ocorrer.

Do ponto de vista algébrico, a regressão logística fornece-nos os coeficientes estimados para as variáveis independentes. A natureza não-linear da transformação logística requer o uso da máxima verossimilhança para encontrar os coeficientes estimados mais prováveis. De acordo com Sharma (1996), pode-se supor que a variável aleatória binária seja Y com valores 0 e 1, sendo a probabilidade $P(Y=1)$ dada por

$$P(Y = 1) = p = \frac{e^{\beta X}}{1 + e^{\beta X}} \quad \text{Equação 6.4}$$

onde β é o vetor dos coeficientes e X o vetor das variáveis independentes. Essa equação pode ser reescrita como

$$\ln \frac{p}{1-p} = \beta X \quad \text{Equação 6.5}$$

Assumindo que n observações são independentes, a função da máxima verossimilhança é dada por

$$L = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \quad \text{Equação 6.6}$$

assim sendo, o log da função de máxima verossimilhança é dado por

$$\ln L = l = \sum_{i=1}^n y_i \left(\frac{e^{\beta X}}{1 + e^{\beta X}} \right) + \sum_{i=1}^n (1 - y_i) \left(\frac{e^{\beta X}}{1 + e^{\beta X}} \right) \quad \text{Equação 6.7}$$

Os parâmetros estimados do vetor são obtidos com a maximização da Equação 6.7. O procedimento mais usual é utilizar as derivadas de primeira ordem com relação a cada uma das variáveis independentes, estabelecendo cada para cada equação valor igual a zero, o que resolve as equações. Entretanto, não existe uma solução analítica para o problema, sendo que os β_s estimados são obtidos utilizando-se o método Newton-Raphson.

De uma maneira geral, o antilogaritmo dos coeficientes β_i nos modelos de regressão logística representam a chance de ocorrer $\{y=1\}$ quando X_i cresce em uma unidade, mantendo-se todos os demais x 's fixos, presentes no modelo. Alguns pesquisadores freqüentemente calculam o coeficiente $[(e^{\beta_i}) - 1]$ com o objetivo de obter o percentual de mudança na probabilidade de ocorrer um evento.

Se os valores de $[(e^{\beta_i}) - 1]$ forem positivos, haverá a seguinte implicação: as chances de o evento ocorrer aumentam; se for negativo, significa o oposto. Ou seja, para β_i positivo existe uma correlação positiva entre as duas variáveis. Ou seja: as probabilidades de um evento ocorrer, dado que o outro ocorreu, são positivas; os valores negativos, por outro lado, apresentam a relação inversa (Sincich, 1996). A regressão logística pode testar ainda a hipótese do valor ser zero, situação na qual a rejeição da hipótese significa que existe relacionamento entre as duas variáveis, sendo que o sinal da expressão discutida anteriormente indicará o sentido. Entretanto, se a solução falhar em rejeitar a hipótese de que os coeficientes são zero, as variáveis serão consideradas independentes. A regressão logística utiliza o teste estatístico *Wald*.

6.2.7

Analise da variância - ANOVA

A análise da variância é utilizada para comparar as médias de grupos de respondentes. No caso específico desta tese, ela será utilizada para comparar benefícios alcançados, concordância com relação aos processos internos e conjuntos. O objetivo é utilizar os dados coletados na amostra para fazer inferências com relação à população.

A forma para determinar se existe diferença entre as médias das populações μ_1 e μ_2 é examinar a dispersão (ou variância) entre as médias amostrais y_1 e y_2 , comparando-a com a variabilidade dentro das amostras. Quanto maiores as variações, maiores serão as evidências de que existe diferença nas médias das duas populações μ_1 e μ_2 . Para realizar o teste, a variação dentro da amostra e a variação entre amostras devem ser quantificadas. Na seqüência, o teste é feito utilizando-se a distribuição F com graus de liberdade $\mathbf{v}_1=1$ e $\mathbf{v}_2 = n_1 + n_2 - 2$. Tem-se então uma MST (variação entre as amostras) e uma MSE (variação dentro da amostra) (Sincich, 1996).

$$MSE = \frac{\sum_{i=1}^{n_1} \left(y_{i1} - \bar{y}_1 \right)^2 + \sum_{i=1}^{n_2} \left(y_{i2} - \bar{y}_2 \right)^2}{n_1 + n_2 - 2} \quad \text{Equação 6.1}$$

onde

y_{i1} é a *i*ésima observação na amostra 1

y_{i2} é a *i*ésima observação na amostra 2.

