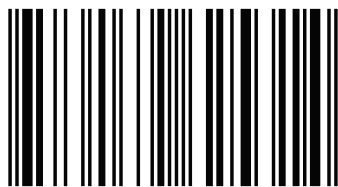


## Um Procedimento Inferencial para Análise Fatorial

A análise fatorial é a denominação atribuída às técnicas estatísticas paramétricas multivariadas utilizadas para estudar o inter-relacionamento entre um conjunto de variáveis observadas. É um processo destinado essencialmente à redução e à sumarização dos dados, tornando-se em vários campos da pesquisa científica uma boa opção para um melhor gerenciamento de informações reais, gerando variáveis remanescentes mais significativas e fáceis de serem trabalhadas. Ainda assim, uma possível limitação da análise fatorial é que não existem testes estatísticos conclusivos regularmente empregados para a sua significância. Conseqüentemente, é difícil saber, seguramente, se os resultados são meramente acidentais, ou realmente refletem algo significativo. Por esse motivo, este trabalho propõe-se a estabelecer um procedimento inferencial para a análise fatorial utilizando-se de técnicas CIS (Computer Intensive Statistics), tais como o bootstrap e o jackknife, que permitam que a análise fatorial saia do terreno puramente descritivo e ladeando a insuficiência da teoria da distribuição de amostragem que se faz sentir em técnicas multivariadas.

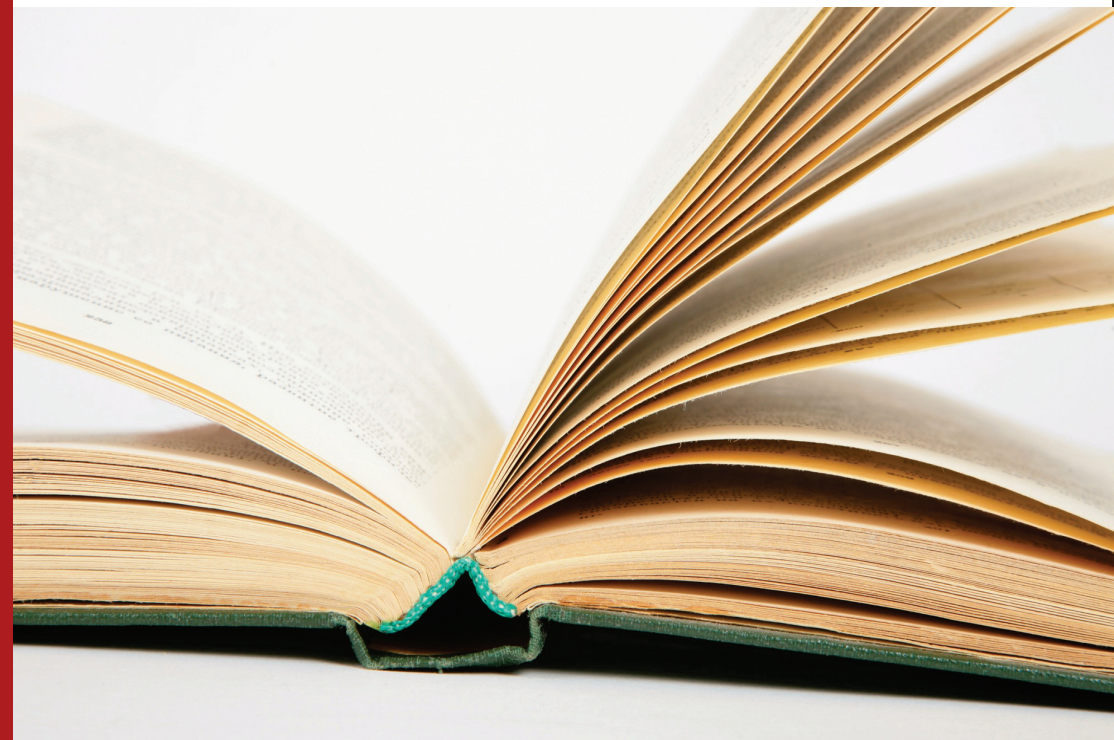
Pós-doutor em Sociologia (Análise Multivariada) pelo Instituto Universitário de Pesquisas do Rio de Janeiro. Doutor em Engenharia Elétrica (Análise Multivariada) pela Pontifícia Universidade Católica do Rio de Janeiro (2006). Professor Associado 3 da Universidade Federal Rural do Rio de Janeiro.



978-613-9-62531-4

Valor-p para Análise Fatorial

Costa



Giovani Glaucio Costa

## Um Procedimento Inferencial para Análise Fatorial

Construção de Intervalos de Confiança e Testes de Hipóteses

 Novas Edições Acadêmicas

**Giovani Glaucio Costa**

**Um Procedimento Inferencial para Análise Fatorial**



**Giovani Glaucio Costa**

**Um Procedimento Inferencial para  
Análise Fatorial**

**Construção de Intervalos de Confiança e Testes de  
Hipóteses**

**Novas Edições Acadêmicas**

## **Imprint**

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: [www.ingimage.com](http://www.ingimage.com)

Publisher:

Novas Edições Acadêmicas

is a trademark of

International Book Market Service Ltd., member of OmniScriptum Publishing Group

17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page

**ISBN: 978-613-9-62531-4**

Zugl. / Aprovado/a pela/pelo: Rio de Janeiro, Universidade Federal Rural do Rio de Janeiro, Tese de Doutorado, 2018

Copyright © Giovani Glaucio Costa

Copyright © 2018 International Book Market Service Ltd., member of OmniScriptum Publishing Group

All rights reserved. Beau Bassin 2018

"O pensamento lógico puro não pode nos proporcionar qualquer conhecimento do mundo empírico; todo conhecimento da realidade parte da experiência e nela termina".

As proposições a que se chega por meios ilógicos exclusivamente são completamente desprovidas de realidade".

-Albert Einstein-

Dedico esta tese com muito amor a minha mãe, Oneida Barreto de Campos Costa, aos meus irmãos e a minha sobrinha Juliana Paula da Costa Lima.

## AGRADECIMENTOS

### Meus agradecimentos

- Ao professor Reinaldo de Castro Souza, orientador da tese e meu grande amigo , pelo apoio, dedicação, atenção, motivação e confiança depositada;
- Ao professor Vitor Hugo de Carvalho Gouvêa, co-orientador da tese, pela inestimável colaboração no aprimoramento das metodologias e conclusões realizadas;
- Ao professor Basílio de Bragança Pereira pelas sugestões competentes dadas ao aprimoramento da qualidade metodológica desta tese de doutorado e pela atenção dispensada;
- Ao Departamento de Pós-graduação em Engenharia Elétrica da PUC - Rio, por ter facilitado a execução deste trabalho e disseminado os conhecimentos necessários à execução do presente esforço intelectual.



## AGRADECIMENTOS ESPECIAIS

### Meus agradecimentos especiais

- Ao Luiz Flávio Autran Monteiro Gomes, grande amigo e orientador do mestrado, pelo grande incentivo que vem dando para que eu avance em meus conhecimentos científicos;
- Ao Wilson Lins Morgado, grande amigo, e colaborador na formulação dos programas SAS apresentados nesta tese de doutorado;
- À Maria Alcina Portes, Márcia Guerra e Ana Maria Pascoal, secretárias do Departamento de Pós-Graduação em Engenharia Elétrica, pela dedicação, atenção, carinho e amizade depositados em mim nestes quatro anos de doutorado.

## Resumo

A análise fatorial é a denominação atribuída às técnicas estatísticas paramétricas multivariadas utilizadas para estudar o inter-relacionamento entre um conjunto de variáveis observadas. É um processo destinado essencialmente à redução e à sumarização dos dados, tornando-se em vários campos da pesquisa científica uma boa opção para um melhor gerenciamento de informações reais, gerando variáveis remanescentes mais significativas e fáceis de serem trabalhadas. Ainda assim, uma possível limitação da análise fatorial é que não existem testes estatísticos conclusivos ou satisfatoriamente eficazes e que possam ser regularmente empregados, portanto, para a sua significância. Conseqüentemente, é difícil saber se os resultados são meramente acidentais, ou realmente refletem algo significativo. Por esse motivo, esta tese de doutorado visa estabelecer um procedimento inferencial para a análise fatorial utilizando-se de técnicas CIS (Computer Intensive Statistics), tais como o *bootstrap* e o *jackknife*, que permitam que a análise fatorial saia do terreno puramente descritivo e ladeando a insuficiência da teoria da distribuição de amostragem que se faz sentir em técnicas multivariadas.

## Abstract

Factor analysis is the denomination attributed to the multivariate parametric statistical techniques used to study the inter-relationship between a set of observed variables. It is a process essentially intended to reduce and summarize data, thus becoming a good option for a better management of real information, generating remainder variables that are more significant and easier to work with, in various fields of scientific research. However, a possible limitation of factor analysis is that there are no conclusive statistical tests regularly employed in testing the hypotheses. Consequently, it is difficult to know if the results are merely accidents, or indeed, reflect something of significance. For this reason, this study intends to establish an inferential procedure for factor analysis, using CIS (Computer Intensive Statistics) techniques, such as the *bootstrap* and *jackknife*, which allow factor analysis to pass out of the purely descriptive, solving the problem of the insufficiency of sample distribution theory as seen in multivariate techniques.

# SUMÁRIO

INTRODUÇÃO.....	9
CAPÍTULO 1:METODOLOGIA DA PESQUISA .....	13
CAPÍTULO 2:REVISÃO DE LITERATURA.....	16
CAPÍTULO 3:ANÁLISE FATORIAL.....	22
CAPÍTULO 4:REAMOSTRAGEM .....	43
CAPÍTULO 5:TESTE INFERÊNCIAL PARA ANÁLISE FATORIAL.....	54
CAPÍTULO 6:TESTE DE NORMALIDADE DAS CARGAS FATORIAIS.....	71
CAPÍTULO 7:TESTE DE CONVERGÊNCIA DOS RESULTADOS <i>BOOTSTRAP</i> .....	88
CAPÍTULO 8:ESTUDOS DE CASOS:APLICAÇÕES DO TIAF.....	99
CAPÍTULO 9:COMPARAÇÃO DO TIAF COM O MÉTODO TRADICIONAL DE HAIR E ANDERSON .....	157
CONCLUSÃO .....	167
BIBLIOGRAFIA E REFERÊNCIAS BIBLIOGRÁFICAS .....	171
ANEXO 1-OS PROGRAMAS <i>BOOTSTRAP</i> E <i>JACKKNIFE</i> EM R 2.1.1.....	174
ANEXO 2-O PROGRAMA <i>BOOTSTRAP</i> EM SAS V.8 .....	183
ANEXO 3-O PROGRAMA <i>JACKKNIFE</i> EM SAS V.8 .....	204

## INTRODUÇÃO

A presente pesquisa tem por objetivo estabelecer um critério para significância de cargas fatoriais com metodologias *CIS* (Computer Intensive Statistics).

Mais precisamente, trata-se do desenvolvimento de um procedimento que consiste na modelagem estatística inferencial para os resultados de um estudo por análise fatorial, envolvendo a matriz principal e a matriz varimax de cargas fatoriais de componentes principais geradas pela análise e utilizando-se dos procedimentos *jackknife* e *bootstrap*, isto é, um critério para significância de cargas fatoriais com metodologias *CIS*.

A análise fatorial é utilizada para identificar constructos ou dimensões básicas nos dados e reduzir o número de variáveis por meio da eliminação da redundância.

A análise fatorial carece de testes estatísticos que possam ser regularmente empregáveis para a significância de um determinado resultado encontrado. Portanto, é difícil saber se os resultados são meramente acidentais ou refletem realmente algo significativo.

A presente tese de doutorado pode trazer uma contribuição suficientemente original a respeito do tema pesquisado. Ela representa um progresso para a aplicação da análise fatorial em situações reais e práticas de pesquisa, uma vez que, propõe análises multivariadas com estudos além dos exploratórios, ou seja, o conhecimento do erro padrão e do modelo de probabilidade permitem construir intervalos de confiança e sair, conseqüentemente, do terreno puramente descritivo.

Espera-se que com o sucesso na especificação de um procedimento estatístico para inferência dos resultados das análises fatoriais por

componentes principais, o pesquisador possa ter dado sua contribuição no aprimoramento do uso da análise fatorial por Componentes Principais nos mais diversos campos de aplicação.

Estão em curso inúmeras investigações sobre as teorias *bootstrap* e *jackknife*, nomeadamente no que toca à validade assintótica e à aplicação na construção de intervalos de confiança. A importância do tema do estudo ora proposto pode avaliar-se pela quantidade de artigos que, nos últimos dez anos, estão aparecendo em todas as revistas da especialidade.

Esta tese visa motivar a discussão e o embasamento necessário para que se crie uma teoria sólida sobre a questão da indução estatística, não somente na técnica específica da análise fatorial, como também em outras metodologias multivariadas, tais como, análise discriminante, análise de conglomerados, análise de correlação canônica e em muitas outras metodologias multivariadas.

O capítulo 1 trata do detalhamento das metodologias aplicadas para se chegar ao objetivo deste trabalho.

No capítulo 2, é realizada a fundamentação teórica do problema proposto, a revisão de literatura, para um teste de originalidade do tema, bem como para avaliação de sua relevância científica social. A revisão de literatura é útil também para se evidenciar que o problema tratado não é fruto de interesse e esforços somente do autor deste trabalho. Ele pode ser objeto de preocupação de outros estudiosos da área da análise estatística de dados.

O capítulo 3 embasa de forma sintética o leitor com a técnica da análise fatorial exploratória, apresentando um estudo de caso real à

aplicação da análise fatorial. Este capítulo é importante para aqueles leitores que não estão familiarizados com este método multivariado.

O capítulo 4 discute de forma resumida o que é reamostragem e sua importância dentro da questão estatístico-inferencial. Detalham as técnicas de reamostragens jackknife e bootstrap, fornecendo os procedimentos para a obtenção das respectivas amostras.

O capítulo 5 estabelece o Método TIAF (Teste Inferencial para Análise Fatorial) em detalhes: suas metodologias, os passos de sua construção, vantagens e desvantagens e sua importância dentro da questão indutiva na metodologia da análise fatorial.

O capítulo 6 realiza testes de significância de aderência para a normalidade das cargas fatoriais, para verificar se este importante pressuposto é garantido para se construir intervalos de confiança e calcular-se valores-p baseados na premissa gaussiana.

O capítulo 7 realiza análise da convergência, ensaiando diferentes valores para o número de reamostragens (B) no método bootstrap para verificar em que estágio se produz indicadores de qualidade das distribuições amostrais em seu ponto ótimo.

O capítulo 8 aplica o “Método TIAF”, estabelecido no capítulo 5, a casos reais de pesquisas (bases de dados) procurando evidenciar as potencialidades do método.

O capítulo 9 compara o “Método Tradicional de Hair e Anderson”, baseado na teoria assintótica, com o “Método TIAF”, baseado em técnicas de reamostragens.

A conclusão comporta idéias, tais como: uma síntese interpretativa dos principais argumentos do desenvolvimento do trabalho, o ponto de

vista do pesquisador, até aqui não revelado, os aspectos do tema discutido que deveriam ser mais aprofundados em pesquisas anteriores como a expectativa de trabalhos futuros. A conclusão dará, então, uma visão sintética do assunto.

# CAPÍTULO 1

## METODOLOGIA DA PESQUISA

Como comentado na introdução, esta tese de doutorado objetiva propor um processo inferencial para análise fatorial por componentes principais.

A idéia é utilizar técnicas *CIS* (Computer *Intensive Statistics*), que cogitam o modelo de densidade de probabilidade que explica o comportamento aleatório da estatística observada e seus parâmetros característicos.

As técnicas *CIS* dispõem-se principalmente de dois métodos que serão empregados no estudo referido: o *bootstrap* e o *jackknife*.

Em uma análise fatorial, os resultados mais importantes são as cargas fatoriais, os escores fatoriais e as porcentagens de variância explicada. As cargas fatoriais, ou seja, as correlações entre os fatores e as variáveis, são utilizadas para interpretar os fatores. Um analista pode tomar uma ou duas variáveis que têm alta carga sobre um fator para representá-lo em uma subsequente análise. Este procedimento será utilizado nesta tese para inferência estatística envolvendo a análise fatorial.

Esta tese especifica também a variância, o viés e o erro médio quadrático (EMQ) da variável aleatória ‘carga fatorial’ das variáveis dos fatores determinantes de uma análise fatorial (os fatores 1 e 2, por opção neste trabalho) para os procedimentos *bootstrap* e *jackknife*. Com estes resultados, pode-se obter um procedimento computacional, um algoritmo,



para a construção de intervalos de confiança e testes de hipóteses para as estimativas obtidas.

Para aplicar as doutrinas expostas, considerar-se-ão bases de dados diversas, onde é relevante reduzir o número de variáveis a um conjunto mais fácil de manipular. As estatísticas em questão são as ‘cargas fatoriais’ das componentes principais 1 e 2 da matriz rodada de fatores pelo método varimax.

A interpretação dos fatores baseia-se nas cargas fatoriais, que são as correlações entre os fatores e as variáveis originais. As cargas fatoriais oferecem, assim, uma indicação de quais variáveis originais estão correlacionadas com cada fator, bem como a extensão dessa correlação. Essa informação pode ser então utilizada para identificar e rotular subjetivamente os fatores não observáveis e esse resultado caracteriza uma determinada rodada de análise fatorial para uma dada amostra selecionada. Ao interpretar fatores, é preciso tomar a decisão sobre quais cargas fatoriais valem a pena considerar. A discussão passa por questões relativas à significância estatística e prática, bem como o número de variáveis que afetam a interpretação de cargas fatoriais.

Os fatores que explicam a maior parte da variação total do problema são nomeados e classificados com grande frequência em função das variáveis originais com maior carga fatorial, mas sempre a nível descritivo, já que o desconhecimento da distribuição por amostragem torna inviável fazer acompanhar as estimativas do respectivo erro padrão, para não falar na construção de intervalo de confiança ou na realização de testes de significância.

A opção de se usar as metodologias *CIS* surge quando não se conhece o enviesamento ou o desvio padrão teórico das estimativas ou quando o estudo da distribuição por amostragem assume caráter por demais complexo que dificulta a dedução do erro padrão e do não enviesamento do estimador considerado, o que acontece para grande parte das estatísticas paramétricas da análise multivariada, o que inclui o caso da análise fatorial.

Nestes casos, com a aplicação do *bootstrap* e *jackknife* é possível obter, de forma expedita, através das computações maciças, estimativas do desvio padrão e enviesamento da estatística em causa em substituição análise teórica. Com o *bootstrap*, por exemplo, é possível determinar a distribuição por amostragem da estatística e seus parâmetros característicos.

Os métodos *jackknife* e *bootstrap* permitem ladear a insuficiência da teoria da amostragem que se faz sentir no campo da análise multivariada.

# CAPÍTULO 2

## REVISÃO DE LITERATURA

Apesar das tentativas de construção de vários testes para a confiabilidade estatística da análise fatorial, segundo Stewart (1981), nenhum procedimento totalmente comprovado está disponível. A ausência de testes adequados provém da dificuldade de especificação de parâmetros teóricos dos modelos de distribuição por amostragem das estatísticas envolvidas na técnica da análise fatorial. Por isso, é difícil saber se os resultados são meramente acidentais, ou realmente refletem algo significativo, como mesmo comentaram Aaker-Kumar-Day (2001).

Em uma análise fatorial por componentes principais, existem decisões que devem ser tomadas com base nos pares de autovalores-autovetores, estimados na amostra. Esses autovalores e autovetores são diferentes dos respectivos valores populacionais devido às variações amostrais. Derivações a respeito das distribuições amostrais dos autovalores e autovetores são apresentados por Anderson (1963). Esta inferência permite que somente se considere na análise fatorial os fatores cujos autovalores são significativos. Os resultados indutivos propostos por Anderson (1963) só valem para grandes amostras. Um problema aqui é que nas amostras grandes (tamanhos maiores que 200) muitos fatores têm probabilidade de serem significantes, ainda que, sob um ponto de vista prático, muitos deles sejam responsáveis por apenas uma pequena parcela da variância total. O referido teste tem se mostrado pouco conclusivo.

Hair, Anderson, Tatham e Black (2005), ou “Método Hair e Anderson”, para simplificação, apresentam um critério para significância de cargas fatoriais. Segundo estes autores, ao determinar um nível de significância à interpretação de cargas, uma abordagem semelhante à determinação da significância estatística de coeficiente de correlação poderia ser usada. Entretanto, conforme os autores mesmo relatam e confirmado por trabalhos realizados por Cliff e Hamburges (1967), as cargas fatoriais têm substancialmente erros-padrão maiores do que as correlações normais. Assim, as cargas fatoriais devem ser avaliadas em níveis consideravelmente mais restritos. Hair e Anderson (2005) ensinam que o pesquisador pode empregar o conceito de poder estatístico para especificar cargas fatoriais consideravelmente significantes para diferentes tipos de amostras. Especificado um poder estatístico de 80%, um nível de significância de 5% e os erros-padrão estimados pela análise fatorial, a Tabela 1 contém os tamanhos de amostras necessários para cada valor de carga fatorial ser considerada significativa. Hair e Anderson (2005), exemplificam: em uma amostra de 100 respondentes, as cargas fatoriais de 0,55 ou mais são significantes. Por outro lado, segundo este critério, em uma amostra de tamanho 50, somente podem ser consideradas significantes as cargas fatoriais maiores ou iguais a 0,75.

**Tabela 1**

**Orientações para identificação de cargas fatoriais significantes com base no tamanho da amostra**

Carga fatorial	Tamanho necessário da amostra para significância
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

Fonte: cálculos feitos com SOLO Power Analysis, BMDP Statistical Software, Inc., 1993.

Nota: a significância é baseada em um nível de significância de 0,05( $\alpha$ ), um nível de poder de 80% e erros padrão, os quais se pressupõe que sejam o dobro dos de coeficientes de correlação convencionais.

Uma desvantagem desta abordagem relatada também por Hair e Anderson (2005) é que o número de variáveis analisadas e o fator específico em exame não são considerados. Foi mostrado que quando o pesquisador se move do primeiro fator para fatores posteriores, o nível aceitável para uma carga fatorial seja julgado significante deve aumentar. Segundo Kaiser (1970), o fato de que a variância única e a variância do erro começam a surgir em fatores posteriores significa que algum ajuste para cima no nível de significância deve ser incluído. O número de variáveis em análise também é importante na decisão sobre quais cargas são significantes. À medida que o número de variáveis aumenta, o nível aceitável para considerar uma carga significante diminui. O ajuste para o número de variáveis é cada vez mais importante quando se move do primeiro fator extraído para fatores posteriores.

Hair e Anderson (2005) dão as seguintes orientações para o critério de significância exposto por eles:

1. Quanto maior o tamanho da amostra, menor o valor da carga fatorial para ser considerada significativa;
2. Quanto maior o número de variáveis a serem incluídas no problema, menor o valor das cargas fatoriais para serem consideradas significativas;
3. Quanto maior o número de fatores, maior o valor da carga em fatores posteriores a serem consideradas significativas para interpretação.

Contudo, esta abordagem não está ajustada para levar em consideração as observações mencionadas por Hair e Anderson (2005), uma vez que o critério proposto na Tabela 1 é geral, independe do número de variáveis do problema ou do fator em questão. Portanto, este teste de significância é pouco conclusivo.

Ludovic Lebart, Alain Morineau e Merie Piron (1998), pesquisadores franceses, introduziram um conceito de valor do teste a fim de julgar se cargas fatoriais de uma componente principal é significativamente diferente da média geral. Se o tamanho da amostra é grande se compara o valor do teste a uma variável normal padrão utilizando o Teorema Central do Limite. O problema está em afirmar e assumir simplesmente, sem o devido teste de aderência e gráficos de probabilidades, que as cargas fatoriais seguem regularmente o modelo gaussiano.

Já G.Saporta (1996), outro estatístico francês, num estudo paralelo, assume que as cargas fatoriais(C) têm, quando o tamanho da amostra(n) é suficientemente grande, pelo Teorema Central do Limite, distribuição normal com o seguinte erro-padrão:

$$EP = C^2 n. [(N - 1) / (N- n)],$$

Onde: N é o tamanho da população de onde os n elementos foram selecionados aleatoriamente.

Com base na distribuição amostral definida acima, é obtido uma estatística de teste e considera-se, então, como significativa as cargas fatoriais de uma componente principal que tem um valor superior em valor absoluto a 2 (ao nível de significância de 5%). Esta prática permite um resultado rápido quanto à significância estatística de cargas fatoriais de um fator em foco. Mas este procedimento inferencial para as cargas fatoriais incorre no mesmo problema do teste de significância de Ludovic Lebart, Alain Morineau e Merie Piron: assumir simplesmente, sem o devido teste de aderência e gráficos de probabilidades, que as cargas fatoriais seguem regularmente o modelo gaussiano.

Insuficientes referências conclusivas podem ser feitas para delimitar o cenário que cerca o uso de procedimentos indutivos para a análise fatorial. O estudo de procedimentos inferenciais para a matriz rodada da análise fatorial ainda não foi satisfatoriamente desenvolvido por cientistas da área, o que dificulta o seu uso regular.

Dos estudos acima, pode-se inferir que o problema da especificação de um critério de significância para as cargas fatoriais é objeto de preocupação dos estudiosos da área de análise fatorial e análise multivariada em geral.

A revisão de literatura indica que o assunto proposto para pesquisa necessita melhores definições, melhor precisão e clareza do que já existe sobre o mesmo.

Diante do que foi exposto em parágrafos acima, fica clara a necessidade de acrescentar novos conhecimentos (avanços) à questão da inferência estatística para a significância da matriz principal de uma análise fatorial. Portanto, apresentam-se, sem mais considerações, os próximos capítulos desta tese.



# CAPÍTULO 3

## ANÁLISE FATORIAL

O desenvolvimento tecnológico oriundo das descobertas científicas tem alavancado o próprio desenvolvimento científico, ampliando em várias ordens de grandeza a capacidade de obter informações de acontecimentos e fenômenos que estão sendo analisados. Uma grande massa de informação deve ser processada antes de ser transformada em conhecimento. Portanto, cada vez mais se está necessitando de ferramentas estatísticas que apresentem uma visão mais global do fenômeno que aquela possível numa abordagem univariada. A denominação “**Análise Multivariada**” *corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos.*

A análise fatorial é a denominação atribuída às técnicas estatísticas paramétricas multivariadas utilizadas para estudar o inter-relacionamento entre um conjunto de variáveis observadas. Diferentemente da regressão múltipla, em que uma variável é, explicitamente, considerada critério e as demais prognóstico, na análise fatorial todas as variáveis são consideradas simultaneamente.

A análise fatorial é um tipo de processo destinado essencialmente à redução e à sumarização dos dados. Em pesquisas, pode haver um grande número de variáveis, a maioria delas correlacionadas, redundantes, e que devem ser reduzidas a um nível gerenciável. Estudam-se as relações entre

conjuntos de muitas variáveis inter-relacionadas representando-as em termos de alguns fatores fundamentais.

Muitas investigações realizadas na prática envolvem uma quantidade enorme de atributos ou variáveis. Nesse momento, a análise poderia invisibilizar-se. Além disso, muitos dos atributos são redundantes, mensurando, na verdade, o mesmo constructo. Para determinar quais são redundantes, e o que eles medem, o analista pode lançar mão da análise fatorial (factor *analysis*). O resultado seria um conjunto de novas variáveis (ou fatores) criado por meio da combinação de conjuntos de atributos ou variável original.

A premissa mais importante da análise fatorial é a de que existem fatores que dão suporte às variáveis e que estes fatores representam completa e adequadamente as variáveis originais. Em termos práticos, essa premissa significa que a lista de variáveis deve ser completa, ou seja, cada fator entre elas é medido pelo menos uma vez e idealmente, várias vezes, de diferentes perspectivas. Se, por alguma razão, a lista inicial de variáveis for deficiente desde o início, será preciso uma grande dose de sorte para que algo de útil e lógico possa emergir do processo.

### **3.1. ETAPAS DE REALIZAÇÃO DA ANÁLISE FATORIAL**

Há, basicamente, quatro passos na condução da análise fatorial: entrada de dados, cálculo das correlações entre as variáveis, extração inicial dos fatores e a rotação da matriz.

***Entrada de Dados(BASE)***: os dados de entrada da análise fatorial geralmente tomam a forma de um conjunto de valores de variáveis para cada objeto ou indivíduo na amostra. Na verdade, qualquer matriz cujos

componentes ofereçam uma medida de similaridade entre variáveis, pode ser passível de análise fatorial. A medida de similaridade não precisa ser uma correlação, embora, geralmente, o seja.

**Cálculo das Correlações:** para o cálculo da matriz de correlações podem ser utilizadas duas abordagens: análise fatorial R e a análise fatorial Q. Na análise fatorial R, as correlações são calculadas entre variáveis e na análise fatorial Q entre casos. Nos procedimentos da análise fatorial R, procura-se agrupar as diferentes variáveis em alguns fatores específicos. Casos podem ser pessoas, produtos, lojas ou qualquer outros elementos de tal forma que tratados pela análise fatorial possam ser agrupados em fatores específicos. Este procedimento poderá ser usado para encontrar pessoas, produtos, ou lojas, por exemplos, que pertençam a diferentes grupos ou segmentos ou para reagrupar pessoas em grupos específicos. Todas as variáveis devem ser correlacionadas, duas a duas.

**Extração de Fatores Iniciais:** há diferentes métodos de extração de fatores da matriz de correlações. De forma geral, estes métodos são numericamente complexos para serem aqui abordados. Um desses métodos, o das componentes principais, é intensamente utilizado na prática e será o processo adotado para extração dos fatores das análises fatoriais empregadas nesta tese.

O objetivo da extração de fatores é encontrar um conjunto de fatores que formem uma combinação linear das variáveis originais ou da matriz de correlações. Desta forma, se as variáveis  $X_1, X_2, X_3, \dots, X_n$  são altamente correlacionadas entre si, elas serão combinadas para formar um fator, e assim, sucessivamente, com todas as demais variáveis da matriz de correlação. Uma combinação linear entre variáveis pode ser assim definida:

$$F_j = C_{1j}X_1 + C_{2j}X_2 + \dots + C_{nj}X_n$$

Onde  $F_j$  é uma combinação linear das variáveis  $X_1, X_2, \dots, X_n$  e é denominado de componente principal. O método das componentes principais para a análise fatorial envolve a procura de um conjunto de valores de  $C_{ij}$  nesta equação que forme uma combinação linear que explique mais a variância da matriz de correlação que qualquer outro conjunto de valores para  $C_{ij}$ . É chamado de primeiro fator principal. A seguir, a variância explicada pelo primeiro fator é subtraída da matriz de correlações original, resultando-se assim, as matrizes residuais, adotando-se o mesmo procedimento anterior, obtém-se o segundo fator principal, e assim, sucessivamente todos os fatores principais, até que uma variância muito pequena permaneça sem explicação. A natureza deste procedimento permite extrair fatores que não são correlacionados ou que tenham correlação muito pequenas uns com outros. Neste caso, os fatores são chamados de ortogonais.

**Rotação:** os fatores extraídos conforme o item anterior apresentam, freqüentemente, muitas dificuldades para serem interpretados. Para ajudar na interpretação desses fatores a solução inicial deve ser rotada. Há duas formas de procedimento da rotação da matriz: a rotação ortogonal (rotação varimax), que mantém os fatores não correlacionados e a rotação oblíqua, que torna os fatores correlacionados entre si. A idéia básica do giro de fatores é identificar alguns fatores que possuam variáveis que tenham alta correlação e outros com variáveis que possuam baixa correlação. Os cálculos para realização destas rotações são muitos complexos e fugirá do

escopo desta pesquisa. A realização da rotação de matrizes demanda uma quantidade tão grande de cálculos que a utilização da análise fatorial em pesquisas só passou a ser viável com o advento dos computadores e dos programas de análises que possibilitaram a realização dos cálculos de forma rápida, precisa e econômica.

Os resultados mais importantes são: quais e quantos são os fatores extraídos pela análise? e que variáveis originais fazem parte de cada fator extraído pela análise? . A obtenção das respostas a estas perguntas se fará pela interpretação das cargas da matriz rotada (rotação varimax). Além disso, os autovalores associados a cada fator e a informação da porcentagem da variação total explicada por cada fator gerado proporcionam verificar o grau de importância de cada fator na explicação do problema proposto.

### **3.2.O MODELO DE ANÁLISE FATORIAL**

Matematicamente, a análise fatorial é semelhante à análise de regressão múltipla, pelo fato de cada variável ser expressa como uma combinação linear de fatores subjacentes. A quantidade de variância que cada variável compartilha com todas as outras variáveis incluídas na análise é chamada *comunalidade*, isto é, em que grau uma variável está correlacionada com as outras. A covariância entre as variáveis é descrita em termos de um pequeno número de fatores, mais uma componente aleatória ou erro de variância. Esses fatores são não observados abertamente. Se as variáveis são padronizadas, o modelo original fatorial pode ser representado como:

$$X_i = C'_{i1}F_1 + C'_{i2}F_2 + \dots + C'_{in}F_n + \varepsilon_i$$

O modelo fica então,

$$\mathbf{X}_i = \mathbf{C}'_{ij}\mathbf{F}_j + \varepsilon_i$$

Onde:

$X_i$  = i-ésima variável padronizada

$C'_{ij}$  = cargas dos fatores, covariâncias

$F_j = F_1, F_2, F_3, \dots$  = escores padronizados dos fatores

$\varepsilon_i$  = erros de variância

Matricialmente tem-se :

$$\tilde{\mathbf{X}} = \tilde{\mathbf{C}}' \tilde{\mathbf{F}} + \tilde{\boldsymbol{\varepsilon}}$$

com  $\mathbf{F} = (F_1, \dots, F_n)$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_i)$  e

$$\tilde{\mathbf{C}}' = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & \dots & C_{1n} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & \dots & C_{2n} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & \dots & C_{3n} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & \dots & C_{4n} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & \dots & C_{5n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ C_{n1} & C_{n2} & C_{n3} & C_{n4} & C_{n5} & \dots & C_{nn} \end{pmatrix}$$

No modelo original de análise fatorial ortogonal, fazem-se as seguintes suposições sobre  $F$ :  $E(F_j) = 0$  e  $\text{Var}(F_j) = 1$ ,  $i = 1, \dots, n$ . Além disso, num modelo ortogonal, admite-se

a)  $Cov(F) = I_n$ , onde  $I_n$  é a matriz identidade de ordem  $n$

b)  $E(\varepsilon) = 0$

c)  $Cov(\varepsilon) = \Psi = \text{diag}(\Psi_1, \dots, \Psi_t)$

d)  $Cov(F, \varepsilon) = 0$

A partir dessas suposições é possível analisar o modelo proposto e interpretar suas componentes.

No modelo fatorial, especificado acima, assim como no modelo de regressão, existe um pequeno conjunto de variáveis independentes, aqui chamadas de fatores, que hipoteticamente explicam, ou são a causa, da variável dependente. Os coeficientes de regressão, aqui nomeados de cargas fatoriais, fazem a ligação entre os fatores e as variáveis e são utilizadas para ajudar na interpretação dos fatores. Nesse contexto, as cargas fatoriais são as correlações entre os fatores e as variáveis. O termo erro, tanto no modelo de regressão como no fatorial, absorve os erros de mensuração e a variação na variável dependente que não é explicada ou causada pelos fatores. A fonte da variação inexplicada na variável dependente é um conceito importante tanto na análise fatorial (porcentagem de variância explicada e comunalidade), como na análise de regressão ( $R^2$ ).

### 3.3. AS ESTATÍSTICAS-CHAVES ASSOCIADAS À ANÁLISE FATORIAL

As estatísticas existentes e utilizadas regularmente no processo de análise fatorial são:

- **Teste de esfericidade de Bartlett** : uma estatística de teste usada para examinar a hipótese de que as variáveis não sejam correlacionadas na população. Em outras palavras, a matriz de correlação da população é uma matriz identidade; cada variável se

correlaciona perfeitamente com ela própria ( $r=1$ ), mas não apresenta correlação com as outras variáveis ( $r=0$ ).

- **Matriz de correlação:** o triângulo inferior da matriz exibe as correlações simples,  $r$ , entre todos os pares possíveis de variáveis incluídas na análise. Os elementos da diagonal, que são todos iguais a 1, em geral são omitidos.
- **Comunalidade:** porção da variância que uma variável compartilha com todas as outras variáveis consideradas. É também a proporção de variância explicada pelos fatores comuns.
- **Autovalor:** representa a variância total explicada por cada fator.
- **Cargas fatoriais:** correlação simples entre as variáveis e os fatores.
- **Gráfico das cargas dos fatores:** gráfico das variáveis originais utilizando as cargas fatoriais como ordenadas.
- **Matriz de fatores ou matriz principal:** contém as cargas fatoriais de todas as variáveis em todos os fatores extraídos.
- **Escore fatoriais:** escores compostos estimados para cada entrevistado nos fatores derivados.
- **Medida de adequabilidade da amostra de Kaiser-Meyer-Olkin (KMO):** índice usado para avaliar a adequabilidade da análise fatorial. Valores altos (entre 0,5 e 1,0) indicam que a análise fatorial é apropriada. Valores abaixo de 0,5 indicam que a análise fatorial pode ser inadequada.
- **Porcentagem de variância:** porcentagem da variância total atribuída a cada fator.



- **Resíduos:** diferenças entre as correlações observadas, dadas na matriz de correlação de entrada (input) e as correlações reproduzidas, conforme estimadas pela matriz de fatores.
- **Scree plot:** gráfico dos autovalores *versus* número de fatores por ordem de extração.

### 3.4. CONSIDERAÇÕES GERAIS

Na presente pesquisa, as cargas fatoriais serão consideradas correlações, mas na verdade elas só serão correlações apenas quando:

(1) As variáveis de entrada estão padronizadas (cada variável tem sua média subtraída e é dividida pelo seu desvio-padrão)

(2) Os fatores são perpendiculares ou independentes (ortogonais).

Estas duas condições normalmente estão presentes. Além disso, muitos programas de análise fatorial iniciam-se com o cálculo da matriz de correlação, um processo de padronização das variáveis. Se nenhuma destas condições for atendida, as cargas fatoriais, embora não sejam correlações, podem ainda ser interpretadas como indicadores de associação entre variáveis e os fatores. Quando as variáveis estão padronizadas, os coeficientes fatoriais tornam-se “coeficientes beta” no contexto da Análise de Regressão. Diferentemente da Análise de Regressão, em que as variáveis independentes têm algum grau de correlação, os fatores são independentes, mas as variáveis originais devem ser correlacionadas. Por esse motivo é que uma carga fatorial representa aqui uma correlação, enquanto um coeficiente beta, no contexto da Análise de Regressão não é uma correlação.

Uma análise fatorial envolve a estimação de um grande número de parâmetros e, para que isso seja feito com um mínimo de qualidade, é

necessário um tamanho amostral relativamente grande em comparação ao número de variáveis envolvidas. Há na literatura estatística, uma série de sugestões para a escolha desse tamanho de amostra. Em geral, essas opções baseiam-se na experiência pessoal dos diversos autores que, em alguns casos, sugerem um tamanho amostral da ordem de 20 vezes o número de variáveis envolvidas (Hair, 1998). Reis (1997) e Hair (1998) sugerem que o número de observações deva ser de no mínimo 5 vezes o número de variáveis, além disso, indicam que preferencialmente a análise seja feita com pelo menos 100 observações. Hair (1998) enfatiza que ela não deve ser utilizada em amostras inferiores a 50 observações.

O sucesso de uma análise fatorial está diretamente ligado aos objetivos iniciais do pesquisador; por exemplo, se a intenção é a simples redução do número de variáveis, ela será bem sucedida se for possível determinar um pequeno conjunto de fatores que consiga explicar uma parte considerável da variabilidade do conjunto inicial de variáveis. De qualquer modo, há algumas propriedades que são desejáveis a uma solução de uma análise fatorial:

a) Encontrar um número relativamente pequeno de fatores que possuam um alto grau de explicação da variabilidade original dos dados;

b) Encontrar fatores interpretáveis.

Dentre as razões que explicariam o insucesso de uma análise fatorial, destaca-se:

a) Tamanho insuficiente da amostra: uma amostra pequena pode não conseguir refletir de maneira precisa a estrutura de interdependência dos dados;

b) Variáveis com uma fraca interdependência, por exemplo, considere uma escala composta por itens, onde cada item mede um aspecto diferente do constructo de interesse, nesse caso é possível que uma análise fatorial não consiga identificar fatores com um grau razoável de interpretação. Hair (1998) discute que para o sucesso de uma análise fatorial é necessário que exista um número razoável de correlações superiores (em módulo) a 0,30, caso contrário a estrutura de interdependência será muito tênue para produzir resultados satisfatórios;

c) A estrutura de dependência pode não ser homogênea em toda a amostra. Considere, como ilustração, itens de uma escala que se associam diferentemente (possuem estruturas de dependência diferentes) para homens e mulheres, nesse caso, uma análise fatorial aplicada apenas a um dos sexos pode ser bem sucedida, mas aplicada à amostra total, não. Parece razoável que, no caso de insucesso e quando existirem razões teóricas para isso, se faça uma análise fatorial para cada subgrupo de interesse de uma amostra.

Entretanto, a maior limitação da análise fatorial é o fato de ser um processo muito subjetivo. A determinação do número de fatores, sua interpretação e a seleção da rotação (se um conjunto de fatores não agrada ao analista, a rotação pode prosseguir indefinidamente), tudo isso envolve um julgamento subjetivo.

Uma outra limitação relevante é que não existem testes estatísticos regularmente empregados na análise fatorial. Conseqüentemente é difícil saber se os resultados são meramente acidentais ou realmente refletem algo significativo. Por esse motivo, um procedimento padrão após uma rodada de análise fatorial para uma amostra específica deveria ser a realização de

um procedimento inferencial que pudesse testar a significância das estimativas encontradas. Este é o objetivo desta tese.

Para atingir o objetivo descrito no parágrafo acima, um dos procedimentos padrão utilizados para a análise fatorial deveria ser a seleção de um grande número de amostras aleatórias de uma população alvo e submeter cada uma delas a uma análise fatorial independente. Se os mesmo fatores emergirem em cada análise, pode-se ficar mais confiante de que os resultados não representam apenas um acidente estatístico. Contudo, este critério de validação de resultados da análise fatorial ainda está baseado estritamente na análise descritiva. Seria relevante que se estabelecesse um procedimento de significância da matriz de fatores em função da teoria das probabilidades, da teoria da amostragem e da inferência estatística.

