

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/280493625>

# Análise Multivariada com dados binários e sua potencial aplicação na área de Tecnologia da Informação e Comunicação

Conference Paper · June 2015

CITATIONS

4

READS

1,812

4 authors:



**Érica Siqueira**

Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas

29 PUBLICATIONS 34 CITATIONS

[SEE PROFILE](#)



**Marcelo Henrique de Araujo**

Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas

29 PUBLICATIONS 31 CITATIONS

[SEE PROFILE](#)



**Cesar Souza**

University of São Paulo

74 PUBLICATIONS 324 CITATIONS

[SEE PROFILE](#)



**Nicolau Reinhard**

University of São Paulo

84 PUBLICATIONS 546 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Open Data Intermediaries [View project](#)



Digital Divide [View project](#)

## **Análise Multivariada com dados binários e sua potencial aplicação na área de Tecnologia da Informação e Comunicação**

**Autoria:** Erica Souza Siqueira, Marcelo Henrique de Araujo, Cesar Alexandre de Souza, Nicolau Reinhard

### **Resumo**

Este artigo tem como objetivo discutir e demonstrar a aplicação de técnicas estatísticas multivariadas com dados de natureza binária. Para tanto, as análises empreendidas neste trabalho se fundamentaram nos dados secundários da pesquisa “TIC Domicílios” e “TIC Empresas”, ambas coordenadas pelo Centro de Estudos sobre as Tecnologias da Informação e Comunicação (Cetic.br). Inicialmente, demonstrou-se a aplicação da Análise Fatorial Binária para fins exploratórios, visando propiciar redução dimensional dos fenômenos estudados. Em seguida, demonstrou-se a aplicação da modelagem de equações estruturais, com base em dados binários, na construção de um índice referente ao uso das TICs em empresas. Tanto a “redução dimensional”, quanto à “criação de um índice” são objetivos recorrentes dentro de estudos da área de TI/SI. Portanto, demonstrando a aplicabilidade de tais técnicas para o desenvolvimento de investigações científicas nesta área de estudos.

**Palavras Chave:** Análise Fatorial Binária, Modelagem de Equações Estruturais, *Partial Least Squares* (PLS), Análise Fatorial Confirmatória.

### **1. Introdução**

Desenvolver uma pesquisa quantitativa na área de Tecnologia da Informação e Comunicação (TIC) demanda grande esforço por conta das exigências das técnicas de análises de dados que, para garantir a sua validade, requerem grandes amostras e, muitas vezes, pressupõem normalidade de sua distribuição. Além disso, a extrapolação do resultado de tais pesquisas requer critérios bem executados de amostragem que respeitem, entre outros, as estratificações a fim de garantir a representatividade. Quando se toma como objeto de análise o mundo das organizações, essas exigências são ainda mais limitadoras. Como conseguir um grande número de empresas? Como garantir normalidade dos dados? Assim, se observa uma grande quantidade de estudos da área de TIC que desenvolvem investigações de cunho quantitativo com a coleta de dados por meio de questionários, mas que não se atentam em relação aos (i) procedimentos de amostragem (no geral, tende-se a assumir uma amostragem por conveniência), (ii) tamanho da amostra, (iii) averiguação da normalidade dos dados, além da (iv) escolha da técnica estatística, considerando as premissas da técnica e da natureza dos dados que serão analisados.

Diante deste cenário, a utilização de dados secundários surge como uma oportunidade para ampliar a realização de estudos quantitativos na área de TIC. Como exemplo, podem ser observados os microdados da série histórica de pesquisas (*surveys*) sobre a disponibilização e uso das Tecnologias da Informação e Comunicação (TIC) realizadas pelo Centro de Estudos sobre as Tecnologias da Informação e Comunicação (CETIC.br). Tais pesquisas são desenvolvidas com critérios claros e rigorosos de amostragem, tendo grandes amostras que representam a realidade brasileira sobre o uso dos artefatos tecnológicos. Todavia, uma característica de tais dados que poderia reduzir o leque de aplicação de técnicas estatísticas multivariadas é a presença de variáveis de natureza dicotômica.

A partir da supracitada limitação, o presente artigo tem como objetivo discutir a aplicação do ferramental estatístico de cunho multivariado, especificamente para dados de natureza binária. Para tanto, será demonstrado o potencial da aplicação das técnicas de (i) análise fatorial e (ii)

modelagem de equações estruturais tendo como base dados secundários (especificamente de natureza dicotômica). Tais técnicas podem ser aplicadas visando atingir objetivos comuns em investigações da área de Tecnologia da Informação e Comunicação, por exemplo, redução da dimensionalidade de dados e a construção de um índice. Sendo que este último objetivo pode ser especialmente importante quando se deseja apurar a satisfação do consumidor, medir o capital intelectual de uma empresa ou calcular uma pontuação para crédito de um potencial cliente. Para fundamentar a discussão empreendida neste artigo são apresentados e discutidos exemplos de outros estudos, a fim de demonstrar a viabilidade da aplicação das técnicas de Análise Fatorial e Modelagem de Equações Estruturais com dados dicotômicos ou binários.

Sendo assim, a contribuição deste artigo se dá em termos metodológicos, visando compreender o potencial e de que maneira as técnicas estatísticas de análise fatorial e modelagem de equações estruturais – tipicamente utilizadas para dados de natureza métrica – podem ser utilizadas com um conjunto de dados binários.

Posto isso, as demais seções deste artigo estão organizadas da seguinte forma: na seção 2 discute-se como aplicar a técnica estatística de análise fatorial, porém adequando a mesma para lidar com as peculiaridades dos dados binários. A discussão empreendida na seção 2 é exemplificada por meio dos microdados da pesquisa “TIC Domicílios” (CGI, 2012), na qual avaliou-se o uso de serviços de governo eletrônico no Brasil. Em seguida, na seção 3, são discutidos os pressupostos teóricos que fundamentam a construção de um índice de uso das TIC nas pequenas e médias empresas, por meio da técnica de modelagem de equações estruturais. Nas seções 4, 5 e 6, apresenta-se detalhadamente os passos para criação de um índice para avaliação do uso das Tecnologias da Informação e Comunicação por empresas, tendo como base os dados secundários da pesquisa “TIC Empresas” (CGI, 2012). Por fim, na seção 7 são tecidas as considerações finais deste artigo.

