Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística - IBGE

REVISTA BRASILEIRA DE ESTATÍSTICA

volume 72 número 235 janeiro/dezembro 2011

ISSN 0034-7175

R. bras. Estat., Rio de Janeiro, v. 72, n. 235, p. 1-141, jan./dez. 2011

Instituto Brasileiro de Geografia e Estatística - IBGE
Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro - RJ - Brasil

© IBGE. 2012

Revista Brasileira de Estatística, ISSN 0034-7175

Órgão oficial do IBGE e da Associação Brasileira de Estatística - ABE.

Publicação semestral que se destina a promover e ampliar o uso de métodos estatísticos através de divulgação de artigos inéditos tratando de aplicações da Estatística nas mais diversas áreas do conhecimento. Temas abordando aspectos do desenvolvimento metodológico serão aceitos, desde que relevantes para a produção e uso de estatísticas públicas.
Os originais para publicação deverão ser submetidos para o site
http://rbes.submitcentral.com.br/login.php
Os artigos submetidos às RBEs não devem ter sido publicados ou estar sendo considerados para publicação em outros periódicos.

A Revista não se responsabiliza pelos conceitos emitidos em matéria assinada.

Editor Responsável
Francisco Louzada (USP - São Carlos)

Editor-Executivo
Pedro Luis do Nascimento Silva (ENCE/IBGE)

Editor de Metodologias
Fernando Moura (UFRJ)

Editor de Estatísticas Oficiais
Denise Britz do Nascimento Silva (DPE/IBGE)

Editores Associados
Dalton Francisco de Andrade (UFSC)
José André de Moura Brito (ENCE/IBGE)
Viviana Giampaoli (IME-USP)
Beatriz Vaz de Melo Mendes (UFRJ)
Thelma Sáfadi (UFLA)
Paulo Justiniano Ribeiro Junior (UFP)
Josmar Mazucheli (UEM)
Luis A Milan (UFSCar)
Cristiano Ferraz (UFPE)
Gleici Castro Perdoná (FMRP-USP)
Ana Maria Nogales Vasconcelos (UNB)
Ronaldo Dias (UNICAMP)
Mário de Castro (ICMC-USP)
Nuno Duarte Bittencourt (ENCE/IBGE)
Solange Trindade Corrêa (DPE/IBGE)

Editoração
Marilene Pereira Piau Câmara - ENCE/IBGE

Impressão
Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2012.

Capa
Renato J. Aguiar - Coordenação de *Marketing*/CDDI/IBGE

Ilustração da Capa
Marcos Balster - Coordenação de *Marketing*/CDDI/IBGE

Revista brasileira de estatística / IBGE, - v.1, n.1 (jan./mar.1940), - Rio de Janeiro : IBGE, 1940. v.

Trimestral (1940-1986), semestral (1987-).
Continuação de: Revista de economia e estatística.
Índices acumulados de autor e assunto publicados no v.43 (1940-1979) e v. 50 (1980-1989).
Co-edição com a Associação Brasileira de Estatística a partir do v.58.
ISSN 0034-7175 = Revista brasileira de estatística.

I. Estatística – Periódicos. I. IBGE. II. Associação Brasileira de Estatística.

Gerência de Biblioteca e Acervos Especiais CDU 31(05)
RJ-IBGE/88-05 (rev.2009) PERIÓDICO

Impresso no Brasil/*Printed in Brazil*

Sumário

Nota do editor 5

Artigos

Classificação de modelos de resposta ao item policotômicos com aplicação ao marketing..... 7

Jorge Luis Bazán
José Afonso Mazzon
Martin Hernani Merino

Uma adaptação do teste de hipótese de Hayter e Tsui para situações de comparações de medidas de vários tratamentos com um grupo controle..... 41

Carolina Silva Pena
Sueli Aparecida Mingoti

Modelagem longitudinal hierárquica: uma aplicação aos dados da PME..... 75

Aline de Carvalho Veiga

O comportamento dos índices de ações em países emergentes: uma análise com dados em painel e modelos hierárquicos..... 97

Luiz Paulo Lopes Fávero
José Elias Feres de Almeida

Nota do Editor

Este volume da RBEs do ano de 2011 é composto por quatro artigos. O primeiro artigo, de autoria de Jorge Luis Bazán, José Afonso Mazzon e Martín Hernani Merino, apresenta procedimentos de classificação de modelos de resposta ao ítem com aplicação ao marketing. O segundo artigo, de autoria de Sueli Aparecida Mingoti, apresenta uma adaptação do teste de hipótese de Hayter e Tsui para situações de comparações de médias de vários tratamentos com um grupo controle. O terceiro artigo, de autoria de Alinne de Carvalho Veiga, apresenta uma aplicação de modelo longitudinal hierárquico aos dados da Pesquisa Mensal de Emprego. O quarto artigo, de autoria de Luiz Paulo Lopes Fávero e Jose Elias Feres de Almeida, analisa o comportamento dos índices de ações em países emergentes através de uma análise de dados em painel e modelos hierárquicos.

Informo que este é o último número da RBEs em que atuo como Editor Responsável, posição esta que será ocupada pela Profa. Lúcia Pereira Barroso do IME-USP, a quem, juntamente com todo o novo Corpo Editorial, apresento os meus votos de estima e de um excelente mandato junto à RBEs.

Aproveito a oportunidade para agradecer a Pedro Luis do Nascimento Silva (Editor Executivo), todos os Editores Associados, revisores do periódico, autores, IBGE e ABE, que durante esses oito anos colaboraram com a RBEs. A todos o meu muito obrigado, foi um grande prazer servir como Editor Responsável pela RBEs.

Uma excelente leitura.

Francisco Louzada
Editor Responsável

Classificação de modelos de resposta ao item policotômicos com aplicação ao marketing

Jorge Luis Bazán[1]
José Afonso Mazzon[2]
Martín Hernani Merino[3]

Resumo

Têm sido comum trabalhos publicados em Administração e particularmente em Marketing utilizarem técnicas estatísticas aplicadas a escalas ordinais assumindo que estas fossem intervalares. Com o propósito de apresentar uma alternativa adequada a essa prática, apresenta-se, neste artigo, uma taxonomia de diversos modelos de teoria de resposta ao item policotômicos (TRIP) baseada nos tipos de logitos de passo de uma etapa do item (funções de resposta). A ênfase da apresentação é introdutória e a metodologia proposta é ilustrada com uma aplicação considerando uma escala de atitudes de consumidores de serviços bancários frente ao mobile banking, uma área de recente interesse em Marketing. Esse exemplo ilustra também o uso de software para implementar a maioria destes modelos.

Palavras-chave: Teoria de Resposta ao Item; Itens Policotômicos; Marketing.

[1] *Professor da Pontificia Universidad Católica del Peru Correspondência: jlbazan@pucp.edu.pe*
[2] *Professor da Faculdade de Economia, Administração e Contabilidade da USPCorrespondência: jamazzon@usp.br*
[3] *Doutorando na Faculdade de Economia, Administração e Contabilidade da USP Correspondência: martinhernani@usp.br*

1. Introdução

O principal método de pesquisa empregado em marketing é o denominado survey, o qual se baseia na aplicação de questionários junto a uma amostra de consumidores de bens e serviços (Franses & Paap, 2001; Malhotra, 2011). Em termos de pesquisa de marketing, a maioria dos questionários aplicados a estudos do comportamento do consumidor em relação a produtos, preços, canais de distribuição, promoção e venda pessoal, satisfação, lealdade a marcas, dentre outros aspectos, usam predominantemente escalas de respostas denominadas policotômicas, no sentido de possibilitar ao consumidor um maior leque de categorias de respostas. É o caso, por exemplo, de escalas do tipo Likert variando de nenhuma importância a total concordância, nenhuma diferenciação a total diferenciação ou de total discordância a total concordância. O uso deste tipo de escalas policotômicas tem trazido contribuições extremamente úteis e são importantes, segundo reconhecem Singh, Howell e Rhoads (1990). Os motivos apontados pelos autores são: 1) muitas das escalas de atitudes em marketing utilizam categorias de resposta do tipo Likert; 2) para um conjunto fixo de itens, o uso de dados obtidos por meio dessa escala é conhecido por produzir mais informação sobre o nível de atitude de um respondente do que seria por dados dicotômicos.

Em geral os dados obtidos mediante estes instrumentos de coleta são analisados usando procedimentos de testes estatísticos clássicos (Pasquali, 2009). Questionamentos também são feitos ao uso de técnicas estatísticas aplicadas a escalas assumidas como intervalares, mas que na realidade são ordinais (Hair, Black, Babin, Anderson & Tatham, 2006; Mazzon, 1981). Uma das formas de superar tal problema reside na utilização de modelos da teoria de resposta ao item (TRI), os quais têm sido usados mais recentemente de forma crescente (Baker & Kim, 2004; Rao & Sinharay, 2006), principalmente nas áreas de educação (Andrade, Tavares & Valle, 2000), saúde (Cook, Taylor, Dodd, Teal & Mchorney, 2007), engenharia (Omar, Ghariebeh, Salazar & Saito, 2007) e psicologia (Tay, Diener, Drasgow & Vermunt, 2011).

As aplicações da TRI em marketing são ainda incipientes, tanto em termos acadêmicos quanto profissionais. Destaca-se aqui o trabalho pioneiro de Balasubramanian e Kamakura (1989), que aplicaram com sucesso, dentro de um contexto de pesquisa de comportamento do consumidor, modelos da TRI a dados dicotômicos (Ex: Sim/Não) e a dados policotômicos de uma escala tipo Likert de 7 pontos, dicotomizada para uma escala de 2 pontos. No ano seguinte, Singh, Howell e Rhoads (1990) apresentaram a TRI como um método alternativo para o tratamento de escalas do tipo Likert na implementação de pesquisas de marketing. Singh (2004) discutiu os princípios, as características e efetuou uma avaliação crítica do uso da TRI em marketing. Destaca-se ainda o trabalho realizado por De Jong, Steenkamp e Fox (2007), os quais aplicaram TRI hierárquico a uma pesquisa nacional de consumidores. Por último, Tay, Diner, Drasgow e Vermunt (2011) efetuaram uma aplicação de modelo TRI multinível para a mensuração de uma escala de emoções.

Mais recentemente, diante de um crescente interesse pela pesquisa do consumidor para testar medidas e teorias no contexto internacional em uma situação de (in)variância de mensurações para comparar nações, mercados e segmentos de consumidores, esse tema tornou-se de enorme importância (De Jong, Steenkamp & Fox, 2007). Esses autores trataram do assunto montando e testando um modelo hierárquico da TRI com o qual comparam países, apesar da falta de invariância de qualquer um dos itens. Na pesquisa, também se reconheceu a natureza ordinal dos dados através das diferenças das nações na escala de medida. Esses autores encontraram uma forte não invariância das métricas da escala e seu uso através dos países para o construto utilizado na pesquisa: suscetibilidade do consumidor para a influência normativa nas decisões de compra. Face à situação de elevado uso nas pesquisas de marketing de escalas de mensuração do tipo Likert, utilizando várias categorias de resposta – usualmente de 5 a 9 pontos – é que se apresentam a seguir alguns exemplos aplicados ao marketing para uma melhor visualização e compreensão de diferentes tipos de categorias de resposta.

Não obstante tais resultados, o uso de modelos TRI para dados policotômicos ainda se encontra quase que ausente na literatura de marketing, embora reconhecidos pesquisadores atestem sua contribuição na mensuração de variáveis latentes (Balasubramanian & Kamakura, 1989; Bechtel, 1985; Singh, 2004; Singh, Rowel e

Rhoads, 1990; Tay, Diner, Drasgow e Vermunt, 2011), sugerindo isso um potencial uso em pesquisas futuras.

Um tema recentemente estudado relaciona-se com atitudes e comportamentos de consumidores brasileiros em relação a um novo serviço financeiro: o mobile banking. Avaliar corretamente o que os consumidores pensam e como se comportam em relação a este serviço é fundamental para uma instituição financeira, pois além de adicionar valor ao cliente, o uso dessa ferramenta reduz significativamente o custo de uma transação, podendo assim gerar benefícios tanto aos clientes quanto aos bancos. Pousttchi e Schuring (2004) explicam que este conceito é uma parte do *e-commerce* e é uma forma de execução de serviços financeiros de forma que – dentro de um procedimento eletrônico sistematizado – o cliente utiliza técnicas de comunicação móvel conjuntamente com aparelhos celulares. Tezza, Bórnia e Andrade (2011) efetuaram uma aplicação de TRI na área de comércio eletrônico. Puschel, Mazzon e Hernandez (2010) elaboraram e testaram um modelo no qual aplicaram diversas teorias da psicologia social, difusão da inovação e adoção de tecnologias para descrever os fatores que influenciam o processo de adoção de mobile banking. Estes autores, após uma exaustiva revisão de literatura, propuseram um modelo integrativo de adoção de mobile banking composto por 14 fatores (variáveis latentes) com seus respectivos indicadores (61 itens) para atingir o propósito esboçado. Esses fatores e respectivos itens originam-se de diferentes teorias, a saber: Theory of Reasoned Action (TRA), Innovation Diffusion Theory (IDT), Technology Acceptance Model (TAM), Theory of Planned Behavior (TPB) e Decomposed Theory of Planned Behavior (DTPB) e são os seguintes: compatibilidade social; vantagem relativa; visibilidade; demonstração de resultados; imagem da instituição financeira; testabilidade do serviço, utilidade percebida do serviço; atitude em relação ao mobile banking; influência de amigos e colegas na decisão de compra; controle percebido sobre o comportamento; autoeficácia; condição facilitadora dos recursos disponíveis; condição facilitadora da tecnologia disponível e intenção de adotar o mobile banking em um dado período de tempo. Para mensurar cada um desses 61 itens, a pesquisa utilizou uma escala do tipo Likert aplicada na forma bi-etápica de concordância baseada em seis posições: discordo totalmente = 1; discordo muito = 2; discordo pouco = 3; concordo pouco = 4; concordo muito = 5 e concordo totalmente = 6.

No entanto, o uso de modelos da TRI para dados policotômicos não tem sido usado com este tipo de escala.

Com base nessas considerações, apresenta-se neste artigo uma caracterização e uma taxonomia dos diversos modelos da teoria de resposta ao item policotômicos (TRIP) baseadas nos tipos de logitos de passo (funções de resposta) de uma etapa do item assim como uma aplicação ao marketing de um desses modelos, especificamente a uma escala de mobile banking proposta por Puschel, Mazzon e Hernandez (2010). A natureza do trabalho é introdutória e semelhante a Andrade, Tavares e Valle (2000), estando focada na modelagem e apresentação de uma aplicação e não nos métodos de estimação. Assim, o propósito do artigo é contribuir para a discussão da escolha do modelo TRIP mais pertinente para uma particular aplicação.

A fundamentação do trabalho está apoiada nos artigos de Hemker (2001), Van der Ark (2001) e Vermunt (2001). A apresentação de alguns modelos da TRIP e de procedimentos de estimação com ênfase no processo de inferência estatística podem ser vistos em Azevedo (2003). Também podem ser consultados Van der Linden e Hambleton (1997) para modelos TRIP paramétricos e Boomsma, Van Duijn e Snijders (2001) para o caso de modelos TRIP não paramétricos.

O artigo está organizado em cinco seções. Na seção 2 é apresentada uma caracterização dos modelos TRIP introduzindo aspectos preliminares da modelagem da resposta policotômica. Na seção 3 apresenta-se uma taxonomia dos modelos TRIP, seus casos particulares e as relações entre eles. Na seção 4 é apresentada uma aplicação baseada no uso de um modelo TRIP a uma escala de avaliação de mobile banking proposta por Puschel, Mazzon e Hernandez (2010), ilustrando o uso de um software específico para implementar a maioria destes modelos. A aplicação corresponde ao uso do modelo de resposta graduada para o estudo de uma escala de atitudes frente ao uso de mobile banking. Finalmente, a seção 5 apresenta a discussão e implicações dos resultados do estudo efetuado.

2. Caracterização dos modelos TRIP

2.1. Possíveis formatos de itens policotômicos

Consideremos o interesse de um pesquisador em avaliar atitudes e comportamentos dos consumidores em relação ao serviço de *mobile banking*. Imaginemos ainda diferentes formatos para mensuração de itens de um questionário, como os apresentados na figura a seguir.

Exemplo 1. Item: Usar o *mobile banking* para fazer transferências bancárias é uma boa idéia. Categorias de resposta: a) Discorda muito b) Discorda pouco c) Nem concorda nem discorda d) Concorda pouco e) Concorda muito	Exemplo 2. Item: Escreva e explique três atributos que caracterizam o *mobile banking* para um consumidor. Categorias de resposta: a) Não escreveu nenhum b) Escreveu um atributo c) Escreveu dois atributos d) Escreveu três atributos
Exemplo 3. Item: Realizar uma transferência bancária usando *mobile banking*. Categorias de resposta: a) Não faz nada b) Entra na plataforma virtual mas não consegue realizar a transação c) Realiza a transação mas não consegue confirmar o resultado d) Realiza a transação corretamente	Exemplo 4. Item: Qualificação do uso pelo consumidor do *mobile banking*. Categorias de resposta: a) Insuficiente b) Regular c) Boa d) Excelente

Figura 1. Exemplos de itens policotômicos de pesquisa sobre mobile banking

Embora todos os itens sejam focados a indagar o uso de *mobile banking* entre as pessoas, eles têm aspectos em comum como também têm aspectos diferenciados. Por exemplo, todos os itens da figura 1 apresentam respostas em mais de duas categorias ordenadas embora esse ordenamento não seja da mesma forma. Assim, no item 1 e 4, as categorias são fixas e exaustivas, mas no item 1 as categorias são estabelecidas pelo

pesquisador (nos estudos variam, em geral, de 5 a 9 categorias). Por outro lado, nos itens 2, 3 e 4, as categorias correspondem a fases parciais da solução do problema estabelecido no item e não podem ser variadas pelo pesquisador.

Estes tipos de formatos de categorias de itens são frequentes em pesquisa de avaliação e são denominados na literatura como itens policotômicos. Embora estes itens possam ser dicotomizados, a preferência por itens policotômicos frente a itens dicotômicos está baseada, segundo Van der Ark (2001), nos seguintes aspectos: a) o fato de poucos itens policotômicos serem necessários para obter o mesmo grau de fidegnidade (confiabilidade); b) alguns traços serem mais fáceis de ser medidos em escalas graduais; c) para algumas variáveis, respostas aos itens são mais bem expressas em uma escala ordinal. No caso dos itens apresentados, é fácil ver as vantagens das respostas ordinais policotômicas ao invés de usar respostas dicotômicas. Nos itens policotômicos, pode-se pensar que passar de uma categoria a outra superior supõe uma etapa maior (assim, por exemplo, nos itens apresentados, passar da categoria "a" à categoria "b" significa completar a primeira etapa). Deste modo, pensar em categorias ou etapas é equivalente, sendo que se o número de etapas é m o número de categorias é $m+1$.

Analisando com maior cuidado as diferentes etapas dos itens policotômicos apresentados, nem sempre é possível pensar que as etapas são equivalentes ou igualmente espaçadas. Por exemplo, no item 3, não é claro que completar a etapa 1 (passar da categoria "a" à categoria "b") corresponde ao mesmo esforço que se precisa para completar a etapa 3 (passar da categoria "c" à "d"). Mas, no caso do item 2 isso é possível de sustentar. A suposição principal é que um processo cognitivo em etapas e ordenado é caracterizado através de um escore crescente, onde o escore da etapa é atingido se o avaliado completou satisfatoriamente até dita etapa e não falhou na etapa seguinte.

2.2. TRIP e logitos de passo

Os modelos da teoria de resposta ao item adotam dois axiomas fundamentais segundo a visão de Pasquali (2009): a) o desempenho do sujeito numa tarefa (item do teste) se explica em função de um conjunto de fatores ou traços latentes (aptidões,

habilidades, etc.). O desempenho é o efeito e os traços latentes são as causas; b) a relação entre o desempenho na tarefa e o conjunto dos traços latentes pode ser descrita por uma equação monotônica crescente, chamada de curva característica do item (CCI). Esse autor argumenta que na TRI o pesquisador apresenta ao examinado um estímulo ou uma série de estímulos (tais como, por exemplo, itens de um teste ou itens motivadores da escolha de uma marca de produto) e ele responde aos mesmos. A partir das respostas dadas pelo sujeito, isto é, analisando as suas respostas aos itens específicos, pode-se inferir sobre o traço latente do examinado, hipotetizando-se múltiplas relações entre as respostas observadas deste sujeito com o nível do seu traço latente. Estas relações podem ser expressas através de uma equação matemática que descreve a forma da função que estas relações assumem.

Os modelos TRIP expressam as probabilidades de acerto para diferentes categorias baseadas em parâmetros relativos às categorias e a uma variável latente que expressa a propriedade que está sendo medida no item. Um acerto é considerado quando a pessoa escolhe uma alternativa qualificada numa categoria específica de resposta dentre as diferentes categorias de resposta disponíveis para o item.

Nos estudos de marketing é usual a inclusão de questões que contêm itens policotômicos que têm diferentes categorias de resposta e que são aplicados a um conjunto de consumidores. Denote como $j = 1, \cdots, I$ os itens do teste ou questionário considerado, e como $i = 1, \cdots, n$ as pessoas pesquisadas com o teste de tamanho I, e seja $h_j = 0, 1, \ldots, m_j$ as categorias de resposta possíveis do item j, que podem ser diferentes para cada item do teste. Sem perda de generalidade, vamos supor que essas categorias estão ordenadas em ordem crescente tal que o total de categorias de resposta é m, omitindo o subscrito j.

Considerando os itens da figura 1 sobre *mobile banking*, é razoável pensar que as diferentes categorias de resposta dos itens policotômicos têm diferentes probabilidades de ser registradas por um mesmo avaliado, mas que a soma dessas probabilidades é *1*. Para modelar isto, denotemos essas probabilidades por P_{ijh} com os subscritos i, j e h para o avaliado, o item e as categorias, respectivamente, e por $\sum_{h=0}^{m_j} P_{ijh} = 1$ quando somadas as probabilidades das diferentes categorias do item j para um mesmo avaliado

i. Considere também $P^{+}{}_{ijh} = \sum_{h}^{m_j} P_{ijh}$ que denota as probabilidades acumuladas a partir da categoria h (à direita de h). Veja a tabela 1, baseada no item do exemplo 3, para um melhor entendimento da notação.

Tabela 1. Notação para as probabilidades individuais e acumuladas à direita das categorias de resposta a um item i respondido por avaliado j usando o item do exemplo 3um

Categorias de resposta	Não faz nada	Entra na plataforma virtual mas não consegue realizar a transação	Realiza a transação mas não consegue confirmar o resultado	Realiza a transação corretamente
H	0	1	2	3
Probabilidades individuais	P_{ij0}	P_{ij1}	P_{ij2}	P_{ij3}
Probabilidades acumuladas à direita	$P^{+}{}_{ij0}= P_{ij0+} P_{ij1+} P_{ij2+} P_{ij3=1}$	$P^{+}{}_{ij1}= P_{ij1+} P_{ij2+} P_{ij3}$	$P^{+}{}_{ij2}= P_{ij2+} P_{ij3}$	$P^{+}{}_{ij3}= P_{ij3}$

Expressou-se que como as $m+1$ categorias de resposta estão ordenadas, as diferenças entre duas categorias adjacentes podem ser pensadas como etapas do item. Assim um item com $m+1$ categorias têm m etapas. Por causa disso o modelo pode enfatizar as etapas das categorias ou de maneira geral as respostas ordinais indistintamente (Agresti, 1990). Nesta forma, é possível re-escrever as probabilidades em função das etapas. Assim, para uma etapa k, pode-se escrever:

$$P_{ijk} = P_{ijh} \quad P^{+}{}_{ijk} = P^{+}{}_{ijh} \qquad k = 1, \cdots, m_j \qquad \sum_{k=1}^{m_j} P_{ijk} = 1$$

É claro na representação da tabela 1, que cada etapa (ou categoria) do item pode ser pensada de maneira binária com duas possibilidades ("acertar" a etapa ou "falhar" a etapa, vale dizer, registrar ou não a categoria como resposta). Assim, para descrever os itens policotômicos, pode-se dividi-los em dicotomias que mantêm a natureza ordinal das categorias de resposta. Essas dicotomias são chamadas etapas de item (*item steps*).

Na TRIP, como será apresentado, três diferentes definições da etapa de um item podem ser predominantes: a acumulativa, a de crédito parcial e a condicional (Hemker, 2001). Cada definição está baseada nas diferentes maneiras de dividir o escore de um

item policotômico. Assim, para a acumulativa se considera P_{ijk}^{+}, para a de crédito parcial se toma $\frac{P_{ijk}}{P_{ijk-1}+P_{ijk}}$ e para a condicional é considerado $\frac{P_{ijk}^{+}}{P_{ijk-1}^{+}}$.