### **3.5. APLICAÇÕES NA INTERNET E EM COMPUTADORES**

Existem programas de computador para implementar a análise fatorial. A estrutura básica e os programas de microcomputador são semelhantes para SPSS, SAS, e BMDP. Nos pacotes SPSS, pode-se utilizar o programa FACTOR não só para análise de componentes principais como para análise fatorial utilizando outros métodos de extração de fatores, como o de máxima verossimilhança. Existem também outros métodos de análise fatorial, com escores fatoriais calculados.

No sistema SAS, o programa PRINCOMP faz análise de componentes principais e calcula escores dessas componentes.

No Minitab pode-se acessar a análise fatorial utilizando-se Multivariate>Factor analysis. Pode-se utilizar a análise de componentes principais ou a máxima verossimilhança para determinar a extração de fatores iniciais.

No caso de se usar a máxima verossimilhança é preciso especificar o número de fatores a serem extraídos. Se não se especifica um número de fatores em uma extração de componentes principais, o programa o fará igual a um número de variáveis no conjunto de dados.

A análise fatorial não está incluída no Excel.

O leitor deve ter notado na apresentação feita da análise fatorial que a atenção foi concentrada na entrada de dados e na interpretação das saídas, e sucintamente em como o processamento é feito. O pesquisador opta por este caminho pelas seguintes razões: a complexidade das formulações e teorizações existentes na análise fatorial é tão extensa que não cabe apresentá-la numa tese de doutorado cuja concentração será à busca de um procedimento inferencial para análise fatorial, utilizando-se de procedimentos *CIS*.

A utilização da análise fatorial na prática da pesquisa é feita essencialmente com o uso de computadores e programas de análises especiais. Não há necessidade de o pesquisador conhecer profundamente suas formulações e teorizações, mas sim, saber como utilizá-los e, principalmente, como interpretar os resultados obtidos.

Discutiram-se, neste capítulo, aspectos essenciais ligados a uma análise fatorial Ortogonal (exploratória). Sugere-se aos interessados na aplicação dessa técnica a leitura de Hair et al. (2005), que descreve a análise fatorial de modo bastante informal. O livro trata, basicamente, da análise fatorial baseada na matriz de correlação (que parece ser a mais utilizada na prática). O livro de Reis (1997) também traz uma interessante introdução à técnica com uma abordagem um pouco mais formal. O texto de Reymont e Joreskog (1996) é indicado para aqueles que já possuem

bons conhecimentos básicos sobre análise fatorial e desejam aprofundar-se no assunto. Para quem busca um texto um pouco mais formal do ponto de vista estatístico, mas com bons exemplos práticos, sugere-se os livros de Johnson e Wichern (1998), Johnson (1998), Dillon e Goldstein (1984), Mardia et al. (1989) e Sharma (1996).

### **3.6. UMA APLICAÇÃO AO MODELO DE ANÁLISE FATORIAL EXPLORATÓRIA**

#### ***“Busca de Fatores Associados à Prática de Atos Infracionais por Adolescentes –Um Estudo Preliminar”***

Para exemplificar o uso da análise fatorial exploratória, considerou-se uma amostra de 212 profissionais que realizam o trabalho multidisciplinar de ressocialização de adolescentes em conflito com a lei no Estado do Rio de Janeiro. A pesquisa visa verificar quais os atributos ou variáveis que na opinião desses profissionais são determinantes da prática de atos infracionais por parte de adolescentes, numa escala de 0 a 10.

Os profissionais tiveram que responder ao questionário abaixo:

**Em sua opinião, qual o grau de participação , numa escala de 0 a 10 , que cada fator abaixo tem tido na prática de atos infracionais por parte de adolescentes?**

<i>Atributos ou Variáveis do Problema (<math>V_i</math>)</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
1-Auto-escolha (V1)											
2-Comércio de drogas (V2)											
3-Desestrutura familiar (V3)											
4-Desemprego no país (V4)											
5-Desigualdade social (V5)											
6-Dependência química (V6)											
7-Formação educacional deficiente (V7)											
8-Formação profissional deficiente (V8)											
9-Falta de limites de comportamento social (V9)											
10-Falta de políticas públicas estaduais (V10)											
11-Falta de políticas públicas federais (V11)											
12-Índole propensa à prática de ato infracional (V12)											
13-Influência negativa dos meios de comunicação (V13)											
14-Influência negativa do meio social em que vive (V14)											
15-Má distribuição de renda no país (V15)											
16-Omissão do governo estadual (V16)											
17-Omissão do governo federal (V17)											
18-Pobreza (V18)											
19-Preconceito social e econômico (V19)											
20-Perspectiva de progressão sócio-econômica fácil (V20)											
21-Personalidade violenta (V21)											
22-Recessão econômica (V22)											
23-Residência em local dominado pelo tráfico de drogas (V23)											

Objetiva-se, mediante a técnica de análise fatorial , substituir o conjunto inicial de características determinantes, por outras de menor número, mas que guardam significativa explicação original do problema, de modo a levantar as dimensões latentes nas variáveis originais do fenômeno , visando dar uma interpretação mais compreensível segundo direções comuns .

Aplicou-se a rotação para transformar os coeficientes das componentes principais retidas numa estrutura mais simplificada. O objetivo é dividir o conjunto inicial de variáveis em subconjuntos com maior grau de independência possível. Optou-se pela rotação **varimax**. Este processo pretende que, para cada componente principal, existam apenas alguns pesos significativos e todos os outros sejam próximos de zero, através da maximização da variância entre os fatores para a rotação das matrizes fatoriais.

### 3.6.1. Teste de Validade da Análise Fatorial

A finalidade deste item é verificar se a aplicação da análise fatorial tem validade para as variáveis escolhidas, sendo justificado pela pouca quantidade de respondentes da pesquisa.

Para efetuar o procedimento referido utilizou-se dos testes de Kaiser-Meyer-Olkin(KMO) e de Esfericidade de Bartlett.

O teste de KMO deve ser interpretado segundo Quadro 1 :

**Quadro 1**

<i>KMO</i>	Análise de componentes principais
<b>1,00- 0,90</b>	<i>Muito boa</i>
<b>0,80-0,90</b>	<i>Boa</i>
<b>0,70-0,80</b>	<i>Média</i>
<b>0,60-0,70</b>	<i>Razoável</i>
<b>0,50-0,60</b>	<i>Má</i>
<b>&lt; 0,50</b>	<i>Inaceitável</i>



### 3.6.2. Análise Fatorial

As cargas das variáveis associadas aos dois fatores mais significativos (de maiores variâncias explicadas) apresentadas na Tabela 2 são analisadas a seguir. A interpretação do Gráfico 1 (*Screen Plot*) também confirma a escolha pela análise dos dois fatores analisados. No teste de esfericidade de Bartlett, deve-se rejeitar a hipótese nula que permite afirmar não haver correlação entre as variáveis iniciais. A estatística de teste definida por Bartlett tem distribuição assintótica de  $\chi^2$ .

O teste de KMO obteve um valor de 0,744, o que indica pelo Quadro 1 que a análise de componentes principais pode ser feita.

O teste de esfericidade de Bartlett dá um valor de  $\chi^2 = 1313,455$  com 253 graus de liberdade. Consultando o Quadro 1 verifica-se que  $\chi^2 > \chi^2_{0,95}$ , pelo que rejeita-se a hipótese nula, ou seja, as variáveis são correlacionadas. Pela análise do valor-p (Sig=0,000), que é inferior a 0,05 conclui-se da mesma forma.

Constata-se que o primeiro fator subjacente ao conjunto de atributos associados à prática de atos infracionais por parte de adolescentes é significativamente representado por quatro variáveis originais: falta de políticas públicas federais, falta de políticas públicas estaduais, omissão do governo estadual e omissão do governo federal. Analisando tais resultados, pode-se concluir que esse fator se refere ao apoio político à questão do adolescente em conflito com a lei.

O fator seguinte é explicado fundamentalmente por seis variáveis: desemprego no país, desigualdade social, má distribuição de renda no país, pobreza, preconceito social e econômico e recessão econômica. Desta forma tais resultados parecem sugerir que a questão econômica constitui-se

em fator determinante para a prática de atos inflacionais por parte de adolescentes no Rio de Janeiro.

**Tabela 2(Matriz de Fatores Rotada)**

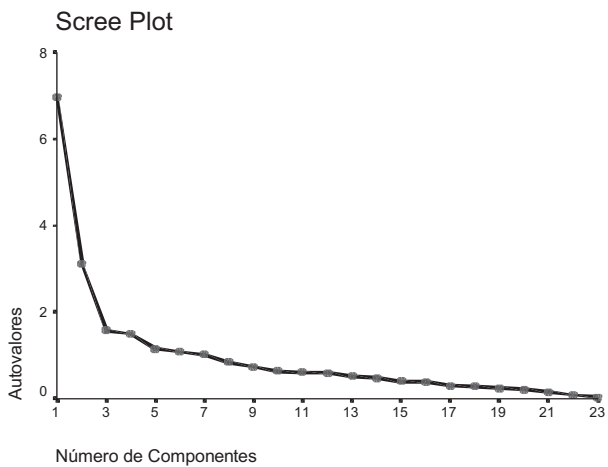
Variáveis	Comunalidades	Cargas Fatoriais						
		1	2	3	4	5	6	7
Auto Escolha	0,757	-0,186	-0,094	0,83	0,014	0,07	-0,033	<b>0,837</b>
Comércio de drogas	0,636	0,069	0,027	0,16	<b>0,751</b>	0,207	0,124	-0,091
Desestrutura familiar	0,475	0,142	0,195	<b>0,397</b>	0,358	0,336	-0,051	0,125
Desemprego no país	0,702	0,200	<b>0,643</b>	0,202	0,383	0,004	0,071	0,239
Desigualdade social	0,686	0,181	<b>0,644</b>	0,337	0,131	0,118	-0,307	0,000
Dependência química	0,652	0,241	0,060	0,376	<b>0,578</b>	0,061	0,332	0,001
Formação educacional deficiente	<b>0,811</b>	0,122	0,226	<b>0,831</b>	0,01	0,189	-0,050	-0,080
Formação profissional deficiente	0,792	0,130	0,288	<b>0,793</b>	0,29	-0,22	0,186	0,166
Falta de limites de comportamento social	0,660	0,298	0,017	<b>0,458</b>	-0,145	0,099	0,357	<b>0,450</b>
Falta de políticas públicas estaduais	<b>0,868</b>	<b>0,889</b>	0,044	0,157	0,079	-0,026	0,151	-0,149
Falta de políticas públicas federais	<b>0,819</b>	<b>0,867</b>	0,034	0,143	0,141	-0,027	0,088	-0,130
Indole propensa à prática de ato infracional	0,783	-0,285	0,034	-0,59	0,031	<b>0,633</b>	0,288	<b>0,461</b>
Influência negativa dos meios de comunicação	0,720	0,079	0,013	0,143	0,098	<b>0,811</b>	0,147	-0,073
Influência negativa do meio em que vivem	0,630	0,197	0,16	0,131	0,363	<b>0,608</b>	-0,046	0,207
Má distribuição de renda no país	0,737	0,656	<b>0,518</b>	0,101	0,140	0,030	-0,087	-0,004

Variáveis	Comunidades	Cargas Fatoriais						
		1	2	3	4	5	6	7
Omissão do governo estadual	0,812	0,856	0,246	0,066	0,068	0,099	0,008	0,001
Omissão do governo federal	0,792	0,830	0,288	0,055	0,015	0,113	-0,060	0,014
Pobreza	0,689	0,173	0,719	0,182	0,003	0,135	0,267	-0,137
Preconceito social e econômico	0,736	0,212	0,616	0,123	-0,037	0,374	0,307	-0,247
Perspectiva de progresso sócio-econômica fácil	0,636	0,010	0,140	0,114	0,244	0,092	0,722	-0,068
Personalidade violenta	0,790	-0,101	0,188	-0,024	0,142	0,378	0,637	0,419
Recessão Econômica	0,669	0,215	0,689	0,126	0,312	-0,105	0,155	0,012
Residência em local dominado pelo tráfico de drogas	0,689	0,030	0,365	-0,41	0,737	0,049	0,052	0,072

**Tabela 3**

Estadísticas	1	2	3	4	5	6	7
Autovalores	7,02	3,1	1,6	1,5	1,2	1,1	1,02
% Var.Explic.	30,5	13,6	7,0	6,6	5,1	4,7	4,4

**Gráfico 1 –Autovalores por componentes principais**



No Quadro 2, estão os resultados da análise efetuada, destacando-se as variáveis principais, caracterizadoras dos fatores.

## Quadro 2

Fatores Determinantes Relacionados à Prática de Atos Infracionais por Parte de Adolescentes

<b>Ordem do Fator</b>	<b>Denominação (Variáveis Latentes)</b>	<b>Variáveis Determinantes</b>
1	<i>Político</i>	<ul style="list-style-type: none"><li>• Falta de políticas públicas federais</li><li>• Falta de políticas públicas estaduais</li><li>• Omissão do governo estadual</li><li>• Omissão do governo federal</li></ul>
2	<i>Econômico</i>	<ul style="list-style-type: none"><li>• Desemprego no país</li><li>• Desigualdade social</li><li>• Má distribuição de renda no país</li><li>• Pobreza</li><li>• Preconceito social e econômico</li><li>• Recessão econômica</li></ul>

O processo de análise fatorial realizado no exemplo acima foi baseado em análises descritivas, exploratórias. Nenhum procedimento inferencial envolvendo a significância das cargas fatoriais da Tabela 2 foi desenvolvido. Mas pergunta-se : os resultados encontrados na matriz de fatores é real, válido para a população como um todo ou somente pode ser atribuído à amostra colhida ? Para responder a esta pergunta teria-se que ter um procedimento conclusivo que pudesse, com eficácia, estabelecer a significância das cargas fatoriais da Tabela 2. É o que vai ser proposto no Capítulo 5. No capítulo 8, Caso 1: Busca de Fatores, será realizado o teste inferencial para a significância dos resultados desta análise fatorial.

# CAPÍTULO 4

## REAMOSTRAGEM

O tipo de estatística não-paramétrica que foi ensinado no passado desempenhou um importante papel na análise de dados que não são contínuos e, portanto, não podem empregar a distribuição normal de probabilidade para fazer estimativas de parâmetros e de intervalo de confiança. Mas existe uma nova perspectiva sobre estimação não-paramétrica que também se relaciona com estimação de parâmetros e de intervalo de confiança para variáveis no mínimo em escala intervalar.

Com isso, não se tem que assumir que o intervalo de confiança para um parâmetro segue a distribuição normal. Pode-se até mesmo gerar intervalos de confiança para parâmetros como a mediana, o que geralmente é difícil de avaliar com as técnicas de inferência paramétrica tradicionais.

Essa abordagem não-paramétrica é conhecida como reamostragem e tem conquistado apoio como uma alternativa aos métodos clássicos de inferência paramétrica.

A reamostragem descarta a distribuição amostral assumida de uma estatística e calcula uma distribuição empírica – a real distribuição da estatística ao longo de centenas ou milhares de amostras.

Com a reamostragem, não se tem que confiar na distribuição assumida nem se tem que ser cuidadoso quanto à violação de uma das suposições inerentes. Pode-se calcular uma real distribuição de estatísticas da amostra e pode-se agora ver onde o 95 ou o 99 percentil estão realmente.

Mas de onde vêm as múltiplas amostras? É necessário reunir amostras separadas, aumentando sensivelmente o custo de coleta de dados? Ao longo dos anos estatísticos desenvolveram diversos procedimentos para criar as múltiplas amostras necessárias para a reamostragem *a partir da amostra original*.

Agora uma amostra pode gerar um grande número de outras amostras que podem ser empregadas para gerar a distribuição amostral empírica de uma estatística de interesse.

#### **4.1. CONCEITOS BÁSICOS EM REAMOSTRAGEM**

Reamostragem, contudo, não usa a distribuição de probabilidades assumida, mas ao invés disso ela calcula uma distribuição empírica de estatísticas estimadas. Criando múltiplas amostras da amostra original, a reamostragem agora precisa apenas do poder computacional para estimar um valor de uma estatística para cada amostra. Logo que eles estejam todos calculados, pode-se realizar o teste de normalidade dos valores e até mesmo construir intervalos de confiança e realizar testes de hipóteses.

A reamostragem engloba diversos métodos. Para esta tese de doutorado, se estudará e aplicará as de *bootstrap* e *jackknife*.

#### **4.2. MÉTODOS DE REAMOSTRAGEM**

Uma diferença chave entre os vários métodos de reamostragem é se as amostras são extraídas com ou sem reposição. A amostragem com reposição obtém uma observação a partir da amostra e então a coloca de volta na amostra para possivelmente ser usada novamente. A amostragem sem reposição obtém observações da amostra, mas uma vez obtidas eles não estão mais disponíveis.

O verdadeiro poder da reamostragem vem de amostragem **com reposição**. Pesquisas têm mostrado que esse método fornece estimativas diretas dos intervalos de confiança e valores-p, apesar de haver modificações nos métodos simples para obtenção dos intervalos de confiança.

#### **4.3. JACKKNIFE VERSUS BOOTSTRAP**

Os métodos *jackknife* e *bootstrap* diferem na maneira como eles obtém a amostra.

O método *jackknife* computa  $n$  subconjuntos ( $n$ =tamanho da amostra) pela eliminação seqüencial de um caso de cada amostra. Assim cada amostra tem um tamanho de  $n - 1$  e difere apenas pelo caso omitido em cada amostra.

Apesar de o método *jackknife* ter sido ultrapassado pelo *bootstrap* como um eficiente estimador de intervalos de confiança e cálculos de significâncias, ele continua como uma medida viável de observações influentes (uma observação que exerce uma influência desproporcional sobre um ou mais aspectos das estimativas e essa influência pode ser baseada em valores extremos das variáveis) e uma opção para muitos pacotes estatísticos.

O método *bootstrap* obtém sua amostra via amostragem com reposição da amostra original. A chave é a substituição das observações após a amostragem, o que permite ao pesquisador criar tantas amostras quanto necessárias e jamais se preocupar quanto à duplicação de amostras, exceto quando isso acontecer ao acaso. Cada amostra pode ser analisada independentemente e os resultados compilados ao longo da amostra. Por



exemplo, a melhor estimativa da média é exatamente a média de todas as médias estimadas ao longo das amostras.

O intervalo de confiança também pode ser diretamente calculado. As duas abordagens mais simples :

- 1) Calculam o erro padrão simplesmente como o desvio padrão das estimativas estimadas;
- 2) Literalmente ordenam as estimativas e definem os valores que contém os 5% extremos (ou 1%) dos valores estimados.

#### **4.4.LIMITAÇÕES**

Apesar de procedimentos de reamostragem não serem restritos por quaisquer suposições paramétricas, eles ainda têm certas limitações :

- 1) A amostra deve ser grande o bastante e obtida (a princípio aleatoriamente) de forma a ser representativa da população completa. Técnicas de reamostragem não podem conter quaisquer enviesamentos que traga como consequência uma amostra não representativa;
- 2) Métodos paramétricos são melhores em muitos casos para fazer estimativas pontuais. Os procedimentos de reamostragem podem completar as estimativas pontuais de métodos paramétricos fornecendo as estimativas de intervalos de confiança;
- 3) As técnicas de reamostragem não são adequadas para identificar parâmetros que têm um domínio amostral muito estreito, como os valores mínimos e máximos. A reamostragem funciona melhor quando a distribuição inteira é considerada para obter o parâmetro em análise.

Os procedimentos de reamostragem que foram discutidos neste capítulo fornecem uma perspectiva alternativa sobre uma das avaliações-chaves feitas na análise de dados: a variabilidade da estatística estudada. Essa é a base do teste de hipótese e da avaliação de significância estatística. Técnicas de reamostragem aumentam a habilidade do pesquisador para examinar a real distribuição dos estimadores tratados, ao invés de confiar plenamente na distribuição assumida. Técnicas como essas fornecem um modo direto para “conhecer seus dados” e evitar a armadilha muito comum de se tornar muito confiante em técnicas estatísticas ao invés de raciocinar sobre o que se sabe a respeito dos dados.

Nas próximas seções, se abordarão com mais detalhes as duas técnicas de reamostragem exemplificadas neste capítulo.

#### **4.5.O MÉTODO *JACKKNIFE***

O *jackknife* é um método não paramétrico destinado a estimar o enviesamento e, portanto reduzi-lo, e a variância de estimadores em condições teoricamente complexas ou em que não se tem confiança no modelo especificado. Foi introduzido por Quenouille em 1949, retomado por Tukey em 1958, e desenvolvido na última década. Tal como o *bootstrap* é um método de reamostragem, pois se baseia na construção de subamostras da amostra original.

##### **4.5.1.O Procedimento para Obtenção da Amostra *Jackknife***

1º) Seleciona-se uma amostra original de tamanho “*n*”:

$$X = \{ X_1, X_2, X_3, \dots, X_{n-1}, X_n \}$$

2º) Defini-se a estatística de interesse:

$$\hat{\theta} = F(x)$$

(3°) Gera-se a amostra *jackknife* 1 :

$$x^{(1)} = \{x_2, x_3, \dots, x_{n-1}, x_n\}$$

$$\hat{\theta}_{(1)} = F[x^{(1)}]$$

4° ) Gera-se a amostra *jackknife* 2 :

$$x^{(2)} = \{x_1, x_3, \dots, x_{n-1}, x_n\}$$

$$\hat{\theta}_{(2)} = F[x^{(2)}]$$

e assim por diante...

5° ) Gera-se a amostra *jackknife* n-1 :

$$x^{(n-1)} = \{x_1, x_2, \dots, x_{n-2}, x_n\}$$

$$\hat{\theta}_{(n-1)} = F[x^{(n-1)}]$$

6° ) Gera-se a amostra *jackknife* n :

$$x^{(n)} = \{x_1, x_2, \dots, x_{n-1}\}$$

$$\hat{\theta}_{(n)} = F[x^{(n)}]$$

(7°) Estima-se o erro padrão da estatística do 2° passo através da expressão:

$$\hat{S}_{\text{Jack}} = \frac{n-1}{n} \left\{ \sum_{i=1}^n [\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 \right\}^{1/2}, \text{ sendo } \hat{\theta}_{(.)} = \sum_{i=1}^n \hat{\theta}_{(i)} / n$$

Onde  $\theta$  é o valor que a estatística ser assume na amostra original.

Este processo multiplica a informação inicial, pois para além das estimativas de  $\theta$ , obtêm-se mais  $n$  valores para a mesma estatística. É a partir destes valores que serão determinadas às estimativas *jackknife*.

É de notar que, para que se preserve uma certa qualidade das estimativas, o *Jackknife* só deve ser aplicado à estatística funcional. Uma estatística funcional é aquela obtida junto a estimadores que são funções dos elementos da amostra. Contudo, também se obtém bons resultados na estimativa do erro padrão para estimativas não funcionais, desde que sejam definidas simetricamente nos seus  $n$  argumentos (invariantes a qualquer permutação das variáveis).

As propriedades do enviezamento *jackknife*, tal como do seu erro padrão, dependem da estatística considerada. Sabe-se que quando a estatística é uma funcional quadrática, a estimativa do enviezamento funcional é uma estimativa centrada do verdadeiro enviezamento.

#### 4.6.O MÉTODO *BOOTSTRAP*

O *bootstrap*, introduzido por Efron no final dos anos 70, vem historicamente na linha do *jackknife* e pode-se dizer que é uma técnica não paramétrica que procura substituir a análise estatística teórica (insuficiente em muitos casos como se exemplifica na utilização da análise fatorial) pela força bruta da computação, cada vez mais acessível e menos dispendiosa.

A terminologia, introduzida por Efron (1979), é basicamente uma técnica de reamostragem, que permite aproximar a distribuição de uma função das observações pela distribuição empírica dos dados baseada em uma amostra de tamanho finita. A amostragem é feita, com reposição, da distribuição da qual os dados são obtidos, se esta é conhecida (*bootstrap* Paramétrico) ou da amostra original (*bootstrap* não-paramétrico). Neste último caso supõe-se que as observações são obtidas da função de distribuição empírica  $F(x)$ , que designa uma massa de probabilidade igual  $1/n$  para cada ponto amostral.

O *bootstrap* aborda o cálculo do intervalo de confiança de parâmetros e cálculos de valores-p, em circunstâncias em que outras técnicas não são aplicáveis, em particular no caso em que o número de amostras é reduzido.

Esta técnica foi extrapolada para resolução de muitos problemas de difícil resolução através de técnicas de análise estatística tradicionais, baseadas na hipótese de um elevado número de amostras.

A técnica *bootstrap* tenta realizar o que seria desejável na prática, se tal fosse possível: *repetir a experiência*.

As observações são escolhidas de forma aleatória e as estimativas recalculadas.

A idéia básica da técnica *bootstrap* é: uma vez que não se dispõe de toda a população de amostras (observações) faça-se o melhor com o que se dispõe, que é o conjunto de amostrado:  $x=(x_1, \dots, x_n)$ .

A técnica *bootstrap* trata a amostra original como se esta representasse exatamente toda a população (conjunto de experiências, realizações).

Até agora o *bootstrap* chega aos mesmos resultados que o processo tradicional, baseado na máxima verossimilhança. A sua grande virtude consiste em apresentar solução para casos em que a dedução da precisão da estimativa, de seu viés e do EMQ aparenta ser impossível ou mesmo demasiado complexa. É especialmente para estes casos que o *bootstrap* se vocaciona.

#### **4.6.1 .O Procedimento para Obtenção da Amostra *Bootstrap***

Seja uma amostra original e a estatística de interesse abaixo:

$$x = \{x_1, x_2, x_3, \dots, x_{n-1}, x_n\}.$$

$$\hat{\theta} = F(x)$$

(1º) Geram-se as amostras *bootstrap*  $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n^*)}$  com reposição de  $x$ .

(2º) Calculam-se as estimativas da estatística de interesse:

$$\hat{\theta}_{(B)} = F[x_{(B)}], \quad B=1, \dots, n(^*)$$

(3º) Calcula-se o erro padrão *bootstrap*,  $S_{boot}^{\wedge}$ , dado por:

$$S_{boot}^{\wedge} = \left\{ \frac{1}{n^{(*)} - 1} \cdot \sum_{B=1}^{n^{(*)}} [\hat{\theta}_B - \hat{\theta}_{(*)}]^2 \right\}^{1/2}, \text{ sendo}$$

$$\hat{\theta}_{(*)} = \frac{\sum_{B=1}^{n^{(*)}} \hat{\theta}_{(B)}}{n^{(*)}}$$

O procedimento acima se aplica ao caso do *bootstrap* não-paramétrico. Opta-se por utilizar o *bootstrap* paramétrico, procede-se da mesma forma, com a única diferença de que cada amostra *bootstrap* é obtida da distribuição paramétrica que originou os dados que se tem em mãos, ao invés de reamostrar-se as observações disponíveis.

Desde o aparecimento do *bootstrap*, vários autores vêm tentando estabelecer confirmação empírica ou teórica da sua validade. Devido ao fato desta técnica atuar como um método de aproximação de distribuições, todas as provas de consistência e precisão dos estimadores são resultados assintóticos, mas sua validade pode ser estendida para espaços amostrais finitos. Para um estudo mais detalhado dos teoremas e demonstrações pode-se consultar Shao e Tu (1995), que oferecem um apanhado geral dos trabalhos que surgiram sobre o assunto. As conclusões a que chegaram estes autores são que a aproximação *bootstrap* é válida para a maioria das estatísticas de interesse e que seus estimadores são consistentes.

Se  $B \rightarrow \infty$ , então as estimativas do erro-padrão, do enviesamento e do EMQ se igualam às estimativas de máxima verossimilhança (Efron, 1982). Para o cálculo das estimativas *bootstrap* geralmente é suficiente um valor de  $B=100$ . Contudo, para se determinar a distribuição por amostragem com precisão deve considerar-se um valor para  $B$  substancialmente mais elevado. Segundo Efron(1982), geralmente  $B=1000$  proporciona bons resultados. E em ambos os casos, convém ensaiar diferentes valores para  $B$  até se verificar a convergência dos resultados.



# CAPÍTULO 5

## TESTE INFERENCIAL PARA ANÁLISE FATORIAL

Ao interpretar fatores, é preciso tomar a decisão sobre quais cargas fatoriais valem à pena considerar. A discussão a seguir detalha formalmente questões relativas à significância estatística e prática, bem como ao número de variáveis que afetam a interpretação de cargas fatoriais.

### 5.1.FORMALIZAÇÃO DO TESTE INFERENCIAL

Seja abaixo uma matriz usual de fatores rodada pelo método varimax, de uma análise fatorial por componentes principais:

	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	....	$F_n$
$X_1$	$c_{11}$	$c_{12}$	$c_{13}$	$c_{14}$	$c_{15}$	....	$c_{1n}$
$X_2$	$c_{21}$	$c_{21}$	$c_{23}$	$c_{24}$	$c_{25}$	....	$c_{2n}$
$X_3$	$c_{31}$	$c_{32}$	$c_{33}$	$c_{34}$	$c_{35}$	....	$c_{3n}$
$X_4$	$c_{41}$	$c_{42}$	$c_{43}$	$c_{44}$	$c_{45}$	....	$c_{4n}$
$X_5$	$c_{51}$	$c_{52}$	$c_{53}$	$c_{54}$	$c_{55}$	....	$c_{5n}$
.....							
$X_n$	$c_{n1}$	$c_{n2}$	$c_{n3}$	$c_{n4}$	$c_{n5}$	....	$c_{nn}$

$X_i$  é o atributo ou a variável considerada no modelo ,  $F_j$  é a componente principal ou fator da rodada da análise fatorial e  $C_{ij}$  é a carga fatorial rodada para  $X_i$  e  $F_j$  .

Considere os fatores 1 e 2 , as duas componentes principais que reúnem a maior porcentagem de variância explicada do problema(a inferência pode envolver mais de 2 fatores, conforme interesse do analista). Seja  $C_{i1}$  , a carga fatorial da componente 1( $F_1$ ), então  $C_{i1}$  é a correlação entre  $F_1$  e  $X_i$  . Seja  $C_{i2}$  , a carga fatorial da componente 2( $F_2$ ), então  $C_{i2}$  é a correlação entre  $F_2$  e  $X_i$  .

No modelo original de análise fatorial, faz-se as seguintes suposições sobre  $F_j$  :  $E(F_j)=0$  e  $Var(F_j)=1$  ,  $j=1, \dots , n$  . Além disso , num modelo ortogonal , sabe-se que :

a)  $Cov(F_j)=I_n$ , onde  $I_n$  é a matriz identidade de ordem  $n$ ,

b)  $E(\varepsilon)=0$

c)  $Cov(\varepsilon)=\Psi=diag(\Psi_1, \dots , \Psi_i)$

d)  $Cov(F, \varepsilon)=0$

A interpretação das cargas fatoriais advém de  $Cov(X_i , F_j)$ . Tem-se :

$$Cov(X_i , F_j) = Cov (C'_{ij}F_1 + \dots + C'_{ij}F_j + \dots + C'_{in}F_n + \varepsilon_i ; F_j) ;$$

$$i,j=1,\dots, n$$

Aplicando as suposições (a) e (d) , vem que

$$Cov(X_i , F_j) = Cov( C'_{ij} F_j , F_j ) = C'_{ij}$$

de onde se conclui que as cargas fatoriais são as covariâncias medidas entre as variáveis observadas e os fatores comuns.

Para facilitar a interpretação de um fator comum, pode-se medir a correlação existente entre o fator e cada variável observada. Dessa forma tem-se :

$$C_{ij} = COV(X_i F_j) / \sqrt{Var(X_i) Var(F_j)}$$

As cargas de  $F_1$  e as cargas de  $F_2$  definem a natureza dos respectivos fatores, os nomeiam, e caracterizam a análise fatorial encontrada e a matriz principal de fatores formada por correlações constitui o modelo usual de análise fatorial.

A inferência estatística proposta nesta tese é testar a significância das cargas fatoriais da matriz principal da análise fatorial, utilizando-se dos intervalos de confiança *bootstrap* e *jackknife*, percentílico principalmente e do **Valor-p** *bootstrap* e *jackknife*. Por simplificação, serão testadas somente as significâncias das cargas dos fatores 1 e 2.

A inferência estatística sugerida, então, é testar a significância de todas as cargas fatoriais de um determinado fator de interesse, os fatores 1 e 2, utilizando-se dos intervalos de confiança *bootstrap* e *jackknife*, percentílico principalmente, e dos **valores-p** *bootstrap* e *jackknife*.

Os intervalos de confiança e os **valores-p**, portanto, são construídos através da distribuição por amostragem avaliada, empírica, das variáveis aleatórias cargas fatoriais de uma componente principal de interesse. A distribuição por amostragem assim obtida tem por objetivo estabelecer um modelo adequado à interpretação do comportamento regular da estatística investigada e de seus parâmetros característicos. A experiência nesta investigação é a base para se montar o modelo, ou para ajustá-lo ao modelo

ideal (teórico). A distribuição por amostragem e os seus respectivos intervalo de confiança e valor-p são obtidos a posteriori, com base na experiência, e não da maneira clássica, a priori, através de informações teóricas existentes sobre a distribuição por amostragem e seu respectivo erro-padrão.

Para realizar o teste de significância supra mencionado, se utilizará a idéia de que o intervalo de confiança pode ser usado imediatamente, sem qualquer outro cálculo para testar qualquer hipótese: o intervalo de confiança pode ser considerado como um conjunto de hipóteses aceitáveis.

Qualquer hipótese nula que esteja fora do intervalo de confiança deve ser rejeitada. Por outro lado, qualquer hipótese que esteja dentro do intervalo de confiança deve ser aceita.

Nesta tese, a hipótese nula é de que não existe correlação entre a variável  $X_i$  e o fator  $F_j$ , contra a hipótese alternativa de que existe tal correlação:

$$H_0: C_{ij} = 0,00.$$

$$H_1: C_{ij} \neq 0,00.$$

Na verdade, a hipótese alternativa que está sendo testada é:

- $C_{ij} > 0,00$ , se o sentido da correlação do fator for direto com a variável testada ou.
- $C_{ij} < 0,00$ , se o sentido da correlação do fator for inverso com a variável testada.

Matematicamente, pode-se exprimir por  $C_{ij} = 0,00$  a hipótese de  $X_i$  e  $F_j$  serem não-correlacionados. Para testar esta hipótese, basta ver se o valor 0 está contido no intervalo de confiança.

Portanto, se o zero estiver contido no intervalo de confiança, aceita-se a hipótese nula de que não existe correlação, resultado não significativo para a carga fatorial  $C_{ij}$ . Por outro lado, se o zero estiver fora do intervalo de confiança considera-se a hipótese alternativa de que existe correlação, resultado significativo para  $C_{ij}$ . Este teste passa a se chamar “**Teste Inferêncial para Análise Fatorial pelo Intervalo de Confiança -TIAFIC**”.

Em lugar de aceitar ou rejeitar simplesmente , uma forma de teste mais adequada é o cálculo do valor-p. Os passos para se calcular o valor-p *bootstrap* e *jackknife* são :

1º)Localizar as distribuições amostrais das cargas fatoriais por variável:

2º)Calcular a **distribuição G da variável 1( $G_1$ )** da seguinte forma: diminuindo cada valor constante da coluna  $V_1$  de sua média e dividindo de seu desvio-padrão;

- Realizar a mesma operação com as outras variáveis da base de dados;
- Este passo gerará os arquivos das distribuições amostrais com as variáveis normalizadas.

3º)Calcular a estatística **g** do teste da variável 1 da seguinte forma:

$$g_1 = \frac{C_1}{S_1}$$

**Onde :**

$C_1$  = a carga fatorial da variável 1( $V_1$ ), com o fator 1, obtida da rotação da **amostra original**;

$S_1$  = o desvio-padrão da variável 1( $V_1$ ) com o fator 1.

- Realizar a mesma operação com as outras variáveis da base de dados.

4º) Calcular o **VALOR-P TIAF** da seguinte forma:

**Se  $g_1 \geq 0$  :**

**VALOR-P =  $P(G \geq g_1) = (\text{n}^\circ \text{ de valores da coluna } G_1 \geq g_1) \setminus \text{n}^\circ \text{ de reamostragens}$**

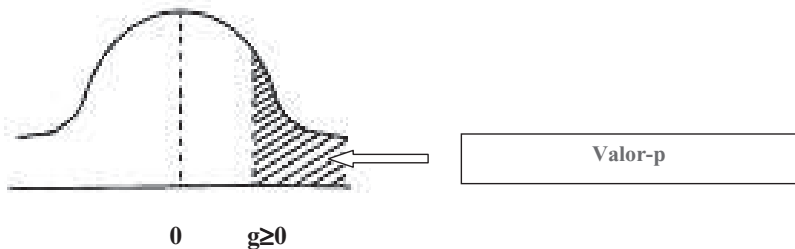
**Se  $g_1 \leq 0$ :**

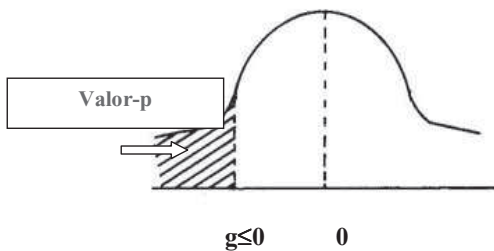
**VALOR-P =  $P(G \leq g_1) = (\text{n}^\circ \text{ de valores da coluna } G_1 \leq g_1) \setminus \text{n}^\circ \text{ de reamostragens}$**

Procura-se , então , a probabilidade na cauda, além deste valor observado de  $g$  : este é o **valor-p**.

- Realizar a mesma operação com as outras variáveis da base de dados.

**Visualização das áreas correspondentes aos VALORES-P calculados:**





Este teste passa a se chamar “*Teste Inferencial para Análise Fatorial pelo Valor-p -TIAFVP*”.

Cargas não significantes indicam que suas respectivas variáveis não participam ou não devem participar estatisticamente da nomeação/interpretação do fator em foco. Em contrapartida, cargas significantes pressupõem que a variável correspondente contribui estatisticamente para a formação do fator e esta contribuição é proporcional à sua significância ordinária ou magnitude.

Hair e Anderson(2005) sugerem como uma das etapas para a realização de uma boa análise fatorial, além da significância estatística, a verificação da significância prática. Vejam no trecho abaixo o que os referidos autores relatam sobre a questão:

*“A força da análise multivariada é sua forma aparentemente mágica de ordenar um grande número de possíveis alternativas e encontrar as que têm significância estatística. Muitos pesquisadores ficam míopes ao se concentrar somente na significância alcançada dos resultados sem compreender suas interpretações, sejam boas ou ruins. Ao invés disso, o pesquisador deve olhar não apenas a significância estatística dos resultados, mas também sua significância prática. A significância prática faz a pergunta : ‘E daí?’ para*

*qualquer aplicação administrativa, os resultados devem ter um efeito demonstrável que justifique a ação. Em termos acadêmicos, a pesquisa está se concentrando não apenas em resultados estatisticamente significantes, mas também em suas implicações substantivas e teóricas, as quais são muitas vezes extraídas de sua significância prática”.*

Uma regra ou sugestão utilizada Hair e Anderson(2005) para se decidir se uma determinada variável pode ser considerada na nomeação de um fator sob investigação (significância prática) seria decidir pela sua significância ordinária toda vez que a variável de carga significativa estatisticamente estiver entre / 0,3 a 1,0 /, pelo motivo indicado no Quadro3:

### Quadro 3

#### Verificação da Significância Prática

Valores $C_{ij}$ de um Determinado Fator ( $F_i$ )	Correlação ou Relação
0,00  — 0,30	<i>Ausência de correlação ou grau de correlação não tolerável</i>
0,30  — 1,00	<i>Existência de correlação ou grau de correlação tolerável</i>

Matematicamente, a hipótese desta tese consiste que se uma determinada carga de uma variável,  $C_{ij}$ , não for significativa estatisticamente, então a correlação da variável  $X_i$  com o fator  $F_j$ ,  $C_{ij}$ , é nula, e da natureza ou nomeação de  $F_j$  não deve participar a variável  $X_i$ .Conseqüentemente, a análise fatorial será interpretada de acordo com um processo inferencial estatístico.



Denomina-se de vetor  $\theta_1$  o parâmetro populacional definido como as cargas fatoriais da primeira componente principal da matriz  $\text{varimax}(F_1)$ , obtido da amostra original e denomina-se de vetor  $\theta_2$  o parâmetro populacional definido como as cargas fatoriais da segunda componente principal da matriz  $\text{varimax}(F_2)$ , obtido da amostra original.

Adotando-se como estimador de  $\theta_1$ , a matriz aleatória formada pelas estimativas das cargas fatoriais de  $F_1$ , cujas colunas são as variáveis do problema e as linhas, as  $\mathbf{B}$  ou as  $\mathbf{n}$  reamostragens obtidas através das técnicas *bootstrap* e *jackknife* e como de  $\theta_2$ , a matriz aleatória formada pelas estimativas das cargas fatoriais de  $F_2$  cujas colunas são as variáveis do problema e as linhas, as  $\mathbf{B}$  ou as  $\mathbf{n}$  reamostragens obtidas através das técnicas *bootstrap* e *jackknife*; o problema é encontrar a distribuição por amostragem dos estimadores dos vetores  $\theta_1$  e  $\theta_2$ , isto é, das matrizes aleatórias de  $F_1$  e  $F_2$ , ou seja, a sua variância, viés e EMQ. Com essas estimativas, é possível construir intervalos de confiança para as estimativas dos vetores  $\theta_1$  e  $\theta_2$ , calcular valores-p e realizar o *TIAF*.

A relevância deste procedimento sugerido é que em análise fatorial não é tão fácil encontrar pela teoria tradicional as exatas distribuições de amostragem dos estimadores dos vetores  $\theta_1$  e  $\theta_2$ . As técnicas *CIS* surgem originalmente como uma alternativa para este tipo de situação, mas se tornaram mais populares e viáveis devido em grande parte ao avanço computacional das últimas décadas.

Para estimar a distribuição por amostragem dos estimadores dos vetores  $\theta_1$  e  $\theta_2$  para qualquer pesquisa em que se utilize de uma análise fatorial por componente principal com rodada *varimax* serão oferecidas nesta tese as opções *bootstrap* e *jackknife*.

A teoria assintótica dos estimadores dos vetores  $\theta_1$  e  $\theta_2$  deve ser estabelecida, isto é, deve verificar se a aplicação das aproximações *bootstrap* e *jackknife* são válidas para estas estatísticas e se são consistentes. Esta fase será desenvolvida empiricamente com base em estudos de casos no capítulo 8: Estudo de Casos: Aplicações do TIAF.