## **2. Análise Fatorial Exploratória para redução dimensional com dados binários**

O foco desta seção recai em discutir o uso da técnica da Análise Fatorial visando à redução da dimensionalidade dos dados. Embora, tal objetivo seja inerente à referida técnica, o diferencial da discussão empreendida neste artigo se dá na sua aplicação específica em dados de natureza binária.

Segundo Hair et al. (2009) a Análise Fatorial (ou Análise dos Fatores Comuns) consiste de uma técnica de interdependência cujo objetivo é definir a estrutura latente entre as variáveis da análise. Em outras palavras, a premissa subjacente a esta técnica é que seja possível representar um conjunto de variáveis, por meio de um número de fatores intrínsecos. Portanto, um dos objetivos da Análise Fatorial é a simplificação/redução de um grande número de variáveis, de acordo a proposta desta seção (FÁVERO et al., 2009, p.235).

Em consonância a discussão realizada em Hair et al. (2009) e Fávero et al. (2009), um dos requisitos para aplicação da técnica discutida nesta seção é a de que a natureza dos dados analisados seja métrica (i.e. quantitativos). Todavia, no campo de Tecnologia da Informação e Comunicação, em virtude da variedade de objetos de pesquisa estudados, nem sempre é possível atender tal premissa, principalmente quando se verifica a ausência/presença de um determinado fenômeno (dado binário) ou então fenômenos não mensuráveis, mas que seja possível identificar uma ordem entre os mesmos (e.g. classe socioeconômica; escala ordinal).

Entretanto, mesmo quando as supracitadas premissas não puderem ser atendidas, ainda assim é possível utilizar do ferramental da Análise Fatorial, todavia adequando-a para lidar com dados de natureza ordinal e binária. Assim, em consonância aos objetivos deste artigo, nesta

seção busca-se discutir e demonstrar especificamente como se dá a operacionalização da técnica de Análise Fatorial para dados binários, conhecida como *Análise Fatorial Binária* (AFB).

Além da questão da natureza dos dados utilizados para aplicação da técnica de Análise Fatorial, esta ainda pode ser classificada quanto ao seu objetivo, isto é, a referida técnica pode ser aplicada tanto para fins (i) *exploratórios*, quanto para fins (ii) *confirmatórios*. Nesta seção discute-se especificamente o uso da Análise Fatorial Binária para fins exploratórios.

Segundo Bartholomew et al. (2002) existem duas abordagens distintas para adaptar o modelo de Análise Fatorial para lidar com dados binários. Uma delas refere-se à abordagem denominada como “modelo de variáveis latentes” (*latent variables models*), que se fundamenta em um modelo logístico para estimar os fatores latentes. A segunda dessas consiste da “abordagem de variáveis subjacentes” (*underlying variables approach*), que se fundamenta em uma matriz de correlações tetracórica (*tetrachoric correlations*) para estimar os fatores (BARTHOLOMEW et al., 2002). A fim de restringir o escopo deste trabalho, nesta seção buscar-se-á demonstrar, por meio de exemplos extraídos de outras investigações, como operacionalizar a aplicação da Análise Fatorial Binária para fins exploratórios por meio da abordagem de variáveis subjacentes.

Enquanto o algoritmo da Análise Fatorial se utiliza do cálculo da matriz de correlação de Pearson (específica para representar a correlação entre dados quantitativos) para extrair os fatores latentes, a Análise Fatorial Binária – utilizando a abordagem de variáveis subjacentes – utiliza a matriz de correlação tetracórica como entrada para a execução da técnica de Análise Fatorial Binária. Neste sentido, as correlações tetracóricas são usadas especificamente para estimar as correlações entre os dados binários (BARTHOLOMEW et al., 2002).

Visando exemplificar a aplicação da Análise Fatorial Binária (AFB), utilizaremos os resultados de pesquisa empreendida por Araujo (2013), que utilizou dos microdados da *survey* anual coordenada pelo Centro de Estudos sobre as Tecnologias da Informação e Comunicação (CETIC.br), a pesquisa TIC Domicílios. Embora o referido estudo tenha realizado uma análise longitudinal, considerando os dados da edição de 2007, 2009 e 2011, para fins de demonstração, neste artigo analisaremos apenas as análises realizadas no ano de 2011.

A supracitada pesquisa teve como objetivo realizar uma análise dos fatores que influenciam a adoção de serviços de governo eletrônico (e-gov) no Brasil. Nesta investigação, o autor se utilizou do ferramental da Análise Fatorial Binária para fins de redução de dimensionalidade. Conforme ilustrado no Quadro 1, foram analisados dezenove serviços de e-gov, representados por 19 variáveis binárias (que medem o uso/não uso de cada um dos serviços elencados), em conformidade as análises desenvolvidas em Araujo (2013). Em virtude da grande variedade de variáveis analisadas, no Quadro 1 cada uma destas recebe uma codificação única (e.g. G1, G2...), a fim facilitar a interpretação dos resultados.

Código	Tipos de Serviços de Governo eletrônico
G1	Buscar Informações sobre como emitir documentos
G2	Obter certidões negativas, licenças e permissões.
G3	Consultar o CPF – Cadastro de Pessoa Física
G4	Emissão de Documentos
G5	Obter informações sobre impostos e taxas em sites de governo
G6	Fazer Declaração de Imposto de Renda (incluindo declaração de isento)
G7	Fazer Pagamento impostos, multas e taxas (IPVA, IPTU, multas etc.)
G8	Buscar informações sobre previdência social e benefícios sociais.
G9	Solicitar serviços junto à previdência social (salário-maternidade, auxílio doença etc.)
G10	Fazer boletim de ocorrência
G11	Consultar andamento de atos processuais na justiça
G12	Buscar informações sobre direito do consumidor
G13	Buscar informações sobre veículos roubados (Registro de roubo e furto de veículos)
G14	Consultar pontos na carteira de habilitação e multas
G15	Buscar informações sobre serviços públicos de educação.
G16	Fazer matrículas em escolas públicas ou instituições de ensino público
G17	Buscar Informações sobre direitos do trabalhador.
G18	Fazer Inscrição em concursos públicos
G19	Buscar informações sobre empregos em algum canal de comunicação do governo pela Internet