Para as três definições a probabilidade de passar de uma etapa do item a outra pode ser escrita como uma função de um traço latente unidimensional u, onde tais funções são chamadas funções de resposta da etapa de um item (FREI). Consequentemente, existem três definições de FREI e cada definição leva a diferentes famílias de modelos TRIP, tal como será apresentado na seção 2.4.

Segundo Agresti (1990) e Van der Ark (2001), pode-se estabelecer três tipos de logitos de transição da *k*-ésima resposta ordinal do item *j* da pessoa *i* (L_{ijk}) aplicando o logito para as diferentes FREI. Assim, para o caso de FREI acumulativa seu logito é chamado de acumulativo, para a FREI de crédito parcial seu logito é chamado de categorias adjacentes, e para o caso da FREI condicional seu logito é chamado de razão-continuação. Isto é mostrado na tabela 2 a seguir:

Tabela 2. Função de resposta e logitos de passo de uma etapa de um item policotômico (com respostas ordinais)

Funções de resposta de uma etapa do item (FREI) (HEMKER, 2001)	Logitos de passo de uma etapa do item (L_{ijk}) (VAN DER ARK, 2001)
Acumulativo: P_{ijk}^{+}	*Acumulativo:* $L_{ijk}=logito\left(P_{ijk}^{+}\right)=\log\left(\frac{P_{ijk}^{+}}{1-P_{ijk}^{+}}\right)$
De crédito parcial: $\frac{P_{ijk}}{P_{ijk-1}+P_{ijk}}$	*De categorias adjacentes:* $L_{ijk}=logito\left(\frac{P_{ijk}}{P_{ijk-1}+P_{ijk}}\right)=\log\left(\frac{P_{ijk}}{P_{ijk-1}}\right)$
Condicional: $\frac{P_{ijk}^{+}}{P_{ijk-1}^{+}}$	*De razão-continuação:* $L_{ijk}=logito\left(\frac{P_{ijk}^{+}}{P_{ijk-1}^{+}}\right)=\log\left(\frac{P_{ijk}^{+}}{P_{ijk-1}^{+}-P_{ijk}^{+}}\right)$

em que dada uma proporção p o logito é uma função definida como $\text{logito}(p)=\frac{p}{1-p}$.

2.3. Modelagem da TRIP

Na modelagem da TRIP, as probabilidades P_{ijk} introduzidas acima, são probabilidades condicionais. Para isso, são considerados dois tipos de variáveis aleatórias: uma variável aleatória observável x, e uma variável aleatória não observável ou latente u. Na literatura psicométrica é frequente denotar u por θ para enfatizar que se trata de um parâmetro do modelo referido aos respondentes. Consideramos ainda mais apropriado o *status* de variável aleatória, conforme apresentado por Bartholomew e Knoot (1999).

A variável x toma diferentes valores categóricos ordenados, em que x = 0 significa que o avaliado não completou nenhuma etapa e se situa na primeira categoria de resposta do item, enquanto x = mi expressa que o avaliado completou todas as etapas da tarefa e se situa na última categoria de resposta do item.

A variável latente u é suposta contínua, pois se estabeleceu a suposição do continuum dada por Pfanzagl, Baumann e Huber (1971), ou seja, que o item mede uma propriedade do construto (educacional, psicológico, ou de qualquer área de pesquisa) que as pessoas possuem em diferentes graus mas que não é observável diretamente senão sob a forma de x, que neste caso toma valores ordenados finitos, vale dizer, como uma variável policotômica. Este construto é considerado aqui, de modo geral, como uma habilidade ou traço latente.

As probabilidades condicionais individuais P_{ijk} são supostas monotonicamente crescentes a respeito da variável latente u, o que equivale dizer ao maior valor da variável latente, maior probabilidade de obter uma determinada categoria e, portanto, são da forma $P_{ijh} = P_{ik}(u_i) = P(x_{ij} = h / u_i)$, expressando a probabilidade que o avaliado i responda corretamente à etapa k do item i dada sua habilidade latente u. Por outro lado, as probabilidades acumuladas à direita são da forma $P^{+}_{ijh} = P(x_{ij} \geq h / u_i)$, e expressam a probabilidade de que o avaliado i responda corretamente à etapa k ou uma superior do item j dada sua habilidade latente u_i. As probabilidades P_{ijk} correspondem à componente aleatória **observável** do modelo TRIP.

Considere também a seguinte componente $\eta_i = a_{jk}\left(u_i - b_{jk}\right)$ que corresponde a uma função linear relativa à habilidade latente u_i em que a_{jk} é um parâmetro de intercepto da resposta ordinal k do item j, e b_{jk} é um parâmetro de localização da resposta ordinal k do item j definido sobre toda a reta real. Esta componente é uma componente aleatória **latente** do modelo da TRIP.

Finalmente, a componente L_{ijk} definida acima está baseada na probabilidade P_{ijk}, denotando que $L_{ijk}=g(P_{ijk})$ onde g é uma função monótona não decrescente e diferençável com respeito a P_{ijk} e, portanto, de u_i. Ao se conectar a componente aleatória observável do modelo com a componente aleatória latente, tem-se que $L_{ijk}=g(P_{ijk})=\eta_i = a_{jk}\left(u_i - b_{jk}\right)$, considerada uma função de **ligação** do modelo da TRIP que relaciona as componentes aleatórias observáveis e latentes.

Desta maneira pode-se escrever que o modelo da TRIP relaciona uma componente aleatória dada pelas probabilidades das respostas ordinais com uma componente linear a respeito de uma variável latente através de uma função de ligação, que neste caso são diferentes logitos que podem ser estabelecidos. Os modelos que satisfazem estas características são denominados modelos da TRIP logísticos paramétricos. Os modelos não paramétricos não estabelecem a componente aleatória latente da forma linear, mas mantêm a idéia que a componente aleatória latente expressa pela função de ligação é monótona não decrescente e diferençável com respeito a u_i sem estabelecer uma forma funcional específica.

Esta caracterização também pode ser usada para o caso dicotômico ($h=2$ ou $k=1$ nas expressões anteriores e, portanto, este subscrito pode ser omitido). A componente aleatória observável é dada por $P_{ij} = P_j(u_i) = P(x_{ij} = 1/u_i)$, a componente aleatória latente é $\eta_i = a_j\left(u_i - b_j\right)$, e a função de ligação $L_{ij}=g(P_{ij})$ é o logito de P_{ij}.

Assim, tem-se que $g(P_{ij})=a_j\left(u_i - b_j\right)$ corresponde ao $logito\left(P_{ijk}\right) = \log\left(\frac{P_{ijk}}{1-P_{ijk}}\right)=$ $a_j\left(u_i - b_j\right)$ ou da forma conhecida $P_{ij} = \frac{\exp\left[a_j(u_i - b_j)\right]}{1+\exp\left[a_j(u_i - b_j)\right]} = \frac{1}{1+\exp\left[a_j(u_i - b_j)\right]^{-1}}$ que é o modelo logístico de 2 parâmetros em que a expressão anterior é chamada CCI.

Da mesma forma, para o caso policotômico, diferentes modelos da TRIP podem ser definidos relacionando os logitos acumulativos, de razão-continuação e de categorias adjacentes dados na tabela 2 com a função linear da variável latente u_i. O uso dos logitos é justificado pois preservam a relação monótona crescente com respeito à habilidade ou traço latente, o que equivale dizer, a maior habilidade latente com maior chance de resposta correta.

3. Modelos da TRIP

3.1. Famílias de modelos TRIP paramétricos

Considerando os três tipos de funções de ligação dadas pelos logitos para respostas ordinais na seção 2.2, bem como as componentes aleatórias observável e latente definidas na seção 2.3, pode-se estabelecer as seguintes famílias de modelos de resposta ao item policotômicos paramétricos:

a) Modelo de Resposta Gradual (*Graded Response Model*): GRM

Este modelo foi originalmente derivado por Samejima (1969). Considerando o logito acumulativo como função de ligação da resposta ordinal com a componente aleatória latente, tem-se:

$$logito\left(P_{ijk}^{+}\right)=\log\left(\frac{P_{ijk}^{+}}{1-P_{ijk}^{+}}\right)= a_{jk}\left(u_i-b_{jk}\right) \quad \text{para } k=1,\ldots,m_i\ ;\ i=1,\ldots,n \ \text{ e } j=1,\ldots,I$$

Alternativamente:

$$P_{ijk}^{+}=\frac{\exp\left[a_{jk}(u_i-b_{jk})\right]}{1+\exp\left[a_{jk}(u_i-b_{jk})\right]}=\frac{1}{1+\exp\left[a_{jk}(u_i-b_{jk})\right]^{-1}} \quad \text{para } k=1,\ldots,m_i\ ;\ i=1,\ldots,n \ \text{ e } j=1,\ldots,I$$

Este modelo é considerado um "modelo de diferenças", pois claramente a probabilidade de um avaliado *j* receber um escore *k* no item *i* é dada por $P_{ijk}=P^{+}_{ijk} - P^{+}_{ijk+1}$ (Andrade, Tavares & Valle, 2000; Samejima, 1997). Por conveniência de notação é considerada $P^{+}_{ij0}=1$ e $P^{+}_{ijm+1}\equiv 0$. Além disso, os parâmetros de localização devem satisfazer $b_{j1} < b_{j2} < \cdots < b_{jm}$ (Samejima, 1969), ou seja, devem estar perfeitamente ordenados.

A FREI que se considera na família de modelos de resposta gradual (GRM) é a FREI acumulada. Assume-se uma ordem estrita nas etapas do item. Se uma etapa não é atingida (acontece uma falha), todas as etapas seguintes não são atingidas; se uma etapa é atingida (ou passada), as etapas anteriores também são passadas. Isto implica que todos os possíveis escores dizem alguma coisa acerca de todas as outras etapas. As pessoas que passam a etapa k de um item j têm um escore ao menos igual a k. Pessoas com escore do item menores que k falham na etapa k; se uma pessoa tem escore do item igual a k , ela passa os primeiros k itens e falha nos seguintes m-k itens.Estes tipos de modelos são chamados modelos acumulativos (Moleenar, 1983; Samejima, 1969). Eles também são conhecidos como modelos de probabilidade acumulados (Mellenberg, 1995), modelos de diferença (Thissen & Steinberg, 1986), modelos de Thurstone (Andrich, 1995) e modelos de gradação (Van Engelenburg, 1997).

No caso do item do exemplo 1 da escala de mobile banking apresentada na figura 1, as categorias têm uma ordem estrita que pode ser numerada de 1 a 5. É claro que o avaliado precisa analisar todas as categorias antes de escolher uma, e se, por exemplo, ele escolhe a categoria “indistinto” significa que ele não optou por nenhuma das categorias prévias nem por nenhuma das categorias seguintes; portanto, seu escore deve ser 3. Observe que os outros itens apresentados não satisfazem estas características.

b) Modelo de Crédito Parcial (Partial Credit Models): PCM

Este modelo foi originalmente derivado por Masters (1982). Considerando o logito de categorias adjacentes como função de ligação da resposta ordinal com a componente aleatória latente, tem-se:

$$L_{ijk} = logito\left(\frac{P_{ijk}}{P_{ijk-1}+P_{ijk}}\right) = \log\left(\frac{P_{ijk}}{P_{ijk-1}}\right) = a_{jk}\left(u_i - b_{jk}\right)$$

para $k=1,...,m_i$; $i=1,...,n$ e $j=1,...,I$

Alternativamente:

$$\frac{P_{ijk}}{P_{ijk-1}+P_{ijk}} = \frac{\exp\left[a_{jk}(u_i - b_{jk})\right]}{1+\exp\left[a_{jk}(u_i - b_{jk})\right]} = \frac{1}{1+\exp\left[a_{jk}(u_i - b_{jk})\right]^{-1}}$$

Este modelo é também denominado de "modelo dividido pelo total", pois:

$$P_{ijk} = \frac{\exp\left[\sum_{v=0}^{k} a_{jv}(u_i - b_{iv})\right]}{\sum_{v=0}^{m}\exp\left[\sum_{u=0}^{v} a_{jv}(u_i - b_{iu})\right]}$$ para $k=0,...,m_i$; $i=1,...,I$ e $j=1,...,n$

considerando-se, por conveniência de notação, $b_{i0}\equiv 0$.

A FREI que se considera na família de modelos de resposta de crédito parcial (PCM) é a FREI de crédito parcial. Implica que só as pessoas que passam a etapa *k* do item são as que têm escores *k*, e só as que falham a etapa *k* são as que têm escores *k-1*. Só os escores *k* e *k-1* têm influencia direta na FREI, e não a avaliação de todas as etapas, como na FREI acumulativa. Saliente-se que, indiretamente, todas as etapas anteriores e seguintes influem em cada etapa, o que complica a interpretação da dificuldade da etapa do item (Moleenar, 1983; Verhelst & Verstralen, 1997). Esses tipos de modelos são chamados modelos de crédito parcial (Moleenar, 1983). Eles também são conhecidos como modelos de categorias adjacentes (Mellenberg, 1995), modelos de Rasch (Andrich, 1995), ou modelos de partição (Van Engelenburg, 1997).

No caso dos itens dos exemplos 2 e 3 da escala de avaliação de *mobile banking* descrita na figura 1, o avaliado constrói sua solução independentemente de conhecer todas as etapas do problema. É o avaliador que qualifica sua resposta de acordo com o critério definido nas categorias do item, que neste caso pode ser de 0 a 3. O avaliador atribui os escores dependendo se cada etapa é apresentada na solução do avaliado. Para cada etapa da solução presente, o avaliador atribui um "crédito" ou escore igual a 1. Se o avaliador não encontra a etapa (falha), então o avaliado fica com os "créditos

acumulados" ou o escore até a etapa prévia. Observe que aqui o avaliado pode tentar todas as etapas, mas o escore que ele consegue depende de quantas etapas ele completa corretamente.

c) Modelo Sequencial (*Sequential Models*): SM

Este modelo, originalmente proposto por Tutz (1990), considera o logito de razão-continuação como função de ligação da resposta ordinal com a componente aleatória latente, ou seja:

$$L_{ijk} = logito\left(\frac{P^{+}_{ijk}}{P^{+}_{ijk-1}}\right) = \log\left(\frac{P^{+}_{ijk}}{P^{+}_{ijk-1} - P^{+}_{ijk}}\right) = a_{jk}\left(u_i - b_{jk}\right)$$

para $k=1,\ldots,m_i$; $i=1,\ldots,n$ e $j=1,\ldots,I$

Alternativamente:

$$\frac{P^{+}_{ijk}}{P^{+}_{ijk-1}} = \frac{\exp\left[a_{jk}(u_i - b_{jk})\right]}{1+\exp\left[a_{jk}(u_i - b_{jk})\right]}$$

Este modelo pode ser expresso (Akkermans, 1998) por:

$$P_{ijk} = \prod_{r=1}^{k}\frac{P^{+}_{ijr}}{P^{+}_{ijr-1}} - \prod_{r=1}^{k+1}\frac{P^{+}_{ijr}}{P^{+}_{ijr-1}} = \left\{\prod_{r=1}^{k}\frac{\exp\left[a_{jk}(u_i - b_{jk})\right]}{1+\exp\left[a_{jk}(u_i - b_{jk})\right]}\right\}\left[\frac{1}{\exp\left[a_{jk+1}(u_i - b_{jk+1})\right]}\right],$$
$$k=0,1,\ldots,m$$

considerando-se, por conveniência de notação:

$$\prod_{r=1}^{0}\frac{P^{+}_{ijr}}{P^{+}_{ijr-1}} \equiv 1 \quad \text{e} \quad \prod_{r=1}^{m+1}\frac{P^{+}_{ijr}}{P^{+}_{ijr-1}} \equiv 0$$

A FREI que se considera na família de modelos de resposta sequencial (SM) é a FREI condicional. Também se assume uma ordem estrita das etapas do item como no caso da família GRM. Em contraste com a FREI acumulada, falhar numa etapa aqui implica que as etapas seguintes do item nunca são consideradas, isto é, falhou na passagem. Passar uma etapa do item implica que a etapa seguinte é considerada. Assim, a etapa do item *k* é atingida por todas as pessoas com um escore igual ou maior

que *k*, e não atingidas pelas pessoas com escores iguais a *k-1*; as únicas pessoas que falham nesta etapa do item são aquelas que têm um escore total igual a *k*-1. Pessoas com escores do item baixos não tentam na etapa *k* do item, e assim não há falha. Isto significa que só escores iguais ou maiores que *k-1* dizem algo a respeito da etapa *k*, a qual considera que o número de escores que têm implicações na etapa do item depende da etapa do item. Estes tipos de modelos são chamados modelos condicionais (Moleenar, 1983) ou modelos sequenciais (Van Engelenburg, 1997).

No caso do item 4, semelhante ao item 1, as etapas do problema são conhecidas pelo avaliado e, também semelhante ao item 3, é atribuído um escore com o qual se consegue calcular a etapa correspondente. É o avaliador que qualifica sua resposta de acordo com o critério definido nas categorias do item que, neste caso, foi de 0 a 3. O avaliador atribui os escores dependendo se cada etapa é apresentada na solução do avaliado. Observe que a diferença do item 3 em relação ao item 4 é que neste o avaliador não precisa observar a solução das etapas seguintes, pois claramente essas não estão presentes. Se acontecer uma falha numa etapa o processo cessa e as etapas seguintes não são consideradas.

d) Modelos particulares da TRIP paramétricos

As famílias de modelos TRIP são consideradas restritas quando são impostas restrições para os parâmetros de item de intercepto (a_{jk}) e de localização (b_{jk}) em η_i que leva a casos particulares da componente aleatória latente e, portanto, a modelos particulares na família de modelos da TRIP paramétricos. Tem-se assim:

Restrições no parâmetro de intercepto:

1. Os modelos com a_{jk} diferentes e, portanto, $\eta_i = a_{jk}\left(u_i - b_{jk}\right)$, são chamados modelos com parâmetros de intercepto não restrito e são denotados como modelos *2P*(*jk*).
2. Os modelos com $a_{jk} = a_j$ para todo *i*, portanto, $\eta_i = a_j\left(u_i - b_{jk}\right)$, são chamados modelos com parâmetros de intercepto restrito ao item e são denotados como modelos 2P(*j*).
3. Os modelos com $a_{jk} = a_k$ para todo *k*, portanto, $\eta_i = a_k\left(u_i - b_{jk}\right)$, são chamados modelos com parâmetros de intercepto restrito à resposta ordinal e são denotados como modelos *2P*(*k*).

4. Os modelos com $a_{jk}=1$ para todo *k* e todo *i*, portanto, $\eta_i=\left(u_i-b_{jk}\right)$, são chamados modelos sem parâmetros de intercepto ou modelos de Rasch, sendo denotados como modelos 1P.

Restrições no parâmetro de localização:

5. O parâmetro b_{jk} pode ser descomposto em dois componentes aditivos, $b_{jk}=\delta_j+\tau_k$, em que δ_j é um componente do item, e τ_k é um componente da resposta ordinal, de modo que $\sum_{k=1}^{m}\tau_k=0$. Esta decomposição na componente sistemática define modelos de escala gradual (*rating scale*) e, nesse caso, os modelos podem ser denotados adicionando $-$R.

Como exemplo da notação introduzida, o modelo denotado por 1P-SM-R corresponde ao modelo com um parâmetro (sem parâmetro de intercepto) da família sequencial que admite uma escala gradual por decomposição do parâmetro de localização. O modelo denotado por *2P*(*jk*)*-PCM* corresponde ao modelo com dois parâmetros sem restrições no parâmetro de intercepto da família de crédito parcial.

De acordo com as restrições dos parâmetros de intercepto e de localização parece possível considerar até oito modelos particulares (4×2) para cada uma das três famílias de modelos policotômicos. No entanto, esses 24 modelos paramétricos não existem. Os modelos que existem (Van der Ark, 2001) aparecem na tabela a seguir:

Tabela 3. Modelos particulares na família de modelos paramétricos policotômicos, segundo Van der Ark (2001)

Restrições para o parâmetro de intercepto	Famílias de modelos paramétricos					
	Resposta Gradual		Crédito Parcial		Seqüencial	
	Não admite escala gradual	Admite escala gradual	Não admite escala gradual	Admite escala gradual	Não admite escala gradual	Admite escala gradual
Não restrito	-	-	-	-	2P(jk)-SM	-
Restrito ao item	2P(j)-GRM Samejima (1969)	-	2P(j)-PCM Muraki (1992)	-	2P(j)-SM	-
Restrito à resposta ordinal	-	-	2P(k)-PCM	-	2P(k)-SM	-
Sem parâmetro de intercepto	1P-GRM	1P-GRM-R	1P-PCM Masters (1982)	1P-PCM-R Andrich (1978)	1P-SM Tutz (1990)	1P-SM-R Tutz (1990)

Existem diferentes fatos que explicam porque alguns modelos não existem. Por exemplo, para a família de resposta gradual, o parâmetro de intercepto não pode mudar sobre as respostas ordinais do mesmo item e, assim, modelos 2P(jk) e 2P(k) não existem (Mellenberg, 1995).

O modelo de aceleração de Samejima (*Aceleration Models*): AM

Este modelo não é um modelo logístico (Samejima, 1995) e é definido por:

$$\frac{P_{ijk}^{+e}}{P_{ijk-1}^{+}} = \left\{ \frac{\exp(a_{jk}(u_i - b_{jk})}{1 + \exp(a_{jk}(u_i - b_{jk})} \right\}^{\varepsilon_j}$$

em que $P_{ijk}^{+e} = P(x_{ij} > k \mid u_j)$ e $\varepsilon_j \geq 0$

Alternativamente:

$$\text{logito}\left(\frac{P_{ijk}^{+e}}{P_{ijk}^{+}} \right)^{1/\varepsilon_j} = a_{jk}(u_i - b_{jk})$$

Observe que se $\varepsilon_j = 1$, tem-se o modelo sequencial SM. Saliente-se que ε_j é considerado como um novo parâmetro de item e também a expressão acima é considerada uma nova CCI de tipo assimétrico. Para o caso dicotômico veja o trabalho de Bolfarine e Bazán (2010).

3.2 Modelos TRIP não paramétricos

Considerando as funções de ligação definidas na tabela 2 e sem assumir que a componente aleatória latente tem a forma linear dada na seção 2.3, mas mantendo que as funções de ligação implicam que as probabilidades P_{ijk} são não decrescentes em u_i, pode-se ter modelos não paramétricos alternativos para cada família de modelos de resposta policotômicos paramétricos. Esses modelos correspondem ao modelo de resposta gradual não paramétrico (NP-GRM), o modelo de resposta de crédito parcial não paramétrico (NP-PCM) e o modelo sequencial não paramétrico (NP-SM).

Boomsma, A., Van Duijn, M.A.J., & Snijders

A área de pesquisa de modelos não paramétricos tem sido bastante desenvolvida ultimamente (Boomsma, Van Duijn & Snijders, 2001; Hardouin, Bonnaud-Antignac & Sébille, 2011; Sijtsma & Molenaar, 2002; Xu & Douglas, 2006;). Van der Ark (2001) considera alguns modelos NP-GRM que assumem diferentes propriedades a mais para os logitos acumulativos:

a) os logitos não se interceptam (modelo DMM);

b) os logitos são invariantemente ordenados para todos os itens, ou seja, satisfazem logito(P_{ijk}^{+}) < logito(P_{ljk}^{+}) para todo j e todo k (modelo ISOP);

c) os logitos não se interceptam e são invariantemente ordenados para todos os itens (Modelo SDMM).

O modelo DMM ou de dupla monotonicidade é apresentado em Moolenar (1997); o modelo SDMM ou de dupla monotonicidade forte é apresentado por Sijtsma e Hemker (1998), enquanto o modelo ISOP ou probabilístico ordinal isotônico é apresentado em Junker (1998) e Scheiblechner (1995).

a) Relações entre os modelos TRI policotômicos

A figura 2, devida a Van der Ark (2001), apresenta uma estrutura hierárquica entre diversos modelos TRIP. As provas desses resultados estão referidas em Van der Ark (1999).