Deverão ser desenvolvidos para cada base de pesquisa dois algoritmos, um *bootstrap* e outro *jackknife* para o cálculo dos respectivos intervalos de confiança, valores-p e resultados TIAF'S e que foram programados através dos softwares *R 2.1.1* e *SAS Versão 8* para obtenção computacional das estimativas.

Compara-se pelo EMQ, por exemplo, o melhor procedimento e assim pode-se utilizar o intervalo de confiança e o valor-p da melhor técnica para testar a significância de estimativas geradas, dos estimadores dos vetores  $\theta_1$  e  $\theta_2$ , utilizando-se o intervalo de confiança e o valor-p.

## **5.2.ALGORITMO *BOOTSTRAP* PARA REALIZAÇÃO DO TIAF**

Os algoritmos para realização computacional dos procedimentos inferenciais utilizando-se das técnicas *bootstrap* seguem abaixo.

Uma amostra existente de tamanho  $n$  é tomada como uma amostra original:

- 1º)Da amostra original, seleciona-se a primeira amostra *bootstrap*;
- 2º)Com base na amostra selecionada, gera-se a matriz principal de fatores;
- 3º)Da matriz principal de fatores, obtém-se a matriz rodada de fatores ortogonal (rotação *varimax*);

4º) Com base na matriz rodada de fatores, selecionam-se dos fatores 1 e 2 as cargas fatoriais, isto é, as estimativas dos vetores  $\theta_1$  e  $\theta_2$ ;

5º) Repetem-se os passos de 1º a 4º B vezes, por exemplo, 1000 vezes;

6º) Ordenam-se os valores obtidos para as cargas fatoriais, do menor ao maior. Determinam-se limites de confiança para uma especificada probabilidade  $\alpha$ , igual ao nível de significância, de acordo com as expressões abaixo:

$$\hat{\theta}(q_1) = \text{limite inferior, onde } q_1 = B \cdot \alpha / 2$$

$$\hat{\theta}(q_2) = \text{limite superior, onde } q_2 = B - q_1 + 1$$

Se forem 1000 iterações(B) e  $\alpha = 0.05$ , o limite inferior será o valor na 25ª posição ( $q_1 = B \cdot \alpha / 2$ ) e o limite superior o na 976ª posição ( $q_2 = B - q_1 + 1$ ).

Pode-se então afirmar, com uma probabilidade  $\alpha$  de se estar errado, que o intervalo de confiança construído tem alta probabilidade de conter o verdadeiro valor da carga fatorial sobre o qual a estimação foi baseada.

7º) Calculam-se os **valores-p bootstrap** para cada variável da pesquisa baseada na cauda direita ou esquerda da distribuição por amostragem empírica e real da variável em foco, isto é, a probabilidade na cauda, além deste valor observado da estimativa: este é o **valor-p bootstrap**. Este valor corresponde à credibilidade da hipótese nula.

8º)Calculam-se: o valor esperado, a variância, o viés e o EMQ das distribuições amostrais obtidas, como indicado abaixo:

$$\begin{aligned} \text{Viés} &= [ E(\hat{\theta}) - \theta ] \\ \text{EMQ} &= S_{boot}^2 + \text{Viés}^2 \end{aligned}$$

As estatísticas erro-padrão, viés e EMQ são indicadores da eficácia e da qualidade da estimação efetuada.

A distribuição por amostragem de  $\hat{\theta}$  corresponde ao histograma dos B valores determinados para as estimativas de  $\theta_{(1)}, \theta_{(2)}, \theta_{(3)}, \dots, \theta_{(B)}$ .

$\theta$  é o vetor de cargas fatoriais dos fatores  $F_1$  e  $F_2$  da matriz varimax da amostra original.

O algoritmo Jackknife segue procedimentos análogos.

### 5.3.COMPUTAÇÃO DOS ALGORITMOS *BOOTSTRAP* E *JACKKNIFE*

Os programas propostos para os métodos *bootstrap* e *jackknife* com base nos algoritmos elaborados acima para reamostragem das B estatísticas no método *bootstrap* e nas  $n$ , no esquema *jackknife*, com uma confiança de 95% e com base nos softwares *R 2.1.1* e *SAS Versão 8*.

Os programas computacionais elaborados para realizar os procedimentos inferenciais para a análise fatorial utilizando as técnicas *bootstrap* e *jackknife* encontram-se nos Anexos 1 e 2(versão R e versão SAS) e constituem a materialização, o produto concreto de todo o estudo realizado neste trabalho. Nos estudos de caso, utilizou-se dos resultados submetidos ao programa R 2.1.1, que é um programa *free*, mais disponível e de mais fácil execução.

Os programas em R 2.1.1 também estarão em uma biblioteca desenvolvida para o R, disponibilizando-a na rede. A vantagem é que a biblioteca pode ser usada por diferentes pessoas, que irão eventualmente reportar críticas ao procedimento, que pode então ser atualizado e aperfeiçoado posteriormente.

A escolha entre *bootstrap* e *jackknife* não é imediata, depende do usuário. Neste contexto, aponta-se que aquele que fornecer menor estimativa de precisão e viés, isto é, que minimize o erro médio quadrático(EMQ).

#### **5.4.O MÉTODO DO TIAF**

Esta sessão objetiva estabelecer o método científico para realizar o TIAF.

O método científico é a ferramenta colocada à disposição do cientista que, através da pesquisa, pretende penetrar no segredo de seu objeto de estudo.

O método, em sentido amplo, é a ordem que se deve impor aos diversos processos necessários para atingir um fim dado ou um resultado desejado.

Método Científico é um instrumento de que se serve a inteligência para descobrir relações, verdades e leis referentes aos diversos objetos de investigação.

O método científico é um dispositivo ordenado, um conjunto de procedimentos sistemáticos que o pesquisador emprega para obter o conhecimento adequado do problema que se propõe resolver.

O método é constituído de um conjunto de processos ou técnicas que formam os passos do caminho a percorrer na busca da verdade.

Toda investigação nasce da observação cuidadosa de fatos que necessitam de uma maior explicação. Esta é imaginada através da hipótese. Em seguida, procura-se verificar a veracidade da solução sugerida. Nas ciências experimentais, isto é feito através de ensaios e experiências; nas ciências humanas, é feito através de demonstrações racionais e lógicas por meio da argumentação. Descoberta a explicação do fato, achada a relação de causalidade entre os fenômenos ou sua coexistência ou ainda sua finalidade, formula-se a lei. É a tarefa da indução: aplicar a relação necessária descoberta a casos não observados da mesma espécie.

Esta explicação parcial e fracionada de uma realidade não satisfaz a curiosidade científica. Por isso, o cientista reúne as tentativas de explicação, os princípios e leis particulares numa visão unificadora, mais ampla e globalizada, através da teoria ou do sistema.

Em resumo, o desenvolvimento do método científico se faz pelos processos ou técnicas da observação, hipótese, demonstração (experimental ou racional), indução da lei e teoria. Além disso, e simultaneamente com os processos referidos, o pesquisador sempre estará usando as técnicas da análise, da síntese e da indução.

O método científico é, pois, um meio imprescindível com a qual o espírito científico do pesquisador, com ordem e rigor, procura penetrar no sentido dos fatos e fenômenos que pretende conhecer.

Fundamentado em tudo o que foi dito em parágrafos acima e para solucionar o problema proposto nesta tese é necessário estabelecer o método científico para testar a significância de cargas fatoriais de componentes principais. O método científico da investigação deste estudo chama-se “**Método TIAF**” .

O “**Método TIAF**” consiste nas seguintes etapas que devem ser seguidas nesta ordem para garantir o máximo de eficácia na obtenção do objetivo proposto:

*1º) Selecione uma amostra aleatória representativa da população, de tamanho proporcional ao número de variáveis e que reúna um conjunto o mais completo possível de características sobre o problema proposto (amostra original ou base de dados):*

*2º) Efetue os testes de validação da análise fatorial;*

- 3º) Rode a análise fatorial e obtenha a matriz de fatores rodada varimax;
- 4º) Submeta os programas bootstrap e jackknife;
- 5º) Verifique comparativamente qual o melhor método para seus dados (o bootstrap ou o jackknife);
- 6º) Identifique para quais variáveis as cargas dos fatores de interesse são significantes, isto é, onde existe correlação significativa com o fator;
- 7º) Identifique qual das variáveis significante estatisticamente podem participar da nomeação e/ou interpretação dos fatores (significância prática);
- 8º) De acordo com um critério do analista, selecione as variáveis que nomearão os fatores em foco;
- 9º) Nomeie os fatores em estudo;
- 10º) Tomada de decisão administrativa, tendo uma confiança alta de que os resultados amostrais levados em consideração na análise de dados não são acidentais, frutos do acaso.

A superioridade da abordagem inferencial para as cargas fatoriais empregadas neste trabalho, em relação ao apresentado pelo Método Hair e Anderson (2005), é que o número de variáveis contempladas e o fator específico em exame são considerados no processo de significância das cargas fatoriais. As reamostragens são realizadas levando em consideração todas as variáveis do problema (quanto mais completa for a lista de variáveis melhor o teste de significância da análise fatorial) e para um fator específico  $F_j$  gerado em cada amostra *bootstrap* ou *jackknife* conforme seu poder de explicação dentro do modelo fatorial.



No capítulo 9, se discutirá com mais detalhes aspectos comparativos e de desempenho entre o “Método Tradicional de Hair e Anderson” e o “Método TIAF”.

# CAPÍTULO 6

## TESTE DE NORMALIDADE DAS CARGAS FATORIAIS

Até o presente momento os intervalos de confiança e os valores-p para o TIAF são sugeridos com base em métodos não paramétricos, que não exigem a normalidade das distribuições por amostragem das estimativas obtidas: utilizou-se o intervalo de confiança percentílico e o valor-p foi calculado diretamente das distribuições por amostragem empíricas das estimativas . Contudo, sabe-se que as alternativas não paramétricas por não exigirem o modelo gaussiano como pressuposto básico e sim distribuição livre, constituem em testes de significância com poder bem menor que os dos correspondentes paramétricos.

O poder de um teste é a probabilidade de rejeitar a hipótese nula, quando ela é realmente falsa e, por isso mesmo, deve ser rejeitada. Em conseqüência, ao admitir uma hipótese nula como falsa é mais provável que o pesquisador o rejeite pelo uso adequado de testes paramétricos dos que os não paramétricos.

Para minimizar a probabilidade de erros de decisão nos TIAF'S realizados, seria mais prudente “conectar” os intervalos de confiança e os valores-p construídos ao modelo de distribuição de probabilidades das estimativas investigadas. A curva normal por ser a mais freqüente distribuição ajustável à descrição de variáveis aleatórias e por otimizar o poder de um teste de hipótese será utilizada como prova de aderência aos dados.

O pesquisador sempre deve lembrar que os testes de significância são menos úteis em amostras pequenas (menos que 30) e muitos sensíveis em amostras grandes (que excedem a 1000 observações). Logo, o pesquisador deve usar também testes gráficos para avaliar o real grau de desvio da normalidade.

Para realizar o teste de aderência supra citado, isto é, o teste de normalidade aos dados, se aplicará, primeiro, o **Gráfico de Probabilidades** (*P-P-Probability Plots*), que indica graficamente o ajustamento das cargas fatoriais estimadas ao modelo gaussiano. Este tipo de gráfico indica no eixo das abscissas as frequências relativas acumuladas observadas nas reamostragens e no eixo das ordenadas a função distribuição de probabilidades esperadas. A diagonal principal do gráfico representa um ajustamento perfeito das reamostragens a função de distribuição de probabilidades. Quanto mais os pontos se afastam da diagonal, ou se distribuem segundo um determinado padrão, menor é o ajustamento das reamostragens à distribuição teórica em análise. Além do gráfico de ajustamento à normal, um gráfico do ajustamento dos resíduos ( $y_{\text{obs}} - y_{\text{est}}$ ) será indicado conjuntamente. Se as reamostragens são perfeitamente normais, os resíduos distribuir-se-ão segundo uma faixa horizontal em torno do zero, sem denotar qualquer padrão de distribuição. Por simplificação, somente serão mencionados os resultados e interpretações verificadas, pois estes gráficos são bastantes simples e fáceis de serem produzidos.

Para confirmar os resultados encontrados pelo teste do Gráfico de Probabilidades se recorrerá ao **Teste de Kolmogorov-Smirnov** (Siegel, Sidney; 1975). Kolmogorov-Smirnov desenvolveram um método, em geral

mais poderoso que o do Qui-Quadrado ( $\chi^2$ ), para testar a aderência, em que a variável de teste é a maior diferença observada entre a função de distribuição acumulada do modelo e a da amostra.

A função de distribuição acumulada do modelo testado, ou função de repartição, dá as probabilidades acumuladas em cada ponto, ou seja,  $F(x) = P(X \leq x)$ . A função de distribuição acumulada da amostra corresponderá ao gráfico das freqüências relativas acumuladas(ogiva). Designa-se essa segunda função por  $G(x)$ . O teste consta simplesmente da verificação do valor :

$$d = \max |F(x) - G(x)|$$

e da comparação com um valor crítico tabelado em função do nível de significância( $\alpha$ ) e do tamanho da amostra (n). Se d for maior que o valor crítico, rejeita-se  $H_0$ . Os valores críticos para os níveis de significância usuais estão tabelados. Um método mais prático e rigoroso de se testar a hipótese nula é calcular o nível de credibilidade da hipótese nula, isto é, o seu valor-p. O valor-p pode ser então definido como a probabilidade condicional do valor crítico “d” ser igual ou maior que o valor efetivamente observado, dado a hipótese nula do modelo como verdadeira, isto é , procura-se a probabilidade na cauda , além deste valor observado de “d”.

Para realizar o teste de aderência das cargas fatoriais em função das variáveis de cada problema se utilizará uma amostra de três bases de dados do elenco de bases aplicadas neste estudo(seis bases ao todo, ver capítulo 8): a base busca de fatores, a base creme dental e a base modo de vida.

A Base Busca de Fatores é fruto de uma pesquisa realizada junto a 212 servidores de reformatórios de menores infratores, que derem a sua

opinião sobre o grau de influência de 23 atributos na prática de ato infracional praticada por adolescente.

A Base Creme Dental foi gerada pela aplicação de um questionário a 40 respondentes que deram a sua opinião sobre os benefícios que esperam da compra de um creme dental. Foram listados 6 benefícios que constituíram em variáveis do modelo.

A Base Modo de Vida surgiu de um *survey* envolvendo somente 25 respondentes sobre o comportamento de donas de casa na hora da compra de bens e serviços. Foram listados 7 comportamentos que constituíram em variáveis do modelo.

Os gráficos de probabilidades não serão ilustrados no texto do trabalho, mas seus os resultados serão comunicados nas análises e interpretações realizadas em cada base de dados.

## 6.1. TESTE DE NORMALIDADE DA “BASE BUSCA DE FATORES”

### 6.1.1. Teste de Normalidade das Variáveis no Método *Bootstrap*:

Tabela 4

#### Caso 1 : Busca de Fatores

Resultado do Teste de Normalidade para o Caso 1 das Estimativas *Bootstrap* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,50	Normal	0,10	Normal
V2	0,36	Normal	0,58	Normal
V3	0,11	Normal	0,98	Normal
V4	0,95	Normal	0,98	Normal
V5	0,33	Normal	0,89	Normal
V6	0,39	Normal	0,69	Normal
V7	0,99	Normal	0,54	Normal
V8	0,39	Normal	0,97	Normal
V9	0,51	Normal	0,77	Normal
V10	0,04	Nnormal	0,96	Normal
V11	0,05	Nnormal	0,99	Nnormal
V12	0,60	Normal	0,15	Normal
V13	0,51	Normal	0,79	Normal
V14	0,33	Normal	0,79	Normal
V15	0,09	Normal	0,75	Normal
V16	0,25	Normal	0,79	Normal
V17	0,29	Normal	0,85	Normal
V18	0,85	Normal	0,60	Normal

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V19	0,98	Normal	0,96	Normal
V20	0,76	Normal	0,64	Normal
V21	0,98	Normal	0,33	Normal
V22	0,49	Normal	0,68	Normal
V23	0,48	Normal	0,95	Normal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal=rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

No método *bootstrap*, fator 1, o gráfico de ajustamento à normal indica que a maioria das variáveis se ajusta ao modelo normal, pois se observa em muitos atributos um afastamento cíclico fraco em relação à diagonal. Exceções são constatadas nas variáveis  $V_{10}, V_{11}, V_{14}, V_{16}, V_{17}, V_{18}, V_{19}$  e  $V_{23}$ . O gráfico do ajustamento dos resíduos revela um tímido comportamento cíclico em torno de zero para a maioria das variáveis do Caso 1, denotando fraco afastamento em relação à normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal.

No mesmo método, fator 2, o gráfico de ajustamento à normal indica também que a maioria das variáveis se ajusta ao modelo normal, pois se observa, igualmente ao fator 1, que em muitos atributos existe um certo afastamento cíclico fraco em relação à diagonal. Exceções são constatadas nas variáveis  $V_1, V_2, V_{10}, V_{11}, V_{12}, V_{14}, V_{16}, V_{17}, V_{21}$  e  $V_{23}$ . O gráfico do ajustamento dos resíduos revela um tímido comportamento cíclico em torno de zero para a maioria das variáveis do Caso 1, denotando fraco

afastamento em relação à normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal..

O teste de normalidade da Tabela 4 confirma o que o gráfico de probabilidades havia indicado: as estimativas das cargas fatoriais obtidas pelo método *bootstrap* são em geral normalmente distribuídas, o que justificaria a construção alternativa de um intervalo de confiança *bootstrap* normal, isto tanto para o fator 1 quanto para o fator 2. Contudo, como também existem variáveis que não se ajustam à normalidade, adotando-se um critério de uniformidade e comparatibilidade, os intervalos de confiança serão construídos seguindo o procedimento percentílico.



### 6.1.2. Teste de Normalidade das Variáveis no Método *Jackknife*

Tabela 5

#### Caso 1 : Busca de Fatores

Resultado do Teste de Normalidade Para o Caso 1 das Estimativas *Jackknife* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,00	Nnormal	0,00	Nnormal
V2	0,00	Nnormal	0,00	Nnormal
V3	0,00	Nnormal	0,00	Nnormal
V4	0,00	Nnormal	0,02	Nnormal
V5	0,00	Nnormal	0,00	Nnormal
V6	0,00	Nnormal	0,00	Nnormal
V7	0,00	Nnormal	0,00	Nnormal
V8	0,00	Nnormal	0,00	Nnormal
V9	0,00	Nnormal	0,00	Nnormal
V10	0,00	Nnormal	0,00	Nnormal
V11	0,00	Nnormal	0,03	Nnormal
V12	0,00	Nnormal	0,00	Nnormal

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V13	0,00	Nnormal	0,00	Nnormal
V14	0,00	Nnormal	0,00	Nnormal
V15	0,00	Nnormal	0,00	Nnormal
V16	0,00	Nnormal	0,00	Nnormal
V17	0,00	Nnormal	0,00	Nnormal
V18	0,00	Nnormal	0,00	Nnormal
V19	0,00	Nnormal	0,00	Nnormal
V20	0,01	Nnormal	0,00	Nnormal
V21	0,00	Nnormal	0,00	Nnormal
V22	0,00	Nnormal	0,00	Nnormal
V23	0,00	Nnormal	0,00	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal=rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

No método *jackknife*, fator 1, o gráfico de ajustamento à normal indica que a grande maioria das variáveis se não se ajusta ao modelo normal, pois se observa em muitos atributos um afastamento cíclico acentuado em relação á diagonal. Exceções são constatadas nas variáveis  $V_8$  e  $V_{20}$ . O gráfico do ajustamento dos resíduos revela um fraco comportamento aleatório para a maioria das variáveis do Caso 1, isto é, um acentuado padrão de distribuição, denotando algum afastamento em relação à normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal.

No mesmo método, fator 2, o gráfico de ajustamento à normal indica também que a maioria das variáveis não se ajusta ao modelo normal, pois se observa, igualmente ao fator 1, que em muitos atributos existe um certo

afastamento cíclico em relação á diagonal. O gráfico do ajustamento dos resíduos , como no fator 1 , revela um certo padrão de comportamento dos afastamentos para a maioria das variáveis do Caso 1, denotando algum afastamento em relação à normal dos atributos deste problema e confirmando a conclusão tomada pelo gráfico de ajustamento à normal.

Pelo que indica o Teste de Kolmogorov-Smirnov as estimativas das cargas fatoriais obtidas pelo método *jackknife* não se aderem ao modelo normal, conclusão análoga à observada pela investigação da normalidade pelos Gráficos de Probabilidades, o que justificaria a adoção de intervalos de confiança *jackknife* percentílico, isto tanto para o fator 1, quanto para o fator 2.

## 6.2. TESTE DE NORMALIDADE DA “BASE CREME DENTAL”

### 6.2.1. Teste de Normalidade das Variáveis no Método *Bootstrap*:

Tabela 6

#### Caso 2 : Creme Dental

#### Resultado do Teste de Normalidade para o Caso 2 das Estimativas *Bootstrap* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,00	Nnormal	0,00	Nnormal
V2	0,00	Nnormal	0,00	Nnormal
V3	0,00	Nnormal	0,00	Nnormal
V4	0,00	Nnormal	0,00	Nnormal
V5	0,00	Nnormal	0,00	Nnormal
V6	0,00	Nnormal	0,00	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal=rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

O Teste de Kolmogorov-Smirnov revela (ver Tabela 6), ao nível de 5% de significância, que as estimativas geradas pelo método *bootstrap* não seguem o modelo normal de probabilidade. Isto tanto para o fator 1 quanto para o fator 2. Os Gráficos de Probabilidades revelam para a maioria das variáveis um afastamento em relação à diagonal, o que confirma os resultados do teste de aderência de utilizado (*K-S*). Os gráficos de afastamento de resíduos não apresentam um comportamento horizontal em torno de zero, indicando também um afastamento das estimativas da normalidade.

## 6.2.2. Teste de Normalidade das Variáveis no Método *Jackknife*:

**Tabela 7**

**Resultado do Teste de Normalidade Para o Caso 2 das Estimativas *Jackknife* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2**

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,26	Normal	0,46	Normal
V2	0,85	Normal	0,08	Normal
V3	0,05	Nnormal	0,41	Normal
V4	0,41	Normal	0,36	Normal
V5	0,05	Nnormal	0,53	Normal
V6	0,50	Normal	0,03	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal=rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

Pela observação do Gráfico de Probabilidades se observa que para a maioria das variáveis existe um ajustamento razoável dos valores estimados à curva normal, estando os pontos mais próximos da diagonal. Já os gráficos do ajustamento dos resíduos, denotam em algumas variáveis um comportamento cíclico em torno de zero, denotando algum afastamento em relação à normal, isto é menos evidente nos gráficos de ajustamento em que os pontos estão mais próximos da diagonal. Estes resultados coincidem com os testes de hipóteses de aderência obtidos através do Teste de Kolmogorov-Smirnov, ao nível de 5% de significância.

O *TIAF* baseado na normalidade das distribuições por amostragem das estimativas somente teria sentido no método *jackknife*, para o Caso Creme Dental naquelas variáveis onde a normalidade é garantida.

Os intervalos de confiança construídos para as estimativas geradas, tanto no *bootstrap* quanto no *jackknife*, continuarão sendo os percentílicos, para manter a uniformidade de procedimento em todas as variáveis da base e facilitar a análise de comparativa.

### 6.3. TESTE DE NORMALIDADE DA BASE “MODO DE VIDA”

#### 6.3.1. Teste de Normalidade das Variáveis no Método *Bootstrap*:

**Tabela 8**

#### **Caso 3 : Modo de Vida**

**Resultado do Teste de Normalidade para o Caso 3 das Estimativas *Bootstrap* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2**

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,000	Nnormal	0,000	Nnormal
V2	0,000	Nnormal	0,000	Nnormal
V3	0,000	Nnormal	0,000	Nnormal
V4	0,000	Nnormal	0,000	Nnormal
V5	0,000	Nnormal	0,000	Nnormal
V6	0,000	Nnormal	0,000	Nnormal
V7	0,000	Nnormal	0,000	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal=rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

Pela observação do Gráfico de Probabilidades se observa que para todas as variáveis no fator 1 o ajuste à curva normal não pode ser considerado, uma vez evidenciado pontos das variáveis distanciados da diagonal. Os gráficos do ajustamento dos resíduos, no fator 1, confirmam os resultados indicados pelos gráficos de ajustamento à normal: para a grande maioria das variáveis existe um certo comportamento tendência e cíclico para o comportamento dos resíduos, se distanciando do que seria esperado: uma nuvem aleatória e espaçada de pontos. Estes resultados coincidem com os testes de hipóteses de aderência obtidos através do Teste de Kolmogorov-Smirnov, ao nível de 5% de significância.

No fator 2, constatam-se resultados análogos ao fator 1, com a exceção para a variável 3, que diverge do resultado de teste de aderência de Kolmogorov-Smirnov (distribuição não normal, ao nível 5 % de significância), parecendo tanto pelo gráfico de ajustamento à normal quanto no gráfico do Ajustamento dos resíduos haver ajustamento à distribuição gaussiana.

### 6.3.2. Teste de Normalidade das Variáveis no Método *Jackknife*

**Tabela 9**

#### **Caso 3: Modo de Vida**

#### **Resultado do Teste de Normalidade Para o Caso 3 das Estimativas *Jackknife* pelo Método de Kolmogorov-Smirnov dos Fatores 1 e 2**

Variáveis	Fator 1		Fator 2	
	Valor-p	Resultado	Valor-p	Resultado
V1	0,97	Normal	0,15	Normal
V2	0,99	Normal	0,00	Nnormal
V3	0,74	Normal	0,00	NNormal
V4	0,59	Normal	0,01	Nnormal
V5	0,76	Normal	0,03	Nnormal
V6	1,00	Normal	0,00	Nnormal
V7	0,76	Normal	0,04	Nnormal

Nota: Teste ao nível de 5% de significância e Normal =aceita-se a hipótese nula de que as cargas fatoriais estimadas seguem o modelo Normal e Nnormal=rejeita-se a hipótese nula de que as estimativas sejam normalmente distribuídas.

No método *jackknife*, no fator 1, observa-se do Gráfico de Probabilidades que a maioria das variáveis se ajusta à curva normal, uma vez evidenciado pontos das variáveis próximos à diagonal (com exceção



para a variável  $V_1$ ). Os gráficos do ajustamento dos resíduos, no fator 1, confirmam os resultados indicados pelos gráficos de ajustamento à normal: para a grande maioria das variáveis existe um certo comportamento aleatório e espaçado dos pontos. Estes resultados praticamente coincidem com os testes de hipóteses de aderência de Teste de Kolmogorov-Smirnov, ao nível de 5% de significância.

No fator 2, constatam-se resultados opostos ao fator 1: a maioria das variáveis não parece ser ajustar à curva normal, com exceção para a variável  $V_1$ , que apresenta no gráfico de ajustamento à normal pontos toleravelmente próximos à diagonal e comportamento razoavelmente não tendencial e aleatório dos pontos dos resíduos.

O TIAF baseado na normalidade das distribuições por amostragem das estimativas somente teria sentido no método *jackknife* para o Caso 3, no fator 1, mas vale lembrar que a qualidade da distribuição por amostragem das cargas fatoriais nas componentes principais 1 e 2 na técnica *jackknife*, neste estudo de caso, é inferior ao do método *bootstrap*, sendo então recomendável trabalhar com os intervalos de confiança do *bootstrap* ao utilizar o TIAF, ao invés do intervalo de confiança *jackknife*, isto é, utilizar-se do intervalo de confiança percentílico, uma vez que o modelo gaussiano não é garantido à distribuição das estimativas *bootstrap*. Uma outro motivo que se leva a adotar o intervalo de confiança percentílico é que não existe uma uniformidade no ajuste à normal entre as variáveis, entre os fatores e entre os estudos de caso, o que dificultaria a análise comparativa inferencial dos resultados e a homogeneidade do uso de modelos.

#### **6.4.PROCEDIMENTOS METODOLÓGICOS A SEREM SEGUIDOS COM BASE NOS TESTES DE NORMALIDADES**

Verificou-se pelos testes de normalidades realizados que não existe uma regularidade e uniformidade na aderência das distribuições amostrais das cargas fatoriais ao modelo normal: o resultado foi específico para certo estudo de caso, para certo fator e para certa variável. Portanto, para se manter uma uniformidade na construção dos intervalos de confiança para as diversas variáveis e para os diversos problemas, será utilizada a estimação percentílica.

# CAPÍTULO 7

## TESTE DE CONVERGÊNCIA DOS RESULTADOS

### *BOOTSTRAP*

Para o cálculo do EMQ geralmente é suficiente um valor de  $B=100$ . Por outro lado, para se determinar a distribuição por amostragem com qualidade deve considerar-se um valor para  $B$  substancialmente mais elevado. Geralmente  $B=1000$  proporciona bons resultados. Em ambos os casos, convém ensaiar diferentes valores para  $B$  até se verificar a convergência dos resultados. Este é o objetivo deste capítulo.

Trata-se aqui da inferência avaliada empiricamente. Ela tem por objetivo estabelecer um modelo adequado à interpretação da distribuição por amostragem da variável aleatória carga fatorial com base na experiência. Vê-se que, neste caso, a experiência é a base para se montar o modelo inferencial ou ajustá-lo ao modelo ideal (teórico).

Para atingir os objetivos deste capítulo, considera-se somente a técnica do *bootstrap*. Serão ensaiados 4 valores para  $B$ , número de reamostragens :  $B=25$  ,  $B=100$  ,  $B=200$  e  $B=1000$  e verificar a convergência dos resultados. O referido teste de convergência será realizado somente para as seguintes bases estudadas neste texto: a busca de fatores, a creme dental e a modo de vida. Estas bases funcionarão como uma amostra do comportamento da convergência das estatísticas EMQ e INTERVALOS DE CONFIANÇA em função das variáveis dos problemas. Portanto, raciocínio indutivo poderá ser feito com o referido comportamento de outras bases de dados utilizadas nesta tese.

**Tabela 10****Caso 1 : Base Busca de Fatores-EMQ-Fator 1**

<b>Variáveis</b>	<b>B=25</b>	<b>B=100</b>	<b>B=200</b>	<b>B=1000</b>
V1	1,08E-02	1,12E-02	1,12E-02	1,16E-02
V2	8,75E-03	9,00E-03	9,00E-03	9,46E-03
V3	9,93E-03	5,15E-03	5,15E-03	5,43E-03
V4	2,95E-03	1,65E-03	1,65E-03	1,33E-03
V5	2,86E-03	2,02E-03	2,02E-03	1,91E-03
V6	8,55E-03	6,01E-03	6,01E-03	5,58E-03
V7	1,43E-03	2,04E-03	2,04E-03	2,15E-03
V8	4,14E-03	5,11E-03	5,11E-03	4,99E-03
V9	5,95E-03	5,52E-03	5,52E-03	6,72E-03
V10	8,20E-03	6,07E-03	6,07E-03	5,76E-03
V11	5,88E-03	5,45E-03	5,45E-03	5,30E-03
V12	1,23E-02	1,34E-02	1,34E-02	1,51E-02
V13	6,60E-03	7,24E-03	7,24E-03	6,05E-03
V14	9,45E-03	8,11E-03	8,11E-03	7,17E-03
V15	1,42E-03	1,20E-03	1,20E-03	1,31E-03
V16	6,06E-03	3,70E-03	3,70E-03	3,62E-03
V17	6,50E-03	3,88E-03	3,88E-03	3,77E-03
V18	3,11E-03	3,84E-03	3,84E-03	4,64E-03
V19	1,93E-03	1,81E-03	1,81E-03	1,91E-03
V20	6,16E-03	5,18E-03	5,18E-03	5,52E-03
V21	1,32E-02	1,33E-02	1,33E-02	1,61E-02
V22	1,44E-03	1,84E-03	1,84E-03	1,80E-03
V23	6,02E-03	6,25E-03	6,25E-03	1,16E-02

**Tabela 11****Caso 1 : Base Busca de Fatores-EMQ-Fator 2**

<b>Variáveis</b>	<b>B=25</b>	<b>B=100</b>	<b>B=200</b>	<b>B=1000</b>
V1	7,79E-03	8,80E-03	8,44E-03	6,93E-03
V2	8,14E-03	1,18E-02	1,28E-02	8,06E-03
V3	1,11E-02	9,81E-03	1,02E-02	1,31E-02
V4	5,75E-03	5,58E-03	5,28E-03	8,98E-03
V5	6,97E-03	7,77E-03	7,76E-03	5,87E-03
V6	1,05E-02	8,81E-03	9,18E-03	7,20E-03
V7	6,75E-03	9,41E-03	8,06E-03	9,07E-03
V8	1,44E-02	1,28E-02	1,20E-02	8,34E-03
V9	1,80E-02	1,42E-02	1,49E-02	1,11E-02
V10	8,14E-03	8,07E-03	7,28E-03	1,41E-02
V11	7,72E-03	8,38E-03	7,62E-03	6,87E-03
V12	6,53E-03	3,55E-03	3,00E-03	7,20E-03
V13	1,40E-02	9,47E-03	8,75E-03	3,30E-03
V14	9,86E-03	6,61E-03	7,13E-03	9,37E-03
V15	5,57E-03	4,05E-03	4,86E-03	7,80E-03
V16	1,13E-02	8,64E-03	7,32E-03	4,74E-03
V17	1,32E-02	1,13E-02	1,04E-02	7,70E-03
V18	7,09E-03	8,09E-03	8,04E-03	9,47E-03
V19	8,71E-03	7,49E-03	7,12E-03	8,17E-03
V20	1,20E-02	1,03E-02	9,55E-03	6,17E-03
V21	5,90E-03	3,35E-03	2,69E-03	9,67E-03
V22	1,02E-02	7,22E-03	8,07E-03	3,03E-03
V23	8,03E-03	7,49E-03	8,86E-03	8,17E-03

**Tabela 12****Caso 1 : Base Busca de Fatores-Intervalo de Confiança-Fator 1**

Variáveis	B=25		B=100		B=200		B=1000	
	LI	LS	LI	LS	LI	LS	LI	LS
V1	-0,28	0,10	-0,28	0,10	-0,33	0,10	-0,31	0,10
V2	0,11	0,49	0,11	0,46	0,08	0,49	0,10	0,47
V3	0,26	0,62	0,32	0,62	0,29	0,62	0,33	0,62
V4	0,60	0,80	0,60	0,77	0,61	0,78	0,63	0,78
V5	0,58	0,78	0,59	0,76	0,58	0,76	0,59	0,76
V6	0,25	0,65	0,33	0,61	0,31	0,61	0,33	0,61
V7	0,58	0,72	0,55	0,72	0,54	0,72	0,55	0,73
V8	0,46	0,68	0,39	0,68	0,39	0,68	0,41	0,69
V9	0,29	0,57	0,29	0,57	0,25	0,57	0,27	0,58
V10	0,38	0,73	0,46	0,75	0,45	0,76	0,45	0,75
V11	0,47	0,72	0,47	0,75	0,48	0,75	0,46	0,75
V12	-0,07	0,34	-0,16	0,33	-0,18	0,32	-0,17	0,30
V13	0,29	0,59	0,23	0,57	0,21	0,57	0,25	0,55
V14	0,34	0,68	0,31	0,65	0,26	0,64	0,30	0,64
V15	0,70	0,84	0,69	0,84	0,69	0,84	0,70	0,84
V16	0,51	0,83	0,53	0,80	0,56	0,80	0,56	0,79
V17	0,48	0,76	0,52	0,76	0,53	0,77	0,53	0,77
V18	0,51	0,72	0,49	0,72	0,46	0,72	0,45	0,72
V19	0,58	0,73	0,55	0,72	0,55	0,73	0,56	0,73
V20	0,25	0,58	0,24	0,51	0,24	0,52	0,23	0,52
V21	-0,09	0,41	-0,11	0,36	-0,14	0,36	-0,13	0,36
V22	0,61	0,77	0,60	0,77	0,60	0,77	0,60	0,77
V23	0,26	0,55	0,25	0,55	0,22	0,56	0,23	0,56

**Tabela 13****Caso 1 : Base Busca de Fatores-Intervalo de Confiança-Fator 2**

Variáveis	B=25		B=100		B=200		B=1000	
	LI	LS	LI	LS	LI	LS	LI	LS
V1	0,29	0,55	0,24	0,59	0,27	0,60	0,26	0,60
V2	0,13	0,48	0,11	0,50	0,08	0,52	0,08	0,53
V3	-0,04	0,35	-0,01	0,35	-0,02	0,37	-0,02	0,35
V4	-0,13	0,15	-0,13	0,17	-0,11	0,17	-0,12	0,19
V5	-0,27	0,12	-0,22	0,12	-0,22	0,16	-0,18	0,15
V6	0,05	0,49	0,06	0,48	0,06	0,48	0,06	0,43
V7	-0,15	0,17	-0,18	0,23	-0,18	0,21	-0,16	0,22
V8	-0,08	0,40	-0,07	0,40	-0,06	0,40	-0,04	0,38
V9	-0,04	0,40	0,00	0,45	0,00	0,47	-0,01	0,43
V10	-0,74	-0,30	-0,65	-0,31	-0,65	-0,31	-0,67	-0,36
V11	-0,71	-0,29	-0,64	-0,29	-0,63	-0,29	-0,65	-0,32
V12	0,50	0,79	0,58	0,81	0,60	0,80	0,59	0,80
V13	0,08	0,57	0,12	0,50	0,14	0,51	0,12	0,49
V14	0,28	0,63	0,35	0,68	0,35	0,68	0,32	0,67
V15	-0,41	-0,17	-0,41	-0,15	-0,41	-0,12	-0,40	-0,13
V16	-0,67	-0,25	-0,62	-0,25	-0,57	-0,27	-0,59	-0,27
V17	-0,65	-0,20	-0,61	-0,18	-0,58	-0,18	-0,57	-0,20
V18	0,05	0,31	-0,03	0,31	-0,03	0,33	-0,04	0,33
V19	-0,11	0,31	0,00	0,34	0,00	0,34	0,01	0,32
V20	0,01	0,40	0,02	0,40	0,02	0,41	0,00	0,38
V21	0,52	0,81	0,61	0,84	0,62	0,82	0,60	0,82
V22	-0,23	0,23	-0,23	0,12	-0,20	0,14	-0,20	0,17
V23	0,24	0,54	0,23	0,53	0,21	0,59	0,21	0,57

**Tabela 14****Caso 2 : Base Creme Dental-EMQ-Fator 1**

Variáveis	B=25	B=100	B=200	B=1000
V1	9,14E-05	3,65E-05	1,82E-05	3,62E-06
V2	1,23E-02	3,98E-03	1,98E-03	3,94E-04
V3	1,76E-04	4,59E-05	2,28E-05	4,55E-06
V4	1,01E-02	3,22E-03	1,60E-03	3,20E-04
V5	3,46E-04	1,20E-04	5,96E-05	1,19E-05
V6	1,16E-02	3,80E-03	1,89E-03	3,76E-04

**Tabela 15****Caso 2 : Base Creme Dental-EMQ-Fator 2**

Variáveis	B=25	B=100	B=200	B=1000
V1	9,04E-03	2,96E-03	1,47E-03	2,93E-04
V2	2,24E-04	7,85E-05	3,91E-05	7,79E-06
V3	9,42E-03	3,05E-03	1,52E-03	3,02E-04
V4	1,05E-03	3,57E-04	1,78E-04	3,54E-05
V5	7,98E-03	2,57E-03	1,28E-03	2,55E-04
V6	4,06E-04	1,27E-04	6,30E-05	1,26E-05



**Tabela 16****Caso 2 : Base Creme Dental-Intervalo de Confiança-Fator 1**

Variáveis	B=25		B=100		B=200		B=1000	
	LI	LS	LI	LS	LI	LS	LI	LS
V1	0,92	0,95	0,93	0,95	0,94	0,95	0,95	0,95
V2	-0,27	0,20	-0,19	0,15	-0,17	0,06	-0,05	-0,05
V3	0,88	0,94	0,90	0,93	0,92	0,93	0,93	0,93
V4	-0,43	0,00	-0,37	-0,10	-0,36	-0,16	-0,24	-0,24
V5	-0,93	-0,85	-0,92	-0,88	-0,92	-0,88	-0,91	-0,91
V6	-0,14	0,33	-0,09	0,26	-0,05	0,16	0,08	0,08

**Tabela 17****Caso 2 : Base Creme Dental-Intervalo de Confiança-Fator 2**

Variáveis	B=25		B=100		B=200		B=1000	
	LI	LS	LI	LS	LI	LS	LI	LS
V1	-0,16	0,25	-0,08	0,22	-0,01	0,18	0,07	0,07
V2	0,79	0,85	0,81	0,85	0,81	0,84	0,84	0,84
V3	-0,30	0,13	-0,22	0,07	-0,14	0,05	-0,06	-0,06
V4	0,75	0,88	0,78	0,86	0,79	0,86	0,83	0,83
V5	-0,36	0,00	-0,30	-0,03	-0,29	-0,10	-0,18	-0,18
V6	0,80	0,88	0,83	0,88	0,85	0,87	0,87	0,87

**Tabela 18****Caso 3 : Base Modo de Vida-EMQ-Fator 1**

<b>Variáveis</b>	<b>B=25</b>	<b>B=100</b>	<b>B=200</b>	<b>B=1000</b>
V1	1,18E-03	2,86E-04	1,42E-04	2,83E-05
V2	1,03E-02	2,50E-03	1,24E-03	2,48E-04
V3	1,03E-04	2,50E-05	1,24E-05	2,48E-06
V4	7,44E-03	1,80E-03	8,97E-04	1,79E-04
V5	2,36E-03	5,73E-04	2,85E-04	5,68E-05
V6	1,15E-02	2,79E-03	1,39E-03	2,76E-04
V7	1,96E-03	4,75E-04	2,36E-04	4,71E-05

**Tabela 19****Caso 3 : Base Modo de Vida-EMQ-Fator 2**

<b>Variáveis</b>	<b>B=25</b>	<b>B=100</b>	<b>B=200</b>	<b>B=1000</b>
V1	2,15E-02	5,22E-03	2,60E-03	5,18E-04
V2	9,88E-02	2,40E-02	1,19E-02	2,38E-03
V3	4,25E-03	1,03E-03	5,12E-04	1,02E-04
V4	6,05E-02	1,47E-02	7,31E-03	1,46E-03
V5	2,93E-02	7,11E-03	3,54E-03	7,05E-04
V6	7,14E-02	1,73E-02	8,63E-03	1,72E-03
V7	2,59E-02	6,28E-03	3,12E-03	6,22E-04

**Tabela 20****Caso 3 : Base Modo de Vida-Intervalo de Confiança-Fator 1**

Variáveis	B=25		B=100		B=200		B=1000	
	LI	LS	LI	LS	LI	LS	LI	LS
V1	0,75	0,89	0,76	0,87	0,78	0,85	0,82	0,82
V2	0,12	0,50	0,14	0,43	0,16	0,38	0,28	0,28
V3	0,87	0,91	0,87	0,90	0,88	0,89	0,89	0,89
V4	-0,35	0,02	-0,32	-0,03	-0,27	-0,15	-0,20	-0,20
V5	0,60	0,79	0,61	0,78	0,62	0,70	0,66	0,66
V6	-0,18	0,27	-0,09	0,22	-0,04	0,15	0,05	0,05
V7	-0,77	-0,59	-0,74	-0,59	-0,72	-0,64	-0,68	-0,68

**Tabela 21****Caso 3 : Base Modo de Vida-Intervalo de Confiança-Fator 2**

Variáveis	B=25		B=100		B=200		B=1000	
	LI	LS	LI	LS	LI	LS	LI	LS
V1	-0,26	0,50	0,23	0,47	0,31	0,46	0,38	0,38
V2	-0,80	0,80	-0,78	-0,60	-0,74	-0,64	-0,71	-0,71
V3	-0,18	0,09	-0,13	0,08	-0,10	0,02	-0,03	-0,03
V4	-0,54	0,71	0,51	0,69	0,55	0,67	0,63	0,63
V5	-0,24	0,63	0,26	0,62	0,44	0,58	0,51	0,51
V6	-0,72	0,67	-0,69	-0,49	-0,65	-0,53	-0,60	-0,60
V7	-0,34	0,52	0,23	0,49	0,32	0,44	0,38	0,38

Pelos ensaios das convergências dos casos apresentados conclui-se que com um  $B=200$  já se tem resultados bem próximos quando se trabalha com  $B \rightarrow \infty$ , isto é,  $B=1000$ , proporcionando economia de tempo e esforço na computação.