**Quadro 1** – Tipos de serviços de e-gov abordados na pesquisa TIC Domicílios 2011.

**Fonte:** CGI (2012)

Seguindo a análise desenvolvida, o processo de aplicação da AFB para fins exploratórios foi realizado pela seguinte sequência de atividades: (i) Análise da matriz de correlação; (ii) Seleção do método fatorial; (iii) Extração e determinação do número de fatores; (iv) Rotação dos fatores; (v) Interpretação dos mesmos e (vi) Computação dos scores fatorial. Portanto, estando em consonância aos preceitos teóricos apresentados em Fávero et al. (2009), Hair et al. (2009) e Bartholomew et al. (2002). Todas as análises desenvolvidas foram realizadas com o suporte do software *Stata* (versão 11).

Em conformidade a discussão supracitada, para aplicação da AFB como base na “abordagem de variáveis subjacentes”, o primeiro passo consiste do cálculo da matriz de correlação tetracórica. Para fins operacionais, as referidas matrizes foram calculadas por meio de uma biblioteca específica compatível o software *Stata*. Tal biblioteca foi desenvolvida pelo *Institute for Digital Research and Education* (IDRE), centro de pesquisa da Universidade da Califórnia em Los Angeles (UCLA, 2013), cujo conteúdo da referida biblioteca pode ser obtida diretamente no *Stata*, por meio do comando ‘*findit tetrachoric*’.

**Tabela 1** – Matriz de correlação tetracórica dos serviços de governo eletrônico (2011)

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19
G1	1,00																		
G2	0,74	1,00																	
G3	0,73	0,70	1,00																
G4	0,77	0,81	0,70	1,00															
G5	0,58	0,67	0,55	0,65	1,00														
G6	0,48	0,58	0,55	0,57	0,76	1,00													
G7	0,45	0,57	0,47	0,55	0,69	0,73	1,00												
G8	0,50	0,45	0,46	0,52	0,54	0,48	0,41	1,00											
G9	0,48	0,43	0,44	0,50	0,46	0,49	0,43	0,89	1,00										
G10	0,48	0,46	0,47	0,47	0,50	0,43	0,43	0,37	0,36	1,00									
G11	0,46	0,56	0,42	0,52	0,55	0,55	0,48	0,52	0,47	0,55	1,00								
G12	0,53	0,50	0,54	0,55	0,61	0,54	0,44	0,58	0,57	0,49	0,64	1,00							
G13	0,46	0,54	0,50	0,50	0,51	0,49	0,49	0,46	0,45	0,63	0,55	0,62	1,00						
G14	0,43	0,48	0,46	0,51	0,56	0,51	0,59	0,41	0,42	0,47	0,52	0,55	0,72	1,00					
G15	0,44	0,37	0,39	0,47	0,44	0,37	0,30	0,46	0,47	0,32	0,33	0,52	0,36	0,32	1,00				
G16	0,39	0,33	0,36	0,42	0,32	0,27	0,25	0,43	0,47	0,29	0,29	0,41	0,37	0,26	0,63	1,00			
G17	0,49	0,45	0,52	0,53	0,51	0,46	0,39	0,63	0,55	0,34	0,43	0,64	0,46	0,45	0,56	0,47	1,00		
G18	0,40	0,42	0,41	0,47	0,42	0,34	0,30	0,40	0,42	0,34	0,34	0,51	0,38	0,39	0,64	0,50	0,54	1,00	
G19	0,42	0,38	0,41	0,45	0,43	0,32	0,26	0,43	0,43	0,32	0,33	0,52	0,33	0,35	0,57	0,47	0,58	0,68	1,00

**Fonte:** Araujo (2013)

A Tabela 1 sintetiza os resultados do cálculo da matriz tetracórica entre os diferentes serviços de governo eletrônico. Uma premissa essencial para aplicação Análise Fatorial é que os atributos analisados sejam correlacionados, afinal a variável latente (fator) representa um conjunto de variáveis altamente correlacionadas entre si e pouco correlacionadas aos atributos de diferentes fatores. Neste sentido, esta etapa consiste da inspeção da matriz de correlação, a fim de verificar os coeficientes de correlação. Em conformidade aos dados da Tabela 1 é possível notar (visualmente) que diversos destes coeficientes são maiores que 0,3 (HAIR et al., 2009). Portanto, oferecendo indícios iniciais de ajuste das variáveis para aplicação da Análise Fatorial Binária.

Além desta inspeção visual existem estatísticas e testes que permitem melhor avaliar o quão adequado é a base de dados para aplicação da Análise Fatorial, tais como o Teste de Esfericidade de Bartlett ou a estatística KMO (*Kaiser-Meyer-Olkin*) (HAIR et al., 2009; FÁVERO et al., 2009). Portanto, para aplicação da AFB torna-se necessário a execução de tais testes previamente. Conforme descrito anteriormente, a “abordagem de variáveis subjacentes”, utiliza-se da matriz de correlação tetracórica (Tabela 1) como entrada para aplicação da Análise Fatorial. Portanto, evitando que o algoritmo da Análise Fatorial calcule a matriz de correlação, que é baseada na *correlação de Pearson*, específica para representar a associação entre dados quantitativos, mas inadequada para mensurar correlação entre dados binários.

Para analisar a adequação dos dados exemplificados à Análise Fatorial, calculou-se a estatística KMO, cujo resultado foi igual a 0,90. Hair et al. (2009) explicitam que os valores de KMO variam de 0 a 1, de tal forma que quanto mais próximo de 1, maior será a adequação



da amostra à técnica de Análise Fatorial. Logo, o valor da estatística KMO confirma a adequação dos dados analisados à técnica supracitada.

