Um Procedimento Inferêncial para Análise Fatorial Utilizando as Técnicas Bootstrap e Jackknife: Construção de Intervalos de Confiança e Testes de Hipóteses e um Estudo de Caso: Pesquisa Sobre as Percepções de Funcionários Sobre o Nível de Qualidade Total de Empresas

Giovani Glaucio de Oliveira Costa
Pontifícia Universidade Católica do Rio de Janeiro
Rua Marques de São Vicente, 225, Gávea, Rio de Janeiro, CEP 22.453-900, RJ, Brasil.
E-mail: giovaniglaucio@hotmail.com
A quem deve ser dirigida a correspondência da revista

Reinaldo Castro Souza
Departamento de Engenharia Elétrica – PUC-Rio.
Rua Marques de São Vicente, 225, Gávea, Rio de Janeiro, CEP 22.453-900, RJ, Brasil.
E-mail: reinaldo@ele.puc-rio.br

Vítor Hugo de Carvalho Gouvea
Universidade Federal Fluminense
Rua Mário Santos Braga, S/N. Campus do Valonguinho. 7º Andar, Niterói, RJ, CEP 24.020-140, Brasil.
E-mail: getvhcg@vm.uff.br

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 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 artigo se propõe a estabelecer um procedimento inferêncial 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.

Palavras-chaves: análise fatorial, inferência estatística, bootstrap e jackknife

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.

Keywords: factor analysis, inference statistics, bootstrap and jackknife.

1-INTRODUÇÃO

A presente pesquisa tem por objetivo estabelecer um critério para significância de cargas fatoriais: o ***Teste Inferêncial para Análise Fatorial-TIAF.*** Mais precisamente, trata-se do desenvolvimento de um procedimento que consiste na modelagem estatística inferêncial 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.

A investigação ora proposta 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 valores-p e sair, consequentemente, do terreno puramente descritivo.

Esperam-se que com o sucesso na especificação de um procedimento estatístico para inferência dos resultados das análises fatoriais por componentes principais, os pesquisadores possam ter dado sua contribuição inicial 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.

Este artigo visa motivar a discussão e o embasamento inicial 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 e análise de correlação canônica.

2-METODOLOGIA DA PESQUISA

Para atingir os objetivos deste artigo se utilizará da metodologia ***CIS*** (Computer *Intensive Statistics*), que cogitam o modelo de densidade de probabilidade que explica o comportamento aleatório da estatística observada e seus parâmetros característicos de forma empírica.

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

Este artigo trata da especificação da variância, do viés e do erro médio quadrático ***(EMQ)*** da variável aleatória 'carga fatorial' das variáveis dos fatores determinantes de uma análise fatorial (os fatores 1 e 2, neste artigo) 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.

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).

3- REVISÃO DE LITERATURA

A análise fatorial é uma técnica estatística que tem como objetivo descrever a estrutura de dependência de um conjunto de variáveis através da criação de fatores, que são variáveis que, supostamente, medem aspectos comuns.

Historicamente, a origem das técnicas de análise fatorial está ligada a estudos da área de psicologia. Sua criação data do início do século , quando Spearman (Spearman,1904) desenvolveu um método para criação de um índice geral de inteligência (fator g) com base nos resultados da vários testes (escalas) que refletiam essa aptidão. Tratava-se de um primeiro método de análise fatorial, adequado para a estimação de um único fator.

O desenvolvimento inicial de métodos de análise fatorial esteve muito ligado ao problema de avaliação de escalas cognitivas e foi responsabilidade de uma série de pesquisadores da área de psicologia (Spearman, 1904; Thurstone, 1935 e Burt, 1941; por exemplo).

No início, os métodos apresentavam uma característica muito mais empírica do que inferêncial. Em 1940, com Lawley, surge um primeiro trabalho com um rigor matemático, o que fez com que aumentasse a aceitação dessas técnicas, nesse meio (Lawley, 1940).

Uma situação comum em várias áreas do conhecimento é aquela na qual se observam para cada elemento amostral, um grande número de variáveis. Essas variáveis podem ser, por exemplo, características demográficas, um conjunto de itens de uma escala ou mesmo os resultados obtidos por um indivíduo em diferentes escalas de avaliação. Diante de um quadro como esse, o pesquisador enfrenta dois problemas:

a)Como caracterizar a amostra levando-se em conta um conjunto eventualmente grande de variáveis?

b)Como descrever a inter-relação existente entre essas variáveis, eventualmente explicitando uma estrutura de interdependência subjacente aos dados?