Um observação importante é que mesmo com  $B$  pequeno, em torno de 25 reamostragens, os EMQ'S já são bem reduzidos em todos os casos apresentados.

Assim para reamostragens suficientemente grandes ( $B=1000$ ) os resultados do EMQ e dos intervalos de confiança pouco diferem entre si. É o que se chama a ***Regularidade Estatística dos Resultados*** ou a “força” da ***Lei dos Grandes Números***.

O valor hipotético fixo no qual tende a haver uma estabilização dos EMQ'S e dos intervalos de confiança é que efetivamente estabelem-se como EMQ'S e intervalos de confiança obtidos através da distribuição empírica real obtida.

## **7.1.PROCEDIMENTOS METODOLÓGICOS A SEREM SEGUIDOS COM BASE NOS TESTES DE CONVERGÊNCIAS**

Os resultados encontrados para as bases consideradas no teste de convergência poderiam ser generalizados às outras bases de dados desta tese, por indução.

Nas três primeiras bases de dados estudadas, no método *bootstrap*, utilizou-se um número de reamostragens igual a 200 ( $B=200$ ), nas demais bases (são seis ao todo, ver capítulo 8) utilizou-se neste método um  $B=1000$ . Este procedimento permite verificar se realmente os resultados com  $B=200$  é realmente de mesma qualidade que os resultados de  $B=1000$ ,

como os referidos testes de convergência para as três primeiras bases detectaram.

No próximo capítulo, serão realizados os TIAF'S para os casos selecionados neste trabalho com base nos procedimentos assumidos nos capítulos anteriores.

# CAPÍTULO 8

## ESTUDOS DE CASOS: APLICAÇÕES DO TIAF

Como ilustração do desempenho do *bootstrap* e *jackknife* elaborou-se exemplos reais de pesquisa por análise fatorial com componentes principais, onde a matriz de fatores foi rotada, segundo método varimax, o que mantém os fatores ortogonais ou independentes entre si.

O cálculo de estimativas do desvio-padrão e enfiamento destas estatísticas, assim como a determinação da sua distribuição por amostragem, só é possível com métodos como *bootstrap* e o *jackknife*, dado que o desconhecimento das respectivas expressões teóricas invalida a aplicação da estimação paramétrica tradicional.

A aplicação do *bootstrap* e *jackknife* foi feita de acordo com os algoritmos e programas descritos no texto. No *bootstrap*, consideram-se 200 ou 1000 réplicas de cada amostra, isto é,  $B=200$  e  $B=1000$  e nível de significância de 5%.

Neste capítulo, serão aplicados aos estudos de casos o “*Teste Inferencial para Análise Fatorial*”, tanto pelo intervalo de confiança quanto pelo valor-p. No próximo capítulo, será aplicado a alguns estudos de casos o método tradicional “*Teste de Hair-Anderson-Thatam-Black*” e será realizada uma comparação entre os dois métodos de significância de cargas fatoriais: o TIAF e o referido neste parágrafo.

Até o presente estágio da pesquisa foram estudados 6 casos:

- **Caso 1: Base Busca de Fatores;**
- **Caso 2: Base Creme Dental;**

- *Caso 3: Base Modo de Vida;*
- *Caso 4: Base Qualidade Total;*
- *Caso 5: Base Satisfação Body-Up;*
- *Caso 6: Base População de Condados Americanos.*

Por motivo de sintetização, será exemplificada a aplicação do Método TIAF completo somente aos casos 1 e 2. O método TIAF aplicado à Base Busca de Fatores em parte foi desenvolvido no capítulo 3 e em parte neste próprio capítulo. Procedimentos análogos podem ser desenvolvidos para as outras bases de dados. Nos demais casos, a aplicação do método irá somente até a obtenção da significância estatística e da significância prática, tanto pelo intervalo de confiança, quanto para o valor-p.

Nos três primeiros estudos de casos, a análise inferencial será mais detalhada e dividida para que haja uma maior familiarização com o tipo de comportamento da variável aleatória carga fatorial por variáveis dos problemas e por fator ou componente principal nas reamostragens.

A partir do estudo de caso 4, a análise inferencial será realizada com fundamentação em uma única tabela de resultados, tabela esta resultante do *input* direto dos programas em versão **R** ou versão **SAS**. A partir destes casos também o número de reamostragens realizadas será de 1000 ( $B=1000$ ).

## **8.1. CASO 1: BASE BUSCA DE FATORES.**

Detalhes sobre esta base de dados se encontram a partir da página 35 do capítulo 3.

### **8.1.1. Apresentação dos Resultados Computacionais para o *TIAFIC***

A análise fatorial é aplicada pelos pesquisadores sociais com grande freqüência, mas sempre a nível descritivo, já que o desconhecimento da distribuição por amostragem teórica (mesmo quando a distribuição da população é conhecida) torna inviável fazer acompanhar as estimativas do respectivo erro padrão, já para não falar na construção de intervalos de confiança ou na realização de ensaios de significância. Os métodos *CIS* permitem ladear a insuficiência da teoria da amostragem que se faz no presente caso, obtendo empiricamente as estimativas necessárias para a inferência estatística.

Para efeito, obtiveram-se em computador, segundo os programas dos Anexos 1 e 2, os resultados estatísticos que constam das Tabelas 22, 23, 24, 25, 26, e 27.



**Tabela 22**  
**(Caso 1:Busca de fatores)**

**Distribuição por Amostragem *Bootstrap* das Cargas Fatoriais do Fator 1 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Bootstrap</i>			
		Média	Variância	Viés	EMQ
V1	-1,08E-01	-1,24E-01	1,23E-02	1,55E-02	1,25E-02
V2	3,11E-01	3,05E-01	9,82E-03	6,10E-03	9,85E-03
V3	4,98E-01	4,94E-01	5,77E-03	3,17E-03	5,78E-03
V4	7,08E-01	7,06E-01	1,52E-03	2,91E-03	1,53E-03
V5	6,83E-01	6,84E-01	2,21E-03	-1,71E-03	2,21E-03
V6	4,89E-01	4,82E-01	6,28E-03	7,33E-03	6,34E-03
V7	6,46E-01	6,45E-01	2,24E-03	1,83E-03	2,25E-03
V8	5,73E-01	5,67E-01	5,22E-03	6,08E-03	5,25E-03
V9	4,48E-01	4,39E-01	7,39E-03	9,37E-03	7,48E-03
V10	6,30E-01	6,35E-01	5,85E-03	-4,64E-03	5,87E-03
V11	6,38E-01	6,42E-01	5,12E-03	-4,13E-03	5,14E-03
V12	6,87E-02	5,65E-02	1,58E-02	1,22E-02	1,59E-02
V13	4,20E-01	4,12E-01	7,46E-03	8,49E-03	7,53E-03
V14	4,91E-01	4,83E-01	8,30E-03	8,78E-03	8,38E-03
V15	7,83E-01	7,87E-01	1,34E-03	-3,78E-03	1,35E-03
V16	6,97E-01	7,01E-01	3,45E-03	-3,80E-03	3,47E-03
V17	6,69E-01	6,70E-01	3,48E-03	-8,44E-04	3,48E-03
V18	6,01E-01	6,01E-01	4,26E-03	-3,92E-04	4,26E-03
V19	6,53E-01	6,50E-01	1,87E-03	2,59E-03	1,88E-03
V20	3,93E-01	3,80E-01	5,10E-03	1,27E-02	5,26E-03
V21	1,35E-01	1,20E-01	1,56E-02	1,45E-02	1,58E-02
V22	6,91E-01	6,91E-01	1,75E-03	-1,48E-04	1,75E-03
V23	-1,08E-01	-1,24E-01	1,23E-02	4,75E-03	7,00E-03

**Tabela 23**  
**Caso 1: Busca de Fatores**

**Distribuição por Amostragem *Bootstrap* das Cargas Fatoriais do Fator 2 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Bootstrap</i>			
		Média	Variância	Viés	EMQ
V1	6,91E-01	6,91E-01	6,91E-01	-1,48E-04	1,75E-03
V2	4,16E-01	4,11E-01	4,16E-01	4,75E-03	7,00E-03
V3	4,59E-01	4,40E-01	4,59E-01	1,92E-02	8,44E-03
V4	3,18E-01	3,21E-01	3,18E-01	-2,13E-03	1,28E-02
V5	1,56E-01	1,50E-01	1,56E-01	5,50E-03	1,02E-02
V6	4,03E-02	3,94E-02	4,03E-02	9,08E-04	5,28E-03
V7	-1,41E-02	-1,81E-02	-1,41E-02	4,00E-03	7,76E-03
V8	2,49E-01	2,58E-01	2,49E-01	-8,55E-03	9,18E-03
V9	2,37E-02	2,61E-02	2,37E-02	-2,40E-03	8,06E-03
V10	1,62E-01	1,66E-01	1,62E-01	-3,94E-03	1,20E-02
V11	2,22E-01	2,18E-01	2,22E-01	4,06E-03	1,49E-02
V12	-5,21E-01	-5,05E-01	-5,21E-01	-1,63E-02	7,28E-03
V13	-4,94E-01	-4,77E-01	-4,94E-01	-1,65E-02	7,62E-03
V14	7,24E-01	7,13E-01	7,24E-01	1,11E-02	3,00E-03
V15	3,26E-01	3,36E-01	3,26E-01	-9,65E-03	8,75E-03
V16	5,05E-01	5,02E-01	5,05E-01	3,06E-03	7,13E-03
V17	-2,66E-01	-2,48E-01	-2,66E-01	-1,78E-02	4,86E-03
V18	-4,38E-01	-4,16E-01	-4,38E-01	-2,22E-02	7,32E-03
V19	-4,07E-01	-3,86E-01	-4,07E-01	-2,10E-02	1,04E-02
V20	1,36E-01	1,43E-01	1,36E-01	-6,95E-03	8,04E-03
V21	1,65E-01	1,77E-01	1,65E-01	-1,15E-02	7,12E-03
V22	2,05E-01	2,08E-01	2,05E-01	-2,61E-03	9,55E-03
V23	7,43E-01	7,36E-01	7,43E-01	6,87E-03	2,69E-03

**Tabela 24**  
**Caso 1: Busca de Fatores**

**Distribuição por Amostragem *Jackknife* das Cargas Fatoriais do Fator 1 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Jackknife</i>			
		Média	Variância	Viés	EMQ
V1	1,08E-01	-1,08E-01	5,99E-05	0,216102	0,04676
V2	3,11E-01	3,11E-01	4,28E-05	2,84E-05	4,28E-05
V3	4,98E-01	4,98E-01	2,3E-05	1,53E-05	2,3E-05
V4	7,08E-01	7,08E-01	6,57E-06	2,19E-06	6,57E-06
V5	6,83E-01	6,83E-01	9,42E-06	-3,2E-06	9,42E-06
V6	4,89E-01	4,89E-01	2,71E-05	2,36E-05	2,71E-05
V7	6,46E-01	6,46E-01	1,02E-05	2,06E-06	1,02E-05
V8	5,73E-01	5,73E-01	2,35E-05	1E-05	2,35E-05
V9	4,48E-01	4,48E-01	2,92E-05	2,22E-05	2,92E-05
V10	6,30E-01	6,30E-01	2,51E-05	1,38E-05	2,51E-05
V11	6,38E-01	6,38E-01	2,38E-05	1,83E-05	2,38E-05
V12	6,87E-02	6,87E-02	7,43E-05	1,95E-05	7,43E-05
V13	4,20E-01	4,20E-01	2,53E-05	2,23E-05	2,53E-05
V14	4,91E-01	4,91E-01	3,09E-05	2,13E-05	3,09E-05
V15	7,83E-01	7,83E-01	6,22E-06	4,59E-06	6,22E-06
V16	6,97E-01	6,97E-01	1,52E-05	1,19E-05	1,52E-05
V17	6,69E-01	6,69E-01	1,75E-05	1,4E-05	1,75E-05
V18	6,01E-01	6,01E-01	2,24E-05	9,59E-06	2,24E-05
V19	6,53E-01	6,53E-01	8,38E-06	9,54E-06	8,38E-06
V20	3,93E-01	3,93E-01	2,48E-05	2,01E-05	2,48E-05
V21	1,35E-01	1,35E-01	7,49E-05	2,28E-05	7,49E-05
V22	6,91E-01	6,91E-01	8,01E-06	4,92E-06	8,01E-06

Variáveis	Cargas da amostra original	<i>Jackknife</i>			
		Média	Variância	Viés	EMQ
V23	1,08E-01	-1,08E-01	5,99E-05	1,97E-05	3,23E-05

**Tabela 25**  
**Caso 1 : Busca de Fatores**

**Distribuição por Amostragem *Jackknife* das Cargas Fatoriais do Fator 2 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Jackknife</i>			
		Média	Variância	Viés	EMQ
V1	4,16E-01	4,16E-01	3,23E-05	5,23E-05	3,65E-05
V2	4,59E-01	4,59E-01	3,65E-05	1,72E-05	6,18E-05
V3	3,18E-01	3,18E-01	6,18E-05	5,57E-06	4,28E-05
V4	1,56E-01	1,56E-01	4,28E-05	-2,9E-05	2,71E-05
V5	4,03E-02	4,03E-02	2,71E-05	0,028266	0,000835
V6	1,41E-02	-1,41E-02	3,6E-05	3,68E-06	4,19E-05
V7	2,49E-01	2,49E-01	4,19E-05	-2,9E-05	4,1E-05
V8	2,37E-02	2,37E-02	4,1E-05	-2,7E-05	4,88E-05
V9	1,62E-01	1,63E-01	4,88E-05	1,61E-05	6,5E-05
V10	2,22E-01	2,22E-01	6,5E-05	1,042917	1,087706
V11	5,21E-01	-5,21E-01	3,08E-05	0,987886	0,975952
V12	4,94E-01	-4,94E-01	3,31E-05	6,92E-05	1,37E-05
V13	7,24E-01	7,24E-01	1,37E-05	2,82E-05	4,41E-05
V14	3,26E-01	3,26E-01	4,41E-05	2,76E-05	3,6E-05
V15	5,05E-01	5,05E-01	3,6E-05	0,531488	0,282501
V16	2,66E-01	-2,66E-01	2,19E-05	0,875398	0,766356
V17	4,38E-01	-4,38E-01	3,4E-05	0,814166	0,662908
V18	4,07E-01	-4,07E-01	4,05E-05	-2,4E-05	3,59E-05

Variáveis	Cargas da amostra original	Jackknife			
		Média	Variância	Viés	EMQ
V19	1,36E-01	1,36E-01	3,59E-05	-8,6E-06	2,88E-05
V20	1,65E-01	1,65E-01	2,88E-05	5,4E-07	4,11E-05
V21	2,05E-01	2,05E-01	4,11E-05	6,39E-05	1,32E-05
V22	7,43E-01	7,43E-01	1,32E-05	0,046586	0,002206
V23	2,33E-02	-2,33E-02	3,57E-05	1,3E-05	4,64E-05

**Tabela 26**  
**Caso 1 : Busca de Fatores**

**Intervalos de Confiança *Bootstrap* e *Jackknife* Percentílico de 95% e *TIAF* das Cargas Fatoriais por Variáveis do Problema da Primeira Componente Principal (Fator 1)**

Variáveis	Bootstrap		Jackknife	
	Intervalo de Confiança	TIAFIC	Intervalo de Confiança	TIAFIC
V1	-0,33; 0,11	NSIG	-0,13; -0,09	SIG
V2	0,08; 0,49	SIG	0,29; 0,32	SIG
V3	0,29; 0,62	SIG	0,49; 0,51	SIG
V4	0,61; 0,78	SIG	0,70; 0,71	SIG
V5	0,58; 0,76	SIG	0,68; 0,69	SIG
V6	0,31; 0,61	SIG	0,48; 0,50	SI
V7	0,54; 0,72	SIG	0,64; 0,65	SIG
V8	0,39; 0,68	SIG	0,57; 0,58	SIG
V9	0,25; 0,57	SIG	0,44; 0,46	SIG
V10	0,45; 0,76	SIG	0,62; 0,64	SIG
V11	0,48; 0,75	SIG	0,63; 0,65	SIG
V12	-0,18; 0,32	NSIG	0,05; 0,09	SIG
V13	0,21; 0,57	SIG	0,41; 0,43	SIG
V14	0,26; 0,64	SIG	0,48; 0,51	SIG

Variáveis	<i>Bootstrap</i>		<i>Jackknife</i>	
	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>
V15	0,68 ; 0,84	SIG	0,78 ; 0,79	SIG
V16	0,56 ; 0,80	SIG	0,69 ; 0,71	SIG
V16	0,53 ; 0,77	SIG	0,66 ; 0,68	SIG
V18	0,56; 0,72	SIG	0,59 ; 0,61	SIG
V19	0,55; 0,73	SIG	0,65 ; 0,66	SIG
V20	0,24 ; 0,52	SIG	0,38; 0,40	SIG
V21	-0,14; 0,36	NSIG	0,12 ; 0,16	SIG
V22	0,60 ; 0,77	SIG	0,69 ; 0,70	SIG
V23	0,22 ; 0,56	SIG	0,41 ; 0,43	SIG

Nota : SIG=significante e NSIG=não significante

**Tabela 27**  
**Caso 1 : Busca de Fatores**

**Intervalos de Confiança *Bootstrap* e *Jackknife* Percentílico de 95% e *TIAF* das Cargas Fatoriais por Variáveis do Problema da Primeira Componente Principal (Fator 2)**

Variáveis	<i>Bootstrap</i>		<i>Jackknife</i>	
	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>
V1	0,27 ; 0,60	SIG	0,45 ; 0,47	SIG
V2	0,08 ; 0,52	SIG	0,30 ; 0,34	SIG
V3	-0,02 ; 0,37	NSIG	0,14 ; 0,17	SIG
V4	-0,11 ; 0,17	NSIG	0,03 ; 0,05	SIG
V5	-0,22 ; 0,16	NSIG	-0,03 ; 0,00	SIG
V6	0,06 ; 0,48	SIG	0,23 ; 0,26	SIG
V7	-0,17 ; 0,21	NSIG	0,01 ; 0,04	SIG
V8	-0,05 ; 0,40	NSIG	0,14 ; 0,18	SIG
V9	0,00 ; 0,46	SIG	0,20 ; 0,24	SIG
V10	-0,65 ; -0,32	SIG	-0,53 ; -0,51	SIG
V11	-0,63 ; -0,29	SIG	-0,51 ; -0,48	SIG
V12	0,60 ; 0,80	SIG	0,72 ; 0,73	SIG

Variáveis	<i>Bootstrap</i>		<i>Jackknife</i>	
	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>
V13	0,14 ; 0,51	<b>SIG</b>	0,31 ; 0,34	<b>SIG</b>
V14	0,35 ; 0,68	<b>SIG</b>	0,49 ; 0,52	<b>SIG</b>
V15	-0,41 ; -0,12	<b>SIG</b>	-0,27 ; -0,25	<b>SIG</b>
V16	-0,57 ; -0,27	<b>SIG</b>	-0,45 ; -0,42	<b>SIG</b>
V17	-0,57 ; -0,17	<b>SIG</b>	-0,42 ; -0,39	<b>SIG</b>
V18	-0,03 ; 0,33	<b>NSIG</b>	0,12 ; 0,15	<b>SIG</b>
V19	0,00 ; 0,34	<b>NSIG</b>	0,15 ; 0,18	<b>SIG</b>
V20	0,02 ; 0,41	<b>SIG</b>	0,19 ; 0,22	<b>SIG</b>
V21	0,62 ; 0,82	<b>SIG</b>	0,74 ; 0,75	<b>SIG</b>
V22	-0,20 ; 0,14	<b>NSIG</b>	-0,03 ; -0,01	<b>SIG</b>
V23	0,21 ; 0,59	<b>SIG</b>	0,37 ; 0,41	<b>SIG</b>

Nota : SIG=significante e NSIG=não significante

### **8.1.2.Comparação das Distribuições Amostrais entre os Métodos *Bootstrap* e *Jackknife***

Da leitura das Tabelas 24, 25, 26 e 27 se podem fazer as seguintes interpretações:

Tanto as estimativas *bootstrap* quanto as *jackknife* revelaram-se de variância pequena, o que indica que são precisas. Esta informação é válida para ambos os fatores 1 e 2. Contudo, no método *jackknife* a variância se revelou desprezível para todas as variáveis, o que não ocorreu com o método *bootstrap*. Isto indica que o estimador *jackknife* é mais eficiente que o método *bootstrap*.

Uma observação importante é que com base nos estudos de outras estatísticas, as estimativas *jackknife* tendem a superestimar a variância, enquanto que as estimativas *bootstrap* tendem a subestimá-la. Contudo, para o presente caso, isto evidentemente não se verificou em todas as variáveis.

Os envezamentos são desprezíveis para ambos os fatores e para ambos os métodos, e para a maioria das variáveis, o que já é de fato uma informação útil, pois revela que as estimativas *bootstrap* e *jackknife* são acuradas, não envezadas.

O conhecimento do erro-padrão permitiu construir intervalos de confiança e sair, conseqüentemente, do terreno puramente descritivo (ver Tabelas 26 e 27).

Os estimadores *bootstrap* e *jackknife* são precisos e acurados, o que implicou em um EMQ pequeno para a maioria das variáveis, portanto reúnem as qualidades essenciais de um bom estimador. Contudo, as estimativas *jackknife* são nitidamente superiores às estimativas *bootstrap* e



devem ser as escolhidas para inferências e tomadas de decisão posteriores na área de atuação do estudo e para nomeação e interpretação dos fatores.

### **8.1.3. TIAFIC**

O *TIAF* será realizado abordando somente o método *jackknife* que neste caso se mostrou de menor EMQ para ambos os fatores.

Observando as Tabelas 26 e 27, todas as variáveis do problema têm cargas significantes: estas variáveis guardam correlação estatisticamente significativa com as componentes principais em estudo (fatores 1 e 2), mas pelo sugerido nesta tese, somente as variáveis  $V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}, V_{11}, V_{13}, V_{14}, V_{15}, V_{16}, V_{17}, V_{18}, V_{19}, V_{20}, V_{21}, V_{22}$  e  $V_{23}$  podem participar da nomeação e interpretação da componente principal 1. Pelo mesmo motivo, somente as variáveis  $V_1, V_2, V_{10}, V_{11}, V_{12}, V_{13}, V_{14}, V_{16}, V_{17}, V_{21}$  e  $V_{23}$  podem participar da nomeação e interpretação da componente principal 2.

A escolha de variáveis que efetivamente vão participar da nomeação dos fatores fica, então, a cargo do analista e segue um critério subjetivo.

Uma vantagem do Método *TIAF*, evidenciado por este caso, é que mesmo em amostras grandes (tamanhos maiores que 200), as cargas fatoriais têm probabilidade de serem estatisticamente não significantes.

### 8.1.3.TIAFVP

#### 8.1.3.1.TIAFVP *Bootstrap*

A Tabela 28 apresenta os resultados do TIAF pelo Valor-p :

**Tabela 28**

VAR	Fator 1		Fator 2	
	Valor-p	TIAFVP	Valor-p	TIAFVP
V1	0,360	NSIG	0,000	SIG
V2	0,000	SIG	0,505	NSIG
V3	0,000	SIG	0,140	NSIG
V4	0,000	SIG	0,125	NSIG
V5	0,000	SIG	0,140	NSIG
V6	0,000	SIG	0,140	NSIG
V7	0,000	SIG	0,135	NSIG
V8	0,000	SIG	0,315	NSIG
V9	0,000	SIG	0,060	NSIG
V10	0,160	NSIG	0,000	SIG
V11	0,160	NSIG	0,000	SIG
V12	0,065	NSIG	0,015	SIG
V13	0,000	SIG	0,605	NSIG
V14	0,000	SIG	0,510	NSIG
V15	0,000	SIG	0,000	SIG
V16	0,090	NSIG	0,000	SIG
V17	0,075	NSIG	0,000	SIG
V18	0,000	SIG	0,275	NSIG
V19	0,000	SIG	0,190	NSIG
V20	0,000	SIG	0,190	NSIG
V21	0,035	SIG	0,010	SIG
V22	0,000	SIG	0,055	NSIG
V23	0,000	SIG	0,355	NSIG

No fator 1, as variáveis V<sub>2</sub>, V<sub>3</sub>, V<sub>4</sub>, V<sub>5</sub>, V<sub>6</sub>, V<sub>7</sub>, V<sub>8</sub>, V<sub>9</sub>, V<sub>13</sub>, V<sub>18</sub>, V<sub>19</sub>, V<sub>20</sub>, V<sub>21</sub>, V<sub>22</sub> e V<sub>23</sub> são significativas estatisticamente pelo valor-p *bootstrap*. Comparando com o TIAFIC, observa-se divergência nas variáveis V<sub>1</sub>, V<sub>10</sub>, V<sub>16</sub>, V<sub>17</sub> e V<sub>21</sub>.

No fator 2, as variáveis V<sub>1</sub>, V<sub>10</sub>, V<sub>11</sub>, V<sub>12</sub>, V<sub>15</sub>, V<sub>16</sub>, V<sub>17</sub> e V<sub>18</sub> são significativas estatisticamente pelo valor-p *bootstrap*. Comparando com o TIAFIC, observa-se somente divergência na variável V<sub>9</sub>.

Os resultados pelo valor-p tendem a ser mais precisos e confiáveis.

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

### 8.1.3.2. TIAFVP *Jackknife*

A Tabela 29. apresenta os resultados do TIAF pelo Valor-p :

**Tabela 29**

VAR	Fator 1		Fator 2	
	Valor-p	TIAFVP	Valor-p	TIAFVP
V1	0,000	SIG	0,000	SIG
V2	0,000	SIG	0,009	SIG
V3	0,000	SIG	0,000	SIG
V4	0,000	SIG	0,000	SIG
V5	0,000	SIG	0,000	SIG
V6	0,000	SIG	0,000	SIG
V7	0,000	SIG	0,000	SIG
V8	0,000	SIG	0,009	SIG
V9	0,000	SIG	0,000	SIG
V10	0,000	SIG	0,000	SIG
V11	0,000	SIG	0,000	SIG
V12	0,000	SIG	0,000	SIG
V13	0,000	SIG	0,071	NSIG
V14	0,000	SIG	0,000	SIG
V15	0,000	SIG	0,000	SIG
V16	0,000	SIG	0,000	SIG
V17	0,000	SIG	0,000	SIG
V18	0,000	SIG	0,009	SIG
V19	0,000	SIG	0,000	SIG

VAR	Fator 1		Fator 2	
	Valor-p	TIAFVP	Valor-p	TIAFVP
V20	0,000	SIG	0,000	SIG
V21	0,000	SIG	0,000	SIG
V22	0,000	SIG	0,000	SIG
V23	0,000	SIG	0,000	SIG

No fator 1, todas as variáveis são estatisticamente significativas pelo valor-p *jackknife*. Comparando com o TIAFIC, os resultados são coincidentes.

No fator 2, somente a variável  $V_{13}$  não é estatisticamente significativa pelo valor-p *jackknife*. O TIAFIC confirma tal resultado.

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

Os resultados pelo valor-p tendem a ser mais precisos e confiáveis.

No método *jackknife*, os resultados inferenciais se mostraram mais regulares entre os TIAFIC e o TIAFVP, esse resultado já era de ser esperar, uma vez que as distribuições por amostragem geradas no procedimento *jackknife* se revelaram de maior qualidade estatística (menores EMQ'S).

No capítulo 3, mais precisamente no item 3.6, colocou-se que o processo de análise estatística desenvolvido neste estudo de caso ficara incompleto, uma vez que a seguinte pergunta ficava sem resposta: os resultados encontrados na matriz de fatores são reais, válidos para a população como um todo ou somente pode ser atribuído à amostra colhida? Comentou-se, então que para responder a esta pergunta ter-se-ia que ter um procedimento conclusivo que pudesse, com eficácia, estabelecer a significância das cargas fatoriais da Tabela 2. É o que se acabou de fazer

nos itens anteriores deste capítulo: seguindo a lógica do TIAF *jackknife*, todas as cargas fatoriais são estatisticamente significantes no fator 1 e na maioria do fator 2. Para cada fator, somente as cargas significantes devem participar de sua nomeação/interpretação e posteriormente da significância prática. Agora se pode afirmar que o processo de análise fatorial se completou, ou melhor, que a análise estatística de dados atingiu o seu nível mais alto.

Esta lógica segue para todos os estudos de casos que se desenvolverá em seguida. No estudo de caso 2 (Base Creme Dental), além da significância estatística e prática, se processará a nomeação/interpretação dos fatores. Nos demais estudo de casos, por um critério de sintetização de procedimentos, a cada base de dados se apresentará somente o teste inferencial propriamente dito, isto é, as significâncias estatísticas e práticas das estimativas.

## 8.2. CASO 2: BASE CREME DENTAL

Suponha-se que um pesquisador queira avaliar os benefícios que os consumidores esperam da compra de um creme dental. Foi entrevistada em um supermercado uma amostra pré-teste de 40 pessoas, para que indicassem seu grau de concordância com as seguintes afirmações, utilizando uma escala de sete pontos (1=discordância total, 7=concordância total).

V<sub>1</sub>: É importante comprar um creme dental que evite cáries

V<sub>2</sub>: Gosto de um creme dental que clareie os dentes

V<sub>3</sub>: Um creme dental deve fortificar as gengivas

V<sub>4</sub>: Prefiro um creme dental que refresque o hálito

V<sub>5</sub>: Manter os dentes sadios não é uma vantagem importante de um creme dental

V<sub>6</sub>: O aspecto mais importante na compra de um creme dental é tornar os dentes atraentes

## 8.2.1. Apresentação dos Resultados Computacionais para o *TIAFIC*

**Tabela 30**

### **Caso 2 : Creme Dental**

**Distribuição por Amostragem *Bootstrap* das Cargas Fatoriais do Fator 1 para Cada Variável do Problema**

Variáveis	Cargas da Amostra Original	<i>Bootstrap</i>			
		Média	Variância	Viés	EMQ
V1	9,49E-01	9,49E-01	1,76E-05	7,22E-04	1,82E-05
V2	-5,08E-02	-5,01E-02	1,98E-03	-7,18E-04	1,98E-03
V3	9,29E-01	9,28E-01	2,21E-05	8,57E-04	2,28E-05
V4	-2,43E-01	-2,42E-01	1,60E-03	-8,91E-04	1,60E-03
V5	-9,08E-01	-9,08E-01	5,93E-05	-5,61E-04	5,96E-05
V6	8,27E-02	8,32E-02	1,76E-05	-5,67E-04	1,89E-03

**Tabela 31**

### **Caso 2 : Creme Dental**

**Distribuição por Amostragem *Bootstrap* das Cargas Fatoriais do Fator 2 para Cada Variável do Problema**

Variáveis	Cargas da Amostra Original	<i>Bootstrap</i>			
		Média	Variância	Viés	EMQ
V1	7,26E-02	7,18E-02	1,47E-03	7,63E-04	1,47E-03
V2	8,37E-01	8,36E-01	3,78E-05	1,11E-03	3,91E-05
V3	-6,31E-02	-6,37E-02	1,52E-03	6,38E-04	1,52E-03
V4	8,32E-01	8,31E-01	1,77E-04	7,74E-04	1,78E-04

Variáveis	Cargas da Amostra Original	Bootstrap			
		Média	Variância	Viés	EMQ
V5	-1,79E-01	-1,78E-01	1,28E-03	-8,08E-04	1,28E-03
V6	8,68E-01	8,67E-01	6,17E-05	1,16E-03	6,30E-05

**Tabela 32**

## Caso 2 : Creme Dental

### Distribuição por Amostragem *Jackknife* das Cargas Fatoriais do Fator 1 para Cada Variável do Problema

Variáveis	Cargas da Amostra Original	Jackknife			
		Média	Variância	Viés	EMQ
V1	9,49E-01	9,46E-01	7,93E-05	3,61E-03	9,24E-05
V2	-5,08E-02	-4,72E-02	1,01E-02	-3,59E-03	1,01E-02
V3	9,29E-01	9,25E-01	9,77E-05	4,29E-03	1,16E-04
V4	-2,43E-01	-2,38E-01	8,16E-03	-4,46E-03	8,18E-03
V5	-9,08E-01	-9,06E-01	2,96E-04	-2,80E-03	3,04E-04
V6	8,27E-02	8,55E-02	9,63E-03	-2,83E-03	9,64E-03



**Tabela 33**

**Caso 2 : Creme Dental**

**Distribuição por Amostragem *Jackknife* das Cargas Fatoriais do Fator 2 para Cada Variável do Problema**

Variáveis	Cargas da Amostra Original	<i>Jackknife</i>			
		Média	Variância	Viés	EMQ
V1	7,26E-02	6,88E-02	7,50E-03	3,82E-03	7,51E-03
V2	8,37E-01	8,32E-01	1,68E-04	5,54E-03	1,99E-04
V3	-6,31E-02	-6,63E-02	7,74E-03	3,19E-03	7,75E-03
V4	8,32E-01	8,28E-01	8,92E-04	3,87E-03	9,07E-04
V5	-1,79E-01	-1,75E-01	6,51E-03	-4,04E-03	6,53E-03
V6	8,68E-01	8,62E-01	2,87E-04	5,78E-03	3,21E-04

**Tabela 34**

**Caso 2: Creme Dental**

**Intervalos de Confiança *Bootstrap* e *Jackknife* Percentílico de 95% e *TIAF* das Cargas Fatoriais por Variáveis do Problema da Primeira Componente Principal (Fator 1)**

Variáveis	<i>Bootstrap</i>		<i>Jackknife</i>	
	Intervalo de Confiança	<i>TIAFIC</i>	Intervalo de Confiança	<i>TIAFIC</i>
V1	0,94 ; 0,95	SIG	0,92 ; 0,96	SIG
V2	-0,17 ; 0,06	NSIG	-0,26 ; 0,20	NSIG
V3	0,92 ; 0,93	SIG	0,88 ; 0,94	SIG
V4	-0,36; -0,16	SIG	-0,43 ; 0,00	SIG
V5	-0,92 ; -0,88	SIG	-0,93; -0,85	SIG
V6	-0,04; 0,16	NSIG	-0,14 ; 0,33	NSIG

Nota: SIG=significante e NSIG=não significante

**Tabela 35**

**Caso 2 : Creme Dental**

**Intervalos de Confiança *Bootstrap* e *Jackknife* Percentílico de 95% e *TIAF* das Cargas Fatoriais por Variáveis do Problema da Primeira Componente Principal (Fator 2)**

Variáveis	<i>Bootstrap</i>		<i>Jackknife</i>	
	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>	<i>Intervalo de Confiança</i>	<i>TIAFIC</i>
V1	-0,01 ; 0,18	NSIG	-0,16 ; 0,25	NSIG
V2	0,81 ; 0,84	SIG	0,79 ; 0,85	SIG
V3	-0,14 ; 0,04	NSIG	-0,30 ; 0,13	NSIG
V4	0,79 ; 0,86	SIG	0,75 ; 0,88	SIG
V5	-0,29 ; -0,10	SIG	-0,36 ; 0,00	NSIG
V6	0,85 ; 0,87	SIG	0,80 ; 0,90	SIG

Nota : SIG=significante e NSIG=não significante

**8.2.2.Comparação das Distribuições Amostrais entre os Métodos *Bootstrap* e *Jackknife***

Da leitura das Tabelas 30, 31, 32, e 33 se podem fazer as seguintes interpretações:

Tanto as estimativas *bootstrap* quanto as *jackknife* revelaram-se de variâncias praticamente nulas, o que indicam que estas estimativas são altamente precisas. Esta informação é válida para ambos os fatores 1 e 2. Os estimadores *bootstrap* e *jackknife* se mostraram igualmente eficientes.

Vale lembrar que com base nos estudos de outras estatísticas, as estimativas *jackknife* tendem a sobreestimar a variância, enquanto que as estimativas *bootstrap* tendem a subestimá-lo. Contudo, para o presente caso, isto também não se verificou.

Os enfiamentos são desprezíveis para ambos os fatores e para ambos os métodos, e para a maioria das variáveis, o que já é de fato uma informação útil, pois revela que as estimativas *bootstrap* e *jackknife* são acuradas, não enfiadas. O conhecimento do erro-padrão permitiu construir intervalos de confiança e sair, conseqüentemente, do terreno puramente descritivo (ver Tabelas 34 e 35).

Os estimadores *bootstrap* e *jackknife* são altamente precisos e acurados, o que implicou em um EMQ desprezível para a maioria das variáveis, portanto reúnem as qualidades essenciais de um bom estimador. Para a base considerada, as estimativas *bootstrap* e *jackknife* são igualmente eficientes, sendo as estimativas Bootstrap levemente superiores às estimativas *jackknife* no fator 2. Portanto, tanto faz se utilizar as estimativas *bootstrap* ou as estimativas *jackknife* para inferências e tomadas de decisão posteriores na área de atuação do estudo e para nomeação e interpretação dos fatores.

### 8.2.3. TIAFIC

Observando as Tabelas 34 e 35, os resultados do TIAF sugerem que:

- Para o **fator 1**, no método *bootstrap*, somente as variáveis  $V_1$ ,  $V_3$ ,  $V_4$  e  $V_5$  têm cargas fatoriais significantes. Contudo, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_1$ ,  $V_3$  e  $V_5$ ;
- Para o **fator 1**, no método *jackknife*, assim como no método *bootstrap*, somente as variáveis  $V_1$ ,  $V_3$ ,  $V_4$  e  $V_5$  têm cargas fatoriais significantes. Contudo, assim como no *bootstrap*, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_1$ ,  $V_3$  e  $V_5$ ;

- Para o **fator 2**, no método *bootstrap*, somente as variáveis  $V_2$ ,  $V_4$ ,  $V_5$  e  $V_6$  têm cargas fatoriais significantes. Contudo, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_2$ ,  $V_4$  e  $V_6$ ;
- Para o **fator 2**, no método *jackknife*, somente as variáveis  $V_2$ ,  $V_4$  e  $V_6$  têm cargas fatoriais significantes. Contudo, assim como no *bootstrap*, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_2$ ,  $V_4$  e  $V_6$ .

## 8.2.4. TIAFVP

### 8.2.4.1. TIAFVP *Bootstrap*

A Tabela 36 apresenta os resultados do TIAF pelo Valor-p:

**Tabela 36**

VAR	Fator 1		Fator 2	
	Valor-p	TIAFVP	Valor-p	TIAFVP
$V_1$	0,000	SIG	0,455	NSIG
$V_2$	0,385	NSIG	0,000	SIG
$V_3$	0,000	SIG	0,100	NSIG
$V_4$	0,105	NSIG	0,000	SIG
$V_5$	0,000	SIG	0,155	NSIG
$V_6$	0,145	NSIG	0,000	SIG

No fator 1, as variáveis  $V_1$ ,  $V_3$  e  $V_5$  são estatisticamente significantes no valor-p *bootstrap*. Contudo, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_1$ ,  $V_3$  e  $V_5$ . Este resultado é coerente com os obtidos pela inferência envolvendo os intervalos de confiança.

No fator 2, as variáveis  $V_2$ ,  $V_4$  e  $V_6$  são estatisticamente significantes no valor-p *bootstrap*. Contudo, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_2$ ,  $V_4$  e  $V_6$ . Este resultado é coerente com os obtidos pela inferência envolvendo os intervalos de confiança.

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

#### 8.2.4.2. TIAFVP *Jackknife*

**Tabela 37**

VAR	Fator 1		Fator 2	
	Valor-p	TIAFVP	Valor-p	TIAFVP
$V1$	0,000	SIG	0,300	NSIG
$V2$	0,200	NSIG	0,000	SIG
$V3$	0,000	SIG	0,000	SIG
$V4$	0,000	SIG	0,000	SIG
$V5$	0,000	SIG	0,000	SIG
$V6$	0,000	SIG	0,000	SIG

No fator 1, as variáveis  $V_1$ ,  $V_3$ ,  $V_4$  e  $V_5$  são estatisticamente significantes no valor-p *bootstrap*. Contudo, assim como no *bootstrap*, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_1$ ,  $V_3$  e  $V_5$ . Este resultado é coerente com os obtidos pela inferência envolvendo os intervalos de confiança.

No fator 2, as variáveis  $V_2$ ,  $V_3$ ,  $V_4$  e  $V_6$  são estatisticamente significantes no valor-p *bootstrap*. Contudo, assim como no *bootstrap*, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_2$ ,  $V_4$  e  $V_6$ . Este resultado é coerente com os obtidos pela inferência envolvendo os intervalos de confiança.

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

A escolha de variáveis que efetivamente vão participar da nomeação dos fatores fica, então, a cargo do analista ou pesquisador e segue um critério subjetivo.