Em relação ao terceiro passo na aplicação da AFB, utilizou-se da Análise de Componentes Principais (ACP) como método de extração dos fatores. A referida abordagem de extração de variáveis latentes (fatores) considera a variância total dos dados. Segundo Fávero et al. (2009, p. 243) a ACP busca maximizar a variância total explicada, por meio da combinação linear das variáveis observadas. Assim, escolheu-se a ACP como método de extração em virtude do objetivo de redução dimensional dos dados, no intuito de se obter um agrupamento de variáveis que representem os diferentes tipos de serviços de governo eletrônico. Para retenção de fatores, foi utilizado como base o critério *a priori*, no qual se estipula a quantidade de fatores que devem ser retidos, independente do valor do *eigenvalue* e percentual de variância total explicada pelos fatores latentes (c.f. ARAUJO, 2013). Os resultados da aplicação da Análise Fatorial Binária (neste exemplo) estão demonstrados na Tabela 2.

**Tabela 2 – Resultados da Análise Fatorial (2011)**

	1	2	3	4	5	Comunalidade
Buscar Informações sobre como emitir documentos .	0,8258					0,8433
Obter certidões negativas, licenças e permissões.	0,7687					0,8297
Consultar o CPF – Cadastro de Pessoa Física.	0,7559					0,7529
Emissão de Documentos	0,756					0,8273
Obter informações sobre impostos e taxas em sites de governo		0,6996				0,7987
Fazer Declaração de Imposto de Renda		0,7866				0,822
Fazer Pagamento impostos, multas e taxas.		0,8124				0,8096
Buscar informações sobre previdência social e benefícios sociais			0,8749			0,9243
Solicitar serviços junto à previdência social.			0,8617			0,893
Fazer boletim de ocorrência				0,7322		0,6966
Consultar andamento de atos processuais na justiça				0,5719		0,6356
Buscar informações sobre direito do consumidor				0,4854		0,6938
Buscar informações sobre veículos roubados				0,7829		0,8031
Consultar pontos na carteira de habilitação e multas				0,6387		0,7076
Buscar informações sobre serviços públicos de educação					0,7674	0,7262
Fazer matrículas em escolas públicas ou instituições de ensino público					0,6359	0,5806
Buscar Informações sobre direitos do trabalhador					0,524	0,6477
Fazer Inscrição em concursos públicos					0,8053	0,7582
Buscar informações sobre empregos em algum canal de comunicação do governo pela Internet					0,7715	0,7027
<i>Variância Explicada do fator</i>	0,1765	0,1631	0,1447	0,1395	0,1368	
<b>Método de Extração:</b> Principal Components (PCA)						
<b>Rotação:</b> Varimax.						
<b>Variância Total Explicada:</b> 76,07%						
<b>Total de observações:</b> 8661.						

**Fonte:** Araujo (2013)

A partir dos dados da Tabela 2 foi possível evidenciar a formação de cinco fatores que representam os diferentes serviços de governo eletrônico. Esses cinco fatores juntos conseguem explicar 76% da variância original dos dados, portanto a representação das dezenove variáveis por meio dos referidos cinco fatores latentes implica em uma pequena perda de variância de 24%.

Para caracterização das variáveis que compõem os fatores extraídos, foi realizada uma análise das cargas fatoriais (*loadings*), após a aplicação do método de rotação *varimax*. A carga fatorial é uma medida da correlação entre os fatores extraídos e a variável analisada. Portanto, tal medida possibilita explicar qual fator melhor explica a variável original (FÁVERO et al., 2009, p. 256). Logo, uma variável analisada pertencerá ao fator com a qual ela tiver maior

valor de carga fatorial. A fim de facilitar a visualização dos resultados optou-se por apresentar apenas as cargas das variáveis referentes ao fator ao qual ela pertence.

A partir destes resultados apresentados na Tabela 2, foi possível caracterizar cada um dos fatores, identificando as variáveis que os compõem, permitindo assim nomear cada um dos fatores extraídos, conforme evidenciado no Quadro 2.

<b>Fator 1 – Atividades relacionadas à Documentação</b>
Buscar Informações sobre como emitir documentos.
Obter certidões negativas, licenças e permissões.
Consultar o CPF – Cadastro de Pessoa Física.
Emissão de Documentos
<b>Fator 2 – Atividades relacionadas ao Pagamento de Impostos e Taxas.</b>
Obter informações sobre impostos e taxas em sites de governo
Fazer Declaração de Imposto de Renda
Fazer Pagamento impostos, multas e taxas.
<b>Fator 3 – Atividades relacionadas à Assistência Social</b>
Buscar informações sobre previdência social e benefícios sociais
Solicitar serviços junto à previdência social.
<b>Fator 4 – Atividades relacionadas à Justiça e Trânsito.</b>
Fazer boletim de ocorrência
Consultar andamento de atos processuais na justiça
Buscar informações sobre direito do consumidor
Buscar informações sobre veículos roubados
Consultar pontos na carteira de habilitação e multas
<b>Fator 5 – Atividades relacionadas à Educação e Trabalho.</b>
Buscar informações sobre serviços públicos de educação
Fazer matrículas em escolas públicas ou instituições de ensino público
Buscar Informações sobre direitos do trabalhador
Fazer Inscrição em concursos públicos
Buscar informações sobre empregos em algum canal de comunicação do governo pela Internet

**Quadro 2** – Composição e Nomeação dos Fatores (2011).

**Fonte:** Araujo (2013)

O Quadro 2 sintetiza os cinco fatores extraídos. O primeiro fator congrega as variáveis relacionadas aos serviços públicos para obtenção ou solicitação de documentos junto à administração pública, sendo, portanto denominado como “Atividades relacionadas à Documentação”. O fator seguinte representa as tarefas referentes ao pagamento de impostos e taxas. A terceira dimensão latente congrega as atividades relacionadas, tanto à previdência social, quanto a benefícios sociais. O quarto fator, intitulado “Atividades relacionadas à Justiça e Trânsito”, assim como o próprio nome sugere congrega o conjunto de tarefas referentes à (i) segurança e justiça (e.g. realizar boletim de ocorrência, verificar o andamento de processos na justiça) e (ii) trânsito. Em conclusão, o último construto representa as atividades de educação (serviços públicos de educação, matrículas em escolas públicas etc.) e trabalho (emprego e direitos).