A análise fatorial resolve esses dois problemas. Reis(1997) define a análise fatorial como *"um conjunto de técnicas estatísticas cujo objetivo é representar ou descrever um número de variáveis originais a partir de um menor número de variáveis hipotéticas"*. Trata-se de uma técnica estatística multivariada que, a partir da estrutura de dependência existente entre variáveis de interesse(em geral representada pelas correlações ou covariâncias entre essas variáveis), permite a criação de um conjunto menor de variáveis(variáveis latentes ou fatores), obtidas através das originais.Além disso, a técnica permite saber o quanto cada fator está associado a cada variável e o quanto o conjunto de fatores explica da variabilidade total dos dados originais. Note que isso vem ao encontro da resolução do problema (a), haja vista que , quando a análise fatorial é bem sucedida , o pesquisador pode trabalhar com um número reduzido de variáveis sem uma perda muito grande de informações. O problema (b) também é solucionado, já que cada um desses fatores pode representar uma característica subjacente aos dados. Tome por exemplo Spearman(1904),que interpretou o fator "g" como uma medida de inteligência que estaria implicitamente ligada ao desempenho de um conjunto de testes. Esse é o espírito das técnicas que serão abordadas neste trabalho.

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 principiais, 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 ser 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 significante. 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(α), 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 aumentam 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 significante;
2. Quanto maior o número de variáveis a serem incluídas no problema, menor o valor das cargas fatoriais para serem consideradas significantes;
3. Quanto maior o número de fatores, maior o valor da carga em fatores posteriores a serem consideradas significantes 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.

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 inferênciais para a matriz girada da análise fatorial ainda não foi satisfatoriamente desenvolvido por cientistas da área, o que dificulta o seu uso regular.

Ludovic Lebart, Alain Morineau e Merie Piron (Statistique Exploratoire Multidimensionelle, 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 (Probabilités Analyse Des Données Et Statistique), 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 significante 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 inferêncial 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.

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.

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, as próximas seções deste trabalho.

4-DEFINIÇÃO DO TIAF:

O TIAF é um teste de significância para as cargas fatoriais utilizando métodos de reamostragens tais como o bootstrap e jackknife.

5-OBJETIVOS DO TIAF:

- Orientar sobre quais variáveis e cargas fatoriais devem incidir a nomeação de fatores;
- Inferência estatística eficaz para a Análise Fatorial.

6-HIPÓTESES DO TIAF:

Neste estudo, 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:

$C_i > 0{,}\ 00$, se o sentido da correlação do fator for direta com a variável testada ou.
$C_i < 0{,}00$, se o sentido da correlação do fator for inversa com a variável testada

7-REGRAS DE DECISÃO DO TIAF:

• Aceitar H_0, **Não Existe** correlação X_i e F_j;
• Rejeitar H_0, **Existe** correlação X_i e F_j;
• Aceitar H_0, resultado **Não Significante** para C_{ij};
• Rejeitar H_0, resultado **Significante** para C_{ij};
• **Cargas Significantes** indicam que a variável correspondente **deve participar** da interpretação/nomeação do fator em foco;
• **Cargas Não Significantes** indicam que a variável **não deve participar** da interpretação/nomeação do fator em foco.

8-SIGNIFICÂNCIA PRÁTICA:

Regra para identificar quais variáveis **significantes estatisticamente podem participar** da nomeação/interpretação do fator;

Devem participar da nomeação/interpretação do fator, as cargas fatoriais significantes que estiverem no intervalo /*0,3 a 1,0*/.

9-ALGORÍTMO DO TIAF BOOTSTRAP:

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 girada de fatores ortogonal (rotação varimax);
4º)Com base na matriz girada de fatores, selecionam-se dos fatores 1 e 2 as cargas fatoriais, isto é, as estimativas dos vetores θ_1 e θ_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 α , igual ao nível de significância, de acordo com as expressões abaixo:

$\hat{\theta}(q_1)$= *limite inferior , onde* $q_1 = B\ \alpha\backslash 2$

$\hat{\theta}(q_2)$= *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.α/2)e o limite superior o na 976ª posição (q_2=B-q_1+1).

Pode-se então afirmar, com uma probabilidade α 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;

O algorítmo jackknife segue rotinas análogas.

Os programas utilizados para desenvolver os algorítmos acima foram o **R 2.1.1** e o **SAS Versão 8.**

Em breve os programas, na versão R 2.1.1, estarão disponíveis na biblioteca do R..

10-OS DEZ MADAMENTOS DO TIAF:

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 girada 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 quais das variáveis significantes 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.