Fica assim a nomeação\interpretação dos fatores neste estudo de caso:

- Em ambos os métodos (*bootstrap* e *jackknife*),  $V_1$ (prevenção de cáries),  $V_3$ (gengivas saudáveis) e um coeficiente negativo para  $V_5$ (dente sadio não é importante) têm TIAF'S significativos e significâncias práticas para o fator 1. Este fator pode ser rotulado de **fator de benefício à saúde**.

- Em ambos os métodos (*bootstrap* e *jackknife*),  $V_2$ (dentes claros),  $V_4$ (hálito puro) e  $V_6$ (dentes atraentes) têm TIAF'S significativos e significâncias práticas para o fator 2. Este fator pode ser rotulado de **fator de benefício social**.

### **8.3.CASO 3: BASE MODO DE VIDA**

Em um estudo da relação entre comportamento doméstico e comportamento em compras, obtiveram-se dados sobre as seguintes afirmações sobre modo de vida (com base em uma amostra de sete pontos: 1=discordo, 7=concordo):

V<sub>1</sub>=Prefiro ficar em casa a comparecer a uma reunião social

V<sub>2</sub>=Sempre verifico os preços, mesmo em artigos baratos.

V<sub>3</sub>=As lojas são mais interessantes do que os filmes

V<sub>4</sub>=Não compro produtos anunciados em cartazes

V<sub>5</sub>=Sou uma pessoa caseira

V<sub>6</sub>=Economizo e troco cupons

V<sub>7</sub>=As companhias gastam demais com propagandas

A amostra aleatória constituiu-se de 25 entrevistados.

### 8.3.1. Apresentação dos Resultados Computacionais para o *TIAFIC*

**Tabela 38**

#### **Caso 3 : Modo de Vida**

**Distribuição por Amostragem *Bootstrap* das Cargas Fatoriais do Fator 1 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Bootstrap</i>			
		Média	Variância	Viés	EMQ
V1	8,17E-01	8,17E-01	1,42E-04	3,05E-04	1,42E-04
V2	2,79E-01	2,78E-01	1,24E-03	4,58E-04	1,24E-03
V3	8,87E-01	8,87E-01	1,23E-05	3,32E-04	1,24E-05
V4	-2,04E-01	-2,04E-01	8,97E-04	-3,10E-04	8,97E-04
V5	6,64E-01	6,64E-01	2,85E-04	3,47E-04	2,85E-04
V6	5,01E-02	5,00E-02	1,39E-03	1,67E-04	1,39E-03
V7	-6,84E-01	-6,84E-01	2,36E-04	-4,74E-04	2,36E-04

**Tabela 39**

#### **Caso 3 : Modo de Vida**

**Distribuição por Amostragem *Bootstrap* das Cargas Fatoriais do Fator 2 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Bootstrap</i>			
		Média	Variância	Viés	EMQ
V1	3,78E-01	3,75E-01	2,59E-03	3,17E-03	2,60E-03
V2	-7,14E-01	-7,05E-01	1,19E-02	-8,41E-03	1,19E-02
V3	-2,70E-02	-2,64E-02	5,12E-04	-5,46E-04	5,12E-04
V4	6,34E-01	6,27E-01	7,27E-03	6,59E-03	7,31E-03



Variáveis	Cargas da amostra original	<i>Bootstrap</i>			
		Média	Variância	Viés	EMQ
V5	5,05E-01	5,02E-01	3,53E-03	3,47E-03	3,54E-03
V6	-6,04E-01	-5,96E-01	8,58E-03	-7,10E-03	8,63E-03
V7	3,83E-01	3,79E-01	3,11E-03	3,69E-03	3,12E-03

**Tabela 40**

**Caso 3 : Modo de Vida**

**Distribuição por Amostragem *Jackknife* das Cargas Fatoriais do Fator 1 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Jackknife</i>			
		Média	Variância	Viés	EMQ
V1	8,17E-01	8,14E-01	1,17E-03	2,44E-03	1,18E-03
V2	2,79E-01	2,75E-01	1,03E-02	3,66E-03	1,03E-02
V3	8,87E-01	8,84E-01	9,58E-05	2,66E-03	1,03E-04
V4	-2,04E-01	-2,02E-01	7,43E-03	-2,48E-03	7,44E-03
V5	6,64E-01	6,61E-01	2,36E-03	2,78E-03	2,36E-03
V6	5,01E-02	4,88E-02	1,15E-02	1,34E-03	1,15E-02
V7	-6,84E-01	-6,81E-01	1,95E-03	-3,79E-03	1,96E-03

**Tabela 41****Caso 3 : Modo de Vida****Distribuição por Amostragem *Jackknife* das Cargas Fatoriais do Fator 2 para Cada Variável do Problema**

Variáveis	Cargas da amostra original	<i>Jackknife</i>			
		Média	Variância	Viés	EMQ
V1	3,78E-01	3,53E-01	2,09E-02	2,53E-02	2,15E-02
V2	-7,14E-01	-6,46E-01	9,42E-02	-6,72E-02	9,88E-02
V3	-2,70E-02	-2,26E-02	4,23E-03	-4,37E-03	4,25E-03
V4	6,34E-01	5,81E-01	5,77E-02	5,27E-02	6,05E-02
V5	5,05E-01	4,78E-01	2,85E-02	2,77E-02	2,93E-02
V6	-6,04E-01	-5,47E-01	6,82E-02	-5,68E-02	7,14E-02
V7	3,83E-01	3,54E-01	2,50E-02	2,95E-02	2,59E-02

**Tabela 42****Caso 3 : Modo de Vida**

*Intervalos de Confiança Bootstrap e Jackknife Percentílico de 95% e TIAF das Cargas Fatoriais por Variáveis do Problema da Primeira Componente Principal (Fator 1)*

Variáveis	Bootstrap		Jackknife	
	Intervalo de Confiança	TIAFIC	Intervalo de Confiança	TIAFIC
V1	0,78 ;0,85	SIG	0,75 ;0,89	SIG
V2	0,16 ;0,38	SIG	0,12 ;0,50	SIG
V3	0,88 ; 0,89	SIG	0,87 ;0,91	SIG
V4	-0,27 ; -0,15	SIG	-0,35 ;0,02	NSIG
V5	0,62 ;0,70	SIG	0,60 ;0,79	SIG
V6	-0,04; 0,15	NSIG	-0,18 ;0,27	NSIG
V7	-0,72 ;-0,64	SIG	-0,77 ;-0,59	SIG

Nota : SIG=significante e NSIG=não significante

**Tabela 43****Caso 3 : Modo de Vida**

*Intervalos de Confiança Bootstrap e Jackknife Percentílico de 95% e TIAF das Cargas Fatoriais por Variáveis do Problema da Primeira Componente Principal (Fator 2)*

Variáveis	Bootstrap		Jackknife	
	Intervalo de Confiança	TIAFIC	Intervalo de Confiança	TIAFIC
V1	0,31 ;0,46	SIG	-0,26 ;0,50	NSIG
V2	-0,74 ;-0,64	SIG	-0,80 ;0,80	NSIG
V3	-0,10 ;0,02	NSIG	-0,18 ;0,09	NSIG
V4	0,55 ;0,67	SIG	-0,54 ;0,71	NSIG
V5	0,44 ;0,58	SIG	-0,24 ;0,63	NSIG
V6	-0,65 ;-0,53	SIG	-0,72 ;0,67	NSIG
V7	0,32 ;0,44	SIG	-0,34 ;0,52	NSIG

Nota : SIG=significante e NSIG=não significante

### **8.3.2. Comparação das Distribuições Amostrais entre os Métodos *Bootstrap* e *Jackknife***

Da leitura das Tabelas 38 39, 40, e 41 se podem fazer as seguintes interpretações:

Tanto as estimativas *bootstrap* quanto as *jackknife* revelaram-se de variância pequena, o que indica que são precisas. Esta informação é válida para ambos os fatores 1 e 2. Contudo, constata-se uma leve oscilação para o método *jackknife*, com destaque para o fator 2, isto é, o método *jackknife* revelou-se levemente mais impreciso. Portanto, parece que a técnica *bootstrap* constitui num método de estimação mais preciso para este caso. É de se referir que a qualidade da distribuição amostral das cargas fatoriais das componentes principais dos fatores referidos pode estar correlacionada com o número de réplicas utilizado. No *bootstrap*, simulou-se 200 réplicas e enquanto que no *jackknife* 25 réplicas. Isto pode estar implicando em uma distribuição amostral das estimativas *jackknife* de pior qualidade. A distribuição amostral das cargas fatoriais *jackknife* pode ser melhorada ao considerar-se neste método um número maior de entrevistados contidos na base de dados, o que resultará também em um número maior de réplicas *jackknife*.

Com base nos estudos de outras estatísticas, as estimativas *jackknife* tendem a sobreestimar a variância, enquanto que as estimativas *bootstrap* tendem a subestimá-lo. O fato da precisão do *jackknife* ser inferior ao do *bootstrap* já se evidenciou em muitas outras estatísticas analisadas.

Os enfiamentos são desprezíveis para ambos os fatores e para ambos os métodos, e para a maioria das variáveis, o que já é de fato uma informação útil, pois revela que as estimativas *bootstrap* e *jackknife* são acuradas, não enfiadas. Contudo, existe uma maior variação nos valores

*jackknife*, especialmente no fator 2. Este fato, como mencionado anteriormente, pode estar relacionado com o número pequeno de réplicas *jackknife* considerada na simulação. O conhecimento do erro-padrão permitiu construir intervalos de confiança e sair, conseqüentemente, do terreno puramente descritivo (ver Tabelas 42 e 43).

Os estimadores *bootstrap* e *jackknife* são precisos e acurados, o que implicou em um EMQ pequeno para a maioria das variáveis, portanto reúnem as qualidades essenciais de um bom estimador. Contudo, as estimativas *bootstrap* são nitidamente superiores às estimativas *jackknife* e devem ser as escolhidas para inferências e tomadas de decisão posteriores na área de atuação do estudo e para nomeação e interpretação dos fatores.

Vale lembrar que a superioridade da qualidade da distribuição por amostragem da estatística em foco no método *bootstrap* se deve em parte a influência da *Lei dos Grandes Números*, ao realizarmos a experiência com um número suficientemente grande de réplicas, o que não se verificou na experiência com o método *jackknife*, o que resultou em EMQ'S maiores para as estimativas em algumas variáveis.

Sabe-se contudo que geralmente as propriedades do enviezamento *jackknife* são muito inferiores às da variância *jackknife*. Deve-se confiar mais nas estimativas *jackknife* da variância do que nas de enviezamento. Para este caso, as estimativas *bootstrap* parecem ser de maior confiança.

Como foi dito no parágrafo acima, a qualidade da distribuição amostral das cargas fatoriais das primeiras e segundas componentes principais na técnica *bootstrap* revelou-se superior ao da distribuição amostral das cargas fatoriais na simulação *jackknife* em parte pelo fato de que no caso do *bootstrap* utilizou-se de 200 re-amostragens e enquanto que

no *jackknife* de somente 25, resultando em um EMQ maior para algumas variáveis. Portanto, é de se esperar que aumentando o tamanho da amostra *jackknife* as estimativas em estudo possam ser melhoradas.

### 8.3.3. *TIAFIC*

Observando as Tabelas 42 e 43, os resultados do *TIAF* sugerem que:

- Para o **fator 1**, no método *bootstrap*, somente as variáveis  $V_1$ ,  $V_2$ ,  $V_3$ ,  $V_4$ ,  $V_5$  e  $V_7$  têm cargas fatoriais significantes. Contudo, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_1$ ,  $V_3$ ,  $V_5$  e  $V_7$ ;
- Para o **fator 1**, no método *jackknife* assim como no método *bootstrap*, somente as variáveis  $V_1$ ,  $V_2$ ,  $V_3$ ,  $V_4$ ,  $V_5$  e  $V_7$  têm cargas fatoriais significantes. Contudo, assim como no *bootstrap*, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_1$ ,  $V_3$ ,  $V_5$  e  $V_7$ ;
- Para o **fator 2**, no método *bootstrap*, somente as variáveis  $V_1$ ,  $V_2$ ,  $V_4$ ,  $V_5$ ,  $V_6$  e  $V_7$  têm cargas fatoriais significantes. Contudo, sugere-se que podem participar da nomeação/interpretação dos fatores apenas as variáveis  $V_1$ ,  $V_2$ ,  $V_4$ ,  $V_5$ ,  $V_6$  e  $V_7$ .
- Para o **fator 2**, no método *jackknife*, um resultado interessante: nenhuma variável apresentou carga significativa. Isto significa que os resultados encontrados na matriz de fatores para a segunda componente principal são acidentais. Na população, parece não existir correlação entre as variáveis do modelo e o fator 2 na população. Contudo, este resultado deve ser encarado com determinada reserva, uma vez que a qualidade da distribuição por

amostragem das cargas fatoriais no método *jackknife* parece estar comprometida pelo número insuficiente de réplicas simuladas.

Para as tomadas de Decisão administrativas com relação ao comportamento de compras e comportamento doméstico das pessoas é melhor usar os resultados do TIAF *bootstrap*.

Vale lembrar de novo que a escolha de variáveis que efetivamente vão participar da nomeação dos fatores é um processo pessoal, isto é, fica a cargo do analista ou pesquisador e segue um critério subjetivo.

### 8.3.4.TIAFVP

#### 8.3.4.1.TIAFVP *Bootstrap*

A Tabela 44 apresenta os resultados do TIAF pelo Valor-p:

**Tabela 44**

VAR	Fator 1		Fator 2	
	Valor-p	TIAFVP	Valor-p	TIAFVP
V1	0,000	SIG	0,625	NSIG
V2	0,375	NSIG	0,340	NSIG
V3	0,000	SIG	0,165	NSIG
V4	0,305	NSIG	0,030	SIG
V5	0,000	SIG	0,245	NSIG
V6	0,385	NSIG	0,500	NSIG
V7	0,185	NSIG	0,030	SIG

No fator 1, as variáveis  $V_1$ ,  $V_3$  e  $V_5$  são estatisticamente significantes no valor-p *bootstrap*. Reparem que na inferência pelo valor-p menos variáveis se revelaram significantes estatisticamente em comparação com o teste de hipótese pelo intervalo de confiança. Os testes de hipóteses pelo valor-p tendem a ser mais rigorosos dos que os pelos intervalos de confiança. Os resultados pelo valor-p também tendem a ser mais confiáveis.

No fator 2, as variáveis  $V_2$ ,  $V_4$  e  $V_6$  são estatisticamente significantes no valor-p *bootstrap*. Também observam-se neste teste inferencial resultados mais rigorosos quanto à significância do que os encontrados no teste de hipótese pelo intervalo de confiança.

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

#### 8.2.4.2. TIAFVP *Jackknife*

**Tabela 45**

VAR	Fator 1		Fator 2	
	Valor-p	TIAFVP	Valor-p	TIAFVP
$V_1$	0,000	SIG	0,400	NSIG
$V_2$	0,240	NSIG	0,000	SIG
$V_3$	0,000	SIG	0,000	SIG
$V_4$	0,040	SIG	0,000	SIG
$V_5$	0,000	SIG	0,000	SIG
$V_6$	0,080	NSIG	0,200	NSIG
$V_7$	0,000	SIG	0,000	SIG

No fator 1, as variáveis  $V_1$ ,  $V_3$ ,  $V_4$  e  $V_5$  são estatisticamente significantes no valor-p *bootstrap*. Comparando com o TIAFIC, observam-se divergências nas variáveis  $V_2$  e  $V_4$ .

No fator 2, as variáveis  $V_2$ ,  $V_3$ ,  $V_4$  e  $V_6$  são estatisticamente significantes no valor-p *bootstrap*. Na inferência pelo intervalo de confiança todas as variáveis foram não significantes estatisticamente. Os resultados pela inferência pelo valor-p tendem a ser mais confiáveis e precisos.



A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

Pela análise indutiva da base “Modo de Vida” recomenda-se cautela ao utilizá-la para tomadas de decisão. Sugere-se obter uma amostra com representatividade maior.

#### **8.4.CASO 4: BASE QUALIDADE TOTAL**

O objetivo desta pesquisa foi avaliar o nível de aplicação da filosofia da Qualidade Total de produtos e/ou serviços oferecidos pelo conjunto de empresas abaixo :

- ATN
- REDUC
- LOJAS NALIN
- CONTAX
- CASA DA EMPADA
- PETROBÁS
- LOJAS AVON

A Amostra contou com a opinião de 350 funcionários de diversos cargos e diversos níveis hierárquicos das organizações investigadas.

As opiniões foram colhidas através de um questionário com 10 afirmativas que correspondem na verdade aos 10 princípios básicos da Qualidade Total :

- Total satisfação aos clientes ( $V_1$ );
- Gerência participativa ( $V_2$ );
- Desenvolvimento de recursos humanos ( $V_3$ );
- Constância de propósitos ( $V_4$ );
- Aperfeiçoamento contínuo ( $V_5$ );
- Gerência de projetos ( $V_6$ );
- Delegação ( $V_7$ );
- Disseminação de informações ( $V_8$ );
- Garantia da qualidade ( $V_9$ );

- Não aceitação de erros ( $V_{10}$ );

Para cada uma das 10 afirmativas, os colaboradores deram uma nota de 1 a 5 para o grau de sua aplicação prática e real dentro de sua empresa. Quanto mais próximo de 5 mais o princípio referido se desenvolve ou é assumido pela organização.

A Análise Fatorial poderia ser usada para reduzir o número de 10 princípios a um número menor, os fatores 1 e 2, que explicam também satisfatoriamente o nível de qualidade total das empresas observadas e que corresponderiam a opinião geral dos empregados sobre onde está a qualidade total das empresas investigadas.

O TIAF deve ser usado para testar quais variáveis são significativas para cada fator.

#### 8.4.1 .Análise dos Resultados Bootstrap

Utilizou-se neste estudo de caso um processo bootstrap com  $B=1000$ .

No método bootstrap, os resultados gerados pelos programas computacionais se encontram na Tabela 46.

**Tabela 46**

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 1										
V1	2,65E-01	3,06E-01	9,51E-03	1,38E-01	5,16E-01	-4,13E-02	7,81E-03	SIG	0,003	SIG
V2	8,12E-02	2,67E-01	1,07E-01	3,90E-03	9,97E-01	-1,86E-01	7,26E-02	SIG	0,225	NSIG
V3	4,90E-01	4,67E-01	5,53E-03	3,18E-01	5,96E-01	2,31E-02	4,99E-03	SIG	0,000	SIG
V4	1,50E-01	2,09E-01	4,32E-02	-7,55E-02	7,16E-01	-5,90E-02	3,97E-02	NSIG	0,200	NSIG
V5	5,79E-01	4,91E-01	3,97E-02	4,84E-02	6,93E-01	8,78E-02	3,20E-02	SIG	0,000	SIG
V6	5,14E-01	4,68E-01	1,28E-02	1,94E-01	6,09E-01	4,58E-02	1,07E-02	SIG	0,000	SIG
V7	5,25E-01	4,54E-01	2,39E-02	1,01E-01	6,42E-01	7,10E-02	1,89E-02	SIG	0,000	SIG
V8	4,89E-01	4,52E-01	8,48E-03	2,25E-01	5,96E-01	3,69E-02	7,12E-03	SIG	0,000	SIG
V9	4,49E-01	4,28E-01	5,85E-03	2,62E-01	5,58E-01	2,03E-02	5,44E-03	SIG	0,000	SIG
V10	5,45E-01	4,45E-01	4,96E-02	-4,85E-02	6,74E-01	9,99E-02	3,96E-02	NSIG	0,000	SIG

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 2										
V1	3,72E-01	3,16E-01	1,34E-02	2,35E-02	4,70E-01	5,62E-02	1,03E-02	SIG	0,001	SIG
V2	9,94E-01	7,47E-01	1,27E-01	1,50E-02	9,97E-01	2,47E-01	6,61E-02	SIG	0,000	SIG
V3	3,48E-01	3,71E-01	5,41E-03	2,48E-01	5,46E-01	-2,35E-02	4,85E-03	SIG	0,000	SIG
V4	5,16E-01	4,87E-01	5,60E-02	4,36E-03	9,97E-01	2,85E-02	5,52E-02	SIG	0,000	SIG
V5	1,14E-01	2,04E-01	4,04E-02	2,78E-02	6,74E-01	-9,08E-02	3,21E-02	SIG	0,198	NSIG
V6	2,39E-01	2,88E-01	1,12E-02	1,52E-01	5,34E-01	-4,88E-02	8,80E-03	SIG	0,030	SIG
V7	1,44E-01	2,21E-01	2,12E-02	6,64E-02	5,71E-01	-7,66E-02	1,53E-02	SIG	0,189	NSIG
V8	2,58E-01	3,07E-01	8,27E-03	1,74E-01	5,26E-01	-4,86E-02	5,91E-03	SIG	0,007	SIG
V9	2,87E-01	3,23E-01	6,30E-03	1,96E-01	5,00E-01	-3,56E-02	5,03E-03	SIG	0,002	SIG
V10	8,21E-03	1,14E-01	4,38E-02	-1,02E-01	5,96E-01	-1,06E-01	3,27E-02	NSIG	0,218	NSIG

A Tabela 46 contém as estimativas das cargas fatoriais para cada variável e para cada fator ou componente principal.

No fator 1, os enfiamentos e as variâncias são bem reduzidos, isto é, as estimativas são acuradas e precisas, o que implica EMQ'S pequenos. Portanto, as distribuições amostrais possuem as qualidades desejáveis para gerarem bons estimadores e inferências confiáveis.

No fator 1, as variáveis  $V_1, V_2, V_3, V_5, V_6, V_7, V_8$  e  $V_9$  são estatisticamente significantes pela análise do intervalo de confiança. Apenas duas variáveis ( $V_4$  e  $V_{10}$ ) são estatisticamente não significantes. Neste mesmo fator, as variáveis  $V_1, V_3, V_5, V_6, V_7, V_8, V_9$  e  $V_{10}$  são estatisticamente significantes pela análise do valor-p. Somente as variáveis  $V_2$  e  $V_4$  são não significantes estatisticamente pelo valor-p na primeira componente principal. Os resultados das inferências estatísticas pelo intervalo de confiança e valor-p divergiram somente nas variáveis  $V_2$  e  $V_{10}$ .

No fator 2, as médias e as variâncias são bem reduzidas, isto é, as estimativas são acuradas e precisas, o que implica EMQ'S pequenos. Portanto, as distribuições amostrais possuem as qualidades desejáveis para gerarem bons estimadores e inferências confiáveis.

No fator 2, as variáveis  $V_1, V_2, V_3, V_5, V_4, V_5, V_6, V_7, V_8$  e  $V_9$  são estatisticamente significantes pela análise do intervalo de confiança. Apenas uma variável ( $V_{10}$ ) é estatisticamente não significante. Neste mesmo fator, as variáveis  $V_1, V_2, V_3, V_4, V_6, V_8$  e  $V_9$  são estatisticamente significantes pela análise do valor-p. Somente as variáveis  $V_5, V_4$  e  $V_{10}$  são não significantes estatisticamente pelo valor-p na segunda componente principal. Os resultados das inferências estatísticas pelo intervalo de confiança e valor-p divergiram somente nas variáveis  $V_5$  e  $V_7$ .

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

#### 8.4.2 .Análise dos Resultados *Jackknife*

No método *jackknife*, os resultados gerados pelos programas computacionais se encontram na Tabela 47.

**Tabela 47**

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
<b>Fator 1</b>										
V1	2,65E-01	2,65E-01	1,21E-05	2,59E-01	2,73E-01	1,34E-05	1,21E-05	SIG	0,000	SIG
V2	8,12E-02	8,11E-02	6,24E-06	7,60E-02	8,70E-02	2,46E-05	6,24E-06	SIG	0,000	SIG
V3	4,90E-01	4,90E-01	1,04E-05	4,84E-01	4,98E-01	9,24E-06	1,04E-05	SIG	0,000	SIG
V4	1,50E-01	1,50E-01	6,38E-06	1,45E-01	1,56E-01	1,43E-05	6,38E-06	SIG	0,000	SIG
V5	5,79E-01	5,79E-01	7,10E-06	5,74E-01	5,85E-01	4,18E-06	7,10E-06	SIG	0,000	SIG
V6	5,14E-01	5,14E-01	7,18E-06	5,09E-01	5,20E-01	1,47E-05	7,18E-06	SIG	0,000	SIG
V7	5,25E-01	5,25E-01	1,08E-05	5,21E-01	5,34E-01	6,29E-06	1,08E-05	SIG	0,000	SIG
V8	4,89E-01	4,89E-01	8,85E-06	4,83E-01	4,95E-01	9,69E-06	8,85E-06	SIG	0,000	SIG
V9	4,49E-01	4,49E-01	1,06E-05	4,43E-01	4,55E-01	1,28E-05	1,06E-05	SIG	0,000	SIG
V10	5,45E-01	5,45E-01	9,12E-06	5,41E-01	5,53E-01	-1,95E-06	9,12E-06	SIG	0,000	SIG
<b>Fator 2</b>										
V1	3,72E-01	3,72E-01	8,23E-06	3,67E-01	3,81E-01	-4,39E-06	8,23E-06	SIG	0,000	SIG
V2	9,94E-01	9,94E-01	4,15E-08	9,94E-01	9,95E-01	1,14E-06	4,15E-08	SIG	0,000	SIG
V3	3,48E-01	3,48E-01	7,01E-06	3,42E-01	3,54E-01	-8,60E-06	7,01E-06	SIG	0,000	SIG
V4	5,16E-01	5,16E-01	4,75E-06	5,12E-01	5,19E-01	-4,61E-06	4,75E-06	SIG	0,000	SIG
V5	1,14E-01	1,14E-01	5,26E-06	1,10E-01	1,19E-01	-1,42E-05	5,26E-06	SIG	0,000	SIG
V6	2,39E-01	2,39E-01	6,75E-06	2,34E-01	2,45E-01	-1,33E-05	6,75E-06	SIG	0,000	SIG
V7	1,44E-01	1,44E-01	6,32E-06	1,39E-01	1,49E-01	-1,47E-05	6,32E-06	SIG	0,000	SIG
V8	2,58E-01	2,58E-01	6,34E-06	2,53E-01	2,64E-01	-1,20E-05	6,34E-06	SIG	0,000	SIG
V9	2,87E-01	2,87E-01	7,72E-06	2,83E-01	2,95E-01	-1,23E-05	7,72E-06	SIG	0,000	SIG
V10	8,21E-03	8,22E-03	7,14E-06	2,32E-03	1,37E-02	-1,11E-05	7,14E-06	SIG	0,000	SIG

A Tabela 47 contém as estimativas das cargas fatoriais para cada variável e para cada fator ou componente principal.

No fator 1, os enfiamentos e as variâncias são praticamente nulos, isto é, as estimativas são bem acuradas e precisas, o que implica EMQ'S desprezíveis. Portanto, as estimativas possuem as qualidades desejáveis para gerarem inferências confiáveis.

No fator 1, todas as variáveis são estatisticamente significantes pela análise do intervalo de confiança. Resultado idêntico é verificado quando se realiza a inferência pelo valor-p.

No fator 2, os resultados são análogos aos encontrados no fator 1.

Uma observação importante é que os EMQ'S no método *jackknife* são bem menores dos que os do método *bootstrap*. Isto se verifica em ambos os fatores. Portanto, para esta base também o método *Jackknife* tornou um método de estimação de cargas fatoriais mais eficientes.

Após os resultados inferenciais verificados, parte-se para a significância prática das cargas fatoriais significantes por fator (ver quadro 3, capítulo 5) e para a nomeação/interpretação dos mesmos e tomados de decisões estatísticas.

## **8.5.CASO 5: BASE SATISFAÇÃO BODY-UP**

A Base Satisfação Body-Up resultou da aplicação de um questionário numa pesquisa que tem o objetivo de medir o nível de satisfação dos clientes/alunos com a academia.

A investigação contou com 124 alunos da academia de diversas modalidades, idades e gênero. A pesquisa foi aplicada no mês de novembro de 2005.

A análise fatorial é empregada para reduzir o número de variáveis de satisfação da pesquisa a um número menor, gerenciável e para que se entenda que atributos latentes em menor número podem está explicando, com uma variação tolerável, a satisfação dos alunos com a academia.

É uma aplicação simples da análise fatorial e que nem por isso deve prescindir da realização do TIAF.

As variáveis que entraram na análise fatorial por componentes principais estão indicadas no questionário da pesquisa que é apresentado nas páginas seguintes. Elas fazem parte do bloco A do referido instrumento de coleta de dados.

## PESQUISA DE SATISFAÇÃO DA ACADEMIA BODY-UP

Contamos com a sua colaboração para preencher o formulário abaixo para que possamos atendê-lo cada vez melhor. Sua opinião é muito importante para nós.

### BLOCO A: AVALIAÇÃO DA ACADEMIA

1-Indique, por favor, o grau com que você concorda ou discorda das seguintes declarações acerca da academia.

*1-Muito insatisfeito,2-Insatisfeito,3-Neutro4-Satisfeito5-Muito satisfeito*

Declarações	1	2	3	4	5
Horário de Funcionamento(V <sub>1</sub> )					
Nível de conhecimento dos professores e funcionários(V <sub>2</sub> )					
Profissionalismo da equipe de professores e funcionários(V <sub>3</sub> )					
Qualidade dos serviços prestados(V <sub>4</sub> )					
Rapidez do atendimento de professores e funcionários aos alunos(V <sub>5</sub> )					

2-Avalie a academia quanto aos critérios abaixo:

*1-Péssimo,2-Ruim,3-Regular,4-Bom,5-Ótimo*

Declarações	1	2	3	4	5
Pré-atendimento (recepção)- Apresentação(V <sub>6</sub> )					
Pré-atendimento (recepção)- Simpatia(V <sub>7</sub> )					
Pré-atendimento(recepção)- Eficiência(V <sub>8</sub> )					
Pré-atendimento(recepção)- Organização(V <sub>9</sub> )					
Pré-atendimento(recepção)- Rapidez(V <sub>10</sub> )					
Equipe de professores- Apresentação(V <sub>11</sub> )					
Equipe de professores- Simpatia(V <sub>12</sub> )					
Equipe de professores- Eficiência(V <sub>13</sub> )					
Equipe de professores- Organização(V <sub>14</sub> )					
Equipe de professores- Rapidez(V <sub>15</sub> )					
Equipe de manutenção-Apresentação(V <sub>16</sub> )					
Equipe de manutenção-Simpatia(V <sub>17</sub> )					
Equipe de manutenção-Eficiência(V <sub>18</sub> )					
Equipe de manutenção-Organização(V <sub>19</sub> )					
Equipe de manutenção-Rapidez(V <sub>20</sub> )					



<b>Declarações</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Ambiente-Limpeza(V <sub>21</sub> )					
Ambiente-Conforto(V <sub>22</sub> )					
Ambiente-Conservação(V <sub>23</sub> )					
Ambiente-Alimentação(V <sub>24</sub> )					
Ambiente-Localização(V <sub>25</sub> )					

### **BLOCO B : OUTROS**

1-É a primeira vez que você é aluno da academia ?

1-( ) Sim

2-( ) Não

2-Você pretende utilizá-la continuamente ?

1-( ) Sim

2-( ) Não

3-Por que optou por esta academia ?

1-( ) Indicação de “profissional da área”

2-( ) Indicação de amigos/parentes

3-( ) Localização

4-( ) Forma de pagamento

5-( ) Preço

6-( ) Qualidade do serviço

7-( ) Lista telefônica/internet

8-( ) Falta de Opção

9-( ) Outros: \_\_\_\_\_

4-Dê uma nota de 1 a 10 para a academia em geral : \_\_\_\_\_

5-Modalidade de atividade que pratica na academia: \_\_\_\_\_

6-Gênero : 1-( ) Masculino      2-( ) Feminino

7-Idade( em anos) : \_\_\_\_\_ anos

**Muito obrigado pela sua colaboração.**

### 8.5.1. Análise dos Resultados *Bootstrap*

Utilizou-se neste estudo de caso um processo *bootstrap* com B=1000.

No método *bootstrap*, os resultados gerados pelos programas computacionais se encontram na Tabela 48.

**Tabela 48**

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
<b>Fator 1</b>										
V1	1,63E-01	1,92E-01	1,55E-02	-2,43E-02	4,37E-01	-2,94E-02	1,47E-02	NSIG	0,101	NSIG
V2	5,72E-01	5,22E-01	3,19E-02	1,19E-01	7,76E-01	5,00E-02	2,94E-02	SIG	0,000	SIG
V3	5,86E-01	5,74E-01	2,31E-02	1,92E-01	8,53E-01	1,15E-02	2,30E-02	SIG	0,000	SIG
V4	4,78E-01	4,82E-01	1,56E-02	2,16E-01	7,40E-01	-3,86E-03	1,56E-02	SIG	0,000	SIG
V5	5,75E-01	5,43E-01	3,03E-02	1,34E-01	8,49E-01	3,22E-02	2,93E-02	SIG	0,000	SIG
V6	7,37E-01	5,32E-01	8,28E-02	-4,98E-02	8,40E-01	2,06E-01	4,05E-02	NSIG	0,000	SIG
V7	5,57E-01	4,10E-01	4,91E-02	-3,05E-02	6,85E-01	1,47E-01	2,75E-02	NSIG	0,000	SIG
V8	7,69E-01	6,13E-01	6,19E-02	9,41E-02	8,68E-01	1,56E-01	3,75E-02	SIG	0,000	SIG
V9	8,13E-01	6,36E-01	6,78E-02	9,68E-02	8,84E-01	1,77E-01	3,63E-02	SIG	0,000	SIG
V10	7,18E-01	5,70E-01	5,26E-02	7,04E-02	8,07E-01	1,48E-01	3,08E-02	SIG	0,000	SIG
V11	5,22E-01	5,27E-01	1,58E-02	2,38E-01	7,32E-01	-4,92E-03	1,58E-02	SIG	0,000	SIG
V12	4,88E-01	4,97E-01	1,48E-02	2,49E-01	7,06E-01	-8,64E-03	1,48E-02	SIG	0,000	SIG
V13	5,55E-01	5,58E-01	1,45E-02	3,06E-01	7,68E-01	-2,15E-03	1,45E-02	SIG	0,000	SIG
V14	4,34E-01	4,89E-01	1,53E-02	2,14E-01	7,15E-01	-5,49E-02	1,23E-02	SIG	0,000	SIG
V15	5,58E-01	5,93E-01	1,01E-02	4,13E-01	8,33E-01	-3,53E-02	8,84E-03	SIG	0,000	SIG
V16	2,16E-01	3,69E-01	4,88E-02	6,58E-02	8,12E-01	-1,53E-01	2,54E-02	SIG	0,230	NSIG
V17	2,66E-01	4,01E-01	4,14E-02	1,33E-01	7,78E-01	-1,35E-01	2,33E-02	SIG	0,173	NSIG
V18	1,65E-01	3,44E-01	7,23E-02	1,45E-02	8,71E-01	-1,80E-01	4,00E-02	SIG	0,300	NSIG
V19	1,49E-01	3,68E-01	9,11E-02	2,29E-02	9,24E-01	-2,19E-01	4,30E-02	SIG	0,302	NSIG
V20	-4,52E-02	1,90E-01	1,19E-01	-1,42E-01	8,74E-01	-2,35E-01	6,38E-02	NSIG	0,686	NSIG
V21	2,25E-01	3,04E-01	3,47E-02	2,26E-02	6,89E-01	-7,91E-02	2,85E-02	SIG	0,164	NSIG
V22	2,88E-01	3,09E-01	2,38E-02	1,22E-02	6,02E-01	-2,04E-02	2,34E-02	SIG	0,026	SIG
V23	2,95E-01	3,06E-01	1,54E-02	7,42E-02	5,39E-01	-1,10E-02	1,53E-02	SIG	0,006	SIG
V24	2,10E-01	2,71E-01	3,02E-02	-2,26E-02	6,15E-01	-6,17E-02	2,64E-02	NSIG	0,151	NSIG
V25	3,37E-01	3,43E-01	1,31E-02	9,20E-02	5,55E-01	-5,70E-03	1,31E-02	SIG	0,002	SIG
<b>Fator 2</b>										
V1	1,77E-01	1,50E-01	1,31E-02	-7,26E-02	3,82E-01	2,74E-02	1,23E-02	NSIG	0,062	NSIG
V2	2,20E-01	2,65E-01	3,08E-02	3,32E-02	7,44E-01	-4,47E-02	2,88E-02	SIG	0,129	NSIG
V3	3,82E-01	3,92E-01	2,36E-02	1,39E-01	8,75E-01	-9,92E-03	2,35E-02	SIG	0,036	SIG
V4	3,59E-01	3,53E-01	1,53E-02	1,22E-01	6,77E-01	5,89E-03	1,52E-02	SIG	0,018	SIG
V5	3,01E-01	3,20E-01	3,28E-02	2,50E-02	8,20E-01	-1,97E-02	3,24E-02	SIG	0,067	NSIG
V6	2,50E-02	2,19E-01	9,49E-02	-1,08E-01	8,45E-01	-1,94E-01	5,72E-02	NSIG	0,302	NSIG
V7	5,44E-02	1,88E-01	5,25E-02	-7,04E-02	6,87E-01	-1,34E-01	3,47E-02	NSIG	0,261	NSIG
V8	1,74E-01	3,00E-01	7,05E-02	1,52E-02	8,54E-01	-1,26E-01	5,46E-02	SIG	0,217	NSIG
V9	1,74E-01	3,27E-01	7,95E-02	4,54E-02	8,93E-01	-1,53E-01	5,61E-02	SIG	0,226	NSIG

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 2										
V10	1,56E-01	2,88E-01	5,73E-02	1,79E-02	7,77E-01	-1,32E-01	3,98E-02	SIG	0,247	NSIG
V11	3,49E-01	3,51E-01	1,65E-02	1,37E-01	6,51E-01	-1,95E-03	1,65E-02	SIG	0,016	SIG
V12	2,94E-01	2,82E-01	1,55E-02	6,85E-02	5,61E-01	1,19E-02	1,54E-02	SIG	0,019	SIG
V13	3,27E-01	3,18E-01	1,46E-02	1,12E-01	5,69E-01	8,99E-03	1,45E-02	SIG	0,005	SIG
V14	4,48E-01	3,93E-01	1,53E-02	1,49E-01	6,77E-01	5,49E-02	1,23E-02	SIG	0,000	SIG
V15	4,84E-01	4,28E-01	1,65E-02	4,70E-02	6,08E-01	5,65E-02	1,33E-02	SIG	0,000	SIG
V16	6,79E-01	5,27E-01	5,24E-02	6,38E-02	8,03E-01	1,52E-01	2,92E-02	SIG	0,000	SIG
V17	6,88E-01	5,24E-01	4,92E-02	1,01E-01	7,93E-01	1,63E-01	2,26E-02	SIG	0,000	SIG
V18	7,99E-01	6,03E-01	8,99E-02	6,65E-02	9,25E-01	1,96E-01	5,15E-02	SIG	0,000	SIG
V19	8,83E-01	6,51E-01	1,03E-01	4,66E-02	9,45E-01	2,32E-01	4,97E-02	SIG	0,000	SIG
V20	8,18E-01	5,87E-01	1,28E-01	-1,06E-01	9,13E-01	2,31E-01	7,50E-02	NSIG	0,000	SIG
V21	5,32E-01	4,82E-01	4,14E-02	7,25E-02	8,63E-01	5,03E-02	3,88E-02	SIG	0,000	SIG
V22	3,54E-01	3,58E-01	2,86E-02	8,63E-02	7,64E-01	-3,27E-03	2,86E-02	SIG	0,045	SIG
V23	3,99E-01	3,98E-01	1,81E-02	1,63E-01	7,13E-01	7,51E-04	1,81E-02	SIG	0,004	SIG
V24	4,35E-01	3,93E-01	3,53E-02	5,42E-02	7,79E-01	4,17E-02	3,36E-02	SIG	0,013	SIG
V25	3,08E-01	3,17E-01	1,57E-02	1,23E-01	6,73E-01	-9,75E-03	1,56E-02	SIG	0,038	SIG

No fator 1, os envezamentos e as variâncias são bem reduzidos, isto é, as estimativas são acuradas e precisas, o que implica EMQ'S pequenos. Portanto, as distribuições amostrais possuem as qualidades desejáveis para gerarem bons estimadores e inferências confiáveis.

No fator 1, somente as variáveis  $V_1, V_6, V_7, V_{20}$  e  $V_{24}$  não são estatisticamente significantes pela análise do intervalo de confiança. Neste mesmo fator, as variáveis  $V_1, V_{16}, V_{17}, V_{18}, V_{19}, V_{20}, V_{21}$ , e  $V_{24}$  não são estatisticamente significantes pela análise do valor-p. Os resultados das inferências estatísticas pelo intervalo de confiança e valor-p coincidem somente nas variáveis  $V_1, V_{20}$  e  $V_{24}$ .

No fator 2, as médias e as variâncias são bem reduzidas, isto é, as estimativas são acuradas e precisas, o que implica EMQ'S pequenos. Portanto, as distribuições amostrais possuem as qualidades desejáveis para gerarem bons estimadores e inferências confiáveis.

No fator 2, somente s variáveis  $V_1, V_6, V_7$  e  $V_{20}$  não são estatisticamente significantes pela análise do intervalo de confiança. Neste mesmo fator, as variáveis  $V_1, V_2, V_5, V_6, V_7, V_8, V_9$  e  $V_{10}$  não são estatisticamente significantes pela análise do valor-p. Os resultados das inferências estatísticas pelo intervalo de confiança e valor-p convergem somente nas variáveis  $V_1$  e  $V_6$  e  $V_7$ . Os resultados inferenciais pelo valor-p tendem a ser mais precisos.

Recomenda-se somente considerar para significância prática e nomeação/interpretação dos fatores as variáveis com significância estatística.