Neste sentido, o exemplo extraído do trabalho Araujo (2013) demonstra a viabilidade da aplicação da Análise Fatorial para lidar com dados binários, portanto se caracterizando uma ferramenta útil para pesquisas que se utilizem de dados secundários na área de TIC.

### 3. A construção de um índice utilizando modelagem de equações estruturais

Esta seção abordará o uso de modelagem de equações estruturais (MEE) para cálculo de um índice, exemplificando a aplicação da MEE na criação do índice de uso das TIC nas organizações.

Tal índice foi proposto e construído por um dos autores desse trabalho como alternativa para analisar e comparar o uso das TIC nas pequenas e médias empresas brasileiras. Antes de adentrar na técnica e no exemplo, algumas definições precisam ser feitas para melhor entendimento do texto.

A primeira definição é para o termo “constructo” que, no âmbito desse artigo, é o mesmo que um conceito ou, ainda, uma variável latente, fator ou dimensão. Considera-se, então, que um constructo possa ser unidimensional (representado por apenas uma variável latente, de primeira ordem) ou multidimensional (representado por várias variáveis latentes ou por uma variável latente de ordem superior, quando indicadores são outras variáveis latentes).

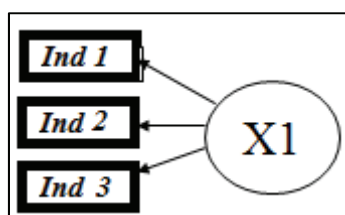
Variável latente é aquela que pode ser definida em termos teóricos e não pode ser observada diretamente, embora possa ser medida por meio de indicadores. (HAIR Jr. et al., 2010).

Os indicadores, também chamados de variável ou item, utilizados para medir, indiretamente, os constructos, serão aqueles considerados reflexivos, ou seja, manifestam variações quando de sua ocorrência na variável latente (HSU et al. 2006). Para um estudo mais aprofundado sobre indicadores reflexivos, bem como os indicadores formativos, consultar Jarvis et al. (2003).

Por meio da modelagem de equações estruturais (MEE) é possível estimar o modelo de mensuração (foco desse artigo) e também o modelo estrutural.

Para exemplificar, na figura abaixo estão presentes os indicadores reflexivos: Ind1, Ind2 e Ind3 que medem a variável latente de primeira ordem X1. A variável latente X1 é o constructo não observável. A figura abaixo também representa um modelo de mensuração, uma vez que relaciona o constructo medido e seus indicadores.

**Figura 1** – Representação de variável latente



**Fonte:** inspirado em Jarvis et al. (2006)

Enquanto no modelo estrutural a preocupação são as correlações entre o constructo e seus indicadores utilizando para tanto a análise fatorial confirmatória, no modelo estrutural busca-se compreender as relações (causais) entre os constructos e para tanto utiliza a técnicas de regressão como dos mínimos quadrados parciais (PLS) e modelagem de caminhos (PM).

Observa-se, assim, que a MEE, como destacou Hair Jr. et al. (2010), é uma técnica de análise multivariada que combina aspectos da regressão linear múltipla e da análise de fatores comuns.

Há ainda que se considerar diferenciação entre LISREL ou PLS, o primeiro roda os modelos utilizando covariância, enquanto o PLS utiliza variância. No âmbito dessa pesquisa, que preocupa-se apenas com o modelo de mensuração, há ainda uma diferença no que tange a técnica de análise fatorial, no PLS os constructos são combinações lineares de seus indicadores, sendo possível estimar os escores fatoriais, que aqui serão utilizados para compor o índice. Para explorar mais as diferenças entre Lisrel e PLS recomenda-se o artigo de Ringle Silva e Bido (2014) sobre a MEE utilizando do software SmartPLS.

Como desvantagens ao uso do PLS, conforme lembrado por Starkweather (2014), destaca-se a questão do viés, que consiste no aumento das correlações locais mais do que as globais, entretanto, a medida que a amostra cresce, diminui o peso do viés.

Entende-se que o pesquisador, quando faz uso da MEE, já conheça a dimensionalidade do constructo que está sendo medido e quais são os indicadores que o mensuram, buscando nesse momento garantir relevância maior da teoria em detrimento dos dados.

A análise fatorial confirmatória (AFC), no contexto da MEE, pode ser utilizada para calcular os scores fatoriais, permitindo analisar a validade discriminante e convergente do modelo de mensuração.

A confiabilidade composta é outra validação do modelo de mensuração, ela pode ser mais interessante do que a medida de Alpha de Cronbach por considerar que cada um dos itens pode ter diferentes importâncias na formação do fator (BARCLAY et al., 1995; FORNELL; LARCKER, 1981).

Vale ressaltar que, objetivando apenas a criação o índice, a aplicação da MEE será utilizada apenas para estimar o modelo de mensuração, uma vez não se busca estudar relações de causa e efeito. Em consonância aos objetivos deste artigo, as seções 4.1 e 4.2 demonstram e discutem, respectivamente, a aplicação da técnica de análise fatorial binária exploratória (AFB) e a da modelagem de equações estruturais (MEE).

#### **4. A construção do índice de uso das TICs**

A construção do índice foi feita utilizando os microdados da pesquisa TIC Empresas (um dos diversos *Surveys* coordenados pelo Centro de Estudos sobre as Tecnologias da Informação e Comunicação) e passou por duas etapas: a) primeiro a análise fatorial binária e, posteriormente b) análise fatorial confirmatória, no âmbito da Modelagem de equações estruturais usando PLS-PM.

A Modelagem de Equações Estruturais é de acordo com o Instituto de Pesquisa e Educação Digital da Universidade da Califórnia (IDRE), a forma mais geral de análise fatorial confirmatória. (IDRE, 2014).

#### 4.1 Aplicação da AFB

Para construir o índice de uso das TIC foi utilizada a técnica de AFB (já descrita na seção 2) com o intuito de buscar os fatores comuns e verificar, primeiramente, se os indicadores selecionados, a partir de uma base de dados secundários, TIC Empresas, tinham validade, uma vez que os indicadores selecionados não foram construídos a partir da teoria.

Esta etapa é o primeiro passo e de suma importância, pois a AFB, de caráter exploratório, vai permitir encontrar fatores comuns, variáveis latentes, bem como seus indicadores, dando maior peso aos dados, neste primeiro momento, em detrimento à teoria.