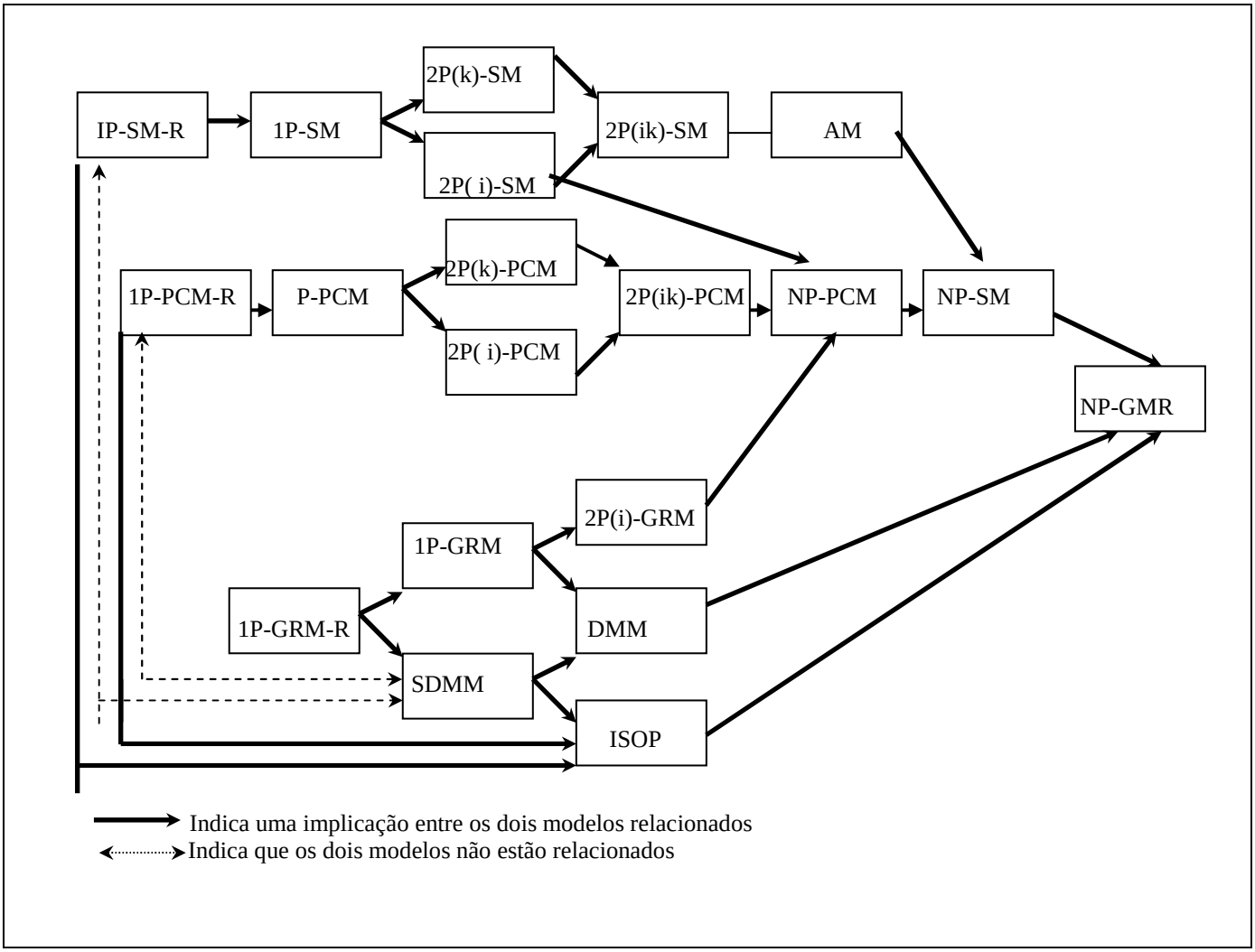


Figura 2. Estrutura hierárquica de modelos TRI policotômicos, segundo Van der Ark (2001)

Uma flecha sólida retrata uma implicação, enquanto uma flecha de duas vias pontilhada indica que dois modelos não estão relacionados – salientando-se que a existência de um não implica a exclusão do outro. Os modelos 1P-PCM-R, 1P-SM-R e 1P-GRM-R são os modelos mais restritivos. O modelo NP-GRM é o modelo mais geral. A importância dos resultados apresentados na figura 2 é que quando se pesquisa sobre modelos alternativos para analisar dados deve-se começar com o modelo mais restritivo e, se ele não se ajusta, trata-se o próximo modelo na ordem hierárquica.

O fato das famílias de modelos de resposta gradual, de crédito parcial e sequencial aparecerem separadas significa que os parâmetros de item que pertencem ao logito de uma classe de modelos não podem ser escritos como função dos parâmetros das outras famílias (Mellenbergh, 1995; Thissen & Steinberg, 1986, para as famílias PCM e GRM). Isto também deve ser considerado no momento de decidir as análises a serem implementadas para um conjunto particular de dados policotômicos. Antes de considerar outra família deve-se optar por outro modelo dentro da mesma família com outra especificação de parâmetros.

4. Aplicação da TRI a uma escala de atitudes em Marketing

O estudo de *mobile banking* de Puschel, Mazzon e Hernandez (2010) apresenta uma escala composta de 14 dimensões. Para o propósito da presente ilustração, escolheu-se a dimensão de atitudes em que os itens são mostrados na tabela 4. Os dados correspondem a 333 não usuários de *mobile banking*, mas que possuem ao menos uma conta bancária e serviço de celular, obtidos de uma amostra intencional. Para detalhes, veja Puschel, Mazzon e Hernandez (2010).

Para estes dados, ajustou-se um modelo de resposta gradual considerando o pacote *ltm* em R (Rizopoulos, 2006). Para uma discussão dos *softwares* que podem ser usados na estimação de modelos TRIP veja o Apêndice. O *ltm* foi usado para ajustar o modelo de resposta graduada de Samejima (1969). A decisão para o uso deste modelo está baseada no tipo de itens considerados e os logitos de passo que podem ser assumidos para estes itens. A mostra de itens aparece na tabela 4 e corresponde a quatro itens de uma escala do tipo Likert de concordância com base em seis posições, anteriormente apresentada.

Estimativas dos parâmetros dos itens da escala foram obtidas considerando a teoria clássica dos testes (TCT) e TRI. Considerando a TCT foram obtidas a média (dificuldade clássica) e desvio padrão das respostas policotômicas, as correlações item/total (índice de discriminação). Considerando o modelo GRM da TRI foi obtida a dificuldade e discriminação dos itens e a função de informação dos itens e da escala, a qual proporciona a quantidade de informação em cada ponto do traço latente. Por fim,

uma análise fatorial exploratória (AFE) considerando o eixo principal, residual mínimo e mínimos quadrados ponderados foi obtida por meio do pacote *psych* de R.

Tabela 4. Itens, estatísticas de itens baseada na TCT, resultados da análise fatorial exploratório (AFE) e estimativas dos parâmetros dos itens usado o modelo GRM da TRI

Itens	TCT			Análise Fatorial	TRI					
	Média dos escores	Desvio Padrão dos scores	CITC	FL	b1	b2	b3	b4	b5	A
Usar o mobile banking para fazer transferências bancárias é uma boa idéia	4.1	1.5	0.70	0.77	-1.625	-1.024	-0.553	0.188	1.084	2.746
Utilizar o mobile banking é uma idéia inteligente	4.7	1.3	0.74	0.81	-2.154	-1.746	-1.065	-0.269	0.671	3.042
Gosto da idéia de utilizar o mobile banking	4.1	1.5	0.82	0.91	-1.411	-0.944	-0.565	0.210	0.955	4.251
É/Seria muito divertido utilizar o mobile banking	3.6	1.6	0.63	0.69	-1.370	-0.860	-0.163	0.708	1.561	2.002

Nota: CITC = correlação item total corrigido; FL = carga fatorial; bj = parâmetro de dificuldade; a = parâmetro de discriminação.

Baseados na AFE identificou-se a soma de quadrados de cargas (2,56), a qual representa a proporção de variância explicada de .64. O teste de hipótese indica que um fator é suficiente para explicar os itens analisados. Os graus de liberdade para o modelo nulo é 6 e a função objetivo foi de 2,07 com qui-quadrado de 683,71. Os graus de liberdade do modelo foram 2 e a função objetivo foi 0,01. Adicionalmente a RMSR foi de 0,01. O Tucker Lewis Index of Factoring Reliability foi 0,996 e o índice RMSEA foi 0,036. A correlação dos escores com o fator foi de 0,95. Desta forma temos uma escala de atitudes frente ao *mobile banking* que comportasse de modo unidimensional considerando os quatro itens apresentados.

Embora não exista um critério de corte para o parâmetro de discriminação a, Zickar, Russel, Smith, Bohle e Tilley (2002) sugerem que todos os parâmetros que são maiores que 1 indicam uma discriminação aceitável entre pessoas. Hafsteinsson, Donovan e Breland (2007) sugerem que quando existem poucos itens em uma escala

(foram usadas três escalas de 8, 8 e 5 itens), um padrão alto de qualidade do item pode requerer um valor 2,0 ou mais, para que se atinja uma alta qualidade de medição. Na escala utilizada como exemplo encontrou-se nos quatro itens de atitudes valores acima de 2, atingindo-se portanto estes padrões de qualidade.

Os valores do parâmetro de discriminação a da TRI, mostram grande diferenciação na indicação da qualidade do item em relação ao expresso pelo índice de discriminação na TCT. Por exemplo, o item 1 e 2 tem CITC 0,70 e 0,74, com os valores do parâmetro a sendo 2,746 e 3,042, respectivamente. Informação adicional é proporcionada pelas funções de informação dos itens na figura 3, a qual revela que a maioria dos itens provê uma boa informação para os valores do traço latente; no entanto, o item 4 (É/Seria muito divertido utilizar o *mobile banking*) apresenta menos quantidade de informação em relação aos valores comuns do traço latente: atitude frente ao *móbile banking*.

Por considerando o sinal dos valores dos parâmetros b identificou se as "dificuldades" de passo entre as categorias de resposta. Assim no caso do item 2 (Utilizar o mobile banking é uma idéia inteligente) é mais fácil atingir a ultima categoria (concordância total) do que nos outros itens. O contrário acontece no item 4.

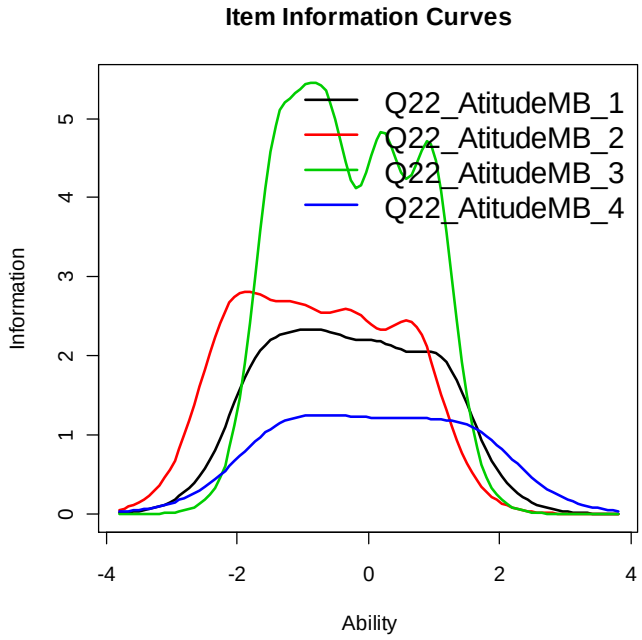


Figura 3. Funcão de informação dos itens para os quatro itens que compõem a escala de atitudes frente ao *mobile banking*.

5. Discussão dos resultados

A taxonomia apresentada das famílias dos modelos de resposta ao item policotômicos na base dos tipos de logitos ou as FREI para as etapas do item, é uma caracterização semelhante a dos modelos lineares generalizados. Moustaki e Knott (2000) introduzem o nome de modelos de traços latentes generalizados para o caso em que a distribuição da variável aleatória manifesta pertence à família exponencial. Eles denominam também a componente aleatória latente como componente sistemática na tradição dos modelos lineares generalizados, sem notar que uma distribuição aleatória é considerada para a variável latente e, portanto, não são valores conhecidos. Note-se que os logitos introduzidos são função de ligação e, desse modo, outras funções de ligação que preservem a monotonia não decrescente em relação à variável latente são possíveis de serem introduzidas (Hemker, 2001). Veja, por exemplo, os trabalhos de Bazán (2005) e Bolfarine e Bazán (2010).

Esta caracterização pode ser útil no desenvolvimento da inferência estatística dos modelos TRI policotômicos. Métodos de estimação, propriedades assintóticas, análises de pressupostos e testes de hipótese relativos a estes modelos poderão utilizar a notação introduzida neste artigo. Outros aspectos da taxonomia apresentada que foram explorados referem-se à correspondência da escala com a realidade psicológica medida, a interpretação dos parâmetros para cada tipo de família e as propriedades de medição que caracterizam estas três famílias.

Uma comparação teórica dos três tipos de famílias de itens policotômicos leva a pensar no paralelismo da medida psicológica com a realidade (Samejima, 1995), isto é, se o tipo de itens ou de regras de formação de escore pode ser aplicado à situação a ser modelada. Duas classificações similares, mas que têm diferentes ênfases, foram introduzidas para avaliar essa questão. A primeira focaliza na classificação dos modelos por meio das características da tarefa (Van Engelenburg, 1997); a segunda classificação focaliza o procedimento ou regras de formação de escore (Akkermans, 1998). Destaque-se ainda a possibilidade das tarefas dos itens serem diferentes se as regras de formação de escore forem diferentes, e vice-versa, além da existência de uma relação direta entre as duas. Ambas as classificações estão baseadas em várias questões.

A primeira questão refere-se ao processo de etapas: são as etapas do item simultâneas ou separadas? As etapas são tomadas simultaneamente se o avaliado atribui um juízo completo numa escala (características da tarefa), ou se o avaliado é considerado num juízo completo (regra de escore). Este processo ou tipo de itens é descrito pela família de modelos com logitos acumulados.

No caso em que as etapas dentro de um item são tomadas separadamente, o segundo critério de classificação é a regra de continuação, com a seguinte questão relacionada: a formação de escore ou a tarefa é continuada depois de falhar a etapa, ou a formação do escore ou tarefa é parada depois de ser falhada? Se o sujeito tenta todas as etapas e o escore é a soma de escores dessas tarefas separadas (binárias), isto é chamado formação de escore paralela. Este tipo de processo do item é descrito pela família de modelos com logitos de crédito parcial.

A terceira questão relaciona-se ao caso da formação de escore ser sequencial, isto é, se o processo ou regra de parar é aplicado quando acontece uma falha. Nesta situação, a questão é a seguinte: é o ordenamento das etapas completamente fixo ou não? Unicamente se uma formação de escore fixa é assumida, os modelos da família de logitos sequências descrevem estes itens (ou o processo). Se não existe ordenamento fixo das etapas, não existe um tipo específico de modelos relativos ao processo. Registre-se que se a reversibilidade acontece, as etapas podem ser invertidas, o que significa que a ordem das etapas não é completamente fixa. Assim, o resultado da reversibilidade não acontece para os modelos sequenciais em concordância com o pressuposto de ordenamento fixo das etapas para este tipo de modelo.

Conclui-se, assim, que é fundamental conhecer as características que distinguem os modelos de modo a serem consideradas pelos construtores de itens e pelos avaliados que atribuem o escore de uma tarefa.

Com respeito à interpretação dos parâmetros dos modelos considere-se, por exemplo, que em alguns testes, falhar na k-ésima etapa leva a um escore de $k-1$. Passar a k-ésima etapa significa que a categoria $k+1$ pode ser tentada, levando a um escore de ao menos k. Tais itens com escore sequencial podem ser analisados pelos modelos TRI-SM. Nessas situações, os parâmetros de localização podem ser interpretados como parâmetros de dificuldade para cada etapa do item. Isto é, se os avaliados com valor u menor ou igual que β_{ik} têm uma probabilidade de passar o item maior que 0,5. Por outro

lado, existe uma literatura recente sobre a taxonomia dos modelos TRI que usa as propriedades de medição (Van der Ark, 2001). Esta literatura deve ser revisada para complementar os resultados apresentados neste artigo, entre elas Sijtsma e Hemker (2000) e Van der Ark (2001).

A hierarquização dos modelos da teoria de resposta ao item policotômicos, dada pela figura 2, pode ser usada na avaliação dos modelos para um mesmo conjunto de dados identificando o modelo seguinte na hierarquia. Esta estratégia pode ser usada para a verificação dos modelos a serem implementados pelos avaliadores. Na base desta caracterização podem ser revisados os distintos *softwares* disponíveis para identificar os modelos que estes estimam e as re-parametrizações que usam.

A aplicação apresentada mostra o potencial de uso destes modelos no campo do marketing. Áreas de conhecimento como saúde, educação, psicologia e análise política, dentre outras, também podem ser beneficiadas com a discussão a respeito do uso de diferentes modelos TRIP, tanto do ponto de vista da construção de escalas considerando os diferentes formatos de itens discutidos, assim como quanto à decisão do modelo estatístico a ser usado para os itens considerados. Ao respeito veja por exemplo o trabalho de De Jong, Steenkamp, Fox e Baumgartner (2008) e De Jong, Steenkamp e Veldkamp (2009).

Finalmente, os modelos da TRI descritos aqui têm sido de natureza cumulativa. Eles sugerem que níveis mais elevados do traço latente (ie, níveis mais elevados de capacidade, maiores níveis de satisfação) deve, com toda a probabilidade, levar a maior pontuação do item, que, por sua vez, deve levar a maior pontuação total do teste. No entanto, este não é freqüentemente o caso para algumas respostas graduadas de concordância (ou seja, discordo totalmente, discordo, Discordo ligeiramente, Ligeiramente Concordo, concordo, concordo totalmente). Essas respostas podem ser consistentes com um modelo diferente - um modelo referido como modelos ordinais de desdobramento ou não acumulativos (*unfolding models* em inglês). Para maiores detalhes veja por exemplo Roberts, Donoghue e Laughlin (2000), Samartini (2006) e Bortolotti (2010).

Referências bibliográficas

Agresti, A. (1990). Categorical data analysis. New York: Wiley.

Akkermans, W. (1998). Studies on statistical models for polytomous items. Unpublished doctoral dissertation. The Netherlands: Twente University.

Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). Introdução a Teoria da Resposta ao Item: Conceitos e Aplicações. In: 14º. SIMPÓSIO DE PROBABILIDADE E ESTATÍSTICA. Caxambu.

Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43, 561-73.

Andrich, D. (1995). Models for measurement, precision, and the nondichotomization of graded responses. Psychometrika, 60, 7-26.

Azevedo, C. L. N. (2003). Métodos de estimação na teoria de resposta ao item. Dissertação de Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.

Baker, F. B., & Kim, S. H. (2004). Item response theory: parameter estimation techniques (2a ed.). New York: Marcel Dekker.

Balasubramanian, S., & Kamakura, W. A. (1989). Measuring consumer attitudes toward the marketplace with tailored interviews. Journal of Marketing Research, 26, 311-26.

Bartholomew, D. J., & Knoot, M. (1999). Latent Variable Models and Factor Analysis. Great Britain: Arnold.

Bazán, J. L. (2005). Uma família de modelos de resposta ao item normal assimétrica. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.

Bechtel, G. G. (1985). Generalizing the Rash Model for consumer rating scales. Marketing Science, 4, 62-73.

Bolfarine, H.; Bazán, J. L. (2010). Bayesian estimation of the logistic positive exponent irt model. Journal of Educational Behavioral Statistics, 35(6), 693-713.

Boomsma, A., Van Duijn, M.A.J., & Snijders, T.A.B. (Ed.). (2001). Essays on item response theory. New York: Springer-Verlag.

Bortolotti, S.L.V. ()2010. Resistência à Mudança Organizacional: Medida de Avaliação por Meio da Teoria da Resposta ao Item, Tese de doutorado, Departamento de Engenharia da Produção, UFSC, Florianópolis.

Childs, R. A., & Chen, W. H. (1999). Obtaining Comparable Item Response Parameter Estimates in MULTILOG and PARSCALE for Two Polytomous IRT Models. Applied Psychological Measurement, 23, 371-79.

Cook, K. F., Taylor, P. W., Dodd, B. G., Teal, C. R., & Mchorney, C. A. (2007). Evidence-based practice for equating health status items: sample size and IRT model. J Appl Meas, 8(2), 175-189.

Curtis, S. M. (2010). BUGS Code for Item Response Theory. Journal of Statistical Software, 36. Recuperado em 20 de janeiro de 2011, de http://www.jstatsoft.org/v36/c01/paper

De Jong, M. G., Steenkamp, J., & Fox, J. P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. Journal of Consumer Research, 24(2), 260-278.

De Jong, M. G., Steenkamp, J., Fox, J. & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. Journal of Marketing Research, 45, 104-115.

De Jong, M. G., Steenkamp, J. & Veldkamp, B. (2009). A model for the construction of country-specific yet internationally comparable short-form marketing scales. Marketing Science, 28(4), 674-689.

Franses, P. H., & Paap, R. (2001). Quantitative Models in Marketing Research. Cambridge: University Press,

Hafsteinsson, L. G., Donovan, J. J., & Breland, B. T. (2007). An item response theory examination of two popular goal orientation measures. Educational and Psychological Measurement, 67, 719–39.

Hair, J. F. JR, Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate Data Analysis. (6a ed.). Upper Saddle River: Prentice-Hall.

Hardouin J.B., Bonnaud-Antignac A., & Sébille V. (2011). Non parametric Item Response Theory using Stata. The Stata Journal, 11(1), 30-51.

Heinen, T. (1996). Latent class and discrete latent trait models: Similarities and differences. Thousand Oaks CA: Sage.

Hemker, B. T. (2001). Reversibility revisited and other comparison of tree types of polytomous IRT models. In: A. Boomsma, M. A. J. Van Duijn, & T. A. B. Snijders (Ed.). Essays on item response theory (pp. 277-96). New York: Springer-Verlag.

Junker, B. W. (1998). Some remarks on Scheiblenchner's treatment of ISOP models. Psychometrika, 63, 73-85.

Malhotra, N. K. (2011). Pesquisa de Marketing: Foco na Decisão. (3a ed.). São Paulo: Pearson.

Masters, G. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-74.

Mazzon, J. A. (1981). Avaliação do Programa de Alimentação do Trabalhador sob o Conceito de Marketing Social. Tese de Doutorado, Faculdade de Economia Administração e Contabilidade, Universidade de São Paulo, São Paulo.

Mellenberg, G. H. (1995). Conceptual notes on models for discrete polytomous item responses. Applied Psychological Measurement, 19, 91-100.

Moleenar, I. W. (1983). Item steps (Heymans Bulletins HB-83-630-EX). Groningen, The Netherlands: University of Groningen.

Moolenar, I. W. (1997). Nonparametric models for polytomous responses. In: W. J. Van der Linden, & R. K. Hambleton (Ed). Handbook of Modern Item Response Theory (pp. 369-380). New York: Springer.

Moustaki, I., & Knott, M. (2000). Generalized Latent Trait Models. Psychometrika, 65, 391-411.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16, 159-77.

Muraki, E., & Bock, R. D. (1996). PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance task. Version 3. Chicago: Scientific Software.

Omar, M. A., Ghariebeh, B., Salazar, A., & Saito, K. (2007). IRT and UVF fluorescence for marine coatings inspection. International Journal of nondestructive testing and evaluation, 40, 62-70.

Pasquali, L. (2009). Psicometria: teoria dos testes na psicologia e na educação (3a ed.). Petrópolis, RJ: Vozes.

Pfanzagl, J., Baumann, V., & Huber, H. (1971). Theory of Measurement. Physica-Verl.

Pousttchi, K., & Schurig, M. (2004). Assessment of today's mobile banking applications from the view of customer requirements. In: 37th ANNUAL HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES. HAWAII (pp. 1-10). Hawaii.

Puschel, J., Mazzon, J. A., & Hernandez, J. M. (2010). Mobile banking: proposition of an integrated adoption intention framework. International Journal of Bank Marketing, v. 28, n. 5, p. 389-409.

Rao, C. R., & Sinharay, S. (2006). Handbook of Statistics 26: Psychometrics. Amsterdam, The Netherlands: Elsevier.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. Journal of Statistical Software, 17, 1–25.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. Applied Psychological Measurement, 24, 3-32.

Samartini, A. L. S. (2006). Modelos com Variáveis Latentes Aplicados à Mensuração de Importância de Atributos, Tese de doutorado, Escola de Administração de Empresas de São Paulo.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph, 17.

Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. Psychometrika, 60, 549-72.

Samejima, F. (1997). Graded Response Model. In: W. J. Van der Linden, & R. K. Hambleton (Ed.). *Handbook of Modern Item Response Theory* (pp. 85-100). New York: Springer.

Scheiblechner, H. (1995). Isotonic ordinal probabilistics models (ISOP). Psychometrika, 60, 281-304.

Sijtsma, K., & Hemker, B. T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. Psychometrika, 63, 183-200.

Sijtsma, K., & Hemker, B. T. (2000). A taxonomy of IRT models for ordering persons and items using simple sum scores. Journal of Educational and Behavioral Statistics, 25, 391-415.

Sijtsma, K,, & Molenaar, I. W. (2002). Introduction to nonparametric item response theory. Thou-sand Oaks, CA: Sage.

Singh, J., Howell, R. D., & Rhoads, G. K. (1990). Adaptive designs for likert-type data: an approach for implementing marketing surveys. Journal of Marketing Research, 27(3), 304-21.

Singh, J. (2004). Tackling measurement problems with Item Response Theory: principles, characteristics, and assessment, with an illustrative example. Journal of Business Research, 57, 184-208.

Tay, L., Diener, E., Drasgow, F., & Vermunt. J. K. (2011). Multilevel mixed-measurement IRT analysis: An explication and application to self-reported emotions across the world. Organizational Research Methods, 14, 177-207.