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

### 8.5.2. Análise dos Resultados *Jackknife*

Os resultados da rodada *jackknife* para este estudo de caso são apresentados na Tabela 49

**Tabela 49**

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
<b>Fator 1</b>										
V1	1,63E-01	1,64E-01	1,17E-04	1,37E-01	1,89E-01	-6,60E-04	1,16E-04	SIG	0,000	SIG
V2	5,72E-01	5,70E-01	7,74E-04	5,49E-01	5,98E-01	1,94E-03	7,71E-04	SIG	0,000	SIG
V3	5,86E-01	5,85E-01	3,22E-04	5,65E-01	6,16E-01	1,01E-03	3,21E-04	SIG	0,000	SIG
V4	4,78E-01	4,78E-01	1,45E-04	4,57E-01	4,98E-01	4,48E-04	1,45E-04	SIG	0,000	SIG
V5	5,75E-01	5,74E-01	4,56E-04	5,55E-01	6,05E-01	1,36E-03	4,54E-04	SIG	0,000	SIG
V6	7,37E-01	7,31E-01	3,88E-03	7,17E-01	7,53E-01	5,84E-03	3,85E-03	SIG	0,000	SIG
V7	5,57E-01	5,52E-01	1,98E-03	5,34E-01	5,76E-01	4,12E-03	1,96E-03	SIG	0,000	SIG
V8	7,69E-01	7,65E-01	2,66E-03	7,44E-01	7,85E-01	4,70E-03	2,63E-03	SIG	0,000	SIG
V9	8,13E-01	8,08E-01	3,02E-03	7,99E-01	8,25E-01	5,00E-03	2,99E-03	SIG	0,000	SIG
V10	7,18E-01	7,14E-01	2,21E-03	7,06E-01	7,38E-01	4,14E-03	2,20E-03	SIG	0,000	SIG
V11	5,22E-01	5,22E-01	2,38E-04	4,96E-01	5,55E-01	6,48E-04	2,37E-04	SIG	0,000	SIG
V12	4,88E-01	4,87E-01	2,86E-04	4,55E-01	5,16E-01	9,16E-04	2,85E-04	SIG	0,000	SIG
V13	5,55E-01	5,54E-01	3,44E-04	5,25E-01	5,82E-01	1,03E-03	3,43E-04	SIG	0,000	SIG
V14	4,34E-01	4,35E-01	1,47E-04	4,07E-01	4,76E-01	-6,05E-04	1,46E-04	SIG	0,000	SIG
V15	5,58E-01	5,58E-01	7,58E-05	5,35E-01	5,77E-01	1,16E-04	7,57E-05	SIG	0,000	SIG
V16	2,16E-01	2,19E-01	1,67E-03	1,99E-01	2,54E-01	-3,59E-03	1,66E-03	SIG	0,008	SIG
V17	2,66E-01	2,69E-01	1,32E-03	2,52E-01	2,83E-01	-3,34E-03	1,31E-03	SIG	0,008	SIG
V18	1,65E-01	1,70E-01	3,10E-03	1,51E-01	1,92E-01	-5,23E-03	3,07E-03	SIG	0,008	SIG
V19	1,49E-01	1,55E-01	4,31E-03	1,36E-01	1,76E-01	-6,21E-03	4,27E-03	SIG	0,008	SIG
V20	-4,52E-02	-3,82E-02	5,73E-03	-5,53E-02	-1,73E-02	-7,04E-03	5,68E-03	SIG	0,000	SIG
V21	2,25E-01	2,27E-01	8,94E-04	1,99E-01	2,72E-01	-2,54E-03	8,88E-04	SIG	0,008	SIG
V22	2,88E-01	2,89E-01	2,23E-04	2,66E-01	3,29E-01	-1,10E-03	2,22E-04	SIG	0,000	SIG
V23	2,95E-01	2,97E-01	2,85E-04	2,69E-01	3,32E-01	-1,53E-03	2,83E-04	SIG	0,000	SIG
V24	2,10E-01	2,12E-01	5,44E-04	1,89E-01	2,56E-01	-1,95E-03	5,40E-04	SIG	0,008	SIG
V25	3,37E-01	3,37E-01	1,05E-04	3,20E-01	3,71E-01	-5,00E-04	1,05E-04	SIG	0,000	SIG
<b>Fator 2</b>										
V1	1,77E-01	1,77E-01	1,24E-04	1,38E-01	2,10E-01	6,48E-04	1,24E-04	SIG	0,000	SIG
V2	2,20E-01	2,22E-01	8,47E-04	1,99E-01	2,44E-01	-1,80E-03	8,44E-04	SIG	0,008	SIG
V3	3,82E-01	3,83E-01	3,15E-04	3,58E-01	4,07E-01	-9,90E-04	3,14E-04	SIG	0,000	SIG
V4	3,59E-01	3,59E-01	1,43E-04	3,36E-01	3,83E-01	-4,20E-04	1,43E-04	SIG	0,000	SIG
V5	3,01E-01	3,02E-01	5,27E-04	2,73E-01	3,25E-01	-1,28E-03	5,26E-04	SIG	0,000	SIG
V6	2,50E-02	3,06E-02	4,32E-03	1,42E-02	3,91E-02	-5,63E-03	4,29E-03	SIG	0,008	SIG
V7	5,44E-02	5,84E-02	2,18E-03	3,87E-02	7,04E-02	-3,99E-03	2,17E-03	SIG	0,008	SIG
V8	1,74E-01	1,79E-01	2,98E-03	1,61E-01	1,87E-01	-4,58E-03	2,96E-03	SIG	0,008	SIG
V9	1,74E-01	1,79E-01	3,40E-03	1,65E-01	1,88E-01	-4,87E-03	3,37E-03	SIG	0,008	SIG
V10	1,56E-01	1,60E-01	2,48E-03	1,35E-01	1,72E-01	-4,00E-03	2,47E-03	SIG	0,008	SIG
V11	3,49E-01	3,49E-01	2,35E-04	3,29E-01	3,69E-01	-4,69E-04	2,35E-04	SIG	0,000	SIG

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 2										
V12	2,94E-01	2,95E-01	3,16E-04	2,72E-01	3,20E-01	-7,76E-04	3,16E-04	SIG	0,000	SIG
V13	3,27E-01	3,28E-01	3,81E-04	3,05E-01	3,51E-01	-8,64E-04	3,80E-04	SIG	0,000	SIG
V14	4,48E-01	4,48E-01	1,36E-04	4,14E-01	4,74E-01	7,24E-04	1,36E-04	SIG	0,000	SIG
V15	4,84E-01	4,84E-01	7,23E-05	4,62E-01	5,13E-01	-9,64E-05	7,23E-05	SIG	0,000	SIG
V16	6,79E-01	6,76E-01	1,68E-03	6,47E-01	7,05E-01	3,54E-03	1,67E-03	SIG	0,000	SIG
V17	6,88E-01	6,84E-01	1,37E-03	6,78E-01	7,05E-01	3,26E-03	1,36E-03	SIG	0,000	SIG
V18	7,99E-01	7,94E-01	3,28E-03	7,86E-01	8,12E-01	5,06E-03	3,25E-03	SIG	0,000	SIG
V19	8,83E-01	8,77E-01	4,51E-03	8,72E-01	8,91E-01	6,05E-03	4,47E-03	SIG	0,000	SIG
V20	8,18E-01	8,11E-01	6,03E-03	8,09E-01	8,35E-01	6,88E-03	5,98E-03	SIG	0,000	SIG
V21	5,32E-01	5,30E-01	8,95E-04	5,10E-01	5,51E-01	2,52E-03	8,89E-04	SIG	0,000	SIG
V22	3,54E-01	3,53E-01	1,84E-04	3,24E-01	3,75E-01	1,14E-03	1,83E-04	SIG	0,000	SIG
V23	3,99E-01	3,98E-01	2,34E-04	3,67E-01	4,17E-01	1,60E-03	2,32E-04	SIG	0,000	SIG
V24	4,35E-01	4,33E-01	5,59E-04	4,08E-01	4,57E-01	2,00E-03	5,55E-04	SIG	0,000	SIG
V25	3,08E-01	3,07E-01	4,79E-05	2,89E-01	3,21E-01	5,85E-04	4,76E-05	SIG	0,000	SIG

Nos fatores 1 e 2, os enfiamentos e as variâncias são bem reduzidos, isto é, as estimativas são acuradas e precisas, o que implica EMQ'S pequenos. Portanto, as distribuições amostrais possuem as qualidades desejáveis para gerarem bons estimadores e inferências confiáveis.

No fator 1 e também no fator 2 todas as variáveis são estatisticamente significantes pela análise do intervalo de confiança. Os resultados das inferências estatísticas pelo valor-p coincidem os do intervalo de confiança.

A significância prática é obtida de forma análoga a realizada no TIAFIC e com base no quadro 3, capítulo 5.

Comparando neste estudo de caso pelos EMQ'S, observa-se que o método *jackknife* é visualmente superior ao método *bootstrap*.

## 8.6.CASO 6: BASE POPULAÇÃO DE CONDATOS AMERICANOS

Os dados que serviram de base para este estudo de caso é formada de 3141 condatos americanos onde foram registradas a população absoluta por gênero e idade e que constam do Quadro 4.

**Quadro 4**

Variáveis	Descrição
V <sub>1</sub>	População masculina de 0 a 4 anos
V <sub>2</sub>	População masculina de 5 a 9 anos
V <sub>3</sub>	População masculina de 10 a 14 anos
V <sub>4</sub>	População masculina de 15 a 19 anos
V <sub>5</sub>	População masculina de 20 a 24 anos
V <sub>6</sub>	População masculina de 25 a 29 anos
V <sub>7</sub>	População masculina de 30 a 34 anos
V <sub>8</sub>	População masculina de 35 a 39 anos
V <sub>9</sub>	População masculina de 40 a 44 anos
V <sub>10</sub>	População masculina de 45 a 59 anos
V <sub>11</sub>	População masculina de 50 a 54 anos
V <sub>12</sub>	População masculina de 55 a 59 anos
V <sub>13</sub>	População masculina de 60 a 64 anos
V <sub>14</sub>	População masculina de 65 a 69 anos
V <sub>15</sub>	População masculina de 70 a 74 anos
V <sub>16</sub>	População masculina de 75 a 79 anos
V <sub>17</sub>	População masculina de 80 a 84 anos
V <sub>18</sub>	População masculina com mais de 85 anos
V <sub>19</sub>	População feminina de 0 a 4 anos
V <sub>20</sub>	População feminina de 5 a 9 anos
V <sub>21</sub>	População feminina de 10 a 14 anos
V <sub>22</sub>	População feminina de 15 a 19 anos
V <sub>23</sub>	População feminina de 20 a 24 anos
V <sub>24</sub>	População feminina de 25 a 29 anos

Variáveis	Descrição
V <sub>25</sub>	População feminina de 30 a 34 anos
V <sub>26</sub>	População feminina de 35 a 39 anos
V <sub>27</sub>	População feminina de 40 a 44 anos
V <sub>28</sub>	População feminina de 45 a 49 anos
V <sub>29</sub>	População feminina de 50 a 54 anos
V <sub>30</sub>	População feminina de 55 a 59 anos
V <sub>31</sub>	População feminina de 60 a 64 anos
V <sub>32</sub>	População feminina de 65 a 69 anos
V <sub>33</sub>	População feminina de 70 a 74 anos
V <sub>34</sub>	População feminina de 75 a 79 anos
V <sub>35</sub>	População feminina de 80 a 84 anos
V <sub>36</sub>	População feminina com mais de 85 anos

O objetivo da aplicação da análise fatorial é tentar entender como se distribui a população dos condatos americanos investigados reduzindo a dimensão do problema por um menor número de variáveis ocultas e que guardam um quantidade razoável de informação das variáveis originais.

Os resultados da inferência estatística constam das páginas abaixo.



### 8.6.1. Análise dos Resultados *Bootstrap*

O *bootstrap* foi realizado com um B=1000 e os resultados constam da Tabela 50.

**Tabela 50**

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 1										
V1	8,11E-01	8,16E-01	2,22E-04	7,85E-01	8,44E-01	-4,46E-03	2,02E-04	SIG	0,000	SIG
V2	8,06E-01	8,12E-01	2,64E-04	7,79E-01	8,43E-01	-5,78E-03	2,30E-04	SIG	0,000	SIG
V3	7,97E-01	8,03E-01	2,87E-04	7,70E-01	8,36E-01	-6,52E-03	2,45E-04	SIG	0,000	SIG
V4	8,02E-01	8,07E-01	2,35E-04	7,75E-01	8,37E-01	-5,30E-03	2,07E-04	SIG	0,000	SIG
V5	8,13E-01	8,16E-01	2,07E-04	7,86E-01	8,43E-01	-2,61E-03	2,00E-04	SIG	0,000	SIG
V6	8,10E-01	8,16E-01	2,07E-04	7,87E-01	8,43E-01	-6,06E-03	1,70E-04	SIG	0,000	SIG
V7	8,03E-01	8,10E-01	2,36E-04	7,80E-01	8,41E-01	-7,41E-03	1,81E-04	SIG	0,000	SIG
V8	7,94E-01	8,02E-01	2,56E-04	7,72E-01	8,35E-01	-8,58E-03	1,82E-04	SIG	0,000	SIG
V9	7,81E-01	7,92E-01	3,14E-04	7,59E-01	8,28E-01	-1,12E-02	1,90E-04	SIG	0,000	SIG
V10	7,74E-01	7,85E-01	3,05E-04	7,53E-01	8,21E-01	-1,15E-02	1,74E-04	SIG	0,000	SIG
V11	7,55E-01	7,66E-01	2,81E-04	7,35E-01	7,99E-01	-1,17E-02	1,45E-04	SIG	0,000	SIG
V12	7,26E-01	7,39E-01	2,67E-04	7,07E-01	7,70E-01	-1,23E-02	1,15E-04	SIG	0,000	SIG
V13	6,74E-01	6,87E-01	2,33E-04	6,57E-01	7,15E-01	-1,26E-02	7,43E-05	SIG	0,000	SIG
V14	6,33E-01	6,44E-01	2,11E-04	6,18E-01	6,71E-01	-1,12E-02	8,54E-05	SIG	0,000	SIG
V15	5,85E-01	5,95E-01	3,75E-04	5,56E-01	6,32E-01	-1,04E-02	2,68E-04	SIG	0,000	SIG
V16	5,53E-01	5,62E-01	6,92E-04	5,10E-01	6,13E-01	-9,09E-03	6,09E-04	SIG	0,000	SIG
V17	5,34E-01	5,44E-01	9,55E-04	4,82E-01	6,03E-01	-1,01E-02	8,54E-04	SIG	0,000	SIG
V18	5,49E-01	5,58E-01	9,50E-04	4,95E-01	6,16E-01	-8,88E-03	8,71E-04	SIG	0,000	SIG
V19	8,12E-01	8,16E-01	2,26E-04	7,85E-01	8,45E-01	-4,43E-03	2,06E-04	SIG	0,000	SIG
V20	8,06E-01	8,12E-01	2,67E-04	7,79E-01	8,43E-01	-5,83E-03	2,33E-04	SIG	0,000	SIG
V21	7,96E-01	8,02E-01	2,91E-04	7,69E-01	8,35E-01	-6,61E-03	2,48E-04	SIG	0,000	SIG
V22	7,91E-01	7,98E-01	2,54E-04	7,66E-01	8,29E-01	-6,48E-03	2,12E-04	SIG	0,000	SIG
V23	7,94E-01	8,00E-01	2,26E-04	7,70E-01	8,29E-01	-6,07E-03	1,89E-04	SIG	0,000	SIG
V24	7,93E-01	8,01E-01	2,63E-04	7,70E-01	8,35E-01	-8,65E-03	1,88E-04	SIG	0,000	SIG
V25	7,89E-01	7,98E-01	2,82E-04	7,66E-01	8,32E-01	-9,26E-03	1,96E-04	SIG	0,000	SIG
V26	7,83E-01	7,93E-01	2,97E-04	7,60E-01	8,28E-01	-1,03E-02	1,91E-04	SIG	0,000	SIG
V27	7,72E-01	7,84E-01	3,47E-04	7,49E-01	8,22E-01	-1,23E-02	1,96E-04	SIG	0,000	SIG
V28	7,58E-01	7,70E-01	3,23E-04	7,36E-01	8,06E-01	-1,26E-02	1,63E-04	SIG	0,000	SIG
V29	7,34E-01	7,46E-01	3,06E-04	7,12E-01	7,81E-01	-1,27E-02	1,45E-04	SIG	0,000	SIG
V30	7,03E-01	7,16E-01	2,64E-04	6,84E-01	7,47E-01	-1,27E-02	1,02E-04	SIG	0,000	SIG
V31	6,56E-01	6,69E-01	2,18E-04	6,40E-01	6,96E-01	-1,26E-02	5,88E-05	SIG	0,000	SIG
V32	6,21E-01	6,33E-01	2,00E-04	6,08E-01	6,59E-01	-1,21E-02	5,31E-05	SIG	0,000	SIG
V33	5,78E-01	5,91E-01	2,86E-04	5,59E-01	6,25E-01	-1,29E-02	1,20E-04	SIG	0,000	SIG
V34	5,63E-01	5,77E-01	3,33E-04	5,41E-01	6,12E-01	-1,36E-02	1,49E-04	SIG	0,000	SIG
V35	5,62E-01	5,77E-01	3,34E-04	5,42E-01	6,13E-01	-1,51E-02	1,07E-04	SIG	0,000	SIG
V36	5,91E-01	6,03E-01	4,14E-04	5,62E-01	6,41E-01	-1,23E-02	2,62E-04	SIG	0,000	SIG
V1	5,80E-01	5,71E-01	5,10E-04	5,25E-01	6,15E-01	8,48E-03	4,38E-04	SIG	0,000	SIG
V2	5,85E-01	5,74E-01	6,37E-04	5,22E-01	6,21E-01	1,10E-02	5,15E-04	SIG	0,000	SIG
V3	5,98E-01	5,86E-01	6,41E-04	5,33E-01	6,31E-01	1,16E-02	5,06E-04	SIG	0,000	SIG

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 2										
V4	5,93E-01	5,83E-01	5,07E-04	5,36E-01	6,26E-01	9,40E-03	4,19E-04	SIG	0,00	SIG
V5	5,61E-01	5,55E-01	5,12E-04	5,12E-01	6,00E-01	5,90E-03	4,77E-04	SIG	0,000	SIG
V6	5,79E-01	5,70E-01	4,76E-04	5,24E-01	6,12E-01	9,37E-03	3,88E-04	SIG	0,000	SIG
V7	5,92E-01	5,81E-01	4,90E-04	5,35E-01	6,23E-01	1,09E-02	3,71E-04	SIG	0,000	SIG
V8	6,05E-01	5,93E-01	4,87E-04	5,46E-01	6,34E-01	1,20E-02	3,43E-04	SIG	0,000	SIG
V9	6,20E-01	6,05E-01	5,54E-04	5,56E-01	6,49E-01	1,48E-02	3,34E-04	SIG	0,000	SIG
V10	6,29E-01	6,14E-01	5,26E-04	5,65E-01	6,55E-01	1,50E-02	3,00E-04	SIG	0,000	SIG
V11	6,52E-01	6,37E-01	4,28E-04	5,94E-01	6,75E-01	1,44E-02	2,19E-04	SIG	0,000	SIG
V12	6,83E-01	6,69E-01	3,44E-04	6,32E-01	7,04E-01	1,40E-02	1,47E-04	SIG	0,000	SIG
V13	7,35E-01	7,22E-01	2,21E-04	6,94E-01	7,51E-01	1,24E-02	6,71E-05	SIG	0,000	SIG
V14	7,71E-01	7,61E-01	1,50E-04	7,38E-01	7,83E-01	9,77E-03	5,42E-05	SIG	0,000	SIG
V15	8,07E-01	7,99E-01	1,95E-04	7,72E-01	8,26E-01	7,90E-03	1,33E-04	SIG	0,000	SIG
V16	8,28E-01	8,22E-01	2,93E-04	7,86E-01	8,55E-01	6,15E-03	2,56E-04	SIG	0,000	SIG
V17	8,36E-01	8,30E-01	3,45E-04	7,93E-01	8,65E-01	6,39E-03	3,04E-04	SIG	0,000	SIG
V18	8,24E-01	8,18E-01	3,66E-04	7,80E-01	8,57E-01	6,00E-03	3,30E-04	SIG	0,000	SIG
V19	5,79E-01	5,71E-01	5,20E-04	5,24E-01	6,15E-01	8,48E-03	4,49E-04	SIG	0,000	SIG
V20	5,86E-01	5,75E-01	6,43E-04	5,22E-01	6,21E-01	1,11E-02	5,20E-04	SIG	0,000	SIG
V21	5,99E-01	5,87E-01	6,44E-04	5,34E-01	6,33E-01	1,17E-02	5,08E-04	SIG	0,000	SIG
V22	6,06E-01	5,95E-01	5,24E-04	5,46E-01	6,38E-01	1,08E-02	4,06E-04	SIG	0,000	SIG
V23	5,98E-01	5,88E-01	4,90E-04	5,41E-01	6,30E-01	1,01E-02	3,88E-04	SIG	0,000	SIG
V24	6,06E-01	5,93E-01	5,19E-04	5,44E-01	6,35E-01	1,25E-02	3,63E-04	SIG	0,000	SIG
V25	6,13E-01	6,00E-01	5,17E-04	5,53E-01	6,42E-01	1,28E-02	3,52E-04	SIG	0,000	SIG
V26	6,20E-01	6,06E-01	5,18E-04	5,58E-01	6,48E-01	1,38E-02	3,28E-04	SIG	0,000	SIG
V27	6,32E-01	6,16E-01	5,70E-04	5,66E-01	6,60E-01	1,58E-02	3,22E-04	SIG	0,000	SIG
V28	6,49E-01	6,33E-01	4,95E-04	5,85E-01	6,73E-01	1,56E-02	2,51E-04	SIG	0,000	SIG
V29	6,76E-01	6,61E-01	4,12E-04	6,19E-01	6,99E-01	1,47E-02	1,95E-04	SIG	0,000	SIG
V30	7,07E-01	6,94E-01	2,99E-04	6,60E-01	7,27E-01	1,35E-02	1,16E-04	SIG	0,000	SIG
V31	7,52E-01	7,41E-01	1,85E-04	7,14E-01	7,66E-01	1,17E-02	4,77E-05	SIG	0,000	SIG
V32	7,82E-01	7,72E-01	1,35E-04	7,51E-01	7,92E-01	1,01E-02	3,33E-05	SIG	0,000	SIG
V33	8,15E-01	8,05E-01	1,54E-04	7,79E-01	8,28E-01	9,53E-03	6,29E-05	SIG	0,000	SIG
V34	8,24E-01	8,14E-01	1,60E-04	7,89E-01	8,38E-01	9,68E-03	6,68E-05	SIG	0,000	SIG
V35	8,23E-01	8,12E-01	1,59E-04	7,86E-01	8,35E-01	1,10E-02	3,85E-05	SIG	0,000	SIG
V36	7,97E-01	7,87E-01	2,07E-04	7,60E-01	8,14E-01	1,04E-02	9,81E-05	SIG	0,000	SIG

Da Tabela 50 constata-se que como em todos os outros estudos de casos, as estimativas são acuradas e precisas, o que implicam em EMQ'S bem reduzidos. Estes resultados corroboram a teoria de que as técnicas *bootstrap* geram estimativas ou distribuições por amostragem de cargas fatoriais de alta qualidade, possibilitando inferências seguras e precisas.

Tanto no fator 1, quanto no fator 2, em todas as variáveis as cargas fatoriais são estatisticamente significantes. Isto tanto para inferência pelo intervalo de confiança quanto pelo valor-p.

Uma observação importante é que se trabalharam neste estudo de caso com uma amostra muito grande de condatos americanos(3141 condatos) e um número de reamostragem suficiente grande, o que pode ter influenciado nos resultados estatísticos significantes na amostra.

### 8.5.1. Análise dos Resultados *Jackknife*

Os resultados da rodada *jackknife* constam da Tabela 51.

**Tabela 51**

VAR	Cargas	Média	Variância	LI	LS	Vies	EMQ	SIGIC	Valor-p	SIGVP
Fator 1										
V1	8,11E-01	8,12E-01	5,05E-06	8,11E-01	8,12E-01	-3,47E-04	4,93E-06	SIG	0,000	SIG
V2	8,06E-01	8,07E-01	5,15E-06	8,06E-01	8,07E-01	-3,51E-04	5,02E-06	SIG	0,000	SIG
V3	7,97E-01	7,97E-01	5,38E-06	7,97E-01	7,97E-01	-3,59E-04	5,25E-06	SIG	0,000	SIG
V4	8,02E-01	8,02E-01	5,30E-06	8,02E-01	8,02E-01	-3,55E-04	5,18E-06	SIG	0,000	SIG
V5	8,13E-01	8,14E-01	4,85E-06	8,13E-01	8,13E-01	-3,33E-04	4,73E-06	SIG	0,000	SIG
V6	8,10E-01	8,10E-01	5,06E-06	8,10E-01	8,10E-01	-3,46E-04	4,94E-06	SIG	0,000	SIG
V7	8,03E-01	8,03E-01	5,28E-06	8,03E-01	8,03E-01	-3,55E-04	5,16E-06	SIG	0,000	SIG
V8	7,94E-01	7,94E-01	5,53E-06	7,94E-01	7,94E-01	-3,64E-04	5,40E-06	SIG	0,000	SIG
V9	7,81E-01	7,82E-01	5,83E-06	7,81E-01	7,82E-01	-3,74E-04	5,69E-06	SIG	0,000	SIG
V10	7,74E-01	7,74E-01	5,99E-06	7,74E-01	7,74E-01	-3,80E-04	5,84E-06	SIG	0,000	SIG
V11	7,55E-01	7,55E-01	6,44E-06	7,55E-01	7,56E-01	-3,94E-04	6,28E-06	SIG	0,000	SIG
V12	7,26E-01	7,27E-01	7,11E-06	7,26E-01	7,27E-01	-4,13E-04	6,94E-06	SIG	0,000	SIG
V13	6,74E-01	6,75E-01	8,28E-06	6,74E-01	6,75E-01	-4,45E-04	8,08E-06	SIG	0,000	SIG
V14	6,33E-01	6,34E-01	9,21E-06	6,33E-01	6,34E-01	-4,67E-04	8,99E-06	SIG	0,000	SIG
V15	5,85E-01	5,85E-01	1,03E-05	5,85E-01	5,85E-01	-4,89E-04	1,00E-05	SIG	0,000	SIG
V16	5,53E-01	5,54E-01	1,10E-05	5,53E-01	5,54E-01	-5,01E-04	1,07E-05	SIG	0,000	SIG
V17	5,34E-01	5,34E-01	1,13E-05	5,34E-01	5,34E-01	-5,07E-04	1,11E-05	SIG	0,000	SIG
V18	5,49E-01	5,49E-01	1,11E-05	5,49E-01	5,49E-01	-4,98E-04	1,09E-05	SIG	0,000	SIG
V19	8,12E-01	8,12E-01	5,05E-06	8,12E-01	8,12E-01	-3,47E-04	4,93E-06	SIG	0,000	SIG
V20	8,06E-01	8,06E-01	5,16E-06	8,06E-01	8,06E-01	-3,52E-04	5,03E-06	SIG	0,000	SIG
V21	7,96E-01	7,96E-01	5,41E-06	7,96E-01	7,96E-01	-3,60E-04	5,28E-06	SIG	0,000	SIG
V22	7,91E-01	7,92E-01	5,54E-06	7,91E-01	7,92E-01	-3,63E-04	5,41E-06	SIG	0,000	SIG
V23	7,94E-01	7,94E-01	5,45E-06	7,94E-01	7,94E-01	-3,58E-04	5,32E-06	SIG	0,000	SIG
V24	7,93E-01	7,93E-01	5,55E-06	7,93E-01	7,93E-01	-3,64E-04	5,41E-06	SIG	0,000	SIG
V25	7,89E-01	7,90E-01	5,67E-06	7,89E-01	7,90E-01	-3,69E-04	5,54E-06	SIG	0,000	SIG
V26	7,83E-01	7,83E-01	5,82E-06	7,83E-01	7,83E-01	-3,74E-04	5,68E-06	SIG	0,000	SIG
V27	7,72E-01	7,72E-01	6,08E-06	7,72E-01	7,72E-01	-3,82E-04	5,94E-06	SIG	0,000	SIG
V28	7,58E-01	7,58E-01	6,40E-06	7,58E-01	7,59E-01	-3,92E-04	6,24E-06	SIG	0,000	SIG
V29	7,34E-01	7,34E-01	6,96E-06	7,34E-01	7,35E-01	-4,09E-04	6,79E-06	SIG	0,000	SIG
V30	7,03E-01	7,04E-01	7,65E-06	7,03E-01	7,04E-01	-4,29E-04	7,47E-06	SIG	0,000	SIG
V31	6,56E-01	6,56E-01	8,73E-06	6,56E-01	6,56E-01	-4,56E-04	8,53E-06	SIG	0,000	SIG
V32	6,21E-01	6,22E-01	9,50E-06	6,21E-01	6,22E-01	-4,75E-04	9,27E-06	SIG	0,000	SIG
V33	5,78E-01	5,79E-01	1,04E-05	5,78E-01	5,79E-01	-4,95E-04	1,02E-05	SIG	0,000	SIG
V34	5,63E-01	5,63E-01	1,07E-05	5,63E-01	5,63E-01	-5,01E-04	1,04E-05	SIG	0,000	SIG
V35	5,62E-01	5,62E-01	1,06E-05	5,62E-01	5,62E-01	-5,01E-04	1,04E-05	SIG	0,000	SIG
V36	5,91E-01	5,91E-01	1,01E-05	5,91E-01	5,91E-01	-4,83E-04	9,84E-06	SIG	0,000	SIG
V1	5,80E-01	5,79E-01	1,05E-05	5,79E-01	5,80E-01	5,00E-04	1,02E-05	SIG	0,000	SIG
V2	5,85E-01	5,85E-01	1,04E-05	5,85E-01	5,85E-01	4,98E-04	1,01E-05	SIG	0,000	SIG
V3	5,98E-01	5,97E-01	1,01E-05	5,97E-01	5,98E-01	4,93E-04	9,88E-06	SIG	0,000	SIG
V4	5,93E-01	5,92E-01	1,02E-05	5,92E-01	5,93E-01	4,94E-04	1,00E-05	SIG	0,000	SIG

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 2										
V5	5,61E-01	5,61E-01	1,07E-05	5,61E-01	5,61E-01	4,98E-04	1,04E-05	SIG	0,000	SIG
V6	5,79E-01	5,79E-01	1,05E-05	5,79E-01	5,79E-01	4,98E-04	1,02E-05	SIG	0,000	SIG
V7	5,92E-01	5,92E-01	1,03E-05	5,92E-01	5,92E-01	4,95E-04	1,00E-05	SIG	0,000	SIG
V8	6,05E-01	6,05E-01	1,00E-05	6,05E-01	6,05E-01	4,90E-04	9,79E-06	SIG	0,000	SIG
V9	6,20E-01	6,19E-01	9,78E-06	6,19E-01	6,20E-01	4,85E-04	9,55E-06	SIG	0,000	SIG
V10	6,29E-01	6,28E-01	9,56E-06	6,28E-01	6,29E-01	4,80E-04	9,33E-06	SIG	0,000	SIG
V11	6,52E-01	6,51E-01	9,10E-06	6,51E-01	6,52E-01	4,68E-04	8,88E-06	SIG	0,000	SIG
V12	6,83E-01	6,83E-01	8,47E-06	6,82E-01	6,83E-01	4,51E-04	8,26E-06	SIG	0,000	SIG
V13	7,35E-01	7,34E-01	7,36E-06	7,34E-01	7,35E-01	4,20E-04	7,18E-06	SIG	0,000	SIG
V14	7,71E-01	7,71E-01	6,55E-06	7,71E-01	7,71E-01	3,94E-04	6,40E-06	SIG	0,000	SIG
V15	8,07E-01	8,07E-01	5,69E-06	8,07E-01	8,07E-01	3,65E-04	5,55E-06	SIG	0,000	SIG
V16	8,28E-01	8,27E-01	5,18E-06	8,27E-01	8,28E-01	3,45E-04	5,06E-06	SIG	0,000	SIG
V17	8,36E-01	8,36E-01	4,87E-06	8,36E-01	8,36E-01	3,33E-04	4,76E-06	SIG	0,000	SIG
V18	8,24E-01	8,24E-01	5,18E-06	8,24E-01	8,24E-01	3,42E-04	5,06E-06	SIG	0,000	SIG
V19	5,79E-01	5,79E-01	1,05E-05	5,79E-01	5,79E-01	5,00E-04	1,02E-05	SIG	0,000	SIG
V20	5,86E-01	5,85E-01	1,03E-05	5,85E-01	5,86E-01	4,98E-04	1,01E-05	SIG	0,000	SIG
V21	5,99E-01	5,99E-01	1,01E-05	5,98E-01	5,99E-01	4,92E-04	9,86E-06	SIG	0,000	SIG
V22	6,06E-01	6,05E-01	1,00E-05	6,05E-01	6,06E-01	4,89E-04	9,77E-06	SIG	0,000	SIG
V23	5,98E-01	5,98E-01	1,01E-05	5,97E-01	5,98E-01	4,89E-04	9,90E-06	SIG	0,000	SIG
V24	6,06E-01	6,05E-01	1,00E-05	6,05E-01	6,06E-01	4,90E-04	9,81E-06	SIG	0,000	SIG
V25	6,13E-01	6,12E-01	9,94E-06	6,12E-01	6,13E-01	4,88E-04	9,71E-06	SIG	0,000	SIG
V26	6,20E-01	6,20E-01	9,80E-06	6,20E-01	6,20E-01	4,85E-04	9,56E-06	SIG	0,000	SIG
V27	6,32E-01	6,32E-01	9,56E-06	6,31E-01	6,32E-01	4,79E-04	9,33E-06	SIG	0,000	SIG
V28	6,49E-01	6,48E-01	9,20E-06	6,47E-01	6,49E-01	4,71E-04	8,98E-06	SIG	0,000	SIG
V29	6,76E-01	6,75E-01	8,65E-06	6,75E-01	6,76E-01	4,56E-04	8,45E-06	SIG	0,000	SIG
V30	7,07E-01	7,07E-01	7,98E-06	7,07E-01	7,07E-01	4,38E-04	7,79E-06	SIG	0,000	SIG
V31	7,52E-01	7,52E-01	7,00E-06	7,52E-01	7,52E-01	4,09E-04	6,83E-06	SIG	0,000	SIG
V32	7,82E-01	7,82E-01	6,32E-06	7,82E-01	7,82E-01	3,87E-04	6,17E-06	SIG	0,000	SIG
V33	8,15E-01	8,14E-01	5,54E-06	8,14E-01	8,15E-01	3,61E-04	5,41E-06	SIG	0,000	SIG
V34	8,24E-01	8,24E-01	5,26E-06	8,24E-01	8,24E-01	3,52E-04	5,14E-06	SIG	0,000	SIG
V35	8,23E-01	8,22E-01	5,23E-06	8,23E-01	8,23E-01	3,52E-04	5,11E-06	SIG	0,000	SIG
V36	7,97E-01	7,97E-01	5,83E-06	7,97E-01	7,97E-01	3,69E-04	5,69E-06	SIG	0,000	SIG

Pela Tabela 51 , os resultados do método *jackknife* são análogos aos resultados e interpretações do *bootstrap*, mas este último método deve ser o preferido numa análise real indutiva deste problema, uma vez que revelou EMQ'S bem menores do que no outro método.

## 8.7. RESUMO DOS RESULTADOS DOS ESTUDOS DE CASO

O Quadro 5 faz um balanço, para os casos analisados, da comparação entre os métodos *bootstrap* e *jackknife* em função da análise do EMQ para estimação das distribuições por amostragem e construção de *intervalos de confiança* e *valor-p*:

**Quadro 5**

<b>Estudo de Casos</b>	<b><i>Bootstrap</i></b>	<b><i>Jackknife</i></b>
<b>1</b>	<i>Pior desempenho</i>	<i>Melhor desempenho</i>
<b>2</b>	<i>Igual desempenho</i>	<i>Igual desempenho</i>
<b>3</b>	<i>Melhor desempenho</i>	<i>Pior desempenho</i>
<b>4</b>	<i>Pior desempenho</i>	<i>Melhor desempenho</i>
<b>5</b>	<i>Pior desempenho</i>	<i>Melhor desempenho</i>
<b>6</b>	<i>Pior desempenho</i>	<i>Melhor desempenho</i>

Pelo balanço exposto, verifica-se que o método *jackknife* teve um saldo de desempenho superior ao do *bootstrap*.

O Quadro 6 faz-se um balanço, para os primeiros casos analisados, da aderência das estimativas à curva normal segundo os métodos de reamostragem utilizados, pois neste trabalho somente foram apresentados às análises para estas bases de dados, servindo de raciocínio indutivo para o que ocorrerá com os outros estudos de casos.

**Quadro 6**

<b>Estudo de Casos</b>	<i>Bootstrap</i>	<i>Jackknife</i>
<b>1</b>	<i>Normal</i>	<i>Não Normal</i>
<b>2</b>	<i>Não Normal</i>	<i>Normal</i>
<b>3</b>	<i>Não Normal</i>	<i>Normal</i>

Pelo balanço exposto, verifica-se que a normalidade das cargas fatoriais dos fatores foi um pouco mais regular no método *jackknife*, mas verifica-se também que a ocorrência da normalidade depende muito do problema que esteja sendo analisado. Pela irregularidade da aderência das cargas fatoriais ao modelo normal, recomenda-se utilizar intervalos de confiança não paramétrico(percentílicos) e para o cálculo dos valores-p para a inferência das cargas fatoriais.

Esta interpretação pode ser generalizada para os outros estudos de casos realizados nesta tese e para outros que venham a ser pesquisados.

Pelos ensaios da convergência realizados com as bases de dados, fica evidente que na simulação *bootstrap* um B=200 já apresenta bons resultados quanto à qualidade da distribuição por amostragem das cargas fatoriais dos fatores considerados, contudo ainda é recomendável, no método *bootstrap*, trabalhar-se com um B=1000, otimizando-se ao máximo a qualidade das distribuições por amostragens computadas.

# CAPÍTULO 9

## COMPARAÇÃO DO TIAF COM O MÉTODO TRADICIONAL DE HAIR E ANDERSON

O objetivo deste capítulo é realizar a comparação dos resultados do Tiaf com um critério tradicional para a significância de cargas fatoriais, baseados na teoria assintótica, por exemplo, o método apresentado por Hair e Anderson(2005).

A realização deste procedimento permite ratificar a eficácia do método estabelecido nesta tese.

A idéia é apresentar os resultados inferenciais dos dois métodos em dois estudos de casos que se efetuou: um de amostra grande(Caso 6) e um para amostra pequena(Caso 3).

Os resultados e interpretações constam das seções abaixo.

### 9.1. ESTUDO DE CASO 6:

A avaliação de cargas fatoriais significantes pelo Método de Hair e Anderson é baseada em um nível de significância de 5%, um nível de poder de 80% e erros padrão, os quais se pressupõem que seja o dobro dos coeficientes de correlações normais.

Segundo a Tabela 1 do capítulo 2, em uma amostra de 3141 condados, as cargas fatoriais de 0,30 ou mais são significantes, uma vez que essa abordagem considera as cargas de 0,30 como significantes para amostras de 350 ou maiores.



As Tabelas das seções seguintes apresentam os resultados comparativos dos dois métodos.

### 9.1.1- Resultados pelo Método *Bootstrap*

**Tabela 52**

**Resultados Comparativos do Método de Hair e Anderson (MétodoH. A) e o TIAF.**

Variáveis	Cargas	Método		Valor-p TIAF	TIAFVP
		H.A	TIAFIC		
<b>Fator 1</b>					
V1	0,81	SIG	SIG	0,000	SIG
V2	0,81	SIG	SIG	0,000	SIG
V3	0,80	SIG	SIG	0,000	SIG
V4	0,80	SIG	SIG	0,000	SIG
V5	0,81	SIG	SIG	0,000	SIG
V6	0,81	SIG	SIG	0,000	SIG
V7	0,80	SIG	SIG	0,000	SIG
V8	0,79	SIG	SIG	0,000	SIG
V9	0,78	SIG	SIG	0,000	SIG
V10	0,77	SIG	SIG	0,000	SIG
V11	0,76	SIG	SIG	0,000	SIG
V12	0,73	SIG	SIG	0,000	SIG
V13	0,67	SIG	SIG	0,000	SIG
V14	0,63	SIG	SIG	0,000	SIG
V15	0,59	SIG	SIG	0,000	SIG
V16	0,55	SIG	SIG	0,000	SIG
V17	0,53	SIG	SIG	0,000	SIG
V18	0,55	SIG	SIG	0,000	SIG
V19	0,81	SIG	SIG	0,000	SIG
V20	0,81	SIG	SIG	0,000	SIG
V21	0,80	SIG	SIG	0,000	SIG
V22	0,79	SIG	SIG	0,000	SIG
V23	0,79	SIG	SIG	0,000	SIG
V24	0,79	SIG	SIG	0,000	SIG
V25	0,79	SIG	SIG	0,000	SIG
V26	0,78	SIG	SIG	0,000	SIG
V27	0,77	SIG	SIG	0,000	SIG
V28	0,76	SIG	SIG	0,000	SIG
V29	0,73	SIG	SIG	0,000	SIG
V30	0,70	SIG	SIG	0,000	SIG
V31	0,66	SIG	SIG	0,000	SIG
V32	0,62	SIG	SIG	0,000	SIG
V33	0,58	SIG	SIG	0,000	SIG
V34	0,56	SIG	SIG	0,000	SIG
V35	0,56	SIG	SIG	0,000	SIG

Variáveis	Cargas	Método H.A	TIAFIC	Valor-p TIAF	TIAFVP
<b>Fator 1</b>					
V36	0,59	SIG	SIG	0,000	SIG
Variáveis	Cargas	Método H.A	TIAFIC	Valor-p TIAF	TIAFVP
<b>Fator 2</b>					
V1	0,58	SIG	SIG	0,000	SIG
V2	0,59	SIG	SIG	0,000	SIG
V3	0,60	SIG	SIG	0,000	SIG
V4	0,59	SIG	SIG	0,000	SIG
V5	0,56	SIG	SIG	0,000	SIG
V6	0,58	SIG	SIG	0,000	SIG
V7	0,59	SIG	SIG	0,000	SIG
V8	0,61	SIG	SIG	0,000	SIG
V9	0,62	SIG	SIG	0,000	SIG
V10	0,63	SIG	SIG	0,000	SIG
V11	0,65	SIG	SIG	0,000	SIG
V12	0,68	SIG	SIG	0,000	SIG
V13	0,74	SIG	SIG	0,000	SIG
V14	0,77	SIG	SIG	0,000	SIG
V15	0,81	SIG	SIG	0,000	SIG
V16	0,83	SIG	SIG	0,000	SIG
V17	0,84	SIG	SIG	0,000	SIG
V18	0,82	SIG	SIG	0,000	SIG
V19	0,58	SIG	SIG	0,000	SIG
V20	0,59	SIG	SIG	0,000	SIG
V21	0,60	SIG	SIG	0,000	SIG
V22	0,61	SIG	SIG	0,000	SIG
V23	0,60	SIG	SIG	0,000	SIG
V24	0,61	SIG	SIG	0,000	SIG
V25	0,61	SIG	SIG	0,000	SIG
V26	0,62	SIG	SIG	0,000	SIG
V27	0,63	SIG	SIG	0,000	SIG
V28	0,65	SIG	SIG	0,000	SIG
V29	0,68	SIG	SIG	0,000	SIG
V30	0,71	SIG	SIG	0,000	SIG
V31	0,75	SIG	SIG	0,000	SIG
V32	0,78	SIG	SIG	0,000	SIG
V33	0,82	SIG	SIG	0,000	SIG
V34	0,82	SIG	SIG	0,000	SIG
V35	0,82	SIG	SIG	0,000	SIG
V36	0,80	SIG	SIG	0,000	SIG

Pela interpretação da Tabela 52, tem-se que comparação dos resultados inferenciais entre o Método Hair e Anderson e o TIAF indica

que a avaliações de significância estatística são idênticas para ambos os métodos.