Com base no estudo que define o uso das TIC nas organizações, desenvolvido por Zwicker, Souza e Vidal (2006), poder-se-ia ficar em dúvida com respeito à utilização da AFB, uma vez que já se conhece a priori, a partir da teoria, os constructos e as variáveis que os medem. Advoga-se assim que, tal conhecimento prévio à cerca da teoria, permitiria a aplicação direta da MEE, sem o uso da análise fatorial binária exploratória. Contudo, visto que o exemplo discutido nesta seção se baseia em dados secundários e isto significa que, as variáveis medidas na pesquisa TIC Empresas não foram pensadas, construídas e validadas a partir da teoria, optou-se pela aplicação da AFB como uma etapa auxiliar na identificação da dimensionalidade dos construtos.

De acordo com Hatcher (1994) a combinação da análise fatorial exploratória seguida da confirmatória é aplicável quando se conhece a teoria subjacente, entretanto, não se conhece a quantidade de fatores a partir das variáveis medidas.

Para Netemeyer et al. (2003), a análise fatorial exploratória pode ser uma forma de entender o constructo e seus fatores comuns ou ainda de purificar a escala, buscando encontrar aqueles itens com baixa carga fatorial para retirá-los do modelo.

A Tabela 3 contém o resultado da análise fatorial binária exploratória, aplicada sobre os dados secundários, após a rotação *varimax* e tendo como entrada, ao invés das variáveis, a matriz de correlação tetracórica (conforme explicado na seção 2). Assim, foi possível verificar a existência de três constructos, que serão, posteriormente, confirmados na MEE.

**Tabela 3** – Resultado da aplicação da análise fatorial

Questões da Pesquisa TIC Empresa	Variável	Fator 1	Fator 2	Fator 3
Utilizou ERP para integrar dados e processos	A5	0,71		
Utilizou CRM para gerenciar informações de clientes	A6	0,75		
Treinamento e educação	B7_B	0,65		
Recrutar pessoal interno e externo	B7_L	0,65		
Monitoramento Mercado	B7_C		0,83	
Buscar informações sobre produtos ou serviços	B7_E		0,93	
Busca de informação governamental	B7_F		0,57	
Sistema de pedidos ou reserva	B9_B			0,77
Fornecer suporte pós-venda	B9_D			0,72
Personalização ou customização de produtos para clientes	B9_F			0,78

**Fonte:** Siqueira et al. (2014)

Desta forma, observando as cargas fatoriais, os fatores foram assim nomeados:

- Fator 1 – Integração Interna;
- Fator 2 – Suporte à Decisão;
- Fator 3 – Integração Externa;

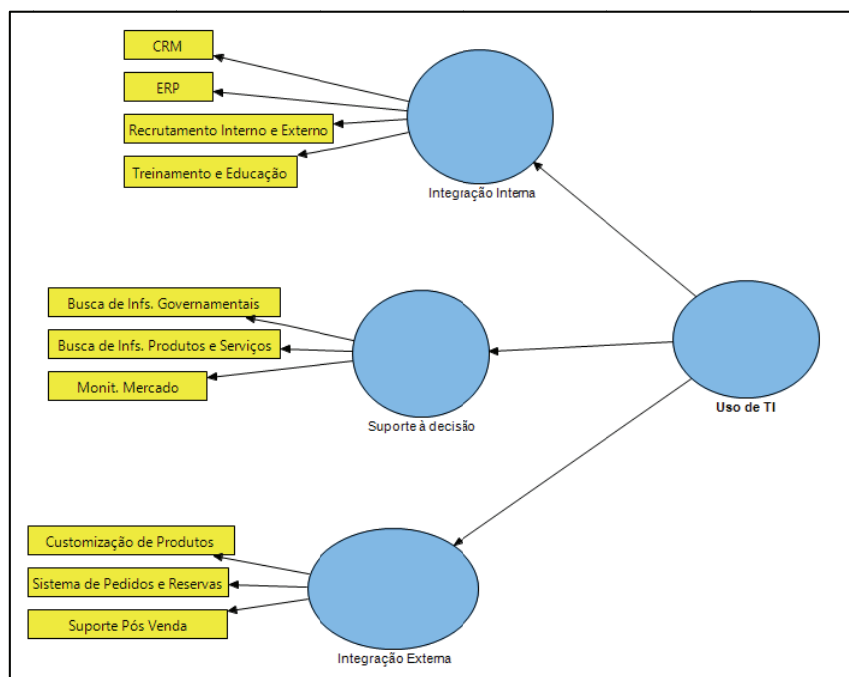
Os três fatores identificados foram denominados dimensões do uso das TIC, ou ainda variáveis latentes de primeira ordem. As variáveis latentes são aquelas não são observáveis diretamente, mas são mensuráveis via indicadores (HAIR Jr. et al., 2010).

#### 4.2 Aplicação da MEE

O constructo “Uso das TIC” é uma variável latente de segunda ordem, uma vez que é formado por outras variáveis latentes de primeira ordem (Integração Interna, Suporte à Decisão e Integração externa).