Tezza, R., Bornia, A. C., & Andrade, D. F. (2011). Measuring web usability using item response theory: Principles, features and opportunities. Interacting with Computers, 23(2), 167-75.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. Psychometrika, 51, 567-77.

Thissen, D. (1991). MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory. Version 6.0. Chicago: Scientific Software.

Tutz, G. (1990). Sequential item response models with an ordered response. British Journal of Mathematical and Statistical Psychology, 43(1), 39-55.

Van der Ark, L. A. (1999). A reference card for the relationships between IRT models for ordered polytomous items and some relevant properties. The Netherlands: Tilburg University. Working Paper No. 99.10.02.

Van der Ark, L. A. (2001). Relationships and properties of polytomous item response theory models. Applied Psychological Measurement, 25, 273-82.

Van der Linden, W. J., & Hambleton, R. K. (Ed). (1997). Handbook of moderns item response theory. New York: Springer.

Van Engelenburg, G. (1997). On psychometric models for polytomous items with ordered categories within the framework of item response theory. Unpublished doctoral dissertation. The Netherlands: University of Amsterdam.

Verhelst, N. D., & Verstralen, H. H. F. M. (1997). Modeling sums of binary responses by the partial credit model. Measurement and Research Department Reports 97-7. Arnhem: CITO.

Vermunt, J. K. (1997). EM: A general program for the analysis of categorical data. Manuscrito não publicado. The Netherlands: Tilburg University. Recuperado em 17 de dezembro de 2005, de http://www.kub.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html

Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. Applied Psychological Measurement, 25, 283-94.

Von Davier, M. (1996). WINMIRA VI.68: A program system for analysis with the Rasch model, with the latent class analysis and with the mixed Rasch model (Software manual). Recuperado em 17 de dezembr de 2005, de http://www.winmira.von-davier.de

Xu, X., & Douglas, J. (2006). Computerized adaptive testing under nonparametric IRT models. Psychometrika, 71,(1), 121-137.

Zickar, M. J., RusseL, S. S., Smith, C. S., Bohle, P., & Tilley, A. J. (2002). Evaluating two morningness scales with item response theory. Personality and Individual Differences, 33, 11–24.

Abstract

It has been common works published in Management and particularly in Marketing using statistical techniques applied to ordinal scales assuming that these were interval. In order to provide a suitable alternative to this practice, it is presented in this article, a taxonomy of different models of item response theory polychotomous (TRIP) based on the types of logit step from one stage of the item (response functions). The emphasis of the presentation is introductory and the proposed methodology is illustrated with an application considering a range of attitudes of consumers towards banking services, with respect to mobile banking, an area of recent interest in Marketing. This example also illustrates the use of software to implement most of these models.

Keywords: Item Response Theory; Polychotomous Items; Marketing.

Apêndice. Softwares para estimar os modelos de resposta ao item policotômicos

Modelos TRI paramétricos são frequentemente estimados usando máxima verossimilhança marginal, na qual quadraturas normais são usadas para resolver as integrais que aparecem na função de verossimilhança. Isto implicitamente assume uma discretização da variável latente (Vermunt, 2001).

Assim, para estimar os parâmetros dos modelos de resposta gradual 2P-GRM(i) e de crédito parcial generalizado 2P(ik)-PCM e, portanto, dos modelos hierarquicamente mais restritivos, podem ser usados os softwares MULTILOG (Thissen, 1991) e PARSCALE (Muraki & Bock, 1996), os quais são parametrizados diferentemente.

Em Childs e Chen (1999) são descritos os procedimentos para obter estimadores comparáveis desses modelos considerados ambos os programas. Tanto MULTILOG como PARSCALE usam o algoritmo EM usando como default uma distribuição normal padrão para a variável latente u; contudo, o número de pontos de quadratura e os critérios de convergência diferem significativamente nos dois programas (Childs & Chen, 1999). Não se dispõe de muitos softwares para estimar os modelos sequenciais; no entanto, o WINMIRA (Von Davier, 1996) pode estimar o modelo 1P-SM.

Um programa que pode ser usado com propósitos acadêmicos para o caso de poucos itens é o LEM (Vermunt, 1997), o qual permite tratar a maioria dos modelos apresentados neste artigo. O programa usa a existência de similaridade entre o modelo de classe latente e a teoria de resposta ao item (Heinen, 1996). Modelos de classe latente de ordem restrita podem ser usados para estimar modelos de resposta ao item policotômicos não paramétricos, enquanto modelos de classe latente log-lineares podem ser usados para estimar modelos de resposta ao item policotômicos paramétricos (Vermunt, 2001). O procedimento de estimação usa o método de máxima verossimilhança e um algoritmo EM é implementado. Recentemente, modelos policotômicos podem ser obtidos em alguns aplicativos no programa R. Um número especial do Journal of Statistical Software foi dedicado à Psicometria utilizando o software R.

Podem ainda ser citados os pacotes erm e ltm. O primeiro lida com modelos tipo Rasch, ou seja, o modelo de escala de classificação (RSM) e sua extensão linear (LRSM), o modelo de crédito parcial (PCM) e sua extensão linear (LPCM) usando estimativa ML condicional. O pacote ltm também ajusta o modelo simples RM. Além disso, implementa funções para estimar os modelos de Birnbaum de dois e três parâmetros com base numa abordagem de máxima verossimilhança marginal, bem como o modelo de resposta graduada para dados policotômicos.

Além disso, sob a perspectiva bayesiana, podem ser citados os pacotes MCMCpack e pscl. Finalmente o programa WinBUGS tem se mostrado uma excelente ferramenta para implementar diversos modelos de resposta ao item como discutidos aqui sob a perspectiva bayesiana, tal como implementado em Curtis (2010).

Uma adaptação do teste de hipótese de Hayter e Tsui para situações de comparações de médias de vários tratamentos com um grupo controle

Carolina Silva Pena[1]
Sueli Aparecida Mingoti[2]

Resumo

Em algumas situações o objetivo é comparar as médias de vários tratamentos independentes através de uma ANOVA e quando a hipótese nula de igualdade entre as médias é rejeitada, comparações dos tratamentos, par a par, são realizadas para averiguar quais desses diferem significativamente. Uma situação particular é aquela em que um dos tratamentos é um grupo controle. Em Estatística Multivariada existem testes estatísticos apropriados para testar o vetor de médias de uma população, como o de Hayter e Tsui. Neste artigo apresenta-se uma adaptação desse teste no contexto de comparações múltiplas com um grupo controle. Resultados da avaliação do desempenho da adaptação proposta são apresentados tendo-se como referência os testes t de Student com a correção de Bonferroni e o de Dunnett. Simulações Monte Carlo foram realizadas para experimentos balanceados e não-balanceados, com três e cinco tratamentos, sob a suposição de normalidade e homocedasticidade considerando-se variâncias conhecidas e desconhecidas.

Palavras-chave: Hayter e Tsui; Comparações múltiplas; Bonferrroni; Dunnett; Poder do teste; Monte Carlo.

[1] Mestrado em Estatística – UFMG 2 Profa. Associada do Depto. de Estatística da UFMG
Universidade Federal de Minas Gerais - Departamento de Estatística da UFMG
End. Av. Antonio Carlos, 6627 – Campus Pampulha Bairro São Luiz – Belo Horizonte – MG - 30161-970
emails: sueliam@est.ufmg.br ou suelimngt@gmail.com

1. Introdução

Existem situações que envolvem a comparação dos efeitos médios de mais de dois tratamentos independentes. Alguns exemplos ocorrem em Ergonomia quando se deseja comparar vários tipos de assentos utilizados por trabalhadores (Werner et al., 2003), em pesquisas de mercado nas quais tem-se amostras de clientes de instituições distintas e deseja-se comparar o perfil desses clientes bem como a opinião desses em relação aos produtos (ou situações) a que foram expostos, ou em estudos de experimentação nos quais se tem processos diferentes de produção e busca-se quais resultam em médias similares, ou médias significativamente maiores, entre outros. No caso de dados com distribuição normal o procedimento mais utilizado é a Análise de Variância (ANOVA) a um fator (VIEIRA, 2005) no qual a igualdade entre as médias de todos os tratamentos é testada através de um teste F de Fisher (Montgomery, 2001). Quando esse teste aponta para a existência de alguma diferença significativa entre as médias dos tratamentos, freqüentemente o pesquisador compara todos os possíveis pares de médias em busca dos tratamentos que contribuiram para a rejeição da hipótese de igualdade entre as médias. Assim, para um conjunto de m tratamentos, é possível realizar até m(m-1)/2 comparações ou (m-1) caso as comparações sejam feitas somente em relação a um grupo controle.

Um teste estatístico muito utilizado para comparações múltiplas é o t de Student quando as variâncias dos tratamentos não são conhecidas, ou o Z da distribuição normal quando as variâncias são conhecidas. Neste caso, um teste de hipóteses é feito para cada comparação de pares de tratamentos, usualmente considerando um nível de significância igual α para cada comparação. Este procedimento, no entanto, deve ser feito com cuidado uma vez que ao analisar todos os testes estatísticos conjuntamente, o nível de significância global do teste conjunto não é preservado no nível de significância nominal pré-estabelecido sendo inflacionado pelo número de comparações efetuadas. Como exemplo, se k testes t de Student independentes forem realizados cada um a um nível de significância α, a probabilidade do erro tipo I considerando-se todos os k testes conjuntamente será dada por: $[1-(1-\alpha)^k]$. Sendo assim, se

para cada comparação das médias for utilizado um valor α igual a 0,05, o nível de significância conjunto do teste será 0,23 para k=5, 0,40 para k=10 e 0,64 para k=20. Para superar tal dificuldade alguns procedimentos estatísticos têm sido propostos para diminuir, ou controlar, a probabilidade de ocorrência do erro tipo I do teste conjunto das comparações múltiplas (também chamada de taxa de erro global do teste). Alguns dos métodos mais utilizados são: o test *t* de Student com a correção de Bonferroni (Dohmen, 2003), o teste de Tukey (1951), o método da diferença mínima significativa de Fisher (1966), o teste de múltipla escala de Duncan (1955), o teste de Newman-Keuls (Steel e Torrie, 1980) e o teste de Dunnett (1955, 1964), dentre outros.

Embora existam vários procedimentos utilizados no contexto de comparações múltiplas não há ainda um consenso geral a respeito de qual deles é o melhor (Girardi et al., 2009). Os métodos que apresentam maior poder de detecção de diferenças reais são os de Fisher (dms) e o de Duncan mas estes procedimentos não mantêm a taxa de erro global no nível nominal α determinado para o teste conjunto (Petersen, 1985). Por outro lado, a correção de Bonferroni e o teste de Tukey controlam a taxa de erro global, o que faz com que muitos pesquisadores optem por sua utilização. O método de Newman-Keuls, é menos poderoso que o teste de Duncan mas apresenta uma probabilidade do erro tipo I global inferior à desse teste. Teoricamente o teste de Dunnett mantém a taxa de erro global no nível α, mas só é aplicável a situações nas quais as comparações são realizadas com um grupo controle.

Em Estatística Multivariada existem alguns testes estatísticos para o vetor de médias de uma população, dentre eles cita-se o teste de Hayter e Tsui (1994) utilizado em controle de qualidade no monitoramento do vetor de médias de processos (ver Timm, 1996). Esse teste mantém a probabilidade do erro tipo I do teste conjunto no nível nominal determinado pelo usuário e identifica automaticamente e com facilidade as variáveis responsáveis pela rejeição da hipótese nula. Estimativas das diferenças das médias das variáveis em relação aos valores postulados na hipótese nula também são obtidas. O teste de Hayter e Tsui leva em consideração a estrutura de correlação entre as variáveis e o seu ponto crítico (que delimita a região de rejeição da hipótese nula) é determinado através de um algoritmo computacional de fácil implementação.

Neste artigo propõe-se uma nova estrutura para utilização do teste de Hayter e Tsui permitindo que ele seja aplicado na comparação de médias de tratamentos par a par, como uma alternativa aos procedimentos de comparações múltiplas univariados já existentes. Uma avaliação do desempenho dessa metodologia será efetuada através de simulações Monte Carlo, considerando-se a comparação de três e cinco tratamentos com um grupo controle, para dados provenientes de distribuições normais com variâncias iguais de (homocedasticidade), sendo essas conhecidas ou não. Os resultados obtidos para o teste de Hayter e Tsui serão comparados aos da correção de Bonferroni aplicada ao teste t de Student (quando a variância da variável resposta for desconhecida) e ao teste Z da distribuição normal (quando a variância for conhecida), bem como ao teste de Dunnett que é um dos mais utilizados em situações em que se tem 5 ou mais tratamentos comparados a um grupo controle.

2. Comparações múltiplas: testes estatísticos univariados

Nesta seção descreve-se brevemente os testes estatísticos Z da distribuição normal e t de Student com as correções de Bonferroni, bem como o teste de Dunnett, utilizados para comparação de médias de grupos independentes, sendo o teste de Dunnett apropriado apenas para a comparação de médias de tratamentos com a de um grupo controle.

2.1 Testes Z da distribuição normal e t de Student com a correção de Bonferroni

Suponha que se tenha apenas dois tratamentos Ti e Tj, e o interesse seja o de comparar as suas médias tendo-se disponível uma amostra aleatória da variável resposta *X* de interesse de cada um dos tratamentos e sendo essas independentes entre si. Suponha que para o tratamento Ti, *X* tenha distribuição normal com média e variância iguais a μ_i e σ_i^2, e que para o tratamento Tj, *X* tenha distribuição normal com média e variância iguais a μ_j e σ_j^2, respectivamente, sendo $\sigma_i^2 = \sigma_j^2 = \sigma^2$. Sejam

$H_0 : \mu_i = \mu_j$ e $H_a : \mu_i \neq \mu_j$ as hipóteses nula e alternativa do teste $i \neq j$. Para o caso em que σ^2 é conhecido a estatística de teste a ser utilizada é definida como em (1) onde $\bar{X}_i$ e $\bar{X}_j$ são as médias amostrais das observações dos tratamentos Ti e Tj e n_i e n_j os tamanhos das amostras dos dois tratamentos, respectivamente.

$$z_0 = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{(\frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_j})}} \qquad (1)$$

Sob a hipótese nula a estatística de teste z_0 dada em (1) tem distribuição normal com média 0 e variância igual a 1. Sendo assim, a hipótese nula é rejeitada para valores de z_0 tais que $|z_0| > z_{\frac{\alpha}{2}}$, onde $z_{\frac{\alpha}{2}}$ é o valor da distribuição normal padronizada tal que $P[Z > z_{\frac{\alpha}{2}}] = \frac{\alpha}{2}$, $0 < \alpha < 1$. Na situação em que a variância σ^2 é desconhecida a estatística de teste em (1) é substituída pela (3) sendo σ^2 estimada como mostrado em (2) onde s_i^2 e s_j^2 são as variâncias amostrais dos tratamentos Ti e Tj, respectivamente.

$$s_c^2 = \frac{(n_i - 1)s_i^2 + (n_j - 1)s_j^2}{n_i + n_j - 2} \qquad (2)$$

Sob a hipótese nula a estatística de teste dada como em (3) tem distribuição *t* de Student com $n_i + n_j - 2$ graus de liberdade, ($t_{n_i+n_j-2}$), e a hipótese nula é rejeitada para valores de *t* tais que $|t| > t_{n_i+n_j-2,\frac{\alpha}{2}}$, em que $P[t_{n_i+n_j-2} > t_{n_i+n_j-2,\frac{\alpha}{2}}] = \frac{\alpha}{2}$, $0 < \alpha < 1$.

$$t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{(\frac{s_c^2}{n_i} + \frac{s_c^2}{n_j})}} \qquad (3)$$

No caso em que se tem amostras observadas de vários tratamentos independentes e o interesse é o de comparar pares de médias desses tratamentos é comum utilizar-se

os testes Z da distribuição normal e o *t* de student na comparação de cada par de médias populacionais, aplicando a correção de Bonferroni no nível de significância de cada teste. Suponha que tenhamos *m* tratamentos independentes, $m>2$, e *k* pares de médias de tratamentos a serem comparadas, sendo as observações de cada tratamento provenientes de uma distribuição normal com variância σ^2. Seja α o nível de significância nominal global do teste, ou seja o nível de significância quando se considera todas as *k* comparações simultaneamente ou conjuntamente, $0<\alpha<1$. Nessa situação para cada um dos pares de tratamentos (Ti e Tj) a serem comparados, $i \neq j$, utiliza-se a estatísticas de teste (1) se a variância σ^2 for conhecida ou a estatística de teste (3) quando a variânca for desconhecida, sendo σ^2 estimado como em (4) que é o Quadrado Médio Residual (*QMR*) proveniente do ajuste da análise de variância paramétrica (ANOVA) para comparação das médias dos *m* tratamentos, sendo s_i^2 as variâncias amostrais dos tratamentos Ti, $i=1,2,\ldots,m$. Sob a hipótese nula de igualdade entre as médias dos tratamentos comparados, a estatística de teste em (3) terá distribuição *t* de Student com *(N-m)* graus de liberdade, sendo $N=\sum_{i=1}^{m} n_i$ o número total de observações amostrais entre todos os tratamentos. No entanto, para que o nível global de significância (α) seja mantido de acordo com o pré-especificado pelo pesquisador, cada comparação de médias deverá ser feita a um nível de significância nominal igual a α/k. Pela desigualdade de Bonferroni (Bain e Engelhart, 2000) o uso dessa correção assegura que a probabilidade do erro tipo I do teste conjunto é no máximo igual ao valor α pré-especificado inicialmente. Se o interesse é a comparação de todos os pares de tratamentos o valor de *k* será igual a $m(m-1)/2$. Para $m=2$ o estimador da variância σ^2 dado em (2) é um caso particular do estimador dado por (4).

$$QMR=\frac{\sum_{i=1}^{m}(n_i-1)s_i^2}{(\sum_{i=1}^{m} n_i)-m} \qquad (4)$$

Quando o número de comparações (*k*) de médias de tratamentos é elevado o

uso dos testes Z da distribuição normal ou *t* de Student com a correção de Bonferroni pode levar a conclusões equivocadas uma vez que, como o nível de significância nominal α é repartido igualmente entre o número de comparações, o nível de significância individual (α/k) para cada par de tratamentos fica muito pequeno e diferenças reais entre as médias de tratamentos passam desapercebidas na análise estatística, ou seja o teste de hipóteses conjunto torna-se menos poderoso.

2.2 Teste de Dunnett

Em algumas situações o pesquisador não está interessado na comparação de médias de todos os pares de tratamentos mas sim, na comparação da média de cada tratamento pesquisado em relação a um tratamento específico, chamado de controle ou padrão. O número de comparações múltiplas (k) passa ser igual ao número de tratamentos menos 1. Embora os testes Z da distribuição normal e *t* de Student descritos na seção anterior ainda possam ser empregados nessa situação, em geral o teste de Dunnett (1955,1964;1980) é o mais utilizado.

Suponha que tenhamos m tratamentos independentes sendo T1 o tratamento (ou grupo) controle e as observações de cada tratamento Ti provenientes de uma distribuição normal com média μ_i e variância σ^2, $i=1,2,...m$. Sejam $H_0 : \mu_2 = \mu_3 = ... = \mu_m = \mu_1$ e $H_a : \mu_i \neq \mu_1$, *para algum tratamento i*, $i=2,3,...,m$, as hipóteses nula e alternativa, e α o nível de significância nominal global do teste, $0 < \alpha < 1$. Considere t_i, $i=2,3,...m$, as estatísticas definidas em (5)

$$t_i = \frac{\overline{X}_i - \overline{X}_1}{\sqrt{QMR\,(\frac{1}{n_i} + \frac{1}{n_1})}} \qquad (5)$$

De acordo com o teste de Dunnett a hipótese nula será rejeitada se $|t_i| > d_{\alpha,m-1,N-m}$, para algum *i*, *i=2,3,...m*, onde N-m é o número de graus de liberdade dos resíduos da ANOVA ajustada para comparação das médias dos *m* tratamentos, $N = \sum_{i=1}^{m} n_i$. A constante $d_{\alpha,m-1,N-m}$ foi tabelada por Dunnett (1955,1964;1980) e está

relacionada com a distribuição *t* de Student multivariada (ver Hsu, 1996). Quando os grupos comparados são balanceados, isto é $n_i = n$, *i=1,2,...,*m, a estatística em (5) torna-se

$$t_i = \frac{\overline{X}_i - \overline{X}_1}{\sqrt{\frac{2\ QMR}{n}}}$$

sendo QMR definido como em (4). As médias dos tratamentos Ti para os quais $|\ t_i\ | > d_{\alpha,m-1,N-m}$ são consideradas significativamente diferentes da média do tratamento denominado controle.

É interessante observar que quando aplicado ao caso de comparações múltiplas com um grupo controle, os valores numéricos da estatística do teste *t* de Student (em valor

absoluto) serão os mesmos obtidos no teste de Dunnett. No entanto, para um mesmo nível de significância global α, $0 < \alpha < 1$, os valores críticos para rejeição da hipótese de igualdade entre as médias de tratamentos diferem.

3. Teste de Hayter e Tsui para comparações múltiplas

Nesta seção apresentamos uma proposta de adaptação do teste de Hayter e Tsui (1994) para o caso em que se tem comparações múltiplas de médias de tratamentos com a de um grupo controle. A idéia básica é transformar o problema univariado de comparações de médias de vários grupos independentes num problema multivariado de comparação do vetor de médias de uma população, no qual os valores do respectivo vetor representam as diferenças entre as médias dos respectivos tratamentos que estão sendo comparados. Para um melhor entendimento da proposta a ser apresentada, vamos introduzir inicialmente o teste convencional de Hayter e Tsui utilizado para testar o vetor de médias de uma população normal *p*-variada, $p \geq 2$.

3.1 Teste de Hayter e Tsui convencional para o vetor de médias

Seja $\boldsymbol{X} = (X_1, X_2, ..., X_p)'$ um vetor aleatório com distribuição normal p-variada com vetor de médias $\boldsymbol{\mu}$ e matriz de correlação $\boldsymbol{P_{pxp}}$, $p \geq 2$. Considere a hipótese nula $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ contra a alternativa $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$, sendo $\boldsymbol{\mu_0} = (\mu_1^0, \mu_2^0, ..., \mu_p^0)'$. Seja $C_{R,\alpha}$ uma constante tal que

$$P\left(\left| \frac{X_i - \mu_i^0}{\sigma_i} \right| \leq C_{R,\alpha},\ para\ todo\ i = 1,2,..., p \right) = 1 - \alpha\ ,\ 0 < \alpha < 1 \qquad (6)$$

onde μ_i^0 e σ_i denotam a média e o desvio padrão, respectivamente, da variável aleatória X_i, $i = 1,2,...,p$. Seja $X_1, X_2, ..., X_n$, uma amostra aleatória do vetor $\boldsymbol{X}$, onde $\boldsymbol{X_j} = (X_{j1}, X_{j2}, ..., X_{jp})'$, $j = 1,2,...,n$, $n > p$. Considere a estatística M definida como:

$$M = max_{1 \leq i \leq p} \left| \frac{\overline{X}_{.i} - \mu_i^0}{\sigma_i / \sqrt{n}} \right| \qquad (7)$$

sendo $\overline{X}_{.i}$ a média amostral da variável aleatória i, $i=1,2,..,p$. A hipótese nula será rejeitada quando o valor amostral observado da estatística M for maior que a constante crítica $C_{R,\alpha}$.

Hayter e Tsui (1994) mostram que este teste estatístico mantém a taxa de erro global exatamente no nível de significância nominal α pré-estabelecido para o teste conjunto e que é um bom competidor ao conhecido teste estatístico multivariado T^2 de Hotelling (Montgomery, 2004). O ponto crítico do teste de Hayter e Tsui ($C_{R,\alpha}$) não está tabelado para $p \geq 3$, e seu valor é determinado através de um algoritmo de simulação, cujos passos são descritos a seguir:

Passo 1. Gera-se um grande número de observações (por exemplo, $N^*=100.000$) provenientes de uma distribuição normal p-variada com vetor de médias zero e matriz de covariâncias $\boldsymbol{P}$, isto é, $\boldsymbol{Z}^{1}, \boldsymbol{Z}^{2}, \ldots, \boldsymbol{Z}^{N^*}$.

Passo 2. Calcula-se a estatística M para cada vetor de observações, $\boldsymbol{Z}^{j} = (Z_1^j, Z_2^j, \ldots, Z_p^j)'$, isto é, $M^j = max_{1 \leq i \leq p} \left| Z_i^j \right|$, para $j=1,2,\ldots,N^*$.

Passo 3. A estimativa do ponto crítico $C_{R,\alpha}$ é dada pelo valor correspondente ao percentil de ordem $(1-\alpha)$ da distribuição empírica da estatística M construída com os N^* valores amostrais $(M^1, M^2, \ldots, M^{N^*})$.