\bar{y}_1 é a média da amostra 1

\bar{y}_2 é a média da amostra 2

$n_1 + n_2$ é o total da amostra

$$MST = \frac{n_1 \left(\bar{y}_1 - \bar{y} \right)^2 + n_2 \left(\bar{y}_2 - \bar{y} \right)^2}{2 - 1} \quad \text{Equação 6.2}$$

onde

\bar{y}_1 é a média da amostra 1

\bar{y}_2 é a média da amostra 2

\bar{y} é a média para a amostra completa (1+2)

tem-se então

$$F = \frac{MST}{MSE} \quad \text{Equação 6.3}$$

Com o cálculo do F, compara-se o F calculado com o F tabelado. Ou seja, se o F for suficientemente grande significa que existe diferença entre as médias das populações; caso contrário, ela não existe. A hipótese nula **H₀**: *não existe diferença entre as médias populacionais*. Ou seja, se a significância for baixa (F maior que o tabelado para um determinado nível de significância) rejeita-se a hipótese nula, pois existe diferença entre as médias das duas amostras.

De acordo com Hair *et al* (1998), recomenda-se que os grupos tenham no mínimo vinte observações. A pré-condição inicial para ANOVA é que a variável

dependente deve ser normalmente distribuída, com igual variância em cada um dos grupos. Entretanto, o autor afirma que há evidências⁵ de que o teste F é robusto o suficiente para suportar a hipótese sem que tais pré-condições sejam cumpridas, exceto em casos muito extremos. No SPSS, o teste de variância entre grupos é feito utilizando-se o teste de Levene; já o teste de normalidade utiliza-se do histograma, *boxplot* e, principalmente, do teste Kolmogorov-Smirnov.

6.2.8

Regressão linear simples

A regressão linear simples é um procedimento utilizado para verificar o comportamento de uma variável em relação a outra. Em outras palavras, tenta-se determinar se uma variável serve como elemento de predição de outra. Nesse caso, uma variável é chamada de “variável dependente”, passando a outra variável a ser denominada “variável independente”. Para o cálculo do coeficiente, utiliza-se o método dos mínimos quadrados. A Equação 6.1 representa a fórmula geral da regressão linear simples.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \text{Equação 6.1}$$

onde

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \quad \text{Equação 6.2}$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} \quad \text{Equação 6.3}$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad \text{Equação 6.4}$$

Para realizar a regressão linear simples, três pré-condições devem ser respeitadas: normalidade, linearidade e homocedasticidade das variáveis.

⁵ Meyers, J.L. (1975) *Fundamentals of Experimental Design*, Bolston, Allyn & Bacon. Winer, B.J. (1962) *Statistical principles in experimental design*, New York, Mc.Graw-Hill.

Normalidade: essa pré-condição requer que as variáveis envolvidas na regressão tenham comportamento normal. Os testes necessários para se verificar a normalidade de uma variável estão descritos na Seção 6.2.7.

Linearidade: essa pré-condição é na realidade implícita ao teste. A relação entre a variável dependente e independente deve ser linear. A partir daí faz-se um gráfico *scartterplot* e verifica-se a linearidade da relação.

Homocedasticidade: essa pré-condição parte do princípio de que a variância da variável dependente ao longo do eixo da variável independente dever ser aleatória. Ela é verificada após executar-se a regressão plotando os resíduos ao longo do modelo estimado na regressão. Tais resíduos não podem ter qualquer comportamento que os faça dependentes ao longo da variável no eixo x. Caso isso ocorra, será constatada a heterocedasticidade, e a pré-condição não estará satisfeita.

6.3

Quadro Resumo.

No Quadro 6.1 são apresentadas as hipóteses e os respectivos testes estatísticos.

Hipótese	Teste Aplicado
<i>H_p: Os fatores-chaves de sucesso (processos internos e processos conjuntos) e os benefícios alcançados na cadeia de frutas e vegetais no Reino Unido coincidem com aqueles identificados pela literatura</i>	Análise Fatorial
<i>H₁: as empresas que estão investindo <u>mais capital</u> nos relacionamentos com seus compradores intermediários chaves estão obtendo mais benefícios que as empresas que estão investindo pouco.</i>	Teste Qui-quadrado, ANOVA
<i>H₂: as empresas que estão investindo <u>mais tempo gerencial</u> nos relacionamentos com seus compradores intermediários chaves estão obtendo mais benefícios que as empresas que estão investindo pouco</i>	Teste Qui-quadrado, ANOVA e Regressão Logística,
<i>H₃: as empresas que utilizam o conceito de longo prazo em relação aos seus compradores intermediários chaves estão conseguindo benefícios mais consistentes dentro dos seus relacionamentos do que as empresas que utilizam o conceito de curto prazo</i>	Teste Qui-quadrado, ANOVA e Regressão Logística
<i>H₄: as empresas que têm um alto nível de receita vindo dos seus compradores intermediários chaves têm melhor percepção da importância do desenvolvimento de processo conjuntos que os de baixo nível de dependência</i>	Teste Qui-quadrado, ANOVA
<i>H₅: as ações de melhoria dos processos internos estão positivamente relacionados com os benefícios alcançados.</i>	Regressão linear Simples

Quadro 6.1: Relação entre hipótese e testes estatísticos utilizados