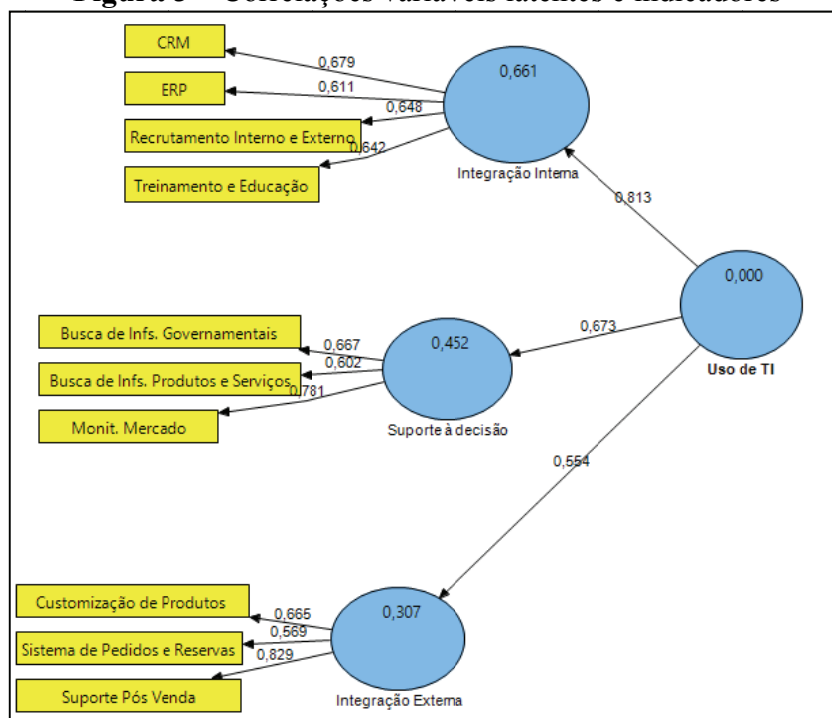
Trata-se, portanto, de um modelo hierárquico contendo variáveis observadas, variáveis latentes de primeira ordem e variável latente de segunda ordem.

O modelo pode ser conferido na Figura 2, desenvolvido no Software *SmartPLS*, que é específico para aplicação de MEE com a técnica PLS-PM. No modelo, os indicadores observáveis são indicados por retângulos e as variáveis latentes são indicadas por círculos.

**Figura 2 – Modelo Hierárquico para mensuração do Uso das TICs.**

Fonte: Siqueira et al. (2014)

Já na Figura 3 é possível analisar o modelo hierárquico para cálculo do uso das TIC no *SmartPLS*, bem como os scores fatoriais calculados utilizando análise fatorial confirmatória no âmbito da MEE:

**Figura 3 – Correlações variáveis latentes e indicadores**

Fonte: Siqueira et al. (2014)

## 5. Validações do modelo de mensuração

O modelo para mensuração do uso das TIC foi assim validado:

- Validade convergente: cargas fatoriais apresentaram valores iguais ou superiores a 0,5 em seus constructos.
- Validade discriminante: observou-se baixas as correlações entre as variáveis latentes ou constructos.
- Confiabilidade composta (*composite reliability*) que foi superior a 0,7 indicando adequação dos indicadores.

Desta forma, por meio do *SmartPLS* foi possível atribuir um *score* para as três dimensões, bem como para o Uso das TIC, em cada uma das observações do banco de dados, conforme a Tabela 4, que contém 10 observações selecionadas da amostra apenas a título de exemplo:

**Tabela 4** – Observações contendo *scores* para as dimensões e Uso das TIC

Identificador	Integração Externa	Integração Interna	Suporte à Decisão	Uso das TIC
2	0,31	0,25	0,24	0,27
4	0,37	1,00	1,00	0,87
6	0,00	0,00	0,49	0,16
8	0,00	1,00	0,73	0,66
9	0,69	0,51	1,00	0,73
13	0,00	0,25	0,49	0,26
15	0,00	0,00	0,00	0,00
22	0,00	0,00	1,00	0,37
24	0,00	0,27	0,76	0,38
28	0,68	0,51	0,73	0,63

Fonte: Siqueira et al. (2014)

## 6. MEE e os dados binários

Há uma série de estudos sobre MEE e dados binários, verificando que é possível rodar análise fatorial confirmatória utilizando como entrada a matriz de correlação tetracórica, e como método de estimação máxima verossimilhança (ML) ou mínimos quadrados ponderados (WLS), entre outros. Entretanto, esses estudos são, com frequência, abordagem de modelagens de equações estruturais baseados em covariância, LISREL, nos quais o dado binário pode interferir na covariância e viola o pressuposto de normalidade, requerido por essa técnica. Mais sobre esse assunto ver os estudos de Muthén (1984), Luo (2011) ou a pesquisa de Yang-Wallentin, Jöreskog e Luo (2010) que explora os vários métodos de estimação com variáveis categóricas e binárias.

No presente artigo, utilizou-se MEE-PLS, baseada em variância, para a qual não há o pressuposto de normalidade das variáveis. O modelo de mensuração obteve as validades esperadas e por isso acredita-se que a técnica possa ser utilizada para dados binários. De acordo com [1], a MEE-PLS pode minimizar os vieses de dados binários e categóricos.



## 7. Conclusão

O desenvolvimento de pesquisas quantitativas na área de Tecnologia da Informação e Comunicação (TIC) pode-se tornar um grande desafio devido à necessidade de atender a uma série rigorosas premissas inerentes das técnicas estatísticas (normalidade, homocedasticidade dos resíduos, dados métricos etc.) e/ou a construção de grandes amostras (problemas de amostragem). Neste sentido, dados secundários podem ser fontes ricas de informação no desenvolvimento de investigações científicas por pesquisadores da área de TIC.

Tradicionalmente as principais técnicas estatísticas multivariadas têm como premissa o uso de dados de natureza métrica. Porém, devido à variedade de fenômenos explorados no supracitado campo de pesquisa, muitas vezes torna-se necessário lidar com fenômenos expressos de maneira dicotômica (e.g. ausência/presença). Consequentemente, reduzindo a possibilidade de aplicação das técnicas estatísticas multivariadas para este tipo de dados, sendo este o cerne da discussão realizada neste trabalho.

Neste sentido, o presente artigo discutiu algumas alternativas para aplicação de técnicas estatísticas multivariadas em dados binários. Perante a variedade de aplicações possíveis, optou-se por discutir e demonstrar o uso da análise fatorial exploratória, como ferramenta para redução dimensional de variáveis. Conforme discutido na Seção 2, o ferramental da análise fatorial binária permitiu identificar os cinco fatores latentes, que representam os diferentes tipos de serviços de governo eletrônico explorados pela pesquisa TIC Domicílios (fonte dos dados secundários).

Além deste objetivo, nas seções 4, 5 e 6 demonstrou-se o passo-a-passo para aplicação da técnica de modelagem de equações estruturais (utilizando o *Partial Least Square*) em dados binários, a fim de criar um índice referente ao uso das Tecnologias da Informação nas empresas. Novamente, os dados utilizados para aplicação da referida técnica foram extraídos do conjunto de pesquisas desenvolvidas pelo Centro de Estudos sobre as Tecnologias da Informação e Comunicação (Cetic.br), mais especificamente, a pesquisa TIC Empresas.