### 9.1.1- Resultados pelo Método *Jackknife*

**Tabela 53**

**Resultados Comparativos do Método de Hair e Anderson (MétodoH. A) e o TIAF.**

VAR	Cargas	Método H.A	TIAFIC	Valor-p TIAF	TIAFVP
<b>Fator 1</b>					
V1	0,81	SIG	SIG	0,000	SIG
V2	0,81	SIG	SIG	0,000	SIG
V3	0,80	SIG	SIG	0,000	SIG
V4	0,80	SIG	SIG	0,000	SIG
V5	0,81	SIG	SIG	0,000	SIG
V6	0,81	SIG	SIG	0,000	SIG
V7	0,80	SIG	SIG	0,000	SIG
V8	0,79	SIG	SIG	0,000	SIG
V9	0,78	SIG	SIG	0,000	SIG
V10	0,77	SIG	SIG	0,000	SIG
V11	0,76	SIG	SIG	0,000	SIG
V12	0,73	SIG	SIG	0,000	SIG
V13	0,67	SIG	SIG	0,000	SIG
V14	0,63	SIG	SIG	0,000	SIG
V15	0,59	SIG	SIG	0,000	SIG
V16	0,55	SIG	SIG	0,000	SIG
V17	0,53	SIG	SIG	0,000	SIG
V18	0,55	SIG	SIG	0,000	SIG
V19	0,81	SIG	SIG	0,000	SIG
V20	0,81	SIG	SIG	0,000	SIG
V21	0,80	SIG	SIG	0,000	SIG
V22	0,79	SIG	SIG	0,000	SIG
V23	0,79	SIG	SIG	0,000	SIG
V24	0,79	SIG	SIG	0,000	SIG
V25	0,79	SIG	SIG	0,000	SIG
V26	0,78	SIG	SIG	0,000	SIG
V27	0,77	SIG	SIG	0,000	SIG
V28	0,76	SIG	SIG	0,000	SIG
V29	0,73	SIG	SIG	0,000	SIG
V30	0,70	SIG	SIG	0,000	SIG
V31	0,66	SIG	SIG	0,000	SIG
V32	0,62	SIG	SIG	0,000	SIG
V33	0,58	SIG	SIG	0,000	SIG

VAR	Cargas	Método H.A	TIAFIC	Valor-p TIAF	TIAFVP
<b>Fator 1</b>					
V34	0,56	SIG	SIG	0,000	SIG
V35	0,56	SIG	SIG	0,000	SIG
V36	0,59	SIG	SIG	0,000	SIG
VAR	Cargas	Método H.A	TIAFIC	Valor-p TIAF	TIAFVP
<b>Fator 2</b>					
V1	0,58	SIG	SIG	0,000	SIG
V2	0,59	SIG	SIG	0,000	SIG
V3	0,60	SIG	SIG	0,000	SIG
V4	0,59	SIG	SIG	0,000	SIG
V5	0,56	SIG	SIG	0,000	SIG
V6	0,58	SIG	SIG	0,000	SIG
V7	0,59	SIG	SIG	0,000	SIG
V8	0,61	SIG	SIG	0,000	SIG
V9	0,62	SIG	SIG	0,000	SIG
V10	0,63	SIG	SIG	0,000	SIG
V11	0,65	SIG	SIG	0,000	SIG
V12	0,68	SIG	SIG	0,000	SIG
V13	0,74	SIG	SIG	0,000	SIG
V14	0,77	SIG	SIG	0,000	SIG
V15	0,81	SIG	SIG	0,000	SIG
V16	0,83	SIG	SIG	0,000	SIG
V17	0,84	SIG	SIG	0,000	SIG
V18	0,82	SIG	SIG	0,000	SIG
V19	0,58	SIG	SIG	0,000	SIG
V20	0,59	SIG	SIG	0,000	SIG
V21	0,60	SIG	SIG	0,000	SIG
V22	0,61	SIG	SIG	0,000	SIG
V23	0,60	SIG	SIG	0,000	SIG
V24	0,61	SIG	SIG	0,000	SIG
V25	0,61	SIG	SIG	0,000	SIG
V26	0,62	SIG	SIG	0,000	SIG
V27	0,63	SIG	SIG	0,000	SIG
V28	0,65	SIG	SIG	0,000	SIG
V29	0,68	SIG	SIG	0,000	SIG
V30	0,71	SIG	SIG	0,000	SIG
V31	0,75	SIG	SIG	0,000	SIG
V32	0,78	SIG	SIG	0,000	SIG
V33	0,82	SIG	SIG	0,000	SIG
V34	0,82	SIG	SIG	0,000	SIG
V35	0,82	SIG	SIG	0,000	SIG
V36	0,80	SIG	SIG	0,000	SIG

Como o Método Tradicional Hair e Anderson(2005) não depende das reamostragens e é baseado nas cargas fatoriais da amostra original, os resultados de significância *bootstrap* é igual ao do *jackknife* e como

também os resultados dos TIAF'S *bootstrap* e *jackknife* foram idênticos para ambos os métodos(*bootstrap* e *jackknife*) as interpretações desta seção são as mesmas da seção anterior.

## 9.2. ESTUDO DE CASO 3

Segundo a Tabela 1 do capítulo 2, em uma amostra de 25 respondentes, as cargas fatoriais de 0,75 ou mais são significantes, uma vez que essa abordagem considera as cargas de 0,75 como significantes para amostras de 50 ou menos.

As Tabelas das seções seguintes apresentam os resultados comparativos dos dois métodos.

### 9.2.1. Resultados pelo Método *Bootstrap*

**Tabela 54**

**Resultados Comparativos do Método de Hair e Anderson (Método H. A) e o TIAF.**

VAR	Cargas	Método H.A	TIAFIC	Valor-p	TIAFVP
<b>Fator 1</b>					
V1	0,85	SIG	NSIG	0,000	SIG
V2	0,04	NSIG	NSIG	0,375	NSIG
V3	0,72	NSIG	NSIG	0,000	SIG
V4	0,13	NSIG	NSIG	0,305	NSIG
V5	0,80	SIG	NSIG	0,000	SIG
V6	-0,08	NSIG	NSIG	0,385	NSIG
V7	-0,36	NSIG	NSIG	0,185	NSIG
<b>Fator 2</b>					
V1	-0,04	NSIG	NSIG	0,625	NSIG
V2	-0,31	NSIG	NSIG	0,340	NSIG
V3	-0,42	NSIG	NSIG	0,165	NSIG
V4	0,66	NSIG	NSIG	0,030	SIG
V5	0,25	NSIG	NSIG	0,245	NSIG
V6	-0,11	NSIG	NSIG	0,500	NSIG
V7	0,78	SIG	NSIG	0,030	SIG

Neste estudo de caso, os resultados entre os métodos tradicionais Hair e Anderson(2005) e o TIAF apresentam divergências. O TIAFIC foi bem mais rigoroso e considerou todas as cargas como não significantes. Já o TIAFVP apresentou resultados mais coerentes indicando como significantes as cargas que logicamente se esperaria que fossem. O método

tradicional de Hair e Anderson divergiu do TIAFVP na variável V<sub>3</sub>, no fator 1, onde a carga 0,72 foi considerada não significativa naquele método e significativa neste método. No fator 2, a divergência ocorreu na variável V<sub>4</sub>, onde a carga 0,64 foi considerada não significativa no método tradicional de Hair e Anderson e significativa no TIAFVP. O método TIAFVP apresentou resultados mais realísticos e coerentes.

Os resultados do Método Tradicional de Hair e Anderson foram mais rigorosos na indicação da significância pois somente sofreram influência do pequeno tamanho da amostra considerado neste estudo de caso.

### 9.2.2. Resultados pelo Método *Jackknife*

**Tabela 55**

Resultados Comparativos do Método de Hair e Anderson (Método H. A) e o TIAF.

VAR	Cargas	Método H.A	TIAFIC	Valor-p	TIAFVP
<b>Fator 1</b>					
V1	0,85	SIG	SIG	0,000	SIG
V2	0,04	NSIG	NSIG	0,240	NSIG
V3	0,72	NSIG	SIG	0,000	SIG
V4	0,13	NSIG	NSIG	0,040	SIG
V5	0,80	SIG	SIG	0,000	SIG
V6	-0,08	NSIG	NSIG	0,080	NSIG
V7	-0,36	NSIG	SIG	0,000	SIG
<b>Fator 2</b>					
V1	-0,04	NSIG	NSIG	0,400	NSIG
V2	-0,31	NSIG	NSIG	0,000	SIG
V3	-0,42	NSIG	NSIG	0,000	SIG
V4	0,66	NSIG	NSIG	0,000	SIG
V5	0,25	NSIG	NSIG	0,000	SIG
V6	-0,11	NSIG	NSIG	0,200	NSIG
V7	0,78	SIG	NSIG	0,000	SIG

Os resultados entre os métodos tradicionais Hair e Anderson(2005) e o TIAF também apresentam divergências na técnica *jackknife*. O método tradicional de Hair e Anderson divergiu do TIAFIC na variável V<sub>3</sub>, no fator

1, onde a carga 0,72 foi considerada não significativa naquele método e significativa neste método. Com o TIAFVP encontrou-se o mesmo resultado. Na variável  $V_4$ , ainda no fator 1, a carga fatorial 0,13 foi considerada não significativa no método tradicional e significativa nos TIAF'S pelo intervalo de confiança e pelo valor-p. Na variável  $V_7$ , ainda no fator 1, a carga fatorial -0,36 foram consideradas não significantes no método tradicional e significativa nos TIAF'S pelo intervalo de confiança e pelo valor-p. Os resultados pelo TIAF estão mais coerentes, pois sofreram influência não somente do tamanho da amostra, mas também do número de variáveis do problema e do fator específico a quem pertencem.

No fator 2, como era de se esperar, pois se trata de um fator com um grau de variação explicada menor do que o fator 1, houve uma divergência maior entre os resultados. Na variável  $V_2$ , a carga fatorial -0,31 foi considerada não significativa no método tradicional e não significativa no TIAFIC, mas significativa pelo valor-p. O resultado apresentado pelo valor-p parece ser o mais aceitável. Na variável  $V_3$ , a carga fatorial -0,42 foi considerada não significativa no método tradicional e não significativa no TIAFIC, mas também significativa pelo valor-p. O resultado apresentado pelo valor-p parece novamente ser o mais racional. Na variável  $V_4$ , a carga fatorial 0,66 foi considerada não significativa no método tradicional e não significativa no TIAFIC, mas significativa pelo valor-p. O resultado apresentado pelo valor-p parece novamente ser o mais racional. Na variável  $V_5$ , a carga fatorial 0,25 foi considerada não significativa no método tradicional e não significativa no TIAFIC, mas significativa pelo valor-p. O resultado apresentado pelo valor-p parece novamente ser o mais racional. Na variável  $V_7$ , a carga fatorial 0,78 foi considerada significativa no método tradicional



e não significativa no TIAFIC, e significativa pelo valor-p. Os resultados apresentados pelos métodos tradicionais e TIAFVP são igualmente aceitáveis.

As análises realizadas neste capítulo serviram para corroborar a tese de que o procedimento inferencial proposto neste trabalho tem eficácia bem satisfatória, podendo-se até afirmar que supera principalmente no uso do valor-p, a eficácia do método tradicional.

## CONCLUSÃO

O objetivo desta tese foi o de sugerir um procedimento inferencial para análise fatorial, mas precisamente estabelecer um critério para a significância para as cargas fatoriais, utilizando-se das técnicas de reamostragem *bootstrap* e *jackknife*, isto é, um critério para a significância de cargas fatoriais.

A análise fatorial é um tipo de processo destinado essencialmente a redução e à sumarização dos dados.

O *bootstrap* e o *jackknife* constituem técnicas **CIS** (Computer Intensive Statistics), que especificam o modelo da distribuição por amostragem empírica das estatísticas consideradas no estudo.

Estas metodologias de reamostragem não paramétricas são particularmente úteis quando é inviável pela metodologia tradicional se estabelecer a distribuição teórica de amostragens das estimativas investigadas, bem como seus parâmetros característicos. Esta dificuldade reside no fato de em muitas situações o comportamento regular dos parâmetros descritivos da distribuição por amostragem da variável aleatória em foco não se evidenciar tal facilmente ao pesquisador, variando de caso para caso ou quando a complexidade da natureza da estimativa analisada é um complicador para o estudo de sua regularidade via métodos tradicionais, como análise assintótica e por máxima verossimilhança.

As técnicas **CIS** funcionam como processos alternativos de estudos do comportamento probabilístico de variáveis aleatórias de interesse apelando para especificação das distribuições por amostragem através da força bruta da computação, simulando o universo de todas as estimativas possíveis para o estimador e calculando seu erro-padrão, seu viés e seu

EMQ com base na experiência. Com estas estatísticas, é possível se obter intervalos de confiança e testes de hipótese saindo, então, do terreno puramente descritivo no que se refere à análise fatorial.

Para realizar o *link* da análise fatorial com as técnicas *bootstrap* e *jackknife*, o pesquisador elaborou dois algoritmos e destes algoritmos estabeleceu um programa em linguagem *R 2.1.1* e dois programas em linguagem de programação *SAS Versão 8*.

Os resultados *bootstrap* e *jackknife* se mostraram a princípio muito eficazes, pois revelaram erro-padrão e viés muito baixo para os estudos de casos considerados.

Os principais achados da pesquisa ou o balanço geral dos resultados fornecidos pelo **Método TIAF** com as bases testadas são:

- *A regularidade no comportamento geral das distribuições por amostragem das variáveis aleatórias cargas fatoriais de componentes principais reamostradas nos estudos de casos não se evidenciou: para cada exemplo se mostrou um resultado próprio;*
- *As teorias existentes sobre o comportamento do erro-padrão e do viés entre o bootstrap e o jackknife não se sustentaram em todos os estudos das estatísticas cargas fatoriais das componentes principais 1 e 2;*
- *Em todas as bases testadas, a precisão, o viés e o **EMQ** se revelaram satisfatórios, em torno de zero;*
- *As cargas fatoriais têm probabilidade de serem estatisticamente não significantes, mesmo se tratando de grandes amostras, nas estatísticas cargas fatoriais de componentes principais 1 e 2;*

- *A vantagem do critério de significância de cargas fatoriais exposto nesta tese é que o número de variáveis analisadas e o fator específico em exame são considerados no processo de significância das cargas fatoriais. As reamostragens são realizadas levando em consideração todas as variáveis do problema (quanto mais completa for a lista de variáveis melhor o teste de significância da análise fatorial) e para um fator específico  $F_j$  gerado em cada amostra bootstrap ou jackknife conforme seu poder de explicação dentro do modelo fatorial.*
- *Não houve uma regularidade no ajuste ou na aderência do modelo normal às estimativas de cargas fatoriais analisadas. O método jackknife foi onde à aderência ao modelo de probabilidade normal mais se evidenciou;*
- *O método jackknife foi o que teve, também, um saldo positivo maior com relação ao desempenho em relação ao método bootstrap para os casos analisados neste estudo;*
- *A análise da convergência no uso da técnica bootstrap indicou que com um  $B=200$  já se tem a **Regularidade Estatística dos Resultados**. Contudo, para otimizar ao máximo a qualidade das distribuições amostrais resultantes ainda sugere-se um  $B=1000$ .*
- *Os resultados comparativos dos Métodos Tradicional Hair e Anderson e o TIAF serviram para corroborar a tese de que o procedimento indutivo aqui proposto **tem eficácia bem satisfatória**, podendo-se até afirmar que supera, principalmente no uso do valor-p, a eficácia do método tradicional.*

Seguindo a mesma metodologia deste trabalho-aplicação de técnicas de reamostragem: bootstrap e jackknife-, pretende-se como trabalho futuro desenvolver um procedimento inferencial para a análise de conglomerados. Segundo Aaker (2001), é difícil avaliar a qualidade do processo de agrupamento. Não existem testes estatísticos padrão para garantir que o resultado não seja puramente aleatório. Apesar das tentativas de construção de vários testes para a confiabilidade estatística dos conglomerados, nenhum procedimento totalmente comprovado está disponível. A ausência de testes adequados provém da dificuldade de especificação de hipóteses nulas realísticas. O valor do critério de medida, a legitimidade do resultado, a aparência de uma hierarquia natural -quando for empregado um método não hierárquico- e a confiabilidade de testes de divisão de amostra, tudo isto oferece informação útil. Entretanto, é difícil saber exatamente quais conglomerados são muitos parecidos e quais objetos são difíceis de serem inseridos. Geralmente, é difícil selecionar um critério e programa de agrupamento por meio de outra referência que não a disponibilizada.

Fica sugerido o desenvolvimento de estudos para identificação de procedimentos inferências para as demais técnicas de análise multivariadas.

## **BIBLIOGRAFIA E REFERÊNCIAS BIBLIOGRÁFICAS**

- Affi**, A. A. e **Clark**, V. (1984).Computer – Aided Multivariate Analysis. Lifetime Learning Publications. Belm. California.
- Anderson**, T.W. (1984).An Introduction to Multivariate Statistical Analysis. 2ed.New York : John Wiley & Sons.
- Barroso**, L . P. , (2003).Tópicos de Análise Multivariada . Universidade de São Paulo.
- Burt**, C .(1941).The Vector of the Mind : An Introduction to Factor Analysis in Psychology. New York : Mc.Millan.
- Cazar**, R. A.(2003).An Exercise on Chemometrics for a Quantitative Analysis Course.Madison:Journal of Chemical Education.
- Chatfield**, C. e **Collins**, A. J.(1980).Introduction to Multivariate Analysis. Chapman and Hall. New York.
- Cliff**, N., e **Hamburge**, C. D. (1967).The Study of Sampling Errors in Factor Analysis by Means of Artificial Experiments. Psychological Bulletin 68: 430-45.
- Cooley**, W. W. e **Lohnes**, P. R.(1971).Multivariate Data Analysis. John Wiley. New York.
- Costa**, Giovani Glaucio de O.(2003).Busca de Fatores Associados à Prática de Atos Infracionais por Parte de Adolescentes no Estado do Rio de Janeiro: Um Estudo Preliminar, Estudo Orientado, PUC-RIO.
- David**, A.Aaker , **Kumar**, V;**George**,S. Day.(1984).Marketing Research.
- Dillon**, W. R. e **Goldstein**, M.(1984).Multivariate Analysis : Methods and Applications . New York : John Wiley & Sons.
- Efron**,B(1979).Bootstrap Methods:Another Look at the Jackknife, The Annals of Statistic,7,1-26.
- Efron**,B.(1980).Computer Intensive Methods in Statistics” in Some Recent Advance in Statistic, Ed. J. Tiago de Oliveira e B.Epstein , Academia das Ciências de Lisboa, Lisboa.
- Efron**,B.(1982).The Jackknife, the Bootstrap , and other Resampling Methods, CBNS 38,SIAM-NSF
- Everitt**, B. S.(1978).Graphical Techniques for Multivariate Data. Heinemann Educational Books. London.
- Ferreira**, D. F. Análise Multivariada . Minas Gerais : Universidade Federal de Lavras.
- Gnanadesikan**, R.(1997).Methods for Statitical Data Analysis of Multivariate Observations. John Wiley. New York.

- Hair**, J. F. Jr. ; **Anderson**, R.E. ; **Tathan**, R. L. e **Black**, W. C.(2005). Trad. **Sant'Anna**, Adonai Schlup ; **Neto**, Anselmo Chaves. Análise Multivariada de Dados. 5. ed. Porto Alegre : Bookman.
- Hair**, J. F. Jr. ; **Anderson**, R.E. ; **Tathan**, R. L. e **Black**, W. C.(1998). Multivariate Data Analysis. 5th ed. Upper Saddle River : Prentice Hall.
- Harman**, Harry H.(1967).Modern Factor Analysis . 2 ed. Chicago : University of Chicago .
- Hawkins**, D.M., Topics in Multivariate Analysis.Cambridge University Press : Cambridge.
- Hompson**, B.(1984).Canonical Correlation Analysis : Uses and Interpretation . Beverly Hills : Sage Pub.
- Jobson**,J.(1992) D.Applied Multivariate Data Analysis. Vol II: Categorical and Multivariate Methods. Springer Verlag. New York.
- Johnson** , D. E .(1998).Applied Multivariate Methods for Data Analysis. Pacific Grove : Duxbury Press..
- Johnson** , R. A. e **Wichern** , D.W .(1998).Applied Multivariate Statistical Analysis. 4ed . Upper Saddle River: Prentice Hall.
- Johnson**, R. A. e **Wichern**, D. W.(1982).Applied Multivariate Statistical Analysis. Prentice Hall, Inc. Englewood Cliffs, New Jersey.
- Kaiser** , H . F .(1958).The Varimax Criterion for Analytic Rotation in Factor Analysis . Psychometrika , 23 , 187-200.
- Kaiser** , H . F .(1974).A Second-Generation Little Jiffy. Psychometrika 35:401-15
- Kaufman** , L. e **Rousseeuw**, P. J.(1990).Finding Groups in Data : An Introduction to Cluster Analysis. New York . Jonh Wile & Sons.
- Lebart**, Ludovic ; **Morineau**,Alain e **Piron**, Marie(1998).Statistique Exploratoire Multidimensionnelle.França.
- Long**, J.S.(1983).Confirmation Factor Analysis . Beverly Hills : Sage Pub.
- Mardia** , K. V. ; **Kent** , J. T. e **Bibby** , J . M.(1989).Multivariate Analysis. London : Academic Press.
- Moita Neto**, J.M. e **Moita**,Graziela Ciaramella.(1998).Uma Introdução à Análise de Dados Multivariados.São Paulo : Química Nova.
- Morrison**, D. (3rd ed.).(1982). Multivariate Statistical Methods. McGraw-Hill. New York.
- Prado**,P.I.; **Lewwinsohn**,Thomas Michael; **Carmo**, R. L; **Hogan**,D.J.(2002).Ordenação Multivariada na Ecologia e seu Uso em Ciências Ambientais.São Paulo:Ambiente e Sociedade.

- Proença**, Isabel Maria D.(1988).Estimativas Jackknife e Bootstrap para o Enviesamento e Desvio-padrão do índice de Gini. Doc.Trabalho N° 67, Cemapre, I.S.E, Lisboa.
- Reis**, E.(1997).Estatística Multivariada Aplicada. Lisboa : Edições Sílabo.
- Reyment**, R. e **Joreskog**, K . G.(1996).Applied Factor Analysis in the Natural Science . Cambridge. Cambridge University Press.
- Rosa**, P . T . M.(2000).Modelos de “Credit Scoring”: Regressão Logística, CHAID e REAL . Dissertação de Mestrado . IME –Universidade de São Paulo.
- Rummel**, R.J.(1970) Applied Factor Analysis . Evanston : Northwestern University Press.  
SAS Institute Inc.
- SAS Procedures Guide**, Version 8.Cary , N.C.(1999). SAS Institute Inc.
- Saporta**, G. (1996). Probabilités Analyse Des Données Et Statistique. França.
- Seber**, G. A. F. (1984).Multivariate Observations. John Wiley. New York.
- Siegel,Sidney**(1975). Estatística Não-paramétrica Para Ciências do Comportamento. São Paulo:  
McGraw-Hill Editora
- Sharma** , S.(1996).Applied Multivariate Techniques . New York : Jonh Wile & Sons..
- Spearman**, C.(1940).General Intelligence Objectively Determined and Measured . American Journal of Psychology , 15 : 201-293.
- Stewart**, David W.(1981).The applications and misapplications of factor analysis in marketing research. Journal of Marketing Research, 18, p.51-62,Feb.
- Tabachnick**, B . G . ; Fidel , L . S .(2001).Using Multivariate Statistics . 4th ed. Allyn 7 Bacon , Boston.
- Thurstone** , L . L.(1993).Multiple Factor Analysis . Chicago : Univ. Chicago Press .
- Thurstone**, L . L.(1993).The Vector of Mind . Chicago : Univ. Chicago Press.



## ANEXO 1-OS PROGRAMAS *BOOTSTRAP* E *JACKKNIFE* EM R 2.1.1

O programa abaixo referê-se ao usado com a base de dados “Modo de Vida”.

```
#=====
#=====
# PROGRAMA FEITO EM 29/SETEMBRO/2005 - ÚLTIMA MODIFICAÇÃO
08/OUTUBRO/2005
#=====
#=====
# Usando a ajuda
help("factanal")
# Lendo os dados
# Informe o número de variáveis
m      <- 11
# Número de variáveis
x      <- matrix(scan("c:/Tese
Gcosta/basepaises.txt"),,m,T)
# Análise Fatorial - guardando os dois fatores
cargas <- factanal(x, factor=2)$loadings[,1:2]
#=====
#=====
# INÍCIO DO BOOTSTRAP
#=====
#=====
n      <- length(x[,1])                                #
Tamanho da amostra original (40)
N      <- length(x[,1])                                #
Tamanho da re-amostra          (40)
M      <- 1000                                         #
Número de replicações          (1000)
#=====
#=====
# Função auxiliar para gerar números de 1 até o tamanho
da amostra
rp <- function(N,n){
y <- rep(NA,N)
for(i in 1:N){
  aux <- runif(1)
  lb <- 0
```

```

ub <- 1/n
for(j in 1:n){
  if(lb<aux && aux <= ub ) y[i] <- j
  lb <- lb + 1/n
  ub <- ub + 1/n
}
}
return(y)
}
#=====
=====
F1 <- matrix(NA,length(x[1,]),M)
F2 <- matrix(NA,length(x[1,]),M)
# Fazendo ...
for(i in 1:M){
  w <- rp(N,n)
# Gerando valores da reamostra
  z <- x[w,]
# Alocando a reamostra
  y <- factanal(z,factor=2)$loading[,1:2]
# Análise Fatorial
  F1[,i] <- y[,1]
  F2[,i] <- y[,2]
}
F1.MEAN <- mean(F1[1,])
F2.MEAN <- mean(F2[1,])
F1.VAR <- var(F1[1,])
F2.VAR <- var(F2[1,])
alpha <- 0.05
q1 <- round(M*alpha/2)
q2 <- M - q1 + 1
F1.LI <- sort(F1[1,])[q1]
F1.LS <- sort(F1[1,])[q2]
F2.LI <- sort(F2[1,])[q1]
F2.LS <- sort(F2[1,])[q2]
for(i in 2:m){
F1.MEAN <- c(F1.MEAN,mean(F1[i,]))
F2.MEAN <- c(F2.MEAN,mean(F2[i,]))
F1.VAR <- c(F1.VAR,var(F1[i,]))
F2.VAR <- c(F2.VAR,var(F2[i,]))
}

```

```

F1.LI    <- c(F1.LI,sort(F1[i,])[q1])
F1.LS    <- c(F1.LS,sort(F1[i,])[q2])
F2.LI    <- c(F2.LI,sort(F2[i,])[q1])
F2.LS    <- c(F2.LS,sort(F2[i,])[q2])
}
F1.VIES  <- cargas[,1] - F1.MEAN
F2.VIES  <- cargas[,2] - F2.MEAN
F1.EMQ   <- F1.VAR - F1.VIES^2
F2.EMQ   <- F2.VAR - F2.VIES^2
F1.SISIC<- rep(NA,m)
F2.SISIC<- rep(NA,m)
for(i in 1:m){
  if(F1.LI[i]<= 0 && 0 <= F1.LS[i]){
    F1.SISIC[i] <- "NSIG"
  }
  else{
    F1.SISIC[i] <- "SIG"
  }
  if(F2.LI[i]<= 0 && 0 <= F2.LS[i]){
    F2.SISIC[i] <- "NSIG"
  }
  else{
    F2.SISIC[i] <- "SIG"
  }
}
}
# Salvando os dados num arquivo - distribuição amostral
options(digits=5)
w <- cbind(t(F1),t(F2))
write(t(w),file="c:/Tese Gcosta/dist-boot-
summary.txt",ncolumn=(2*m))
# Padronização ...
for(i in 1:m){
  F1[i,] <- (F1[i,] - F1.MEAN[i])/sqrt(F1.VAR[i]) #
Padronização
  F2[i,] <- (F2[i,] - F2.MEAN[i])/sqrt(F2.VAR[i]) #
Padronização
}
# Salvando os dados num arquivo - distribuição amostral
padronizada
options(digits=5)

```

```

w <- cbind(t(F1),t(F2))
write(t(w),file="c:/Tese Gcosta/dist-boot-padro-
summary.txt",ncolumn=(2*m))
# Calculando G
g <- cargas/sqrt(cbind(F1.VAR,F2.VAR))
p.valor <- matrix(NA,m,2)
for(j in 1:m){
  cont1 <- 0
  cont2 <- 0
  for(i in 1:M){
    # Fator 1
    if(F1[j,i] >= g[j,1] && g[j,1] >= 0 ) cont1 <- cont1 +
1 # Se g >= 0 ele só fará este
    if(F1[j,i] <= g[j,1] && g[j,1] < 0 ) cont1 <- cont1 +
1 # Se g < 0 ele só fará este
    # Fator 2
    if(F2[j,i] >= g[j,2] && g[j,2] >= 0 ) cont2 <- cont2 +
1 # Se g >= 0 ele só fará este
    if(F2[j,i] <= g[j,2] && g[j,2] < 0 ) cont2 <- cont2 +
1 # Se g < 0 ele só fará este
  }
  p.valor[j,1] <- cont1/M
  p.valor[j,2] <- cont2/M
}
F1.SIGVP <- rep(NA,m)
F2.SIGVP <- rep(NA,m)
for(i in 1:m){
  if( p.valor[i,1] > 0.05 ){
    F1.SIGVP[i] <- "NSIG"
  }else{
    F1.SIGVP[i] <- "SIG"
  }
  if( p.valor[i,2] > 0.05 ){
    F2.SIGVP[i] <- "NSIG"
  }else{
    F2.SIGVP[i] <- "SIG"
  }
}
p.valor
F1.VP <- p.valor[,1]

```

```

F2.VP <- p.valor[,2]
F1.C <- cargas[,1]
F2.C <- cargas[,2]
F1.summary <-
data.frame(cbind(F1.C,F1.MEAN,F1.VAR,F1.LI,F1.LS,F1.VIES,
F1.EMQ,F1.SISIC,F1.VP,F1.SIGVP))
F2.summary <-
data.frame(cbind(F2.C,F2.MEAN,F2.VAR,F2.LI,F2.LS,F2.VIES,
F2.EMQ,F2.SISIC,F2.VP,F2.SIGVP))
options(digits=5)
F1.summary          # RESUMO DO FATOR 1
F2.summary          # RESUMO DO FATOR 2
# Salvando os dados num arquivo - resumo estatístico
options(digits=5)
F.summary <-
rbind(cbind(F1.C,F1.MEAN,F1.VAR,F1.LI,F1.LS,F1.VIES,F1.EM
Q,F1.SISIC,F1.VP,F1.SIGVP),

cbind(F2.C,F2.MEAN,F2.VAR,F2.LI,F2.LS,F2.VIES,F2.EMQ,F2.S
ISIC,F2.VP,F2.SIGVP))
write(t(F.summary),file="c:/Tese Gcosta/boot-
summary.txt",ncolumn=10)
#=====
# FIM DO BOOTSTRAP
#=====
#=====
# Para aqui
#=====
#=====
# INÍCIO DO JACKKNIFE
#=====
n <- length(x[,1]) # Tamanho da amostra original (40)
N <- length(x[,1]) # Tamanho da re-amostra (40)
M <- length(x[,1]) # Número de replicações (40)

```

```

#=====
=====
F1 <- matrix(NA,length(x[1,]),M)
F2 <- matrix(NA,length(x[1,]),M)
# Fazendo ...
for(i in 1:M){
  if(i==1) w <- seq(2,N,1)
# Gerando valores da reamostra
  if(i==N) w <- seq(1,N-1,1)
# Gerando valores da reamostra
  if(i!=1 && i!=N) w <- c(seq(1,(i-1),1),seq((i+1),N,1))
# Gerando valores da reamostra
  z <- x[w,]
# Alocando a reamostra
  y <- factanal(z,factor=2)$loading[,1:2]
# Análise Fatorial
  F1[,i] <- y[,1]
  F2[,i] <- y[,2]
}
F1.MEAN <- mean(F1[1,])
F2.MEAN <- mean(F2[1,])
F1.VAR <- var(F1[1,])
F2.VAR <- var(F2[1,])
alpha <- 0.05
q1 <- round(M*alpha/2)
q2 <- M - q1 + 1
F1.LI <- sort(F1[1,])[q1]
F1.LS <- sort(F1[1,])[q2]
F2.LI <- sort(F2[1,])[q1]
F2.LS <- sort(F2[1,])[q2]
for(i in 2:m){
F1.MEAN <- c(F1.MEAN,mean(F1[i,]))
F2.MEAN <- c(F2.MEAN,mean(F2[i,]))
F1.VAR <- c(F1.VAR,var(F1[i,]))
F2.VAR <- c(F2.VAR,var(F2[i,]))
F1.LI <- c(F1.LI,sort(F1[i,])[q1])
F1.LS <- c(F1.LS,sort(F1[i,])[q2])
F2.LI <- c(F2.LI,sort(F2[i,])[q1])
F2.LS <- c(F2.LS,sort(F2[i,])[q2])
}

```

```

F1.VIES <- cargas[,1] - F1.MEAN
F2.VIES <- cargas[,2] - F2.MEAN
F1.EMQ <- F1.VAR - F1.VIES^2
F2.EMQ <- F2.VAR - F2.VIES^2
F1.SISIC<- rep(NA,m)
F2.SISIC<- rep(NA,m)
for(i in 1:m){
  if(F1.LI[i]<= 0 && 0 <= F1.LS[i]){
    F1.SISIC[i] <- "NSIG"
  }
  else{
    F1.SISIC[i] <- "SIG"
  }
  if(F2.LI[i]<= 0 && 0 <= F2.LS[i]){
    F2.SISIC[i] <- "NSIG"
  }
  else{
    F2.SISIC[i] <- "SIG"
  }
}
# Salvando os dados num arquivo - distribuição amostral
options(digits=5)
w <- cbind(t(F1),t(F2))
write(t(w),file="c:/Tese Gcosta/dist-jack-
summary.txt",ncolumn=(2*m))
# Padronização ...
for(i in 1:m){
  F1[i,] <- (F1[i,] - F1.MEAN[i])/sqrt(F1.VAR[i]) #
Padronização
  F2[i,] <- (F2[i,] - F2.MEAN[i])/sqrt(F2.VAR[i]) #
Padronização
}
# Salvando os dados num arquivo - distribuição amostral
padronizada
options(digits=5)
w <- cbind(t(F1),t(F2))
write(t(w),file="c:/Tese Gcosta/dist-jack-padro-
summary.txt",ncolumn=(2*m))
# Calculando G
g <- cargas/sqrt(cbind(F1.VAR,F2.VAR))

```

```

p.valor <- matrix(NA,m,2)
for(j in 1:m){
  cont1 <- 0
  cont2 <- 0
  for(i in 1:M){
    # Fator 1
    if(F1[j,i] >= g[j,1] && g[j,1]>= 0 ) cont1 <- cont1 +
1 # Se g >= 0 ele só fará este
    if(F1[j,i] <= g[j,1] && g[j,1] < 0 ) cont1 <- cont1 +
1 # Se g < 0 ele só fará este
    # Fator 2
    if(F2[j,i] >= g[j,2] && g[j,2]>= 0 ) cont2 <- cont2 +
1 # Se g >= 0 ele só fará este
    if(F2[j,i] <= g[j,2] && g[j,2] < 0 ) cont2 <- cont2 +
1 # Se g < 0 ele só fará este
  }
  p.valor[j,1] <- cont1/M
  p.valor[j,2] <- cont2/M
}
F1.SIGVP <- rep(NA,m)
F2.SIGVP <- rep(NA,m)
for(i in 1:m){
  if( p.valor[i,1] > 0.05 ){
    F1.SIGVP[i] <- "NSIG"
  }else{
    F1.SIGVP[i] <- "SIG"
  }
  if( p.valor[i,2] > 0.05 ){
    F2.SIGVP[i] <- "NSIG"
  }else{
    F2.SIGVP[i] <- "SIG"
  }
}
}
p.valor
F1.VP <- p.valor[,1]
F2.VP <- p.valor[,2]
F1.C <- cargas[,1]
F2.C <- cargas[,2]

```



```

F1.summary <-
data.frame(cbind(F1.C,F1.MEAN,F1.VAR,F1.LI,F1.LS,F1.VIES,
F1.EMQ,F1.SISIC,F1.VP,F1.SIGVP))
F2.summary <-
data.frame(cbind(F2.C,F2.MEAN,F2.VAR,F2.LI,F2.LS,F2.VIES,
F2.EMQ,F2.SISIC,F2.VP,F2.SIGVP))
options(digits=5)
F1.summary          # RESUMO DO FATOR 1
F2.summary          # RESUMO DO FATOR 2
# Salvando os dados num arquivo - resumo estatístico
options(digits=5)
F.summary <-
rbind(cbind(F1.C,F1.MEAN,F1.VAR,F1.LI,F1.LS,F1.VIES,F1.EM
Q,F1.SISIC,F1.VP,F1.SIGVP),

cbind(F2.C,F2.MEAN,F2.VAR,F2.LI,F2.LS,F2.VIES,F2.EMQ,F2.S
ISIC,F2.VP,F2.SIGVP))
write(t(F.summary),file="c:/Tese Gcosta/jack-
summary.txt",ncolumn=10)
#=====
=====
# FIM DO JACKNIFE
#=====
=====

```

## ANEXO 2-O PROGRAMA BOOTSTRAP EM SAS V.8

O programa abaixo referê-se ao usado com a base de dados “Creme Dental”.

```
*libname in 'C:\Tese Giovani';
libname in2 'C:\Tese Giovani\Terceira Fase';
options nodate ls=max ps=max;

PROC IMPORT OUT= IN2.basecremedental
            DATAFILE= "C:\Tese Giovani\Terceira
Fase\basecremedental.xls"
            DBMS=EXCEL2000 REPLACE;
            GETNAMES=YES;
RUN;

data base(drop=F7);
  set in2.basecremedental;
run;

%macro sel(i,j);
data base&i;
  set base;
  if _n_=&i then delete;
run;

proc factor data=base&i
            method=principal
            rotate=Varimax
            outstat=perfil&i
            nfactors=2 noprint;
  var V1-V6;
run;

data perfil_f1_&i(keep=i f1_v01-f1_v06)
  perfil_f2_&i(keep=i f2_v01-f2_v06);
  set perfil&i;
  if _type_='UNROTATE' and _name_ in ('Factor1'
'Factor2');
```

```

if _name_='Factor1' then do;
    f1_v01= v1 ;    f1_v02= v2 ;    f1_v03= v3 ;    f1_v04=
v4 ;    f1_v05= v5 ;    f1_v06= v6 ;
end;
else if _name_='Factor2' then do;
    f2_v01= v1 ;    f2_v02= v2 ;    f2_v03= v3 ;    f2_v04=
v4 ;    f2_v05= v5 ;    f2_v06= v6 ;
end;
    i=1;

    if _name_='Factor1' then output perfil_f1_&i;
else if _name_='Factor2' then output perfil_f2_&i;
run;

data amostra&i(drop=i);
merge perfil_f1_&i perfil_f2_&i;
by i;
amostra=&i;
run;

proc delete data=perfil&i;
proc delete data=perfil_f1_&i;
proc delete data=perfil_f2_&i;
proc delete data=base&i;
run;
%mend sel;
%sel ( 1 , 87575 ); %sel ( 2 , 42631 );
%sel ( 3 , 98398 ); %sel ( 4 , 82401 );
%sel ( 5 , 75217 ); %sel ( 6 , 23436 );
%sel ( 7 , 78850 ); %sel ( 8 , 99497 );
%sel ( 9 , 75267 ); %sel ( 10 , 12243 );
%sel ( 11 , 55216 ); %sel ( 12 , 90000 );
%sel ( 13 , 10855 ); %sel ( 14 , 20001 );
%sel ( 15 , 44804 ); %sel ( 16 , 31726 );
%sel ( 17 , 93948 ); %sel ( 18 , 74881 );
%sel ( 19 , 20262 ); %sel ( 20 , 13439 );
%sel ( 21 , 3817 ); %sel ( 22 , 6551 );
%sel ( 23 , 54403 ); %sel ( 24 , 11361 );
%sel ( 25 , 97085 ); %sel ( 26 , 83919 );
%sel ( 27 , 85594 ); %sel ( 28 , 35212 );

```

```

%sel ( 29 , 32651 ); %sel ( 30 , 84167 );
%sel ( 31 , 23953 ); %sel ( 32 , 56974 );
%sel ( 33 , 59788 ); %sel ( 34 , 91566 );
%sel ( 35 , 84992 ); %sel ( 36 , 20942 );
%sel ( 37 , 84190 ); %sel ( 38 , 50512 );
%sel ( 39 , 98086 ); %sel ( 40 , 23271 );
%sel ( 41 , 59386 ); %sel ( 42 , 49980 );
%sel ( 43 , 71145 ); %sel ( 44 , 29931 );
%sel ( 45 , 64912 ); %sel ( 46 , 12436 );
%sel ( 47 , 49793 ); %sel ( 48 , 33565 );
%sel ( 49 , 57161 ); %sel ( 50 , 94119 );
%sel ( 51 , 63066 ); %sel ( 52 , 88110 );
%sel ( 53 , 44405 ); %sel ( 54 , 13065 );
%sel ( 55 , 25214 ); %sel ( 56 , 47196 );
%sel ( 57 , 29784 ); %sel ( 58 , 22568 );
%sel ( 59 , 60846 ); %sel ( 60 , 83724 );
%sel ( 61 , 72102 ); %sel ( 62 , 27420 );
%sel ( 63 , 5148 ); %sel ( 64 , 77389 );
%sel ( 65 , 48909 ); %sel ( 66 , 52879 );
%sel ( 67 , 39849 ); %sel ( 68 , 72711 );
%sel ( 69 , 14754 ); %sel ( 70 , 16441 );
%sel ( 71 , 81171 ); %sel ( 72 , 91869 );
%sel ( 73 , 64291 ); %sel ( 74 , 16099 );
%sel ( 75 , 28189 ); %sel ( 76 , 58487 );
%sel ( 77 , 16035 ); %sel ( 78 , 93280 );
%sel ( 79 , 23064 ); %sel ( 80 , 28573 );
%sel ( 81 , 32756 ); %sel ( 82 , 12601 );
%sel ( 83 , 71664 ); %sel ( 84 , 31319 );
%sel ( 85 , 3285 ); %sel ( 86 , 73965 );
%sel ( 87 , 30113 ); %sel ( 88 , 56699 );
%sel ( 89 , 59394 ); %sel ( 90 , 24559 );
%sel ( 91 , 8943 ); %sel ( 92 , 49173 );
%sel ( 93 , 43540 ); %sel ( 94 , 22720 );
%sel ( 95 , 49001 ); %sel ( 96 , 51072 );
%sel ( 97 , 91797 ); %sel ( 98 , 54853 );
%sel ( 99 , 31344 ); %sel ( 100 , 44768 );
%sel ( 101 , 80496 ); %sel ( 102 , 71547 );
%sel ( 103 , 83316 ); %sel ( 104 , 64796 );
%sel ( 105 , 79490 ); %sel ( 106 , 88229 );
%sel ( 107 , 15576 ); %sel ( 108 , 83319 );