É importante observar que o algoritmo para a determinação da constante crítica $C_{R,\alpha}$ é fundamentado na simulação de valores de uma distribuição normal multivariada na qual o vetor de médias de $\boldsymbol{Z}^{j}$ é zero e a matriz de covariâncias $\boldsymbol{P}$ de $\boldsymbol{Z}^{j}$ é a matriz de correlações do vetor aleatório original $\boldsymbol{X}$. Deste modo, a correlação entre as variáveis originais é levada em consideração na determinação da região crítica do teste. No caso em que o desvio padrão da variável X_i, $i=1,2,\ldots,p$, for desconhecido o valor de σ_i em (7) é substituído por sua estimativa amostral e a matriz de correlações P é estimada pela matriz de correlações amostrais R (Johnson e Wichern, 2002). Assim para a determinação da constante $C_{R,\alpha}$ a matriz R será usada no algoritmo apresentado anteriormente. O teste de Hayter e Tsui tem sido explorado em vários trabalhos publicados na literatura, dentre eles cita-se Kalgonga e Kulkarni (2004) e Jiang et al. (2006).

3.2 Proposta de adaptação do teste de Hayter e Tsui para comparações múltiplas com um grupo controle

Nesta seção apresenta-se uma proposta de adaptação do teste de Hayter e Tsui para o caso em que o objetivo é a comparação de médias de vários tratamentos com a de um grupo (tratamento) controle.

Seja X a variável aleatória de interesse e suponha que X tenha sido medida em m grupos independentes (m tratamentos). Denote por X_i a variável X avaliada no tratamento T_i e por $\boldsymbol{X}^{\boldsymbol{i}} = (X_{i1}, X_{i2}, ..., X_{in_i})'$ as observações amostrais da variável X_i, sendo n_i o tamanho da amostra correspondente, $i=1,2,...,m$. Seja T_1 o tratamento controle e suponha que para cada tratamento T_i, X_i tenha distribuição normal com parâmetros (μ_i, σ_i^2), $i= 1, 2,...,m$. Sejam as variáveis aleatórias W_{im} definidas por:

$$W_{im} = X_1 - X_i \quad , \; i = 2,...,m \tag{8}$$

A partir de (8) o teste de Hayter e Tsui será definido para a comparação das médias dos tratamentos com a de um grupo controle (o grupo que gera os valores da variável X_1). Suponha que $\sigma_i^2 = \sigma^2$, para todo $i = 1,2,...,m$. Neste caso, pode-se proceder a comparação do tratamento T_1 em relação aos (m-1) restantes usando o teste de Hayter e Tsui da forma como descrito nos passos (A) e (B) a seguir.

Passo (A) inicialmente considera-se o problema como se tivéssemos o vetor aleatório $\boldsymbol{W_m} = (W_{2m}, W_{3m}, ..., W_{mm})'$, sendo W_m com distribuição normal $(m-1)$ variada com vetor de médias $\boldsymbol{\mu_m^*} = (\mu_1 - \mu_2, \mu_1 - \mu_3, ..., \mu_1 - \mu_m)'$ e matrizes de covariâncias e correlações dadas respectivamente por:

$$\Sigma = \begin{bmatrix} 2\sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & 2\sigma^2 & \cdots & \sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \cdots & \sigma^2 & 2\sigma^2 \end{bmatrix} \quad \text{e} \quad P = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \cdots & \frac{1}{2} \\ \vdots & \cdots & \ddots & \vdots \\ \frac{1}{2} & \frac{1}{2} & \cdots & 1 \end{bmatrix} \tag{9}$$

pois, $var(W_{im}) = Var(X_1) + Var(X_i) - 2Cov(X_1, X_i) = 2\sigma^2$

e

$$cov(W_{im}, W_{km}) = cov(X_1 - X_i, X_1 - X_k) = Var(X_1) - Cov(X_1, X_k) - Cov(X_i, X_1) + Cov(X_i, X_k)$$
$$= Var(X_1) \ , \ i \neq k,$$

uma vez que $Cov(X_1, X_i) = Cov(X_1, X_k) = Cov(X_i, X_k) = 0$, pois os tratamentos T_1, T_i e T_k são independentes, $k \neq i \neq 1$.

Passo (B): considere as hipóteses nula e alternativa como: $H_0 : \boldsymbol{\mu_m^* = 0}$ e $H_a : \boldsymbol{\mu_m^* \neq 0}$. De acordo com o procedimento de Hayer e Tsui (1994) descrito na seção 3.1 a estatística de teste para testar a hipótese nula será dada por

$$M = max(M_{im}, i = 2,3,...,m) \tag{10}$$

sendo $M_{im} = \left| \dfrac{\overline{W}_{im} \ -0}{\sqrt{Var\left(\overline{W}_{im}\right)}} \right|$ e $\overline{W}_{im} = \overline{X}_1 - \overline{X}_i, i = 2,...,m$.

Seja α o nível de significância estabelecido para o teste conjunto. Se o valor da estatística de teste *M* ultrapassar o valor crítico $C_{R,\alpha}$ estipulado para o teste, a hipótese nula de igualdade entre as médias dos vários tratamentos com a do grupo controle é rejeitada sendo as diferenças responsáveis pela rejeição as que resultaram em valores de M_{im} maiores que o valor de $C_{R,\alpha}$. O valor dessa constante crítica é obtido por meio de simulação (ver algoritmo descrito na seção 3.1) de uma distribuição normal *(m-1)* variada com vetor de médias nulo e matriz de covariâncias ***P*** de dimensão *(m-1)x(m-1)* dada como em (9). A variância de $\overline{W}_{im}$ é dada respectivamente por

$$Var(\overline{W}_{im}) = Var(\overline{X}_1 - \overline{X}_i) = Var(\overline{X}_1) + Var(\overline{X}_i) - 2cov(\overline{X}_1, \overline{X}_i)$$
$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_i^2}{n_i}$$

(11)

pois $Cov(\overline{X}_1, \overline{X}_i) = 0$, para todo *i=2,3,...,m*. Quando os tratamentos tem a mesma variância a estatística em (11) torna-se

$$Var(\overline{W}_{im}) = \sigma^2 (\frac{1}{n_1} + \frac{1}{n_i})$$

Em situações nas quais os valores das variâncias dos tratamentos são desconhecidos a variância de $\overline{W}_{im}$ pode ser estimada por

$$\hat{V}ar(\overline{W}_{im}) = \frac{s_1^2}{n_1} + \frac{s_i^2}{n_i}$$

em que s_1^2 representa a variância amostral da variável X_1 (grupo controle) e s_i^2 representa a variância amostral da variável X_i (tratamento *i*), $i \neq 1$. No caso em que as variâncias dos *m* tratamentos são iguais mas desconhecidas, um estimador melhor para a variância de $\overline{W}_{im}$ é dado por

$$\hat{V}ar(\overline{W}_{im}) = QMR(\frac{1}{n_1} + \frac{1}{n_i}) \qquad (12)$$

sendo *QMR* o quadrado médio residual definido como em (4) (seção 2.1).

É importante notar que a adaptação do teste de Hayter e Tsui proposta nos passos (A) e (B) é um teste conjunto para as diferenças de médias de todos os tratamentos com a de um grupo controle e mantém o nível de significância global no nível nominal pré-estabelecido para o teste. Pela estrutura da estatística de teste definida em (10), além de testar conjuntamente as diferenças de interesse entre as médias de tratamentos, essa adaptação também fornece os resultados da comparação individual entre as médias de cada par de tratamentos indicando automaticamente quais pares foram responsáveis pela rejeição da hipótese nula, quando isso ocorrer.

4. Exemplo de aplicação

Apresenta-se nesta seção um exemplo de aplicação da adaptação do teste de Hayter e Tsui proposta neste artigo em um banco de dados real. Os dados analisados são referentes ao conteúdo de açúcar (em gramas) observado em 6 marcas de cereais matinais sendo o cereal da marca 1 considerado como referência em termos nutricionais. Oito pacotes de cada marca foram selecionados aleatoriamente. Os valores observados de conteúdo de açúcar (a cada 30 gramas) estão apresentados na Tabela 1. Pelos testes estatísticos de Anderson Darling (1952) e Bartlet (1937) verificou-se que os dados atendiam a suposições de normalidade (p-valores acima de 0,15) e homocedasticidade (p-valor = 0,866). Como o verdadeiro valor da variância populacional não é conhecido os valores das estatísticas de teste foram obtidos utilizando-se a variância estimada de acordo com (4), seção 2.1.

Seja X_{ij} o conteúdo de açúcar observado na amostra *j* da marca *i*, e seja μ_i o conteúdo médio de açúcar da marca *i* , *i = 1,2,...,6* e *j=1,2,..,8*. A probabilidade de significância obtida na ANOVA a um fator foi próxima de zero indicando a existência de diferença significativa entre os conteúdos médios de açúcar das marcas de cereal. Desta forma, realizou-se as comparações mútiplas sendo cada marca comparada com a marca 1 considerada nesse exemplo como grupo controle. O nível de significância global utilizado no teste conjunto foi 5%.

Tabela 1. Conteúdo de açúcar (a cada 30 g) observado em 6 marcas de cereais matinais

	Marcas de Cereal Matinal					
Amostras	M1	M2	M3	M4	M5	M6
1	4,05	8,73	7,18	5,05	10,17	6,54
2	7,00	5,99	6,27	4,67	9,31	10,18
3	2,86	6,56	4,72	8,25	11,46	4,14
4	4,53	7,57	4,76	7,64	10,60	6,95
5	7,15	9,92	7,37	7,56	11,59	6,75
6	7,14	5,82	5,97	6,56	9,78	5,76
7	6,34	7,98	5,39	3,34	9,72	6,98
8	6,15	8,32	7,13	5,43	6,49	8,01
Média	5,653	7,611	6,099	6,062	9,890	6,914
Desvio-padrão	1,631	1,422	1,075	1,713	1,598	1,735
Variância	2,660	2,022	1,156	2,934	2,554	3,010
p-valor (*)	0,156	0,816	0,372	0,664	0,174	0,355

(*) Teste de Anderson-Darling para normalidade.

O primeiro passo para a aplicação do teste de Hayter e Tsui consiste em calcular o ponto crítico $C_{R,\alpha}$ com base no algoritmo apresentado na seção 3.1. Para um conjunto de 6 tratamentos, sendo um deles o grupo controle, o valor de $C_{R,\alpha}$ obtido pelo algoritmo de simulação foi igual a 2,516 considerando-se a matriz de correlações ***P*** de dimensão 5x5 como dada em (9). O segundo passo é determinar os valores das variáveis $\overline{W}_{im}$ para que se possa calcular o valor da estatística de teste *M*. Como ilustração mostramos os cálculos relativos à comparação do conteúdo médio de açúcar da marcas M1 e M2. Tem-se que:

$$\overline{W}_{26} = 5{,}653 - 7{,}611 = -1{,}958$$

$$QMR = \frac{7(\,2{,}660 + 2{,}022 + 1{,}156 + 2{,}934 + 2{,}554 + 3{,}010\,)}{48-6} = 2{,}3893$$

$$\hat{V}ar(\overline{W}_{26}) = \frac{2{,}3893}{8} + \frac{2{,}3893}{8} = 0{,}5973$$

e portanto, $M_{26} = \left| \frac{-1{,}958 - 0}{\sqrt{0{,}5973}} \right| = 2{,}533$.

O método de Bonferroni por sua vez, consiste na aplicação do teste t de Student para comparação de médias de dois grupos independentes com variâncias iguais sendo o nível de significância usado para a comparação de cada par de médias igual a $0{,}05/5 = 0{,}01$, já que neste exemplo tem-se apenas 5 comparações a serem feitas. Para a comparação das médias das marcas M1 e M2 o valor observado da estatística *t* de Student é dado por:

$$t_{obs} = \frac{(\,\bar{x}_1 - \bar{x}_2\,)}{\sqrt{QMR\,(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{(5{,}653 - 7{,}611)}{\sqrt{(2{,}3893)(\frac{2}{8})}} = \frac{-1{,}958}{\sqrt{0{,}5973}} = -\,2{,}533$$

O valor crítico da distribuição *t* de Student para rejeição da hipótese de igualdade entre as duas médias é: $t_{N-m\,;1-\frac{0{,}01}{2}} = t_{42;\,0{,}995} = 2{,}698$, ou seja a hipótese nula é rejeitada para valores absolutos de t_{obs} maiores que 2,698. Como o valor 2,533 não está na região de rejeição

considera-se que não há diferença significativa entre os conteúdos médios de açúcar dos cereais das marcas M1 e M2. Para o teste de Dunnett o valor da estatística de teste é o mesmo do teste de Hayter e Tsui e o valor crítico para região de rejeição da hipótese nula é igual a 2,61. Sendo assim, por esse teste as marcas de cereal M1 e M2 também são consideradas semelhantes em relação ao conteúdo médio de açúcar. É importante observar que o valor numérico da estatística de teste (em valor absoluto) para comparação dos dois tratamentos é o mesmo nos três testes, no entanto as regiões de rejeição da hipótese nula são diferentes. O valor crítico do teste de Hayter e Tsui adaptado é menor que o valor crítico do teste de Dunnett, que por sua vez é menor que o valor crítico do teste de *t* de Student.

Para as demais comparações de interesse, os cálculos foram feitos de forma similar. A Tabela 2 apresenta os resultados obtidos para todas as comparações utilizando os três testes considerados nesse artigo. O valor da estatística *M* do teste de Hayter e Tsui é 5,482.

Tabela 2: Resultados das comparações múltiplas – Testes t-Student, Dunnett e Hayter e Tsui

Marcas Comparadas	Diferença entre médias	Estatística de teste (*)	t de Student - Bonferroni	Dunnett	Hayter e Tsui
M1 vs. M2	-1,958	2,533	Não rejeita H_0	Não rejeita H_0	Rejeita H_0
M1 vs. M3	-0,446	0,577	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
M1 vs M4	-0,409	0,529	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
M1 vs M5	-4,237	5,482	Rejeita H_0	Rejeita H_0	Rejeita H_0
M1 vs. M6	-1,261	1,632	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0

(*) Em valor absoluto; Valores críticos dos testes: $t_{42;0,995} = 2,698; C_{R,0,05} = 2,516;\ Dunnett = 2,61$.

Observa-se que o teste de Hayter e Tsui rejeita a hipótese nula $H_0 : \boldsymbol{\mu_6^* = 0}$, sendo $\boldsymbol{\mu_m^*} = (\mu_1 - \mu_2, \mu_1 - \mu_3, \mu_1 - \mu_4, \mu_1 - \mu_5, \mu_1 - \mu_6)'$, pois o valor observado da estatística de teste (M = 5,482) é maior que o valor crítico $C_{R,0,05} = 2,516$, sendo as comparações da marca M1 com as marcas M2 e M5 as responsáveis pela rejeição. As conclusões obtidas pelos três testes são concordantes na rejeição da igualdade das médias das marcas M1 e M5 mas diferem no que se refere a comparação dos conteúdos médios de açúcar das marcas M1 e M2, sendo que o teste de Hayter e Tsui aponta para a

existência de uma diferença significativa entre as médias enquanto os outros dois testes indicam a não rejeição da hipótese nula de igualdade entre elas. Destaca-se que o valor observado da estatística de teste para a comparação dessas duas marcas foi bem próximo do valor crítico do teste de Dunnett mas menos próximo do valor crítico do teste *t* de Student. Para as demais comparações as conclusões obtidas a partir dos três testes estatísticos são semelhantes.

5. Simulação

Com intuito de avaliar o desempenho da adaptação do teste de Hayter e Tsui proposta neste artigo, realizou-se um estudo via simulação Monte Carlo no qual trabalhou-se com um conjunto de m = 3 e 5 tratamentos sendo o tratamento 1 o grupo controle. Assim, tem-se dois e quatro testes a serem realizados para comparações múltiplas, respectivamente. Seja *X* a variável resposta de interesse e seja X_i a variável *X* medida nas n_i unidades amostrais do tratamento *i*. Suponha que X_i tenha distribuição normal univariada com média μ_i e variância $\sigma_i^2 = \sigma^2$, $i = 1, 2, \ldots, m$. Tem-se portanto, *m* grupos independentes. As hipóteses nula e alternativa consideradas nas simulações são: $H_0 : \boldsymbol{\mu}_{\mathbf{m}}^* = \mathbf{0}$ e $H_a : \boldsymbol{\mu}_{\mathbf{m}}^* \neq \mathbf{0}$, sendo $\boldsymbol{\mu}_{\mathbf{m}}^* = (\mu_1 - \mu_2, \mu_1 - \mu_3, \ldots, \mu_1 - \mu_m)'$. Foram realizadas 10.000 simulações Monte Carlo considerando diversos valores de médias μ_i e tamanhos amostrais n_i, $i = 1, 2, \ldots, m$. Em cada simulação os três testes estatísticos discutidos nesse artigo foram aplicados à amostra observada. Ao final das 10.000 simulações calculou-se, para cada teste, a proporção de vezes em que a hipótese nula havia sido rejeitada. Quando os dados foram gerados da distribuição normal sob a hipótese nula (ou seja, todos os tratamentos com médias iguais à do grupo controle), essa proporção representava uma estimativa da probabilidade do erro tipo I do teste conjunto (erro global) e quando os dados foram gerados sob a hipótese alternativa ou seja, a média do grupo controle, μ_1, era diferente da média de algum outro tratamento $\mu_k, (k \neq 1)$, a proporção representava uma estimativa do poder do teste. O nível de significância nominal utilizado para todos os testes comparados foi fixado em

$\alpha = 0{,}05$. Para um conjunto de (m-1) comparações e matriz de correlações teórica igual a ***P***, o cálculo da constante crítica $C_{R,\alpha}$ foi feito de acordo com o algoritmo descrito na seção 3.1, considerando-se a simulação de $N^* = 100.000$ observações de uma distribuição normal *(m-1)* variada com vetor de médias nulo e matriz de covariâncias ***P*** de dimensão *(m-1)x(m-1)* dada como em (9). O valor $N^* = 100.000$ é o sugerido no artigo de Hayter e Tsui (1994), no entanto Mingoti e Glória (2003, 2005) mostraram que o valor da constante $C_{R,\alpha}$ pode ser determinado com boa precisão a partir de simulações com $N^* = 10.000$, sendo pequena a variabilidade entre os valores de $C_{R,\alpha}$ obtidos a partir de simulações diferentes, para valores elevados de N^*. Desta forma, para o estudo apresentado nessa seção, uma única simulação foi realizada para a determinação da constante $C_{R,\alpha}$, para cada valor de *m* fixo. Após determinado esse valor calculou-se o valor observado da estatística do teste de Hayter e Tsui como em (10) sendo a hipótese nula rejeitada sempre que $M > C_{R,\alpha}$. No caso em que as variâncias dos tratamentos eram conhecidas, a variância de $\overline{W}_{im}$ foi calculada como em (11) sendo estimada como em (12) quando as variâncias eram desconhecidas.

Na seção 3.2 mostramos que quando os grupos são homocedásticos, a matriz de correlações teórica dada em (9) tem a mesma estrutura independente do fato da variância ser ou não conhecida. Sendo assim, fixado o número de testes a serem realizados, o valor do ponto crítico $C_{R,\alpha}$ do teste de Hayter e Tsui não sofre alterações, sendo o mesmo para situações em que se conhece ou não a variância populacional. No caso da correção de Bonferroni, realizou-se o teste *t* de Student ou o teste Z da distribuição normal para comparação de duas médias sendo que cada comparação foi efetuada no nível de significância nominal de $\alpha/(m-1)$, isto é, 0,025 quando $m = 3$ e 0,0125 quando *m*=5. O teste de Dunnett foi realizado apenas para as situações em que as variâncias eram desconhecidas.

5.1 Resultados - Comparação de três tratamentos

Nesta seção são apresentados os resultados obtidos para grupos balanceados e não-balanceados nas situações em que se utiliza o verdadeiro valor da variância no cálculo da estatística de teste (variância populacional conhecida), como também nas situações em que o valor da variância é estimado pelos dados amostrais (variância desconhecida). O valor do ponto crítico $C_{R,\alpha}$ do teste de Hayter e Tsui foi determinado através da geração aleatória de $N^* = 100.000$ valores de uma distribuição normal bivariada com vetor de médias nulo e matriz de covariâncias igual a $\boldsymbol{P} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$. Para cada vetor amostral gerado foi calculado o valor da estatística de teste M construindo-se então, a distribuição empírica dessa estatística para determinação do valor de $C_{R,\alpha}$. A título de ilustração a distribuição empírica da estatística M é apresentada na Figura 1 com o respectivo valor crítico $C_{R,\alpha}$ $(= 2{,}199 \approx 2{,}20)$ determinado para $\alpha = 0{,}05$. Na Tabela 3 são apresentados os valores da constante $C_{R,\alpha}$ para alguns valores de m e alguns níveis de significância, obtidos utilizando-se o algoritmo da seção 3.1 para $N^* = 100.000$.

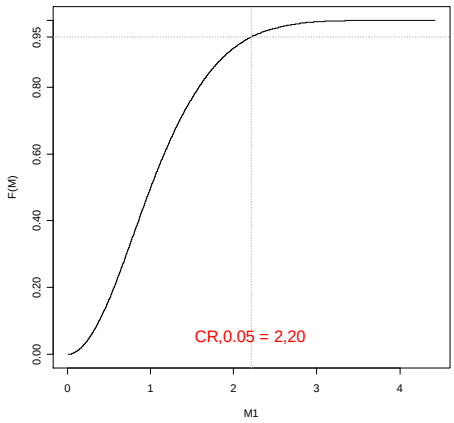


Figura 1: Função de distribuição acumulada empírica da estatística M

Tabela 3: Valores críticos ($C_{R,\alpha}$) do teste de Hayter e Tsui para comparação de m tratamentos

	Nível de significância (%)			
m	**1**	**2,5**	**5**	**10**
3	2,787	2,457	2,199	1,923
4	2,918	2,625	2,361	2,061
5	3,015	2,709	2,455	2,171
6	3,066	2,767	2,516	2,241
7	3,111	2,817	2,579	2,300
8	3,144	2,864	2,629	2,361
9	3,184	2,891	2,655	2,382
10	3,215	2,929	2,675	2,413

O estudo foi feito para situações em que todos os tratamentos apresentavam a mesma média e para situações em que os grupos testados apresentavam algum deslocamento na média em relação ao grupo controle. No primeiro caso a rejeição da hipótese nula configurava o erro tipo I do teste, sendo de interesse avaliar o resultado global do teste, isto é, a proporção de vezes em que a hipótese nula havia sido rejeitada em pelo menos uma das *(m-1)=2* comparações efetuadas entre as médias dos tratamentos. Além disso, calculou-se a proporção de vezes em que a hipótese nula foi rejeitada especificamente na comparação dos tratamentos 1 e 2 (X_1, X_2) e dos tratamentos 1 e 3 (X_1, X_3). Sob a hipótese alternativa avaliou-se a proporção de vezes em que a hipótese nula foi rejeitada especificamente nas comparações em que o tratamento 1 diferia das médias dos outros dois tratamentos, de acordo com o modelo simulado, ou seja o poder do teste. Além disso, calculou-se a razão entre as estimativas de poder do teste de Hayter e Tsui e os poderes dos testes *t* de Student (ou Z-Normal) e Dunnett. Todas as amostras simuladas sob a hipótese nula e alternativa foram procedentes da distribuição normal com variância igual a 1. Embora várias configurações diferentes tenham sido simuladas, neste artigo são mostrados apenas os resultados referentes a situação em que $\mu_1 = \mu_2 = \mu_3 = 0$ sob a hipótese nula e $\mu_1 = 0, \mu_2 = 0{,}25, \mu_3 = 0{,}5$ sob a hipótese alternativa, já que para os outros cenários simulados as conclusões gerais foram semelhantes.

Para grupos balanceados e com variâncias desconhecidas, a Tabela 4 mostra que apenas para amostras pequenas ($n = 5$), o teste de Hayter e Tsui não manteve o erro global próximo ao nível de significância nominal especificado de 5%. Para os

tamanhos amostrais em que os três testes apresentaram estimativas da probabilidade do erro tipo I semelhantes e próximas a *0,05*, observa-se que o teste de Hayter e Tsui foi sempre mais poderoso que os testes t de Student com a correção de Bonferroni e Dunnett, havendo um ganho de poder considerável para $n = 10$ (ganhos de poder > 33% em relação ao *t* de Student e > 20% em relação ao Dunnett) e para $n = 20$ (ganhos de poder entre 14 e 27% em relação ao *t* de Student e de 8 a 15% em relação ao Dunett)e para $n = 20$ (ganhos de poder entre 14 e 27% em relação ao *t* de Student e de 8 a 15% em relação ao Dunnett). Para $n = 50$ os dois testes apresentaram poderes próximos mas ainda com uma certa vantagem para o teste de Hayter e Tsui principlamente quando comparado ao teste *t* de Student. Embora não possamos comparar diretamente os três testes para $n = 5$ já que a estimativa do erro global do teste de Hayter e Tsui está em torno de 8,5%, é interessante notar que o seu ganho de poder em relação ao teste *t* de Student é bem considerável (acima de 88%), o mesmo ocorrendo em relação do teste de Dunnett (ganhos de poder entre 54 e 73%). Os testes de Dunnett e *t* de Student apresentaram estimativas de poder semelhantes com uma ligeira vantagem para o teste de Dunnett.