Por fim, espera-se que a discussão empreendida neste artigo tenha lançado luz a questão da análise estatística multivariada para dados de natureza binária, demonstrando a utilidade e o potencial da aplicação das técnicas de Análise Fatorial Binária e Modelagem de Equações Estruturais (com dados binários) no contexto da área de TIC. Recomenda-se que estudos futuros se aprofundem na referida discussão, demonstrando a variabilidade e adequabilidade de técnicas estatísticas para análise de fenômenos expressos por escala binária, que são comuns na área de Tecnologia da Informação e Comunicação.

## Referências

ARAUJO, M. H. **Análise de Fatores que Influenciam o Uso de Serviços de Governo Eletrônico no Brasil**. Dissertação (Mestrado em Administração) – Faculdade de Economia, Administração e Contabilidade. Universidade de São Paulo, São Paulo, 2013

BARCLAY, D.; HIGGINS, C.; THOMPSON, R. The partial least squares (PLS) approach to causal modelling: personal computer adoption and uses as an illustration. **Technology Studies**, v. 2, p. 285-323, 1995.

BARTHOLOMEW, D.; STEELE, F.; MOUSTAKI, I.; GALBRAITH, J. **The Analysis and Interpretation of Multivariate Data for Social Scientists**. 1st edition. London: Chapman and Hall/CRC Press, 2002.

CGI. **Pesquisa sobre o uso das tecnologias da informação e da comunicação no Brasil 2011: TIC domicílios e TIC empresas**. São Paulo: Comitê Gestor da Internet no Brasil, 2012.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F.L.; CHAN, B.L. **Dados: Modelagem Multivariada para Tomada de Decisões**. Rio de Janeiro: Elsevier, 2009

FORNELL, C. e LARCKER, D. F. Evaluating structural equation models with unobservable variables and measurement error. **Journal of Marketing Research**. v.18, n.1, p. 39-50, 1981.

HAIR Jr., J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate Data Analysis**. 7<sup>th</sup> edition. NJ: Prentice Hall, 2010.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise Multivariada de dados**. 6.ed. Porto Alegre: Artmed, 2009

HATCHER, L. **A step-by-step approach to using the SAS ® System for factor analysis and structural equation modeling**. Cary, NC: SAS Institute Inc. 1994.

HSU, S.-H.; CHEN, W.-H.; HSIEH, M.-J. Robustness testing of PLS, LISREL, EQS and ANN-based SEM for measuring customer satisfaction. **TQM & Business Excellence**, v.17, n.3, pp.355–371., 2006.

IDRE. Statistical Computing Seminars Introduction to Mplus: Featuring Confirmatory Factor Analysis. Disponível em  
<[http://www.ats.ucla.edu/stat/mplus/seminars/IntroMplus\\_CFA/](http://www.ats.ucla.edu/stat/mplus/seminars/IntroMplus_CFA/)> Acessado em 10/10/2014.

JARVIS, C. B.; MACKENZIE, S. B.; PODSAKOFF, P. M. A critical review of construct indicators and measurement model misspecification in Marketing and Consumer Research. **Journal of Consumer Research**, v.30, n. 2, p.199-218, 2003.

LUO, HAO 2011. Some Aspects on Confirmatory Factor Analysis of Ordinal Variables and Generating Non-Normal Data. Acta Universitatis Upsaliensis. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences. Uppsala, 2011

MUTHÉN, B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. **Psychometrika**, v. 49, p. 115-132, 1984

MINTU-WIMSATT, A.; GRAHAM, J. L. Testing a negotiation model on Canadian anglophone and Mexican exporters. **Journal of the Academy of Marketing Science**, v.2, n.3, p. 345-356, 2004.

NETEMEYER, R. G.; BEARDEN, W. O.; SHARMA, S. **Scaling procedures: issues and applications**. Thousand Oaks: Sage Publications, 2003.

RINGLE, C. M.; SILVA, D.; BIDO, D. Modelagem de Equações Estruturais com Utilização do SmartPLS. **Revista Brasileira de Marketing**, v. 13, n. 2, p. 54-71, 2014.

SIQUEIRA, Érica Souza; SOUZA, Cesar Alexandre ; CHAGAS, Vladimir . Uso das TIC Nas Pequenas e Médias Empresas Brasileiras: Construção de Um Índice de Uso Utilizando Dados da Pesquisa TIC Empresas e Modelagem de Equações Estruturais. In: Seminários em Administração - SemeAd, 2014, São Paulo. XVII - SemeAd, 2014.

SHENG-HSUN HSU\_, WUN-HWA CHEN\_\_ & MING-JYH HSIEH\_\_\_ Robustness Testing of PLS, LISREL,EQS and ANN-based SEM for Measuring Customer satisfaction. Total Quality Management Vol. 17, No. 3, 355–371, April 2006

STARKWEATHER, J. Linear Mixed Effects Modeling using R. Disponível em <[http://www.unt.edu/rss/class/Jon/Benchmarks/PLS\\_JDS\\_July2011.pdf](http://www.unt.edu/rss/class/Jon/Benchmarks/PLS_JDS_July2011.pdf)> Acessado em 17/11/2014.

Yang-Wallentin, Fan; Jöreskog Karl G.; Luo Hao. Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models. Structural Equation Modeling: A Multidisciplinary Journal Vol. 17, Iss. 3, 2010.

**UCLA. How can I perform a factor analysis with categorical (or categorical and continuous) variables?** Institute for Digital Research and Education (idre). Disponível em: <[http://www.ats.ucla.edu/stat/stata/faq/efa\\_categorical.htm](http://www.ats.ucla.edu/stat/stata/faq/efa_categorical.htm)>. Acesso em Dez. 2013

ZWICKER, R.; SOUZA, Cesar Alexandre; Vidal, Antonio. An Assessment of the Informatization Level of Brazilian Industrial Companies. Eletronic Journal of Information Systems in Developing Countries. V. 23, n 1. p.1-25, 2006.