```

```

%sel ( 109 , 90206 ); %sel ( 110 , 25221 );
%sel ( 111 , 11345 ); %sel ( 112 , 35133 );
%sel ( 113 , 57730 ); %sel ( 114 , 86329 );
%sel ( 115 , 57344 ); %sel ( 116 , 87914 );
%sel ( 117 , 21095 ); %sel ( 118 , 91279 );
%sel ( 119 , 69902 ); %sel ( 120 , 21540 );
%sel ( 121 , 63059 ); %sel ( 122 , 22079 );
%sel ( 123 , 51355 ); %sel ( 124 , 75010 );
%sel ( 125 , 85215 ); %sel ( 126 , 10590 );
%sel ( 127 , 18821 ); %sel ( 128 , 60801 );
%sel ( 129 , 45243 ); %sel ( 130 , 50163 );
%sel ( 131 , 71707 ); %sel ( 132 , 46959 );
%sel ( 133 , 99526 ); %sel ( 134 , 60669 );
%sel ( 135 , 43433 ); %sel ( 136 , 69226 );
%sel ( 137 , 83372 ); %sel ( 138 , 15105 );
%sel ( 139 , 66160 ); %sel ( 140 , 68847 );
%sel ( 141 , 61005 ); %sel ( 142 , 45044 );
%sel ( 143 , 92248 ); %sel ( 144 , 86235 );
%sel ( 145 , 28617 ); %sel ( 146 , 68294 );
%sel ( 147 , 35647 ); %sel ( 148 , 2813 );
%sel ( 149 , 39138 ); %sel ( 150 , 93455 );
%sel ( 151 , 40786 ); %sel ( 152 , 73002 );
%sel ( 153 , 71325 ); %sel ( 154 , 97954 );
%sel ( 155 , 20338 ); %sel ( 156 , 57102 );
%sel ( 157 , 10793 ); %sel ( 158 , 12959 );
%sel ( 159 , 86666 ); %sel ( 160 , 61205 );
%sel ( 161 , 13028 ); %sel ( 162 , 63430 );
%sel ( 163 , 59274 ); %sel ( 164 , 44090 );
%sel ( 165 , 28600 ); %sel ( 166 , 95663 );
%sel ( 167 , 22556 ); %sel ( 168 , 40501 );
%sel ( 169 , 72080 ); %sel ( 170 , 11798 );
%sel ( 171 , 83621 ); %sel ( 172 , 62901 );
%sel ( 173 , 70967 ); %sel ( 174 , 80505 );
%sel ( 175 , 56130 ); %sel ( 176 , 66572 );
%sel ( 177 , 20738 ); %sel ( 178 , 83718 );
%sel ( 179 , 14546 ); %sel ( 180 , 75444 );
%sel ( 181 , 48545 ); %sel ( 182 , 74441 );
%sel ( 183 , 2142 ); %sel ( 184 , 51781 );
%sel ( 185 , 54372 ); %sel ( 186 , 6390 );
%sel ( 187 , 73222 ); %sel ( 188 , 28451 );

```

```

%sel ( 189 , 31665 ); %sel ( 190 , 93921 );
%sel ( 191 , 13335 ); %sel ( 192 , 75270 );
%sel ( 193 , 39017 ); %sel ( 194 , 97305 );
%sel ( 195 , 46917 ); %sel ( 196 , 86436 );
%sel ( 197 , 4297); %sel ( 198 , 13145 );
%sel ( 199 , 11559 ); %sel ( 200 , 40844 );

```

```

data in2.boot;

```

```

set
amostra1 amostra101
amostra2 amostra102
amostra3 amostra103
amostra4 amostra104
amostra5 amostra105
amostra6 amostra106
amostra7 amostra107
amostra8 amostra108
amostra9 amostra109
amostra10 amostra110
amostra11 amostra111
amostra12 amostra112
amostra13 amostra113
amostra14 amostra114
amostra15 amostra115
amostra16 amostra116
amostra17 amostra117
amostra18 amostra118
amostra19 amostra119
amostra20 amostra120
amostra21 amostra121
amostra22 amostra122
amostra23 amostra123
amostra24 amostra124
amostra25 amostra125
amostra26 amostra126
amostra27 amostra127
amostra28 amostra128
amostra29 amostra129
amostra30 amostra130
amostra31 amostra131

```

amostra32	amostra132
amostra33	amostra133
amostra34	amostra134
amostra35	amostra135
amostra36	amostra136
amostra37	amostra137
amostra38	amostra138
amostra39	amostra139
amostra40	amostra140
amostra41	amostra141
amostra42	amostra142
amostra43	amostra143
amostra44	amostra144
amostra45	amostra145
amostra46	amostra146
amostra47	amostra147
amostra48	amostra148
amostra49	amostra149
amostra50	amostra150
amostra51	amostra151
amostra52	amostra152
amostra53	amostra153
amostra54	amostra154
amostra55	amostra155
amostra56	amostra156
amostra57	amostra157
amostra58	amostra158
amostra59	amostra159
amostra60	amostra160
amostra61	amostra161
amostra62	amostra162
amostra63	amostra163
amostra64	amostra164
amostra65	amostra165
amostra66	amostra166
amostra67	amostra167
amostra68	amostra168
amostra69	amostra169
amostra70	amostra170
amostra71	amostra171

```
amostra72      amostral72
amostra73      amostral73
amostra74      amostral74
amostra75      amostral75
amostra76      amostral76
amostra77      amostral77
amostra78      amostral78
amostra79      amostral79
amostra80      amostral80
amostra81      amostral81
amostra82      amostral82
amostra83      amostral83
amostra84      amostral84
amostra85      amostral85
amostra86      amostral86
amostra87      amostral87
amostra88      amostral88
amostra89      amostral89
amostra90      amostral90
amostra91      amostral91
amostra92      amostral92
amostra93      amostral93
amostra94      amostral94
amostra95      amostral95
amostra96      amostral96
amostra97      amostral97
amostra98      amostral98
amostra99      amostral99
amostra100     amostra200;
run;
```

```
proc delete data=amostral1;
proc delete data=amostra2;
proc delete data=amostra3;
proc delete data=amostra4;
proc delete data=amostra5;
proc delete data=amostra6;
proc delete data=amostra7;
proc delete data=amostra8;
proc delete data=amostra9;
```



```
proc delete data=amostra10;
proc delete data=amostra11;
proc delete data=amostra12;
proc delete data=amostra13;
proc delete data=amostra14;
proc delete data=amostra15;
proc delete data=amostra16;
proc delete data=amostra17;
proc delete data=amostra18;
proc delete data=amostra19;
proc delete data=amostra20;
proc delete data=amostra21;
proc delete data=amostra22;
proc delete data=amostra23;
proc delete data=amostra24;
proc delete data=amostra25;
proc delete data=amostra26;
proc delete data=amostra27;
proc delete data=amostra28;
proc delete data=amostra29;
proc delete data=amostra30;
proc delete data=amostra31;
proc delete data=amostra32;
proc delete data=amostra33;
proc delete data=amostra34;
proc delete data=amostra35;
proc delete data=amostra36;
proc delete data=amostra37;
proc delete data=amostra38;
proc delete data=amostra39;
proc delete data=amostra40;
proc delete data=amostra41;
proc delete data=amostra42;
proc delete data=amostra43;
proc delete data=amostra44;
proc delete data=amostra45;
proc delete data=amostra46;
proc delete data=amostra47;
proc delete data=amostra48;
proc delete data=amostra49;
```

```
proc delete data=amostra50;  
proc delete data=amostra51;  
proc delete data=amostra52;  
proc delete data=amostra53;  
proc delete data=amostra54;  
proc delete data=amostra55;  
proc delete data=amostra56;  
proc delete data=amostra57;  
proc delete data=amostra58;  
proc delete data=amostra59;  
proc delete data=amostra60;  
proc delete data=amostra61;  
proc delete data=amostra62;  
proc delete data=amostra63;  
proc delete data=amostra64;  
proc delete data=amostra65;  
proc delete data=amostra66;  
proc delete data=amostra67;  
proc delete data=amostra68;  
proc delete data=amostra69;  
proc delete data=amostra70;  
proc delete data=amostra71;  
proc delete data=amostra72;  
proc delete data=amostra73;  
proc delete data=amostra74;  
proc delete data=amostra75;  
proc delete data=amostra76;  
proc delete data=amostra77;  
proc delete data=amostra78;  
proc delete data=amostra79;  
proc delete data=amostra80;  
proc delete data=amostra81;  
proc delete data=amostra82;  
proc delete data=amostra83;  
proc delete data=amostra84;  
proc delete data=amostra85;  
proc delete data=amostra86;  
proc delete data=amostra87;  
proc delete data=amostra88;  
proc delete data=amostra89;
```

```
proc delete data=amostra90;
proc delete data=amostra91;
proc delete data=amostra92;
proc delete data=amostra93;
proc delete data=amostra94;
proc delete data=amostra95;
proc delete data=amostra96;
proc delete data=amostra97;
proc delete data=amostra98;
proc delete data=amostra99;
proc delete data=amostra100;
proc delete data=amostra101;
proc delete data=amostra102;
proc delete data=amostra103;
proc delete data=amostra104;
proc delete data=amostra105;
proc delete data=amostra106;
proc delete data=amostra107;
proc delete data=amostra108;
proc delete data=amostra109;
proc delete data=amostra110;
proc delete data=amostra111;
proc delete data=amostra112;
proc delete data=amostra113;
proc delete data=amostra114;
proc delete data=amostra115;
proc delete data=amostra116;
proc delete data=amostra117;
proc delete data=amostra118;
proc delete data=amostra119;
proc delete data=amostra120;
proc delete data=amostra121;
proc delete data=amostra122;
proc delete data=amostra123;
proc delete data=amostra124;
proc delete data=amostra125;
proc delete data=amostra126;
proc delete data=amostra127;
proc delete data=amostra128;
proc delete data=amostra129;
```

```
proc delete data=amostra130;
proc delete data=amostra131;
proc delete data=amostra132;
proc delete data=amostra133;
proc delete data=amostra134;
proc delete data=amostra135;
proc delete data=amostra136;
proc delete data=amostra137;
proc delete data=amostra138;
proc delete data=amostra139;
proc delete data=amostra140;
proc delete data=amostra141;
proc delete data=amostra142;
proc delete data=amostra143;
proc delete data=amostra144;
proc delete data=amostra145;
proc delete data=amostra146;
proc delete data=amostra147;
proc delete data=amostra148;
proc delete data=amostra149;
proc delete data=amostra150;
proc delete data=amostra151;
proc delete data=amostra152;
proc delete data=amostra153;
proc delete data=amostra154;
proc delete data=amostra155;
proc delete data=amostra156;
proc delete data=amostra157;
proc delete data=amostra158;
proc delete data=amostra159;
proc delete data=amostra160;
proc delete data=amostra161;
proc delete data=amostra162;
proc delete data=amostra163;
proc delete data=amostra164;
proc delete data=amostra165;
proc delete data=amostra166;
proc delete data=amostra167;
proc delete data=amostra168;
proc delete data=amostra169;
```

```

proc delete data=amostra170;
proc delete data=amostra171;
proc delete data=amostra172;
proc delete data=amostra173;
proc delete data=amostra174;
proc delete data=amostra175;
proc delete data=amostra176;
proc delete data=amostra177;
proc delete data=amostra178;
proc delete data=amostra179;
proc delete data=amostra180;
proc delete data=amostra181;
proc delete data=amostra182;
proc delete data=amostra183;
proc delete data=amostra184;
proc delete data=amostra185;
proc delete data=amostra186;
proc delete data=amostra187;
proc delete data=amostra188;
proc delete data=amostra189;
proc delete data=amostra190;
proc delete data=amostra191;
proc delete data=amostra192;
proc delete data=amostra193;
proc delete data=amostra194;
proc delete data=amostra195;
proc delete data=amostra196;
proc delete data=amostra197;
proc delete data=amostra198;
proc delete data=amostra199;
proc delete data=amostra200;

proc export data= in2.boot
            outfile= "boot"
            dbms=excel2000 replace;

run;

proc sort data=in2.boot;
  by amostra;
run;

```

```

proc means data=in2.boot mean var max noprint;
  var amostra
  f1_v01  f1_v02  f1_v03  f1_v04  f1_v05
  f1_v06
  f2_v01  f2_v02  f2_v03  f2_v04  f2_v05
  f2_v06
;
  output out=saida(drop=nada1-nada2)
  mean=nada1
  m_f1_v1  m_f1_v2  m_f1_v3  m_f1_v4  m_f1_v5
  m_f1_v6
  m_f2_v1  m_f2_v2  m_f2_v3  m_f2_v4  m_f2_v5
  m_f2_v6
  var =nada2
  v_f1_v1  v_f1_v2  v_f1_v3  v_f1_v4  v_f1_v5
  v_f1_v6
  v_f2_v1  v_f2_v2  v_f2_v3  v_f2_v4  v_f2_v5
  v_f2_v6
  max =B;
  title'Media, Variancia e Tamanho da Amostra - Topicos 2
e 4';
run;

%macro alfa(alfa);
data saidaK(keep=LI LS _type_);
  set saida;
  LI=ROUND(B*&alfa/2); *limite inferior do intervalo
Bootstrap*;
  LS=ROUND((B-LI)+1); *limite superior do intervalo
Bootstrap*;
run;
%mend alfa;
%alfa(0.05); *confiança*;

%macro valor_limite(k,j);
proc sort data=in2.boot;
  by f&k._&j;

data boot&k(keep=n _type_ f&k._&j);

```

```

set in2.boot;
  _type_=0;
  n=_n_;
run;

data limite;
merge saidaK boot&k;
  by _type_;
  if LI=n then LI_Boot_F&k._&j=f&k._&j;
  else      LI_Boot_F&k._&j=.;
  if LS=n then LS_Boot_F&k._&j=f&k._&j;
  else      LS_Boot_F&k._&j=.;
run;

proc means data=limite max noprint;
var LI_Boot_F&k._&j LS_Boot_F&k._&j;
  output out=limites_f&k._&j
         max=LI_Boot LS_Boot;
run;
proc delete data=limite;
proc delete data=boot&k;
run;
%mend valor_limite;
%valor_limite(1,v01 );
%valor_limite(1,v02 );
%valor_limite(1,v03 );
%valor_limite(1,v04 );
%valor_limite(1,v05 );
%valor_limite(1,v06 );
%valor_limite(2,v01 );
%valor_limite(2,v02 );
%valor_limite(2,v03 );
%valor_limite(2,v04 );
%valor_limite(2,v05 );
%valor_limite(2,v06 );

data limites;
  set
  limites_f1_v01  limites_f1_v02  limites_f1_v03
limites_f1_v04  limites_f1_v05

```

```
limites_f1_v06
limites_f2_v01 limites_f2_v02 limites_f2_v03
limites_f2_v04 limites_f2_v05
limites_f2_v06;
```

```
      if _n_=1 then _name_='f1_v01';
else if _n_=2 then _name_='f1_v02';
else if _n_=3 then _name_='f1_v03';
else if _n_=4 then _name_='f1_v04';
else if _n_=5 then _name_='f1_v05';
else if _n_=6 then _name_='f1_v06';
else if _n_=7 then _name_='f2_v01';
else if _n_=8 then _name_='f2_v02';
else if _n_=9 then _name_='f2_v03';
else if _n_=10 then _name_='f2_v04';
else if _n_=11 then _name_='f2_v05';
else if _n_=12 then _name_='f2_v06';
```

```
run;
```

```
proc delete data=limites_f1_v01;
proc delete data=limites_f1_v02;
proc delete data=limites_f1_v03;
proc delete data=limites_f1_v04;
proc delete data=limites_f1_v05;
proc delete data=limites_f1_v06;
proc delete data= limites_f2_v01;
proc delete data=limites_f2_v02;
proc delete data=limites_f2_v03;
proc delete data=limites_f2_v04;
proc delete data=limites_f2_v05;
proc delete data= limites_f2_v06;
run;
```

```
proc factor data=base
            method=principal
            rotate=Varimax
            outstat=perfil
            nfactors=2 noprint;
var V1-V6;
```



```

run;

data perfil_f1(keep=i f1_v01-f1_v06)
  perfil_f2(keep=i f2_v01-f2_v06);
  set perfil;
  if _type_='UNROTATE' and _name_ in ('Factor1'
'Factor2');

if _name_='Factor1' then do;
  f1_v01= v1 ;   f1_v02= v2 ;   f1_v03= v3 ;   f1_v04=
v4 ;   f1_v05= v5 ;
  f1_v06= v6 ;
  end;
else if _name_='Factor2' then do;
  f2_v01= v1 ;   f2_v02= v2 ;   f2_v03= v3 ;   f2_v04=
v4 ;   f2_v05= v5 ;
  f2_v06= v6 ;
  end;
  i=1;

  if _name_='Factor1' then output perfil_f1;
else if _name_='Factor2' then output perfil_f2;
run;

data Total(drop=i);
  merge perfil_f1 perfil_f2;
  by i;
run;

proc delete data=perfil;
proc delete data=perfil_f1;
proc delete data=perfil_f2;
run;

proc transpose data=total out=fator;
  var f1_v01-f1_v06 f2_v01-f2_v06;
run;

data fator;
  set fator;

```

```

    rename coll=cargas;
run;

proc transpose data=saida out=medias;
  var m_f1_v1-m_f1_v6 m_f2_v1-m_f2_v6;
  id _type_;
run;

proc transpose data=saida out=variancias;
  var v_f1_v1-v_f1_v6 v_f2_v1-v_f2_v6;
  id _type_;
run;

data medias;
  set medias;
  rename _0=media;

  if _name_='m_f1_v1' then _name_='f1_v01';
  if _name_='m_f1_v2' then _name_='f1_v02';
  if _name_='m_f1_v3' then _name_='f1_v03';
  if _name_='m_f1_v4' then _name_='f1_v04';
  if _name_='m_f1_v5' then _name_='f1_v05';
  if _name_='m_f1_v6' then _name_='f1_v06';
  if _name_='m_f2_v1 ' then _name_='f2_v01';
  if _name_='m_f2_v2 ' then _name_='f2_v02';
  if _name_='m_f2_v3 ' then _name_='f2_v03';
  if _name_='m_f2_v4 ' then _name_='f2_v04';
  if _name_='m_f2_v5 ' then _name_='f2_v05';
  if _name_='m_f2_v6 ' then _name_='f2_v06';
run;

data variancias;
  set variancias;
  rename _0=variancia;
  if _name_='v_f1_v1' then _name_='f1_v01';
  if _name_='v_f1_v2' then _name_='f1_v02';
  if _name_='v_f1_v3' then _name_='f1_v03';
  if _name_='v_f1_v4' then _name_='f1_v04';
  if _name_='v_f1_v5' then _name_='f1_v05';
  if _name_='v_f1_v6' then _name_='f1_v06';

```

```

if _name_='v_f2_v1 ' then _name_='f2_v01';
if _name_='v_f2_v2 ' then _name_='f2_v02';
if _name_='v_f2_v3 ' then _name_='f2_v03';
if _name_='v_f2_v4 ' then _name_='f2_v04';
if _name_='v_f2_v5 ' then _name_='f2_v05';
if _name_='v_f2_v6 ' then _name_='f2_v06';
run;
proc sort data=limites;
  by _name_;
proc sort data=fator;
  by _name_;
proc sort data=medias;
  by _name_;
proc sort data=variancias;
  by _name_;

data total_5_6(keep=fator_variavel cargas media variancia
                LI_Boot LS_Boot sig_ic vies
erro_medio g);
merge fator medias variancias limites;
  by _name_;
  rename _name_=fator_variavel;
  vies=cargas-media;
  erro_medio=variancia+vies**2;
  zero=0;
  if LI_Boot<=zero<=LS_Boot then sig_ic='Nao_Signif';
  else sig_ic='Signif      ';
  g=cargas/(variancia**(1/2));

run;

proc transpose data=total_5_6
out=g_total_5_6(drop=_name_);
  var g;
  id fator_variavel;
run;

data g_total_5_6;
  set g_total_5_6;
  _type_=0;

```

```

rename f1_v01=g_f1_v01;
rename f1_v02=g_f1_v02;
rename f1_v03=g_f1_v03;
rename f1_v04=g_f1_v04;
rename f1_v05=g_f1_v05;
rename f1_v06=g_f1_v06;
rename f2_v01=g_f2_v01;
rename f2_v02=g_f2_v02;
rename f2_v03=g_f2_v03;
rename f2_v04=g_f2_v04;
rename f2_v05=g_f2_v05;
rename f2_v06=g_f2_v06;
run;

data boot;
  set in2.boot;
  _type_=0;
run;

data boot_norm(keep=norm_f1_v01-norm_f1_v06
                 norm_f2_v01-norm_f2_v06
                 g_f1_v01-g_f1_v06
                 g_f2_v01-g_f2_v06);
merge boot saida g_total_5_6;
  by _type_;

norm_f1_v01=(f1_v01-m_f1_v1)/(v_f1_v1**(1/2));
norm_f1_v02=(f1_v02-m_f1_v2)/(v_f1_v2**(1/2));
norm_f1_v03=(f1_v03-m_f1_v3)/(v_f1_v3**(1/2));
norm_f1_v04=(f1_v04-m_f1_v4)/(v_f1_v4**(1/2));
norm_f1_v05=(f1_v05-m_f1_v5)/(v_f1_v5**(1/2));
norm_f1_v06=(f1_v06-m_f1_v6)/(v_f1_v6**(1/2));
norm_f2_v01=(f2_v01-m_f2_v1)/(v_f2_v1**(1/2));
norm_f2_v02=(f2_v02-m_f2_v2)/(v_f2_v2**(1/2));
norm_f2_v03=(f2_v03-m_f2_v3)/(v_f2_v3**(1/2));
norm_f2_v04=(f2_v04-m_f2_v4)/(v_f2_v4**(1/2));
norm_f2_v05=(f2_v05-m_f2_v5)/(v_f2_v5**(1/2));
norm_f2_v06=(f2_v06-m_f2_v6)/(v_f2_v6**(1/2));

run;

```

```

%macro pvalor(p);

data s_&p(keep=s_&p);
  set boot_norm;
  if (g_&p>0 and norm_&p>g_&p) or
     (g_&p<0 and norm_&p<g_&p) then s_&p=1;
  else
                                     s_&p=0;
run;

proc means data=s_&p sum noprint;
  var s_&p;
  output out=saida_&p
         sum=;
run;

proc delete data=s_&p;

data saida_&p(keep=fator_variavel p_valor);
  set saida_&p;
  fator_variavel="&p";
  p_valor=s_&p/_freq_;
run;

%mend pvalor;
%pvalor(f1_v01);
%pvalor(f1_v02);
%pvalor(f1_v03);
%pvalor(f1_v04);
%pvalor(f1_v05);
%pvalor(f1_v06);
%pvalor(f2_v01);
%pvalor(f2_v02);
%pvalor(f2_v03);
%pvalor(f2_v04);
%pvalor(f2_v05);
%pvalor(f2_v06);

data p_valor;

```

```

    set saida_f1_v01 saida_f1_v02 saida_f1_v03 saida_f1_v04
saida_f1_v05
        saida_f1_v06 saida_f2_v01 saida_f2_v02 saida_f2_v03
saida_f2_v04
        saida_f2_v05 saida_f2_v06;
    if p_valor>0.05 then sig_vp='Nao_Signif';
    else sig_vp='Signif    ';

run;

proc delete data=saida_f1_v01;
proc delete data=saida_f1_v02;
proc delete data=saida_f1_v03;
proc delete data=saida_f1_v04;
proc delete data=saida_f1_v05;
proc delete data=saida_f1_v06;
proc delete data=saida_f2_v01;
proc delete data=saida_f2_v02;
proc delete data=saida_f2_v03;
proc delete data=saida_f2_v04;
proc delete data=saida_f2_v05;
proc delete data=saida_f2_v06;
run;

data total_5_6(drop=g);
merge total_5_6 p_valor;
by fator_variavel;
run;

proc print data=total_5_6;
title'Analise Fatorial - Bootstrap';
run;

```

### ANEXO 3-O PROGRAMA JACKKNIFE EM SAS V.8

O programa abaixo referê-se ao usado com a base de dados “Creme Dental”.

```
libname in2 'C:\Tese Giovani\Terceira Fase';
options nodate ls=max ps=max;
proc import out= in2.basecremedental
             datafile=
"C:\Tese Giovani\Terceira Fase\basecremedental.xls"
             dbms=excel2000 replace;
             getnames=yes;
run;

data base;
  set in2.basecremedental;
run;
%macro sel(i);
data base&i;
  set base;
  if _n_=&i then delete;
run;

proc factor data=base&i
             method=principal
             rotate=Varimax
             outstat=perfil&i
             nfactors=2 noprint;

  var V1-V6;
run;

data perfil_f1_&i(keep=i f1_v01-f1_v06)
      perfil_f2_&i(keep=i f2_v01-f2_v06);
  set perfil&i;
  if _type_='UNROTATE' and _name_ in ('Factor1'
'Factor2');

  if _name_='Factor1' then do;
    f1_v01= v1 ;   f1_v02= v2 ;   f1_v03= v3 ;   f1_v04=
v4 ;   f1_v05= v5 ;   f1_v06= v6 ;
```

```

    end;
    else if _name_='Factor2' then do;
        f2_v01= v1 ;   f2_v02= v2 ;   f2_v03= v3 ;   f2_v04=
v4 ;   f2_v05= v5 ;   f2_v06= v6 ;
    end;
        i=1;

        if _name_='Factor1' then output perfil_f1_&i;
    else if _name_='Factor2' then output perfil_f2_&i;
run;

data amostra&i(drop=i);
merge perfil_f1_&i perfil_f2_&i;
  by i;
  amostra=&i;
run;

proc delete data=perfil&i;
proc delete data=perfil_f1_&i;
proc delete data=perfil_f2_&i;
proc delete data=base&i;
run;

%mend sel;
%sel (1  ); %sel (2  );
%sel (3  ); %sel (4  );
%sel (5  ); %sel (6  );
%sel (7  ); %sel (8  );
%sel (9  ); %sel (10 );
%sel (11 ); %sel (12 );
%sel (13 ); %sel (14 );
%sel (15 ); %sel (16 );
%sel (17 ); %sel (18 );
%sel (19 ); %sel (20 );
%sel (21 ); %sel (22 );
%sel (23 ); %sel (24 );
%sel (25 ); %sel (26 );
%sel (27 ); %sel (28 );
%sel (29 ); %sel (30 );
%sel (31 ); %sel (32 );

```



```
%sel (33 ); %sel (34 );  
%sel (35 ); %sel (36 );  
%sel (37 ); %sel (38 );  
%sel (39 ); %sel (40 );
```

```
data in2.jack;
```

```
set  
amostra1  
amostra2  
amostra3  
amostra4  
amostra5  
amostra6  
amostra7  
amostra8  
amostra9  
amostra10  
amostra11  
amostra12  
amostra13  
amostra14  
amostra15  
amostra16  
amostra17  
amostra18  
amostra19  
amostra20  
amostra21  
amostra22  
amostra23  
amostra24  
amostra25  
amostra26  
amostra27  
amostra28  
amostra29  
amostra30  
amostra31  
amostra32  
amostra33
```

```
amostra34  
amostra35  
amostra36  
amostra37  
amostra38  
amostra39  
amostra40;  
run;
```

```
proc delete data=amostra1;  
proc delete data=amostra2;  
proc delete data=amostra3;  
proc delete data=amostra4;  
proc delete data=amostra5;  
proc delete data=amostra6;  
proc delete data=amostra7;  
proc delete data=amostra8;  
proc delete data=amostra9;  
proc delete data=amostra10;  
proc delete data=amostra11;  
proc delete data=amostra12;  
proc delete data=amostra13;  
proc delete data=amostra14;  
proc delete data=amostra15;  
proc delete data=amostra16;  
proc delete data=amostra17;  
proc delete data=amostra18;  
proc delete data=amostra19;  
proc delete data=amostra20;  
proc delete data=amostra21;  
proc delete data=amostra22;  
proc delete data=amostra23;  
proc delete data=amostra24;  
proc delete data=amostra25;  
proc delete data=amostra26;  
proc delete data=amostra27;  
proc delete data=amostra28;  
proc delete data=amostra29;  
proc delete data=amostra30;  
proc delete data=amostra31;
```

```

proc delete data=amostra32;
proc delete data=amostra33;
proc delete data=amostra34;
proc delete data=amostra35;
proc delete data=amostra36;
proc delete data=amostra37;
proc delete data=amostra38;
proc delete data=amostra39;
proc delete data=amostra40;

proc export data= in2.jack
           outfile= "jack"
           dbms=excel2000 replace;
run;

proc sort data=in2.jack;
  by amostra;
run;

proc means data=in2.jack mean var max noprint;
  var amostra
  f1_v01 f1_v02 f1_v03 f1_v04 f1_v05 f1_v06
  f2_v01 f2_v02 f2_v03 f2_v04 f2_v05 f2_v06;
  output out=saida(drop=nada1-nada2)
         mean=nada1
         m_f1_v1 m_f1_v2 m_f1_v3 m_f1_v4 m_f1_v5 m_f1_v6
         m_f2_v1 m_f2_v2 m_f2_v3 m_f2_v4 m_f2_v5 m_f2_v6

         var =nada2
  v_f1_v1 v_f1_v2 v_f1_v3 v_f1_v4 v_f1_v5 v_f1_v6
  v_f2_v1 v_f2_v2 v_f2_v3 v_f2_v4 v_f2_v5 v_f2_v6
  max =B;
  title'Media, Variancia e Tamanho da Amostra - Topicos 2
e 4';
run;

%macro alfa(alfa);
data saidaK(keep=LI LS _type_);
  set saida;

```

```

    LI=ROUND(B*&alfa/2); *limite inferior do intervalo
    Bootstrap*;
    LS=ROUND((B-LI)+1); *limite superior do intervalo
    Bootstrap*;
run;
%mend alfa;
%alfa(0.05); *confiança*;

%macro valor_limite(k,j);
proc sort data=in2.jack;
  by f&k._&j;

data boot&k(keep=n _type_ f&k._&j);
  set in2.jack;
  _type_=0;
  n=_n_;
run;

data limite;
  merge saidaK boot&k;
  by _type_;
  if LI=n then LI_Jack_F&k._&j=f&k._&j;
  else      LI_Jack_F&k._&j=.;
  if LS=n then LS_Jack_F&k._&j=f&k._&j;
  else      LS_Jack_F&k._&j=.;
run;

proc means data=limite max noprint;
var LI_Jack_F&k._&j LS_Jack_F&k._&j;
  output out=limites_f&k._&j
         max=LI_Jack LS_Jack;
run;
proc delete data=limite;
proc delete data=boot&k;
run;
%mend valor_limite;
%valor_limite(1,v01 );
%valor_limite(1,v02 );
%valor_limite(1,v03 );
%valor_limite(1,v04 );

```

```

%valor_limite(1,v05 );
%valor_limite(1,v06 );
%valor_limite(2,v01 );
%valor_limite(2,v02 );
%valor_limite(2,v03 );
%valor_limite(2,v04 );
%valor_limite(2,v05 );
%valor_limite(2,v06 );

data limites;
  set
    limites_f1_v01  limites_f1_v02  limites_f1_v03
    limites_f1_v04  limites_f1_v05
    limites_f1_v06  limites_f2_v01  limites_f2_v02
    limites_f2_v03  limites_f2_v04  limites_f2_v05
    limites_f2_v06;

    if _n_=1 then _name_='f1_v01';
else if _n_=2 then _name_='f1_v02';
else if _n_=3 then _name_='f1_v03';
else if _n_=4 then _name_='f1_v04';
else if _n_=5 then _name_='f1_v05';
else if _n_=6 then _name_='f1_v06';
else if _n_=7 then _name_='f2_v01';
else if _n_=8 then _name_='f2_v02';
else if _n_=9 then _name_='f2_v03';
else if _n_=10 then _name_='f2_v04';
else if _n_=11 then _name_='f2_v05';
else if _n_=12 then _name_='f2_v06';
run;

proc delete data=limites_f1_v01;
proc delete data=limites_f1_v02;
proc delete data=limites_f1_v03;
proc delete data=limites_f1_v04;
proc delete data=limites_f1_v05;
proc delete data=limites_f1_v06;
proc delete data= limites_f2_v01;
proc delete data=limites_f2_v02;
proc delete data=limites_f2_v03;

```

```

proc delete data=limites_f2_v04;
proc delete data=limites_f2_v05;
proc delete data= limites_f2_v06;

run;

proc factor data=base
            method=principal
            rotate=Varimax
            outstat=perfil
            nfactors=2 noprint;
var V1-V6;
run;

data perfil_f1(keep=i f1_v01-f1_v06)
    perfil_f2(keep=i f2_v01-f2_v06);
set perfil;
if _name_='Factor1' then do;
    f1_v01= v1 ;   f1_v02= v2 ;   f1_v03= v3 ;   f1_v04=
v4 ;   f1_v05= v5 ;
    f1_v06= v6 ;
end;
else if _name_='Factor2' then do;
    f2_v01= v1 ;   f2_v02= v2 ;   f2_v03= v3 ;   f2_v04=
v4 ;   f2_v05= v5 ;
    f2_v06= v6 ;
end;
i=1;

    if _name_='Factor1' then output perfil_f1;
else if _name_='Factor2' then output perfil_f2;
run;

data Total(drop=i);
merge perfil_f1 perfil_f2;
by i;
run;

proc delete data=perfil;
proc delete data=perfil_f1;

```

```

proc delete data=perfil_f2;
run;

proc transpose data=total out=fator;
  var f1_v01-f1_v06 f2_v01-f2_v06;
run;

data fator;
  set fator;
  rename coll=cargas;
run;

proc transpose data=saida out=medias;
  var m_f1_v1-m_f1_v6 m_f2_v1-m_f2_v6;
  id _type_;
run;

proc transpose data=saida out=variâncias;
  var v_f1_v1-v_f1_v6 v_f2_v1-v_f2_v6;
  id _type_;
run;

data medias;
  set medias;
  rename _0=media;

  if _name_='m_f1_v1' then _name_='f1_v01';
  if _name_='m_f1_v2' then _name_='f1_v02';
  if _name_='m_f1_v3' then _name_='f1_v03';
  if _name_='m_f1_v4' then _name_='f1_v04';
  if _name_='m_f1_v5' then _name_='f1_v05';
  if _name_='m_f1_v6' then _name_='f1_v06';
  if _name_='m_f2_v1' then _name_='f2_v01';
  if _name_='m_f2_v2' then _name_='f2_v02';
  if _name_='m_f2_v3' then _name_='f2_v03';
  if _name_='m_f2_v4' then _name_='f2_v04';
  if _name_='m_f2_v5' then _name_='f2_v05';
  if _name_='m_f2_v6' then _name_='f2_v06';
run;
data variâncias;

```

```

set variancias;
rename _0=variancia;
if _name_='v_f1_v1' then _name_='f1_v01';
if _name_='v_f1_v2' then _name_='f1_v02';
if _name_='v_f1_v3' then _name_='f1_v03';
if _name_='v_f1_v4' then _name_='f1_v04';
if _name_='v_f1_v5' then _name_='f1_v05';
if _name_='v_f1_v6' then _name_='f1_v06';
if _name_='v_f2_v1 ' then _name_='f2_v01';
if _name_='v_f2_v2 ' then _name_='f2_v02';
if _name_='v_f2_v3 ' then _name_='f2_v03';
if _name_='v_f2_v4 ' then _name_='f2_v04';
if _name_='v_f2_v5 ' then _name_='f2_v05';
if _name_='v_f2_v6 ' then _name_='f2_v06';
run;
proc sort data=limites;
by _name_;
proc sort data=fator;
by _name_;
proc sort data=medias;
by _name_;
proc sort data=variancias;
by _name_;

data total_5_6(keep=fator_variavel cargas media variancia
LI_Jack LS_Jack sig_ic vies
erro_medio g);
merge fator medias variancias limites;
by _name_;
rename _name_=fator_variavel;
vies=cargas-media;
erro_medio=variancia+vies**2;
zero=0;
if LI_Jack<=zero<=LS_Jack then sig_ic='Nao_Signif';
else sig_ic='Signif ' ;
g=cargas/(variancia**(1/2));

run;

```



```

proc transpose data=total_5_6
out=g_total_5_6(drop=_name_);
  var g;
  id fator_variavel;
run;

data g_total_5_6;
set g_total_5_6;
  _type_=0;
  rename f1_v01=g_f1_v01;
  rename f1_v02=g_f1_v02;
  rename f1_v03=g_f1_v03;
  rename f1_v04=g_f1_v04;
  rename f1_v05=g_f1_v05;
  rename f1_v06=g_f1_v06;
  rename f2_v01=g_f2_v01;
  rename f2_v02=g_f2_v02;
  rename f2_v03=g_f2_v03;
  rename f2_v04=g_f2_v04;
  rename f2_v05=g_f2_v05;
  rename f2_v06=g_f2_v06;
run;

data jack;
set in2.jack;
  _type_=0;
run;

data jack_norm(keep=norm_f1_v01-norm_f1_v06
                 norm_f2_v01-norm_f2_v06
                 g_f1_v01-g_f1_v06
                 g_f2_v01-g_f2_v06);
merge jack saida g_total_5_6;
  by _type_;

norm_f1_v01=(f1_v01-m_f1_v1)/(v_f1_v1**(1/2));
norm_f1_v02=(f1_v02-m_f1_v2)/(v_f1_v2**(1/2));
norm_f1_v03=(f1_v03-m_f1_v3)/(v_f1_v3**(1/2));
norm_f1_v04=(f1_v04-m_f1_v4)/(v_f1_v4**(1/2));
norm_f1_v05=(f1_v05-m_f1_v5)/(v_f1_v5**(1/2));

```

```

norm_f1_v06=(f1_v06-m_f1_v6)/(v_f1_v6**(1/2));
norm_f2_v01=(f2_v01-m_f2_v1)/(v_f2_v1**(1/2));
norm_f2_v02=(f2_v02-m_f2_v2)/(v_f2_v2**(1/2));
norm_f2_v03=(f2_v03-m_f2_v3)/(v_f2_v3**(1/2));
norm_f2_v04=(f2_v04-m_f2_v4)/(v_f2_v4**(1/2));
norm_f2_v05=(f2_v05-m_f2_v5)/(v_f2_v5**(1/2));
norm_f2_v06=(f2_v06-m_f2_v6)/(v_f2_v6**(1/2));

run;

%macro pvalor(p);

data s_&p(keep=s_&p);
  set jack_norm;
  if (g_&p>0 and norm_&p>g_&p) or
     (g_&p<0 and norm_&p<g_&p) then s_&p=1;
  else
     s_&p=0;
run;

proc means data=s_&p sum noprint;
  var s_&p;
  output out=saida_&p
         sum=;
run;

proc delete data=s_&p;

data saida_&p(keep=fator_variavel p_valor);
  set saida_&p;
  fator_variavel="&p";
  p_valor=s_&p/_freq_;
run;

%mend pvalor;
%pvalor(f1_v01);
%pvalor(f1_v02);
%pvalor(f1_v03);
%pvalor(f1_v04);
%pvalor(f1_v05);
%pvalor(f1_v06);

```

```

%pvalor(f2_v01);
%pvalor(f2_v02);
%pvalor(f2_v03);
%pvalor(f2_v04);
%pvalor(f2_v05);
%pvalor(f2_v06);

data p_valor;
  set saida_f1_v01 saida_f1_v02 saida_f1_v03 saida_f1_v04
saida_f1_v05
  saida_f1_v06 saida_f2_v01 saida_f2_v02 saida_f2_v03
saida_f2_v04
  saida_f2_v05 saida_f2_v06;
  if p_valor>0.05 then sig_vp='Nao_Signif';
  else sig_vp='Signif';

run;

proc delete data=saida_f1_v01;
proc delete data=saida_f1_v02;
proc delete data=saida_f1_v03;
proc delete data=saida_f1_v04;
proc delete data=saida_f1_v05;
proc delete data=saida_f1_v06;
proc delete data=saida_f2_v01;
proc delete data=saida_f2_v02;
proc delete data=saida_f2_v03;
proc delete data=saida_f2_v04;
proc delete data=saida_f2_v05;
proc delete data=saida_f2_v06;
run;

data total_5_6(drop=g);
  merge total_5_6 p_valor;
  by fator_variavel;
run;

proc print data=total_5_6;
  title'Analise Fatorial - Jacknife';
run;

```





**More  
Books!** 



**yes**  
**I want morebooks!**

Buy your books fast and straightforward online - at one of the world's fastest growing online book stores! Environmentally sound due to Print-on-Demand technologies.

Buy your books online at  
**[www.get-morebooks.com](http://www.get-morebooks.com)**

Compre os seus livros mais rápido e diretamente na internet, em uma das livrarias on-line com o maior crescimento no mundo! Produção que protege o meio ambiente através das tecnologias de impressão sob demanda.

Compre os seus livros on-line em  
**[www.morebooks.es](http://www.morebooks.es)**

SIA OmniScriptum Publishing  
Brīvības gatve 197  
LV-103 9 Rīga, Latvia  
Telefax: +371 68620455

[info@omniscrptum.com](mailto:info@omniscrptum.com)  
[www.omniscrptum.com](http://www.omniscrptum.com)

OMNI Scriptum