Para amostras não-balanceadas e com variâncias desconhecidas, observa-se pela Tabela 5 que em todos os casos considerados o teste de Hayter e Tsui apresentou estimativas do erro global ao redor de 7%, valor acima do nível de significância nominal estabelecido de 5%. O mesmo foi observado para o teste de Dunnett mas as estimativas foram muito superiores a 5% (valores entre 10 e 18,5%). Em ambos testes os maiores valores ocorreram nas situações em que se tinha dois tratamentos de tamanho 5 e apenas um de tamanho 10. Em relação as estimativas de poder, embora o teste de Dunnett tenha apresentado valores maiores que o teste de Hayter e Tsui esses não podem ser levados em consideração já que as estimativas do erro global do teste de Dunnett estão muito acima de 5%, ou de 7% como apresentado por Hayter e Tsui, e portanto os dois testes não são comparáveis.

O teste *t* de Student com a correção de Bonferroni manteve estimativas do erro global próximas do valor nominal de 5% e resultou em valores estimados de poder próximos aos de Hayter e Tsui.

No caso em que as variâncias dos tratamentos são iguais e conhecidas, para grupos balanceados e não-balanceados (ver Tabelas 6 e 7) o teste de Hayter e Tsui manteve as

estimativas do erro global em torno de 5% e foi ligeiramente mais poderoso que o teste *t* de Student com a correção de Bonferroni em todos os casos apresentados (ganhos de poder entre 2,5 e 8,65%).

É interessante notar que em geral, tanto para experimentos balanceados quanto para não-balanceados com variâncias conhecidas ou desconhecidas, os maiores ganhos de poder do teste de Hayter e Tsui em relação aos testes *t* de Student e Dunnett foram observados em situações nas quais a diferença entre as médias dos dois tratamentos comparados era menor, algo que constitui um ponto positivo para a adaptação do teste de Hayter e Tsui proposta na seção 3.2.

Tabela 4: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui, t de Student com Bonferroni e Dunnett – amostras balanceadas – 3 tratamentos (variância desconhecida)

		Erro do Tipo I			Poder do Teste	
Tamanho dos Grupos	Teste utilizado	Erro Global	Comparação $X_1 X_2$	Comparação $X_1 X_3$	Comparação $X_1 X_2$	Comparação $X_1 X_3$
n =5	Hayter e Tsui	0,0842	0,0475	0,0456	0,0640	0,1140
	t-Bonferroni	0,0457	0,0238	0,0247	0,0340	0,0600
	Razão (t)	----	----	----	1,8820	1,9130
	Dunnett	0,0510	0,0290	0,0279	0,0369	0,0738
	Razão (D)	---	----	----	1,7344	1,5447
n=10	Hayter e Tsui	0,0589	0,0335	0,0318	0,0637	0,1590
	t-Bonferroni	0,0461	0,0257	0,0242	0,0460	0,1190
	Razão (t)	----	----	----	1,3840	1,3340
	Dunnett	0,0477	0,0266	0,0261	0,0488	0,1307
	Razão (D)	---	---	---	1,3053	1,2165
n=20	Hayter e Tsui	0,0552	0,0321	0,0281	0,0886	0,2730
	t-Bonferroni	0,0450	0,0261	0,0218	0,0700	0,2378
	Razão (t)	----	----	----	1,2650	1,1480
	Dunnett	0,0497	0,0290	0,0254	0,0769	0,2512
	Razão (D)	---	----	----	1,1521	1,0868
n=50	Hayter e Tsui	0,0541	0,0282	0,0290	0,1705	0,6156
	t-Bonferroni	0,0481	0,0249	0,0253	0,1550	0,5830
	Razão (t)	----	----	----	1,1000	1,0560
	Dunnett	0,0509	0,0254	0,0285	0,1680	0,6069
	Razão (D)	----	----	----	1,0149	1,0143

*Hipótese nula: : $\mu_1 = \mu_2 = \mu_3 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5.$

Razão (t) denota a razão entre o poder de HT e o poder do teste *t*-Student com a correção de Bonferroni;.
Razão (D) denota a razão entre o poder do teste de Hayter e Tsui e Dunnett

Tabela 5: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui, t de Student com Bonferroni e Dunnett – amostras não-balanceadas – 3 tratamentos – (variância desconhecida)

		Erro do Tipo I			Poder do Teste	
Tamanho dos Grupos	Teste utilizado	Erro Global	Comparação $X_1 X_2$	Comparação $X_1 X_3$	Comparação $X_1 X_2$	Comparação $X_1 X_3$
$n_1 = n_3 = 10$ $n_2 = 5$	Hayter e Tsui	0,0700	0,0359	0,0401	0,0361	0,1148
	t - Bonferroni	0,0470	0,0232	0,0265	0,0360	0,1160
	Razão (t)	----	----	----	1,0028	0,9897
	Dunnett	0,1001	0,0725	0,0352	0,1053	0,1417
	Razão (D)	---	----	----	0,3428	0,8102
$n_1 = 10$ $n_2 = n_3 = 5$	Hayter e Tsui	0,0767	0,0418	0,0400	0,0400	0,0863
	t - Bonferroni	0,0496	0,0260	0,0253	0,0377	0,0799
	Razão	----	----	----	1,0610	1,0801
	Dunnett	0,1587	0,0888	0,0834	0,1171	0,3814
	Razão	---	---	---	0,3219	0,5981
$n_1 = n_3 = 5$ $n_2 = 10$	Hayter e Tsui	0,0712	0,0415	0,0405	0,0422	0,0718
	t - Bonferroni	0,0470	0,0262	0,0247	0,0393	0,0603
	Razão (t)	----	----	----	1,0738	1,1907
	Dunnett	0,1845	0,0884	0,1322	0,1172	0,2453
	Razão	---	----	----	0,3601	0,2927
$n_1 = 5$ $n_2 = n_3 = 10$	Hayter e Tsui	0,0637	0,0353	0,0388	0,0367	0,0783
	t - Bonferroni	0,0415	0,0264	0,0239	0,0366	0,0787
	Razão (t)	----	----	----	1,0027	0,9949
	Dunnett	0,1242	0,0790	0,0755	0,1031	0,1879
	Razão (D)	----	----	----	0,3560	0,4167

*Hipótese nula: : $\mu_1 = \mu_2 = \mu_3 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5.$

Razão (t) denota a razão entre o poder de HT e o poder do teste *t*-Student com a correção de Bonferroni;.
Razão (D) denota a razão entre o poder do teste de Hayter e Tsui e Dunnett

Tabela 6: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui e t-Student com Bonferroni – amostras balanceadas – 3 tratamentos – (variância conhecida)

		Erro do Tipo I			Poder do Teste	
Tamanho dos grupos	Teste estatístico	Erro Global	Comparação $X_1 X_2$	Comparação $X_1 X_3$	Comparação $X_1 X_2$	Comparação $X_1 X_3$
n =5	Hayter e Tsui	0,0487	0,0298	0,0238	0,0393	0,0796
	t – Bonferroni	0,0451	0,0273	0,0218	0,0372	0,0757
	Razão (t)	---	----	----	1,0565	1,0515
n=10	Hayter e Tsui	0,0501	0,0268	0,0266	0,0528	0,1370
	t – Bonferroni	0,0456	0,0248	0,0238	0,0490	0,1306
	Razão (t)	---	----	----	1,0776	1,0490
n=20	Hayter e Tsui	0,0492	0,0278	0,0261	0,0743	0,2682
	t – Bonferroni	0,0444	0,0252	0,0232	0,0703	0,2579
	Razão (t)	---	----	----	1,0569	1,0399
n=50	Hayter e Tsui	0,0507	0,0286	0,0272	0,1704	0,6103
	t - Bonferroni	0,0470	0,0267	0,0247	0,1624	0,5952
	Razão (t)	---	----	----	1,0493	1,0254

*Hipótese nula: : $\mu_1 = \mu_2 = \mu_3 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5.$

Razão (t): Poder de HT/ Poder do teste *t*-Student com a correção de Bonferroni.

Tabela 7: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui e t-Student com Bonferroni – amostras não-balanceadas – 3 tratamentos – (variância conhecida)

		Erro do Tipo I			Poder do Teste	
Tamanho dos Grupos	Teste estatístico	Erro Global	Comparação $X_1 X_2$	Comparação $X_1 X_3$	Comparação $X_1 X_2$	Comparação $X_1 X_3$
$n_1 = n_3 = 10$ $n_2 = 5$	Hayter e Tsui	0,0478	0,0262	0,0243	0,0459	0,1324
	t – Bonferroni	0,0442	0,0243	0,0224	0,0430	0,1276
	Razão (t)	---	----	----	1,0674	1,0376
$n_1 = 10$ $n_2 = n_3 = 5$	Hayter e Tsui	0,0529	0,0272	0,0273	0,0427	0,1006
	t – Bonferroni	0,0482	0,0248	0,0248	0,0393	0,0959
	Razão (t)	---	---	---	1,0865	1,0490
$n_1 = n_3 = 5$ $n_2 = 10$	Hayter e Tsui	0,0518	0,0278	0,0288	0,0421	0,0774
	t – Bonferroni	0,0490	0,0261	0,0272	0,0400	0,0731
	Razão (t)	---	----	----	1,0525	1,0588
$n_1 = 5$ $n_2 = n_3 = 10$	Hayter e Tsui	0,0486	0,0286	0,0263	0,0431	0,0953
	t – Bonferroni	0,0455	0,0271	0,0244	0,0403	0,0919
	Razão (t)	----	----	----	1,0695	1,0370

*Hipótese nula: : $\mu_1 = \mu_2 = \mu_3 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5.$

Razão (t): Poder de HT/ Poder do teste *t*-Student com a correção de Bonferroni.

5.2 Comparação de cinco tratamentos

Nesta seção serão apresentados os resultados obtidos a partir de simulações realizadas com cinco grupos considerando-se $\mu_i = 0, i = 1,2,...,5$, sob a hipótese nula e $\mu_1 = 0$, $\mu_2 = 0,25$, $\mu_3 = 0,5$, $\mu_4 = 1$ e $\mu_5 = 2$, sob a hipótese alternativa. Para $\alpha = 0,05$ a determinação do valor crítico do teste de Hayter e Tsui foi feita através da geração aleatória de N^*=100.000 valores da distribuição normal com vetor de médias nulo e matriz de covariâncias dada por

$$P = \begin{pmatrix} 1 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1 \end{pmatrix}$$

de forma semelhante ao que foi descrito na seção 5.1, sendo o valor crítico determinado $C_{R,0,05}$ = 2,455.

Assim como o observado para o caso de 3 tratamentos, para grupos balanceados com variâncias desconhecidas (ver Tabela 8) as estimativas do erro global do teste de Hayter e Tsui foram próximas a 5% para $n \geq 10$, mas maior para $n = 5$ (estimativa em torno de 7,5%). Em relação ao poder do teste, os 3 testes estatísticos foram similares no que se refere a comparação dos tratamentos 1 e 5, caso em que a diferença entre as médias estabelecidas nas hipóteses nula e alternativa era mais elevada. Nas outras situações consideradas quando comparado ao teste *t* de Student observa-se um ganho expressivo de poder do teste de Hayter e Tsui no caso de grupos com tamanhos $n \geq 10$. Como ilustração na comparação dos tratamentos 1 e 2, caso em que a diferença entre médias é menor, teve-se um ganho de 43% para amostras de tamanho $n = 10$, de 23% para $n = 20$ e 15% para $n = 50$. Em relação ao teste de Dunnett, Hayter e Tsui foi mais poderoso para $n = 10$ (ganhos de poder entre 7,7 e 22,3%) e igualmente poderoso para $n = 20,50$. É interessante observar que embora o teste de Hayter e Tsui tenha apresentado uma estimativa da probabilidade do erro do tipo I acima de 0,05 no caso de $n = 5$, os ganhos de poder em relação ao teste de Dunnett são elevados (entre 33 e 51%, a menos da comparação dos tratamentos 1 e 5).

Para amostras não-balanceadas e variâncias desconhecidas (ver Tabela 9) observa-se que as estimativas do erro global do teste do Hayter e Tsui estão entre 5,7 a 7,9% , enquanto que as estimativas do teste de Dunnett encontram-se entre 8,4 e 19,4%, ou seja valores bem maiores que o nível de 5% estipulado para os testes, fato similar ao visto no caso de 3 tratamentos (ver seção 5.1). Para ambos testes as maiores estimativas ocorreram quando havia 4 tratamentos com 5 observações amostrais e apenas um com 10. O teste *t* de Student por outro lado, apresentou estimativas da probabilidade do erro tipo I próximas a 0,05 em todos os casos mas foi sempre menos poderoso que Hayter e Tsui, mesmo nas situações em que as estimativas do erro global de Hayter e Tsui foram mais próximas a 5% (caso em que se tem 4 tratamentos com 10 observações amostrais e apenas um com 5; estimativas do erro global iguais a 5,7 e 6,8%). Devido ao fato das estimativas do erro global do teste de Dunnett estarem muito acima de 5% não é adequado comparar o seu poder diretamente com os dos outros dois testes, no entanto é importante salientar que para o caso em que sua estimativa do erro global foi mais próxima a 5% (igual a 8,42%), o seu poder foi próximo aos observados para os testes de *t* de Student e Hayter e Tsui, a menos da comparação dos tratamentos 1 e 2.

Os resultados das simulações em que a variância populacional foi considerada conhecida (ver Tabelas 10 e 11) mostram que as estimativas do erro global são próximas a 5% para ambos testes comparados (*t* de Student e Hayter e Tsui) tanto para grupos balanceados quanto para não-balanceados, sendo o teste de Hayter e Tsui o mais poderoso em ambos os casos. De modo similar ao que ocorreu para variância desconhecida, a adaptação do teste de Hayter e Tsui apresentou maiores estimativas de poder que o teste *t* de Student mas valores menos expressivos que nos casos de variância desconhecida (ganhos de poder entre 1,13 e 17%). Para grupos balanceados a razão entre o poder dos dois testes se aproxima de 1 à medida que se aumenta o tamanho da amostra ou a diferença entre as médias dos grupos comparados. Nesta situação o maior ganho observado ocorreu na comparação dos tratamentos 1 e 2 com amostras de tamanho *n*=5, sendo igual a 14,36%. Para grupos não-balanceados o maior ganho percentual de poder ocorreu na comparação dos tratamentos 1 e 2 na configuração em que $n_1 = 10$, $n_i = 5$, *i*=2,3,4,5 (ganho de poder = 17,05%). Em

geral, para grupos balanceados os ganhos de poder foram maiores na situação em que se tem 5 tratamentos em comparação ao caso de 3 tratamentos (ver seção 5.1).

Tabela 8: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui, *t* de Student com a correção de Bonferroni e Dunnett – amostras balanceadas - cinco tratamentos – (variância desconhecida)

			Comparação $X_1 X_2$		Comparação $X_1 X_3$		Comparação $X_1 X_4$		Comparação $X_1 X_5$	
Tamanho dos grupos	Teste utilizado	Erro Global	Erro Tipo I	Poder do teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste
n = 5	Hayter e Tsui	0,0751	0,0218	0,0330	0,0246	0,0668	0,0240	0,2224	0,0247	0,7574
	t - Bonferroni	0,0434	0,0116	0,0177	0,0135	0,0326	0,0127	0,1133	0,0122	0,5263
	Razão (t)	---	----	1,8644	---	2,0491	---	1,9629	---	1,4391
	Teste de Dunnett	0,0499	0,0139	0,0219	0,0162	0,0455	0,0163	0,1669	0,0153	0,6929
	Razão (D)	---	----	1,5068	---		---	1,3325	---	1,0930
n = 10	Hayter e Tsui	0,0581	0,0168	0,0368	0,0175	0,1088	0,0196	0,4277	0,0182	0,9775
	t - Bonferroni	0,0416	0,0116	0,0257	0,0109	0,0725	0,0130	0,3279	0,0121	0,9468
	Razão (t)	----	----	1,4319	---	1,5007	---	1,3044	---	1,0324
	Teste de Dunnett	0,0510	0,0145	0,0318	0,0171	0,0889	0,0156	0,3971	0,0138	0,9692
	Razão (D)	----	----	1,1572	---	1,2238	---	1,0770	---	1,0085
n =20	Hayter e Tsui	0,0531	0,0156	0,0515	0,0157	0,1945	0,0175	0,7552	0,0153	1,0000
	t - Bonferroni	0,0397	0,0109	0,0420	0,0114	0,1566	0,0123	0,6971	0,1090	0,9999
	Razão (t)	----	----	1,2262	---	1,2420	---	1,0833	---	1,0000
	Teste de Dunnett	0,0488	0,0151	0,0492	0,0140	0,1919	0,0131	0,7506	0,0143	1,0000
	Razão (D)	----	----	1,0467	---	1,0135	---	1,0061	---	1,0000
n =50	Hayter e Tsui	0,0498	0,0143	0,1192	0,0152	0,5187	0,0135	0,9935	0,0154	1,0000
	t - Bonferroni	0,0413	0,0119	0,1040	0,0127	0,4815	0,0115	0,9919	0,0124	1,0000
	Razão (t)	----	----	1,1462	---	1,0773	---	1,0016	---	1,0000
	Teste de Dunnett	0,0567	0,0159	0,1155	0,0167	0,5177	0,0173	0,9938	0,0167	1,0000
	Razão (D)	----	----	1,0320	---	1,0019	---	1,0000	---	1,0000

*Hipótese nula: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5; \mu_4 = 1; \mu_5 = 2$.

Razão (t) denota a razão entre o poder de HT e o poder do teste *t*-Student com a correção de Bonferroni;.

Razão (D) denota a razão entre o poder do teste de Hayter e Tsui e Dunnett

Tabela 9: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui, t-Student com a correção de Bonferroni e Dunnett – amostras não-balanceadas - cinco tratamentos – (variância desconhecida)

			Comparação $X_1 X_2$		Comparação $X_1 X_3$		Comparação $X_1 X_4$		Comparação $X_1 X_5$	
Tamanho dos grupos	Teste utilizado	Erro Global	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste
$n_1 = 10$, $n_2 = 5$ $n_3 = n_4 = n_5 = 10$	Hayter e Tsui	0,0680	0,0199	0,0238	0,0201	0,1047	0,0208	0,4268	0,0214	0,9759
	t - Bonferroni	0,0472	0,0141	0,0174	0,0137	0,069	0,0146	0,3182	0,0121	0,9417
	Razão (t)	----	----	1,3678	----	1,5174	----	1,3413	----	1,0363
	Teste de Dunnett	0,0842	0,0445	0,0727	0,0176	0,0976	0,0190	0,4058	0,0185	0,9727
	Razão (D)	----	----	0,3274	----	1,0727	----	1,0517	----	1,0032
$n_2 = n_3 = 5$ $n_1 = n_4 = n_5 = 10$	Hayter e Tsui	0,0709	0,0206	0,0319	0,023	0,0753	0,0195	0,4295	0,0194	0,9734
	t - Bonferroni	0,0453	0,0125	0,0197	0,0132	0,0441	0,0135	0,3249	0,0126	0,9407
	Razão (t)	----	----	1,6193	----	1,7075	----	1,3219	----	1,0348
	Teste de Dunnett	0,1168	0,0512	0,0760	0,0522	0,1474	0,0180	0,4138	0,0176	0,9730
	Razão (D)	----	----	0,4197	----	0,5109	----	1,0379	----	1,0004
$n_1 = n_5 = 10$ $n_2 = n_3 = n_4 = 5$	Hayter e Tsui	0,0747	0,0219	0,0338	0,0221	0,0751	0,0201	0,2921	0,0202	0,9766
	t - Bonferroni	0,0428	0,0131	0,0195	0,0112	0,0439	0,0112	0,1887	0,0114	0,9459
	Razão (t)	----	----	1,7333	----	1,7107	----	1,548	----	1,0325
	Teste de Dunnett	0,1602	0,0561	0,0831	0,0552	0,1571	0,0560	0,4429	0,0211	0,9752
	Razão (D)	----	----	0,4067	----	0,4780	----	0,6595	----	1,0001
$n_1 = 10$, $n_2 = 5$ $n_3 = n_4 = n_5 = 5$	Hayter e Tsui	0,0792	0,0224	0,0341	0,0216	0,0778	0,0242	0,2916	0,022	0,8817
	t - Bonferroni	0,0442	0,0116	0,0174	0,0113	0,0434	0,0134	0,1874	0,0133	0,7542
	Razão (t)	-	-	1,9598	-	1,7926	-	1,556	-	1,1691
	Teste de Dunnett	0,1942	0,0576	0,0852	0,0577	0,1633	0,0604	0,4494	0,0589	0,9462
	Razão (D)	----	----	0,4002	----	0,4764	----	0,6489	----	0,9318
$n_1 = 5$, $n_2 = 10$ $n_3 = n_4 = n_5 = 10$	Hayter e Tsui	0,057	0,0207	0,0321	0,0193	0,0671	0,0197	0,284	0,0207	0,8862
	t - Bonferroni	0,0427	0,0128	0,0208	0,0127	0,0417	0,0142	0,1874	0,0131	0,7576
	Razão (t)	----	----	1,5433	----	1,6091	----	1,5155	----	1,1697
	Teste de Dunnett	0,1290	0,0484	0,0726	0,0466	0,1387	0,0479	0,4216	0,0474	0,9394
	Razão (D)	----	----	0,4421	----	0,4838	----	0,6736	----	0,9434

*Hipótese nula: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5; \mu_4 = 1; \mu_5 = 2$.

Razão (t) denota a razão entre o poder de HT e o poder do teste *t*-Student com a correção de Bonferroni;.

Razão (D) denota a razão entre o poder do teste de Hayter e Tsui e Dunnett

Tabela 10: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui e t de Student com a correção de Bonferroni – amostras balanceadas - cinco tratamentos – (variância conhecida)

			Comparação $X_1 X_2$		Comparação $X_1 X_3$		Comparação $X_1 X_4$		Comparação $X_1 X_5$	
Tamanho dos grupos	Teste utilizado	Erro Global	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste
n = 5	Hayter e Tsui	0,0470	0,0133	0,0215	0,0140	0,0539	0,0134	0,1938	0,0151	0,7583
	t - Bonferroni	0,0418	0,0117	0,0188	0,0123	0,0476	0,0118	0,1821	0,0131	0,7417
	Razão (t)	---	----	1,1436	---	1,1324	---	1,0643	---	1,0224
n = 10	Hayter e Tsui	0,0485	0,0137	0,0319	0,0146	0,0892	0,0153	0,0409	0,0131	0,0979
	t - Bonferroni	0,0418	0,0113	0,0283	0,0125	0,0817	0,0134	0,3905	0,0107	0,9769
	Razão (t)	----	----	1,1272	---	1,0918	---	0,1047	---	0,1002
n =20	Hayter e Tsui	0,0493	0,0131	0,0540	0,0131	0,1928	0,0147	0,7583	0,0167	1,0000
	t - Bonferroni	0,0422	0,0116	0,0480	0,0111	0,1792	0,0122	0,7435	0,0143	1,0000
	Razão (t)	----	----	1,1250	---	1,0759	---	1,0199	---	1,0000
n =50	Hayter e Tsui	0,0500	0,0144	0,1113	0,0141	0,5170	0,0160	0,9952	0,0143	1,0000
	t - Bonferroni	0,0439	0,0125	0,1027	0,0120	0,4988	0,0140	0,9940	0,0129	1,0000
	Razão (t)	----	----	1,0837	---	1,0365	---	1,0012	---	1,0000

*Hipótese nula: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5; \mu_4 = 1; \mu_5 = 2$.

Razão (t): Poder de HT/ Poder do teste t-Student com a correção de Bonferroni.