11-ESTUDO DE CASO:BASE QUALIDADE TOTAL

O objetívo 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 (V1);*
- *Gerência participativa (V2);*
- *Desenvolvimento de recursos humanos (V3);*
- *Constância de propósitos (V4);*
- *Aperfeiçoamento contínuo (V5);*
- *Gerência de projetos (V6);*
- *Delegação (V7);*
- *Disseminação de informações (V8);*
- *Garantia da qualidade (V9);*
- *Não aceitação de erros (V10);*

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 na empresa.

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 a onde está a qualidade total das empresas investigas.

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

12-RESULTADOS DO TIAF APLICADOS AO ESTUDO DE CASO:

Resultados Bootstrap:

Tabela 2

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
					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 2 contém as estimativas das cargas fatoriais para cada variável e para cada fator ou componente principal na simulação bootstrap.

No fator 1, os enviezamentos 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 conforme ao item 8.

Resultados Jackknife:

Tabela 3

VAR	Cargas	Média	Variância	LI	LS	Viés	EMQ	SIGIC	Valor-p	SIGVP
Fator 1										
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
Fator 2										
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 3 contém as estimativas das cargas fatoriais para cada variável e para cada fator ou componente principal.

No fator 1, os enviezamentos 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 Jackkinfe se tornou um método de estimação de cargas fatoriais mais eficientes.

Após os resultados inferênciais verificados, parte-se para a significância prática das cargas fatoriais significantes por fator (ver item 8) e para a nomeação/interpretação dos mesmos e tomados de decisões estatísticas.

13-CONCLUSÕES SOBRE O TRABALHO REALIZADO:

- *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 neste trabalho 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;*
- *A análise da convergência indicou que com um **B=200** já se tem a **Regularidade Estatística dos Resultados,** mas para garantir o máximo de eficácia ainda se recomenda 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 proposto neste trabalho tem eficácia bem satisfatória, podendo-se até afirmar que supera, principalmente no uso do valor-p, a eficácia do método tradicional.*

14-REFERÊNCIAS BILIOGRÁFICAS

[1]Afifi, A. A. e **Clark**, V. (1984).Computer – Aided Multivariate Analysis. Lifetime Learning Publications. Belm. California.

[2]Anderson, T.W. (1984).An Introduction to Multivariate Statistical Analysis. 2ed.New York : John Wiley & Sons.

[3]Cazar, R. A.(2003).An Exercise on Chemometrics for a Quantitative Analysis Course.Madison:Journal of Chemical Education.

[4]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.

[5]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.

[6]Efron,B(1979).Bootstrap Methods:Another Look at the Jackknife, The Annals of Statistic,7,1-26.

[7]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.

[8]Efron,B.(1982).The Jackknife, the Bootstrap , and other Resampling Methods, CBNS 38,SIAM-NSF

[9]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.

[10]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.

[11]Johnson , D. E .(1998).Applied Multivariate Methods for Data Analysis. Pacific Grove : Duxbury Press..

[12]Johnson , R. A .e **Wichern** , D.W .(1998).Applied Multivariate Statistical Analysis. 4ed . Upper Saddle River: Prentice Hall.

[13]Johnson, R. A. e **Wichern**, D. W.(1982).Applied Multivariate Statistical Analysis. Prentice Hall, Inc. Englewood Cliffs, New Jersey.

[14]Kaiser , H . F .(1958).The Varimax Criterion for Analytic Rotation in Factor Analysis . Psychometrika , 23 , 187-200.

[15]Kaiser , H . F .(1974).A Second-Generation Little Jiffy. Psychometrika 35:401-15

[16]Proença, Isabel Maria D.(1988).Estimativas Jackknife e Bootstrap para o Enviesamento e Desvio-padrão do índice de Gini. Doc.Trabalho Nº 67, Cemapre, I.S.E , Lisboa.

[17]Reis, E.(1997).Estatística Multivariada Aplicada. Lisboa : Edições Sílabo.

[18]Reyment, R. e **Joreskog**, K . G.(1996).Applied Factor Analysis in the Natural Science . Cambridge. Cambridge University Press.

[19]SAS Procedures Guide, Version 8.Cary , N.C.(1999). SAS Institute Inc.

[20]Siegel,Sidney(1975)**.** Estatística Não-paramétrica Para Ciências do Comportamento. São Paulo:McGraw-Hill Editora

[21]Sharma , S.(1996).Applied Multivariate Techniques . New York : Jonh Wile & Sons..

[22]Spearman, C.(1940).General Intelligence Objectivelly Determined and Measured . American Journal of Psychology , 15 : 201-293.

[23]Stewart, David W.(1981).The applications and misapplications of factor analysis in marketing research. Journal of Marketing Research, 18, p.51-62,Feb.