Tabela 11: Estimativas da probabilidade do erro tipo I e do poder dos testes de Hayter e Tsui e I de Student com a correção de Bonferroni – amostras não-balanceadas - cinco tratamentos – (variância conhecida)

			Comparação $X_1 X_2$		Comparação $X_1 X_3$		Comparação $X_1 X_4$		Comparação $X_1 X_5$	
Tamanho dos grupos	Teste utilizado	Erro Global	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste	Erro Tipo I	Poder do Teste
$n_1 = 10$, $n_2 = 5$ $n_3 = n_4 = n_5 = 10$	Hayter e Tsui	0,0515	0,0134	0,0261	0,0151	0,0903	0,0157	0,4177	0,0155	0,9803
	t - Bonferroni	0,0451	0,0118	0,0238	0,0129	0,0803	0,0135	0,3991	0,0134	0,9777
	Razão (t)	----	----	1,0966	----	1,1245	----	1,0466	----	1,0027
$n_2 = n_3 = 5$ $n_1 = n_4 = n_5 = 10$	Hayter e Tsui	0,0498	0,0124	0,0258	0,0126	0,0608	0,0159	0,4146	0,0145	0,9771
	t - Bonferroni	0,0434	0,0104	0,0223	0,0113	0,0545	0,0136	0,3951	0,0126	0,9738
	Razão (t)	----	----	1,157	----	1,1156	----	1,0494	----	1,0034
$n_1 = n_5 = 10$ $n_2 = n_3 = n_4 = 5$	Hayter e Tsui	0,0525	0,013	0,0228	0,0135	0,0629	0,0159	0,2663	0,0155	0,9729
	t - Bonferroni	0,0452	0,0113	0,02	0,0118	0,0571	0,0133	0,2497	0,0136	0,9726
	Razão (t)	----	----	1,14	----	1,1016	----	1,0665	----	1,0003
$n_1 = 10$, $n_2 = 5$ $n_3 = n_4 = n_5 = 5$	Hayter e Tsui	0,0404	0,0139	0,0254	0,0136	0,0635	0,0119	0,2735	0,0141	0,8801
	t - Bonferroni	0,0445	0,0124	0,0217	0,0126	0,0572	0,0103	0,2572	0,0125	0,8691
	Razão (t)	----	----	1,1705	----	1,1101	----	1,0634	----	1,0127
$n_1 = 5$, $n_2 = 10$ $n_3 = n_4 = n_5 = 10$	Hayter e Tsui	0,0463	0,0145	0,0221	0,0159	0,0623	0,0167	0,2693	0,0131	0,8826
	t - Bonferroni	0,0409	0,0124	0,0194	0,0141	0,0563	0,0145	0,2561	0,0118	0,8727
	Razão (t)	----	----	1,1392	----	1,1066	----	1,0515	----	1,0113

*Hipótese nula: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$; Hipótese alternativa : $\mu_1 = 0; \mu_2 = 0{,}25; \mu_3 = 0{,}5; \mu_4 = 1; \mu_5 = 2$.

Razão (t): Poder de HT/ Poder do teste *t*-Student com a correção de Bonferroni.

6. Considerações Finais

Neste artigo apresentou-se uma adaptação do teste estatístico multivariado de Hayter e Tsui (1994) ao contexto de comparações múltiplas de tratamentos em problemas univariados. O objetivo foi construir um teste para comparação de médias de todos os tratamentos com um grupo controle que mantivesse a probabilidade do tipo I do teste conjunto (erro global) no nível de significância nominal pré-estabelecido. No estudo foram comparados três e cinco tratamentos homocedásticos, com observações provenientes de distribuições normais com variâncias populacionais iguais conhecidas e desconhecidas. Foram simuladas diferentes situações com o objetivo de estimar as probabilidades do erro tipo I e o poder da adaptação do teste de Hayter e Tsui e compará-los com os respectivos valores obtidos do teste t de Student (ou teste Z da distribuição normal) usando a correção de Bonferroni e do teste de Dunnett.

Nas situações de grupos balanceados em que a variância dos tratamentos foi considerada desconhecida e estimada pelos dados amostrais, a adaptação do teste de Hayter e Tsui apresentou uma estimativa da probabilidade do erro do tipo I acima do nível nominal de 0,05 quando todos os tratamentos tinham tamanhos amostrais iguais a 5. Para amostras com um número maior de observações (a partir de $n = 10$) as estimativas do erro global aproximaram-se de 5% e o poder estimado do teste foi sempre maior ou similar aos observados para os testes *t* de Student com a correção de Bonferroni e Dunnett. Para grupos não-balanceados as estimativas da probabilidade do erro do tipo I do teste de Hayter e Tsui ficaram ao redor de 0,07 e seu poder também foi maior que o do teste *t* de Student com a correção de Bonferroni sendo os ganhos sempre muito consideráveis. O teste de Dunnett por sua vez, foi muito influenciado pelo não-balanceamento dos tratamentos tendo suas estimativas do erro global extremamente inflacionadas (valores entre 10 e 18,4% para 3 tratamentos e entre 8 e 19,5% para 5 tratamentos). Isso mostra que é necessário muita cautela ao utilizar o teste de Dunnett nessas situações sendo o teste de Hayter e Tsui mais indicado. Nos casos em que o verdadeiro valor da variância foi considerado conhecido, as estimativas do erro global do teste de Hayter e Tsui se mantiveram ao redor de 5% mesmo para

pequenas amostras, para grupos balanceados ou não. O seu poder também foi maior que o do teste t de Student no entanto, o ganho comparativo obtido foi menor do que o observado para tratamentos com variâncias estimadas. O teste de Hayter e Tsui adaptado não foi comparado ao de Dunnett na situação de variâncias conhecidas.

É importante ressaltar que ganhos expressivos de poder da adaptação do teste de Hayter e Tsui proposta neste artigo ocorreram em situações em que a diferença entre as médias dos tratamentos comparados era menor.

Sendo assim, os resultados apresentados nesse artigo corroboram positivamente para a indicação da adaptação do teste de Hayter e Tsui no contexto de comparações múltiplas de tratamentos em relação a um grupo controle uma vez que ele se mostrou como uma boa alternativa ao teste t de Student como também ao teste de Dunnett principalmente para grupos balanceados com tamanhos amostrais pequenos e para grupos não-balanceados.

Finalmente destaca-se que o teste de Hayter e Tsui apresentado nesse artigo pode ser extendido para a situação mais geral de comparação de médias de vários tratamentos independentes ou correlacionados.

Referências bibliográficas

ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. Annals of Mathematical Statistics, 23, p.193-212, 1952.

BAIN, L.J.; ENGELHARDT, M. Introduction to Probability and Mathematical Statistics. Boston: Duxbury Press, 2000.

BARTLETT, M. S. Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London, Series A,160, p. 268-282, 1937.

DOHMEN, K. Improved Bonferroni Inequalities with Applications: Inequalities and Identities of Inclusion-Exclusion Type. Berlin: Springer-Verlag, 2003.

DUNCAN, D. B. Multiple Range and Multiple F Tests. Biometrics, 11, p 1- 42, 1955.

DUNNETT, C. W. Pairwise multiple comparisons in the homogeneous variance, unequal sample size case. Journal of American Statistical Association, 75, p. 789-795, 1980.

DUNNETT, C. W. New Tables of Multiple Comparisons with a Control. Biometrics, 20, p. 482-491, 1964.

DUNNETT, C. W. A multiple comparison procedure for comparing several treatments with a control. Journal of American Statistical Association, 50, p. 1096-1121, 1955.

FISHER, R. A. The Design of Experiments. New York: Hafner Publishing Company, 1966.

GIRARDI, L. H.; FILHO, C. A.; STORCK, L. Erro tipo I e poder de cinco testes de comparação múltipla de médias. Revista Brasileira de Biometria, 27, 1, p. 23-36, 2009.

HAYTER, A. J.; TSUI, K-L. Identification and quantification in multivariate quality control problems. Journal of Quality Technology, 26, 3, p.197-208,1994.

HSU, J.C. Multiple Comparisons: Theory and Methods. New York: Chapman and Hill, 1996.

JIANG, W.; SHU, L,; TSUNG, F. A comparative study of point monitoring schemes for ACP processes. Quality and Realiability Engineering International, 22, 8, p. 939-952, 2006.

JOHNSON, R. A.; WICHERN, D.W. Applied multivariate statistical analysis. New Jersey: Prentice Hall, 2002.

Kalgonda, A. A.; Kulkarni, S. R. Multivariate quality control chart for autocorrelated processes. Journal of Applied Statistics, 31,3, p.317-327, 2004.

MINGOTI. S. A.; GLÓRIA, F. A. A. Uma proposta de modificação do índice de capacidade multivariado de Chen. In Anais do XXIII ENEGEP, Ouro Preto, Minas Gerais, 2033 (em cd rom).

MINGOTI. S. A.; GLÓRIA, F. A. A. Comparando os métodos paramétrico e não-paramétrico na determinação do valor crítico do teste estatístico de médias proposto por Hayter e Tsui. Produção, 15,2, p. 251-262, 2005.

MONTGOMERY, D.C. Design and analysis of experiments. New York: John Willey, 2001.

MONTGOMERY, D.C. Introdução ao Controle Estatístico da Qualidade. Rio de Janeiro: Livros Técnicos e Científicos, 2004.

PETERSEN, R. G. Design and analysis of experiments. New York: Marcel Dekker, 1985.

STEEL, R. G. D.; TORRIE, J. H. Principles and Procedures of Statistics. New York: McGraw - Hill,1980.

TIMM, N. H. Multivariate quality control using finite intersection sets. Journal of Quality Technology, 28, p. 233-243, 1996.

TUKEY, J. W. Quick and Dirty Methods in Statistics, Part II, Simple Analysis for Standard Designs. Proceedings of the Fifth Annual Convention, American Society for Quality Control, p. 189-197, 1951.

VIEIRA, S. Análise de Variância. São Paulo: Editora Atlas S.A.,2005.

WERNER, L.; LINDEN, J. C. S. V. D.; RIBEIRO, J. L. D. Análise da percepção sobre assentos de trabalho utilizando técnicas estatísticas multivariadas. Produção, 13, 3, p. 34-49, 2003.

Abstract

Many situations are concerned with the comparison of means of several independent treatments using ANOVA. When the hypothesis of no difference among the means is rejected, comparisons between treatments are carried out with the purpose of identifying which pairs of means are significantly different. One particular situation is which one of the treatments is a control. In Multivariate Statistics there are some statistical tests, such as Hayter and Tsui, appropriated to test the mean vector of one population. In this paper an adaptation of Hayter and Tsui´s test is suggested to be used in the context of multiple comparisons. Results from an evaluation of this methodology are shown taken the t-Student with the Bonferroni correction and the Dunnett test as references. Monte Carlo simulations were performed for balanced and non-balanced experiments, for three and five treatments, under assumptions of normality and homoscedasticity considering known and unknown variances.

Key-words: Hayter e Tsui; Multiple Comparisons; Bonferroni; Dunnett; Power of the test Monte Carlo.

Modelagem longitudinal hierárquica: uma aplicação aos dados da PME

Alinne de Carvalho Veiga[1]

Resumo

Existem diversas abordagens para a análise de dados longitudinais, dados de painel ou de medidas repetidas. A abordagem por modelos multinível é uma das que permite a análise de dados provenientes de pesquisas amostrais complexas também permitindo levar em consideração alguns aspectos importantes do conjunto de dados, tais como: a estrutura hierárquica gerada pelos diferentes estágios da amostragem; a estrutura de correlação existente entre as medidas de um mesmo indivíduo; a estrutura geral dos dados devido ao uso de painéis rotativos, ou à existência de desgaste ou outras formas de não-resposta no painel. Este artigo apresenta uma aplicação de modelos longitudinais multiníveis na PME (Pesquisa Mensal de Emprego) que contém um desenho amostral complexo que inclui painéis rotativos não consecutivos e da forma 4-8-4. Os modelos apresentados para a dinâmica do rendimento do trabalho de chefes de domicílios contêm diferentes níveis de complexidade e incorporam o desenho amostral da PME.

Palavras Chave: Dados Longitudinais; PME; modelos multiníveis; matrizes de correlação.

[1] Universidade do Estado do Rio de Janeiro. E-mail: alinneveiga@gmail.com. Parte da tese de Doutorado pela divisão de Estatística Social, University of Southampton (Inglaterra).

1. Introdução

Os bancos de dados utilizados nas ciências sociais são comumente provenientes de pesquisas amostrais complexas cujos desenhos incorporam múltiplos estágios de seleção das unidades amostrais. Muitas vezes os diferentes estágios da amostragem representam a estrutura hierárquica natural da população em estudo. Quando essas pesquisas são conduzidas de forma repetida, por exemplo, seguindo um mesmo grupo de indivíduos ao longo do tempo, isso gera um banco de dados longitudinal.

Dados longitudinais são também chamados de dados de medidas repetidas ou dados de painel, que são tipicamente formados por um número grande de unidades que são seguidas ao longo de um curto período de tempo (HECKMAN; SINGER, 1985). Como característica, um banco de dados de painel pode ser balanceado ou não balanceado, ou seja, todas as unidades amostrais podem ou não conter o mesmo número de medidas. Além disso, o intervalo de tempo entre as medidas de unidades amostrais diferentes podem ser regulares ou irregulares. Muitas vezes os dados de painel são resultantes de um desenho amostral que inclui painéis rotativos, onde as unidades amostrais são estratificadas em "painéis" que são incluídos na amostra de forma alternada até a total exclusão e a substituição destes por novos painéis. Dados provenientes de painéis rotativos podem ser classificados como dados não-balanceados, uma vez que, dependendo do período de análise, nem todas as unidades amostrais conterão o mesmo número de medidas. Isto pode também ser visto como um problema de não-resposta. Outro tipo bem comum de não-resposta é o desgaste do painel, quando as unidades amostrais deixam a pesquisa de forma prematura.

A análise de dados de painel oferece muitas vantagens, pois permite o estudo de processos dinâmicos ou da estabilidade ao longo do tempo e investigações das variações "entre" e "intra" unidades, além de permitir análises de corte transversal, quando apenas uma ocasião é considerada para estudo. No entanto, é comum que as respostas de uma mesma unidade sejam correlacionadas ao longo do tempo o que requer o uso de ferramentas específicas para a análise de

dados que lidam com as complexidades que dados longitudinais possam conter. Existem diversas abordagens para a análise de dados longitudinais complexos e a abordagem de modelos hierárquicos, ou multiníveis, é uma delas que está descrita na seção 3.

A seção 2 descreve a pesquisa mensal de emprego, a PME, e suas principais características que incluem: a estrutura hierárquica da população; a estrutura de correlação complexa entre as medidas repetidas e o desenho amostral complexo incluindo painéis rotativos. Na seção 4 a dinâmica da renda do trabalho dos chefes de domicílios é investigada através da aplicação de modelos longitudinais com diferentes níveis de complexidade. Esses modelos são comparados e suas aplicabilidades são discutidas na seção 5.

2. Dados em estudo: pesquisa mensal de emprego

Este artigo usa dados da pesquisa mensal de emprego (PME) de 2004 e 2005[2] (IBGE, 2002). A PME é uma pesquisa por amostragem probabilística conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE) que tem como principal objetivo investigar as características do mercado de trabalho brasileiro. Desenhada para ser representativa da população urbana de seis regiões metropolitanas brasileiras (Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre), a PME tem um desenho estratificado e conglomerado em dois estágios. Cada região metropolitana é formada por estratos independentes (constituídos de municípios) de onde as unidades primárias de amostragem (UPA) são selecionadas. As UPAs são os setores censitários que são selecionados por amostragem sistemática proporcional a uma medida de tamanho. As unidades secundárias de amostragem (USA) são os domicílios que são selecionados de cada UPA, representando os conglomerados de onde todos os indivíduos são selecionados e são entrevistados.

A PME é conduzida todo mês e os domicílios selecionados são agrupados em painéis que são incluídos de forma rotativa na amostra, seguindo um padrão conhecido como 4-8-4. De acordo com este padrão rotacional, os domicílios

[2] Estes eram os dados que estavam disponíveis no período em que esse estudo foi iniciado.

permanecem na amostra durante um total de 16 meses, sendo entrevistados nos primeiros quatro meses consecutivos, ficando de fora da pesquisa por oito meses, e retornando nos últimos quatro meses para mais entrevistas. Note, no entanto, que a PME não é uma pesquisa longitudinal por definição, uma vez que esta não segue os unidades de amostragem (indivíduos) e sim os domicílios ao longo do tempo. Portanto se os indivíduos dos domicílios selecionados mudam de moradia, por exemplo, eles deixam de participar da pesquisa e os novos membros passam a fazer parte da amostra. Isto e alguns outros problemas de emparelhamento dos dados mensais da PME, representam uma desvantagem no uso dos dados da PME no sentido longitudinal, uma vez que não se pode garantir a formação do painel individual com uma alta taxa de precisão (ANTONACI; SILVA, 2007).

A forma encontrada para superar os problemas com o emparelhamento dos registros individuais da PME foi a de utilizar um registro por domicílio selecionando apenas os chefes dos domicílios. Esses têm suas oito entrevistas emparelhadas formando um banco de dados longitudinal através do uso de critérios extras para garantir o emparelhamento do mesmo indivíduo ao longo do tempo. Além disso, optou-se por incluir no banco de dados longitudinal apenas aqueles chefes em domicílios iniciados na amostra a partir de Janeiro de 2004 e que, pela definição do desenho amostral, tivessem a seqüência de oito entrevistas completadas até Dezembro de 20053. Como o objetivo é investigar a dinâmica da renda do trabalho dos chefes de domicílios, foram mantidos então apenas os chefes classificados como empregados e com observações válidas para a renda.

Na literatura de mercado de trabalho, a renda do trabalho é a variável mais comumente utilizada por representar o componente mais importante da renda domiciliar (JENKINS, 2000). Modelos de renda são geralmente baseados na equação minceriana (MINCER; POLACHEK, 1974) que, com base na teoria do capital humano, expressa o logaritmo natural da renda como função das características dos indivíduos como educação e experiência. Tentativas anteriores de ajustar a equação minceriana aos dados da PME (BARROS; CORSEUIL; LEITE, 2000) indicaram para modelos mais elaborados que incluem outras características individuais que não fossem apenas a educação e experiência. Além disso, devido a

[3] Desta forma temos painéis iniciados desde Janeiro a Setembro de 2004 e finalizados então de Abril a Dezembro de 2005, respectivamente.

estrutura hierárquica dos dados da PME ainda é possível incorporar as características dos diversos níveis envolvidos na análise como mostrado na Tabela 1. Note que os níveis da hierarquia se referem respectivamente às unidades primárias de amostragem, aos chefes de domicílio e às ocasiões de realização das entrevistas.

Tabela 1: Variáveis Incluídas na Análise

Variáveis	Categorias
Entrevista - Nível 1	
Número da Entrevista (ocasião)	Variando como (0, 1, 2, 3, 12, 13, 14 e 15)
Tipo de Trabalhador	Empregador, Informal, Formal,Militar e Conta própria
Tipo de Atividade	Indústria, construção, comércio, financeira, serviços sociais, serviços domésticos outros serviços e outras atividades
Duração do Emprego	Em meses com termo quadrado (experiência na firma)
Horas Trabalhadas	em log
Proxy para respondente	Indicador para o uso de um outro adulto como proxy para o respondente
Tamanho do Domicílio	Variável contínua
Indiv duos - Nível 2	
Mês da entrevista	De 0 a 23
Gênero	Mulheres como referência
Idade	Idade na primeira entrevista e idade^2
Raça	Indicador para brancos
Educação em anos de estudo	De 0 a 17 anos de estudo na primeira entrevista
UPA - Nível 3	
Região Metropolitana	Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre
Médias para	Idade e educação
Proporções para	Tipo de trabalhador, raça e proxy para respondente

3. Metodologia: modelos multiníveis

A idéia por traz da metodologia de modelos hierárquicos, ou multiníveis, é a de se incorporar a estrutura hierárquica dos dados na análise (GOLDSTEIN, 2003). Como mencionado brevemente acima, muitas vezes essa estrutura representa uma hierarquia natural da população de onde a amostra foi retirada, cujos agrupamentos (aglomerados) selecionados são considerados como uma amostra aleatória de todos os agrupamentos da população. É esperado, no entanto, que as unidades de um mesmo agrupamento não sejam independentes. Essa dependência pode ser, por exemplo, devido a seleção dos indivíduos que compõem a amostra depender da seleção de seus domicílios ou, devido às características compartilhadas entre as pessoas de um mesmo domicílio. Esta abordagem fornece

as ferramentas necessárias para lidar com a possível correlações intra-grupos utilizando o método dos mínimos quadrados generalizados iterativo (MQGI) como principal método de estimação (GOLDSTEIN, 1986). Além disso, cada nível da estrutura hierárquica dos dados pode ser representado no modelo através de uma fonte aleatória de variação nas respostas, o que também permite a inclusão de variáveis para cada um destes níveis. Quando comparados a abordagem clássica de regressão linear, a abordagem multinível mantém as hipóteses de normalidade e linearidade, relaxando a hipótese de independência e homocedasticidade, como será visto adiante.

3.1 Modelo de intercepto aleatório e de coeficientes aleatórios

O Modelo de Intercepto Aleatório é a formulação mais simples dentre a abordagem multinível. Considere dois níveis: pessoas (índice i) agrupadas em domicílios (índice j). O modelo de intercepto aleatório é representado como:

$$y_{ij} = \boldsymbol{x}_{(1)ij}^{T}\boldsymbol{\beta}_1 + \boldsymbol{x}_{2}j^{T}\boldsymbol{\beta}_2 + u_j + e_{ij}. \quad (1)$$

O modelo contém:

- variáveis com características pessoais $(\boldsymbol{x}_{(1)ij})$ e domiciliares $(\boldsymbol{x}_{(2)j})$ associadas aos seus vetores de coeficientes de regressão $\boldsymbol{\beta}_{(1)}$ e $\boldsymbol{\beta}_{(2)}$ respectivamente.
- dois termos de erro, um para cada nível considerado na análise, u_j para o nível 2 e e_{ij} para o nível 1; esses termos são mutuamente independentes e normalmente distribuídos.

O modelo de intercepto aleatório assume que cada grupo contém um intercepto específico, os u_j, ou efeitos aleatórios. Ainda assume-se que u_j são não correlacionados com as variáveis observadas. Para cada termo de erro há um componente da variância seguindo as hipóteses anteriores onde

$$u_j \sim N(0,\ \sigma_u^2) \text{ e } e_{ij} \sim N(0,\ \sigma_e^2). \quad (2)$$

Os componentes de variância σ_u^2 e σ_e^2 representam respectivamente as variâncias entre-grupos e intra-grupos. Através da estimação destes componentes, o coeficiente de correlação intra-grupo ρ, que representa o grau de dependência intra-grupo, pode ser calculado da forma

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}. \qquad (3)$$

Se ρ for aproximadamente zero, o uso da abordagem múltinível não se faz necessário e modelos de regressão clássica, estimados pelo método dos mínimos quadrados ordinários (MQO), podem ser utilizados ao invés (SNIJDERS; BOSKER, 1999). No entanto, valores de $\rho > 0$ justificam o uso de modelos de intercepto aleatório, que assumem que as unidades de um mesmo grupo são correlacionadas condicionalmente.

Se os efeitos de algumas das variáveis independentes variam de forma aleatória entre os grupos, tem-se então o modelo de coeficientes aleatórios expresso da seguinte forma:

$$y_{ij} = \boldsymbol{x}_{(1)ij}^T \boldsymbol{\beta}_1 + \boldsymbol{x}_{(2)j}^T \boldsymbol{\beta}_2 + \boldsymbol{z}_{ij}^T \boldsymbol{u}_j + e_{ij}. \qquad (4)$$

Ainda são mantidos os dois níveis mencionados anteriormente e $\boldsymbol{x}_{(1)ij}$ e $\boldsymbol{x}_{(2)j}$, que ainda representam as variáveis de nível um e dois respectivamente e $\boldsymbol{z}_{ij}$ é o vetor com o subconjunto das variáveis explicativas que têm efeitos aleatórios. Para o caso onde apenas uma variável é considerada como tendo efeito aleatório, este último vetor pode ser representado na forma matricial como

$$Z_j = \begin{pmatrix} 1 & z_{1j} \\ 1 & z_{2j} \\ \vdots & \vdots \\ 1 & z_{n_j j} \end{pmatrix}, \qquad (5)$$

ou seja, como uma matriz com o vetor de $\mathbf{1}$ para o intercepto aleatório e uma variável com efeito aleatório.

O modelo 4 difere do modelo 1 principalmente na quantidade de termos de erros. Ainda considerando o caso onde apenas uma variável independente varia de

forma aleatória entre os grupos, tem-se que $\boldsymbol{u}_j$ representa o vetor de efeitos aleatórios do segundo nível, tal que

$$\boldsymbol{u}_j = \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix}, \tag{6}$$

e supõe-se que $\boldsymbol{u}_j$ tem distribuição normal bivariada com vetor de médias $\mathbf{0}$ e matriz de covariância $\boldsymbol{\Sigma}_u$ tal como,

$$\boldsymbol{u}_j \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix}\right). \tag{7}$$

Os indivíduos de um mesmo grupo ainda são correlacionados condicionalmente, mas esta correlação agora depende dos valores da variável com o efeito aleatório. Portanto, o coeficiente de correlação para o modelo de coeficientes aleatórios não pode mais ser calculado como na equação 3. Desta forma, a hipótese de homocedasticidade é relaxada para o modelo 4, já que a matriz de covariância dos erros compostos V também depende dos valores de $\boldsymbol{z}_{ij}$ (STEELE, 2008). A matriz V é uma matriz bloco-diagonal, onde cada bloco representa os n grupos e é expresso como (HEDEKER; GIBBONS, 1997)

$$V_j = Z_j \Sigma_u Z_j^T + I_{n_j} \sigma^2_e, \tag{8}$$

onde I_{n_j} é uma matriz identidade de tamanho $n_j \times n_j$ e n_j é o número de pessoas (unidades do nível 1) no grupo *j*.

3.2 Modelos longitudinais multiníveis

Modelos longitudinais multiníveis são uma extensão dos modelos multiníveis de corte transversal onde as medidas repetidas tiradas de um mesmo indivíduo representam um nível adicional e o nível mais baixo da estrutura dos dados. Seguindo o contexto anterior, os indivíduos (índice i) passam a representar as unidades de nível dois agrupadas em domicílios (índice j), as unidades de nível três. Neste contexto tem-se uma estrutura em três níveis e os diferentes momentos de tempo, as ocasiões, são as unidades de nível um.

Modelos longitudinais seguindo a abordagem multinível podem ser aplicados em bases de dados longitudinais não balanceadas e com intervalos irregulares de medidas, ainda permitindo a inclusão de variáveis explicativas para cada um dos níveis principalmente aquelas que variam com o tempo. Ainda comparados aos modelos de corte transversal, os modelos longitudinais multiníveis indicam para uma maior variabilidade das observações no nível dois, o dos indivíduos, e por isso este nível deve ser bem representado no modelo através de variáveis com características individuais.

Um dos modelos longitudinais mais conhecidos dentro desta abordagem é o chamado Modelo de Curva de Crescimento (RABE-HESKETH; SKRONDAL, 2005) onde a variável z, com efeito aleatório, representa a medida de tempo *j*, sendo expresso como

$$y_{tij} = \boldsymbol{x}_{(1)tij}^{T}\boldsymbol{\beta}_1 + \boldsymbol{x}_{(2)ij}^{T}\boldsymbol{\beta}_2 + \boldsymbol{x}_{(3)j}^{T}\boldsymbol{\beta}_3 + v_j + \boldsymbol{z}_{tij}^{T}\boldsymbol{u}_{ij} + e_{tij}\,. \tag{9}$$

Comparado ao modelo da equação 4, o modelo de curva de crescimento na equação 8 inclui os conjuntos de variáveis para cada um dos três níveis da estrutura dos dados. O vetor $\boldsymbol{x}_{(1)tij}$ contém as variáveis que variam com o tempo inclusive uma variável que representa a medida do tempo (FREES, 2004). Os outros componentes do modelo 8 são os componentes do erro e o vetor $\boldsymbol{z}_{tij}^{T}$ definido em forma matricial como

$$Z_{ij} = \begin{pmatrix} 1 & z_{1_{ij}} \\ 1 & z_{2_{ij}} \\ \vdots & \vdots \\ 1 & z_{T_{ij}} \end{pmatrix}. \tag{10}$$

A matriz Z_{ij} contém o vetor para o intercepto aleatório e a variável de tempo com efeito aleatório que aqui é chamada de z. O modelo de curva de crescimento mantém as hipóteses do modelo de efeitos aleatórios, logo $v_j \sim N(0,\sigma_v^2)$, $\boldsymbol{u}_{ij} \sim MN(\boldsymbol{0},\Sigma_u)$ e $e_{tij} \sim N(0,\sigma_e^2)$. Seguindo essa formulação, o modelo de curva de crescimento estima uma trajetória linear contínua para cada indivíduo i, uma vez que z é expressa de forma contínua. Esta variável ainda pode ser incluída no modelo como uma função polinomial, com termos quadráticos e cúbicos. Como nos modelos de efeitos aleatórios a hipótese de homocedasticidade é relaxada no

modelo de curva de crescimento e a matriz de covariância dos erros passa a ser expressa em função dos valores da variável de tempo t. No entanto, o modelo ainda assume que as medidas de um mesmo indivíduo são não correlacionadas[4], uma hipótese muitas vezes implausível particularmente quando as medidas são tomadas considerando um intervalo pequeno de tempo entre cada uma delas (HOX, 2000).

3.2.1 Modelos multivariados multiníveis

Uma alternativa para lidar com a possível correlação existente entre as medidas repetidas de um mesmo indivíduo dentro da abordagem multinível é a de se adotar os modelos multivariados multiníveis (Capítulo 6, GOLDSTEIN (2003)). Nesta formulação, as respostas repetidas são consideradas como componentes de um vetor aleatório normal *j*-variado, o que é equivalente a ter um modelo multinível para cada ponto do tempo, todos sendo simultaneamente estimados. A hipótese de independência condicionada entre as medidas repetidas não é mais válida aqui, e ao invés assume-se que existe correlação residual entre as medidas repetidas (STEELE, 2008). Nesta formulação os componentes da matriz de covariância dos erros são parâmetros estimados pelo modelo e diferentes estruturas de correlação podem ser impostas a essa matriz.

O modelo multivariado longitudinal multinível pode ser visto como uma extensão do modelo de curva de crescimento tratando o tempo como uma variável discreta. O modelo então inclui uma variável indicadora para cada momento do tempo (dummies de tempo) e estas variáveis são consideradas como tendo efeitos fixos e aleatórios. Considere o seguinte modelo

$$y_{tij} = \boldsymbol{d}_{tij}^{T}\boldsymbol{\beta} + v_j + \boldsymbol{d}_{tij}^{T}\boldsymbol{u}_{ij}. \qquad (11)$$

onde $\boldsymbol{d}_{tij}^{T}$ são as dummies de tempo associadas ao vetor de efeitos fixos $\boldsymbol{\beta}^{T}$ e ao vetor de efeitos aleatórios no nível dos indivíduos $\boldsymbol{u}_{ij}$. Este é um modelo

[4]Por definição, uma das suposições da abordagem de modelos multiníveis é que os resíduos do nível um, e_{tij}, são mutuamente independentes. Esta hipótese implica na falta de correlação entre os resíduos de nível um de duas medidas repetidas de um mesmo indivíduo.

multivariado de dois níveis (dois termos de erro) onde assume-se que o tempo varia de forma aleatória apenas no nível de indivíduos e portanto

$$v_j \sim N(0,\sigma_v^2) \text{ e } u_{ij} \sim MN\,(0,\Sigma_u). \qquad (12)$$

O nível um para as ocasiões de tempo apenas define a estrutura multivariada dos dados e por isso o modelo não estima o termo de variância para este nível (GOLDSTEIN, 2003).

Nesta seção, a matriz Σ_u contém os termos de variância e covariância para cada ponto do tempo, uma vez que cada dummy de tempo contém um efeito aleatório associado a ela. O tamanho desta matriz vai depender do número de medidas repetidas e os seus parâmetros podem sofrer restrições lineares e não lineares para que se possa expressar uma estrutura de correlação desejada, como por exemplo, as auto-regressiva, uniforme, exponencial ou Toeplitz (Veja SINGER; WILLETT(2003), DIGGLE; LIANG; ZEGER(2002) ou FITZMAURICE; LAIRD; WARE(2004)) para a forma dessas matrizes de correlação). Ou ainda, nenhuma restrição pode ser imposta e os parâmetros da matriz Σ_u serão livremente estimados pelos dados em análise. Note que a matriz de covariância dos resíduos do modelo 10 é dada por $\Sigma_r = \Sigma_v + \Sigma_u = J\sigma_v^2 + \Sigma_u$, onde J é uma matriz com o valor 1 para todos os elementos, portanto Σ_r terá a mesma estrutura imposta sobre Σ_u, e junto com a matriz D_{ij} formada pelas dummies de tempo, o vetor multivariado de respostas é definido como $y_{ij} \sim MN(D_{ij}\beta,\Sigma_r)$.

O modelo multivariado na equação 10 pode ser aplicado a dados de painel não balanceados que pode incluir indivíduos com diferentes números de medidas no tempo determinado tanto pelo desenho da amostra como de forma aleatória. Logo, esse modelo pode ser aplicado aos dados da PME, que são provenientes de um desenho de painel rotativo e os parâmetros da matriz de covariância dos resíduos podem ser restringidos de forma a acomodar o padrão rotacional da amostra seguindo as idéias apresentadas por YANG et al. (2002). No entanto, essa formulação pode não ser adequada para longas séries de tempo.

4. Resultados

Esta seção apresenta os resultados para a modelagem dos dados da PME descritos na seção 2. Foi considerado um banco de dados de painéis balanceados, sem desgaste ou não-resposta, com as oito entrevistas de chefes de domicílios empregados e com algum rendimento do trabalho. Também foram considerados três níveis da estrutura hierárquica dos dados: UPAs, Chefes de Domicílios e Entrevistas. Os modelos aqui apresentados variam no grau de complexidade com o principal objetivo de introduzir num só exercício de modelagem as características do desenho complexo da PME. Como mencionado acima, os dados da PME são gerados através de uma amostra de painéis rotativos simétricos não-consecutivos, caracterizado como 4-8-4. Devido a este padrão rotacional, existe um vão de oito meses entre a quarta e quinta entrevistas, portanto o principal objetivo desta aplicação é o de incorporar este vão na aplicação de um modelo de curva de crescimento e modelos multivariados, o que pode ser visto como um problema de modelar dados longitudinais medidos em intervalos irregulares de tempo.

Como o modelo de curva de crescimento considera o tempo com uma variável contínua a forma mais simples de incorporar o padrão rotacional na análise dos dados da PME, seguindo esta formulação, é a de declarar a variável tempo variando de 0 a 15, dessa forma $t = (0, 1, 2, 3, 12, 13, 14, 15)$, ao invés de considerar o tempo variando de forma consecutiva de 0 a 7. Para este modelo, a matriz de variáveis associados aos efeitos aleatórios é então

$$Z_{ij} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 12 \\ 1 & 13 \\ 1 & 14 \\ 1 & 15 \end{pmatrix}.$$

Os modelos multivariados consideram o tempo como uma variável categórica tornando a inclusão do padrão rotacional uma tarefa menos fácil ou direta quanto para os modelos de curva de crescimento. Uma forma de se alcançar tal objetivo é a de impor restrições aos parâmetros da matriz de covariância dos erros de forma

que seus elementos dependam da distância temporal entre cada par de entrevistas (lags) seguindo o padrão dado por

$$\begin{pmatrix} \mathbf{0} & & & & & & & \\ 1 & \mathbf{0} & & & & & & \\ 2 & 1 & \mathbf{0} & & & & & \\ 3 & 2 & 1 & \mathbf{0} & & & & \\ \mathbf{12} & 11 & \mathbf{10} & 9 & \mathbf{0} & & & \\ 13 & \mathbf{12} & 11 & \mathbf{10} & 1 & \mathbf{0} & & \\ \mathbf{14} & 13 & \mathbf{12} & 11 & 2 & 1 & \mathbf{0} & \\ 15 & \mathbf{14} & 13 & \mathbf{12} & 3 & 2 & 1 & \mathbf{0} \end{pmatrix}.$$

Como se trata de um problema de modelagem de dados longitudinais com intervalos desiguais de medidas, algumas estruturas de correlação não podem ser impostas por não se adequarem a este tipo de dados. Exemplos para isso são as estruturas auto-regressivas ou exponenciais ou a Toeplitz, pois para essas estruturas assume-se que o distanciamento entre as medidas obedecem a uma regularidade. Duas estruturas alternativas desenvolvidas para se adaptarem ao desenho do painel da PME em VEIGA (2010), são aqui aplicadas. A primeira, chama-se Potência Temporal (com base na "Spatial Power" em SAS Institute Inc,Version 8, (1999)) e é uma re-parametrização da correlação exponencial onde o coeficiente de correlação é elevado a potência igual ao lag entre as medidas. Para o padrão rotacional da PME, essa matriz toma a seguinte forma:

$$\begin{pmatrix} \sigma^2 & & & & & & & \\ \sigma^2\rho^1 & \sigma^2 & & & & & & \\ \sigma^2\rho^2 & \sigma^2\rho^1 & \sigma^2 & & & & & \\ \sigma^2\rho^3 & \sigma^2\rho^2 & \sigma^2\rho^1 & \sigma^2 & & & & \\ \sigma^2\rho^{12} & \sigma^2\rho^{11} & \sigma^2\rho^{10} & \sigma^2\rho^9 & \sigma^2 & & & \\ \sigma^2\rho^{13} & \sigma^2\rho^{12} & \sigma^2\rho^{11} & \sigma^2\rho^{10} & \sigma^2\rho^1 & \sigma^2 & & \\ \sigma^2\rho^{14} & \sigma^2\rho^{13} & \sigma^2\rho^{12} & \sigma^2\rho^{11} & \sigma^2\rho^2 & \sigma^2\rho^1 & \sigma^2 & \\ \sigma^2\rho^{15} & \sigma^2\rho^{14} & \sigma^2\rho^{13} & \sigma^2\rho^{12} & \sigma^2\rho^3 & \sigma^2\rho^2 & \sigma^2\rho^1 & \sigma^2 \end{pmatrix}.$$

A segunda estrutura alternativa, chama-se matriz de correlação linear geral lag-dependente (com base na "General Linear" em SAS Institute Inc,Version 8, (1999)) que para o padrão rotacional da PME assume a seguinte forma:

$$\begin{pmatrix} \gamma_0 & & & & & & & \\ \gamma_1 & \gamma_0 & & & & & & \\ \gamma_2 & \gamma_1 & \gamma_0 & & & & & \\ \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 & & & & \\ \gamma_{12} & \gamma_{11} & \gamma_{10} & \gamma_9 & \gamma_0 & & & \\ \gamma_{13} & \gamma_{12} & \gamma_{11} & \gamma_{10} & \gamma_1 & \gamma_0 & & \\ \gamma_{14} & \gamma_{13} & \gamma_{12} & \gamma_{11} & \gamma_2 & \gamma_1 & \gamma_0 & \\ \gamma_{15} & \gamma_{14} & \gamma_{13} & \gamma_{12} & \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 \end{pmatrix}.$$

Essa estrutura de covariância foi motivada pela estrutura Toepliz, de forma que pares de medidas com o mesmo lag terão o mesmo grau de correlação, no entanto, as correlações entre medidas mais distanciadas decaem menos rapidamente do que na estrutura de covariância com potência temporal[5].

4.1 Seleção dos modelos

Na literatura de modelos multiníves, não existe nenhuma regra pré-definida para a seleção de modelos longitudinais. Os modelos expostos nesta seção tiveram como base uma pré-modelagem de corte transversal utilizando apenas a primeira entrevista de cada chefe de domicílio. Tendo como variável dependente contínua o rendimento do trabalho em preços reais e na escala logarítmica, variáveis independentes para cada nível da estrutura dos dados foram incluídas e testadas (veja Tabela 1 para a lista de variáveis). A pré-modelagem também testou os efeitos de interações entre variáveis de um mesmo nível e de níveis diferentes e alguns efeitos aleatórios. Os efeitos foram mantidos nos modelos sempre quando significativos ao nível de 5% e os critérios de qualidade de ajuste AIC e BIC foram utilizados para comparar os diferentes modelos. Uma vez já na modelagem longitudinal, os efeitos foram testados mais uma vez e optou-se por manter apenas o efeito aleatório do tempo para evitar problemas de convergência dos métodos de estimação.

A Tabela 2 apresenta as estatísticas da qualidade do ajuste de quatro modelos: um modelo de curva do crescimento, três modelos multivariados, impondo cada uma das duas matrizes alternativas de covariância descritas acima e

5 Tanto o modelo de curva de crescimento quanto o modelo longitudinal multivariado impondo as estruturas alternativas de covariância, podem ser processados utilizando o pacote SAS e sua Proc Mixed (SINGER, 1998).

um terceiro sem impor qualquer restrição aos parâmetros de Σ_u. O modelo impondo a matriz de potência temporal é o menos adequado pois apresenta os maiores valores para os critérios AIC e BIC. Os modelos com os menores valores, e por isso os mais adequados, são os modelos multivariados com matrizes de covariância impondo a estrutura linear geral lag-dependente e o que não impõe restrições. Este último tem a sua matriz de covariância livremente determinada pelos dados, no entanto, é o menos parcimonioso. Logo, o modelo multivariado que impõe a estrutura linear geral lag-dependente é o que melhor se adéqua aos dados da PME.

Tabela 2: Medidas da Qualidade do Ajuste

Modelo	AIC	BIC (SAS)
Modelo de Curva de Crescimento	21,246	21,273
Sem Impor Estrutura	20,446	20,649
Potência Temporal	30,178	30,195
Linear Geral Lag-Dependente	20,644	20,709

A Tabela 3 apresenta as matrizes de covariância estimadas para cada um dos quatro modelos da Tabela 2. Esses resultados confirmam uma queda muito rápida em direção a zero da correlação para o modelo que impõe a estrutura de potência temporal que nos outros modelos. A matriz para o modelo que não impôs restrições à matriz de covariância Σ_u indica na direção daquela para a estrutura linear geral lag-dependente, ou seja, esta última se mostra bem adequada para os dados em estudo.

A Tabela 4 apresenta os resultados para a parte fixa do modelo multivariado longitudinal que impõe a estrutura linear geral lag-dependente para os dados da PME. Este modelo final foi processado utilizando o método de estimação por MQGI com a opção para erros-padrão robustos que seguem a metodologia do estimador de sanduíche (RABE-HESKETH; SKRONDAL, 2006), que protegem os resultados contra uma má especificação da hipótese de normalidade dos erros, tendo um efeito principal na estimação da parte aleatória do modelo (HOX, 2000). Os gráficos de diagnóstico dos resíduos para o modelo na Tabela 4 se encontram em apêndice.

Note que a Tabela 4 apresenta os resultados para os coeficientes estimados pelo modelo e seus respectivos erros-padrão robustos considerando as variáveis

descritas na Tabela 1 seguindo a uma ordem, de primeiro expor os resultados dos efeitos principais, seguido pelos termos de interação e alguns efeitos de contexto. Um outro ponto importante de se observar é sobre a interpretação dos coeficientes estimados pelo modelo, uma vez que a renda do trabalho foi expressa na escala logarítmica. Portanto, ao invés de interpretar o efeito de cada variável como uma mudança esperada de $\hat{\beta}_k$ sobre a renda para uma mudança de uma unidade em x_k, interpreta-se como um acréscimo ou decréscimo, dependendo do sinal, de $e^{\hat{\beta}_k}$. Ou seja, haverá uma mudança de $b_k\,\% = 100 \times (e^{\hat{\beta}_k}) - 1)\%$ no rendimento do trabalho para uma mudança de uma unidade em x_k (TUFTE, 1974).

De uma forma geral as variáveis incluídas no modelo apresentaram resultados de acordo com o esperado pela literatura existente sobre modelagem da renda do trabalho como em BARROS; CORSEUIL; LEITE(2000). Através de uma interpretação geral do modelo nota-se que o rendimento do trabalho, embora tenha demonstrado uma leve tendência ao crescimento, pareceu ser também bastante estável ao longo do tempo demonstrando uma forte correlação positiva entre as diferentes medidas. As demais variáveis seguem alguns dos resultados tradicionais onde chefes de domicílios homens e brancos têm um maior rendimento; o efeito da experiência sobre os rendimentos, representado pela idade, tem um formato U, sendo os homens os que mais se beneficiam de uma maior experiência, já a experiência na firma apresenta uma relação de formato U invertido, mostrando que os rendimentos, inicialmente, aumentam de forma mais rápida. A Educação dos chefes dos domicílios tem uma relação com os rendimentos em formato J, demonstrando que um maior nível educacional está associado com um maior nível de rendimento. De forma geral, vê-se que o setor da indústria está associado a maiores níveis de rendimento do trabalho; comparados aos empregadores todos os outros tipos de trabalhadores têm menores rendimentos e as regiões do sul e sudeste têm o diferencial de rendimento a seu favor.

Tabela 3: Matrizes de Covariância

Modelos	Σ_r							
Modelo de Curva do Crescimento	0,3237							
	0,2742	0,3155						
	0,2698	0,2665	0,3083					
	0,2655	0,2627	0,2598	0,3022				
	0,2261	0,2280	0,2299	0,2318	0,2938			
	0,2218	0,2242	0,2266	0,2290	0,2506	0,2981		
	0,2174	0,2203	0,2232	0,2262	0,2525	0,2554	0,3034	
	0,2130	0,2165	0,2199	0,2234	0,2544	0,2578	0,2613	0,3098
Sem Impor Estrutura	0,3189							
	0,2708	0,3187						
	0,2646	0,2722	0,3178					
	0,2565	0,2642	0,2696	0,3052				
	0,2254	0,2294	0,2309	0,2285	0,3042			
	0,2198	0,2239	0,2232	0,2249	0,2633	0,2988		
	0,2173	0,2210	0,2208	0,2197	0,2553	0,2592	0,2948	
	0,2165	0,2207	0,2209	0,2211	0,2535	0,2562	0,2588	0,2965
Potência Temporal	0,3199							
	0,2809	0,3199						
	0,2473	0,2809	0,3199					
	0,2182	0,2473	0,2809	0,3199				
	0,0830	0,0908	0,0998	0,1103	0,3199			
	0,0762	0,0830	0,0908	0,0998	0,2809	0,3199		
	0,0704	0,0762	0,0830	0,0908	0,2473	0,2809	0,3199	
	0,0653	0,0704	0,0762	0,0830	0,2182	0,2473	0,2809	0,3199
Linear Geral Lag-Dependente	0,3068							
	0,2655	0,3068						
	0,2600	0,2655	0,3068					
	0,2552	0,2600	0,2655	0,3068				
	0,2228	0,2242	0,2282	0,2284	0,3068			
	0,2207	0,2228	0,2242	0,2282	0,2655	0,3068		
	0,2187	0,2207	0,2228	0,2242	0,2600	0,2655	0,3068	
	0,2162	0,2187	0,2207	0,2228	0,2552	0,2600	0,2655	0,3068

Tabela 4: Modelo Multivariado Multinível: Coeficientes Estimados

Variável	Coef	Rob.EP	Variável	Coef	Rob.EP
Dummy de Tempo			Experiência na firma (x120)	0,053	0,006
d_0	6,290	0,285	termo quadrático	-0,014	0,003
d_1	6,295	0,285	Horas trabalhadas (in Log)	0,251	0,029
d_2	6,293	0,285	Proxy Respondente	0,004	0,008
d_3	6,299	0,285	Região Metropolitana (Recife como ref.)		
d_{12}	6,328	0,285	Salvador	0,037	0,033
d_{13}	6,323	0,285	Belo Horizonte	0,244	0,029
d_{14}	6,322	0,285	Rio de Janeiro	0,211	0,028
d_{15}	6,327	0,285	São Paulo	0,36	0,03
Homens	0,599	0,137	Porto Alegre	0,226	0,033
Brancos	0,226	0,036	***Interação de Homens com:***		
Idade (na entrevista 1)	3,485†	1,429†	Idade (na entrevista 1)	0,004	0,002
termo quadrático	-0,450†	0,056†	Educação (na entrevista 1)	0,057	0,014
Educação (na entrevista 1)	-0,069	0,012	termo quadrático	-0,004	0,001
termo quadrático	0,009	0,001	Horas trabalhadas (in Log)	-0,085	0,034
Tipo de Trabalhador (Empregador como ref.)			Proxy Respondente	-0,042	0,009
Informal	-0,135	0,029	***Interação de Brancos com:***		
Formal	-0,044	0,029	Tipo de Trabalhador (Empregador como ref.)		
Militar	-0,006	0,034	Informal	-0,062	0,035
Conta própria	-0,175	0,024	Formal	-0,105	0,038
Tipo de Atividade (Indústria como ref.)			Militar	-0,146	0,045
Construção	-0,008	0,014	Conta própria	-0,015	0,031
Comercio	-0,041	0,009	***Efeitos de Contexto:***		
Financeira	-0,006	0,012	Prop de Formal	-1,845	0,234
Serviços Sociais	0,033	0,015	Prop de Informal	-2,175	0,268
Serviços domésticos	-0,045	0,023	Prop de Militar	-2,069	0,273
Outros serviços	-0,029	0,011	Prop de Conta Própria	-2,28	0,274
Outras	-0,028	0,034	Media de Educação	0,072	0,007
-2 x *Log-Likelihood*	20.098		Num. de UPA	1.762	
AIC	20.218		Num. de Indivíduos	6.524	
BIC	20.547		Num. de Entrevistas	8	

Nota: † Valores elevados a 10^3.

5. Conclusões

Com o objetivo de se modelar longitudinalmente os dados da pesquisa mensal de emprego (PME), este artigo apresentou uma aplicação de modelos para o rendimento do trabalho de chefes de domicílio. A PME não é uma pesquisa desenhada para servir a análises longitudinais, mas o uso de alguns critérios de emparelhamento de painéis permite a formação de um banco de dados longitudinal um tanto complexo.

Este artigo apresentou uma breve descrição do desenho amostral da PME que inclui uma estrutura hierárquica de diversos níveis e um padrão rotacional da amostra caracterizado como 4-8-4. O exercício de modelagem descrito acima incorporou essas duas principais complexidades através da abordagem de modelos longitudinais multiníveis e multivariados, também descritos neste artigo. A estrutura hierárquica dos dados foi incorporada na análise considerando os níveis das UPAs, dos chefes de domicílios e das medidas repetidas para cada chefe e o padrão rotacional foi incorporado, principalmente, impondo uma estrutura lag-dependente à matriz de covariância dos erros.

Nenhum dos modelos aqui apresentados incorporou o peso longitudinal para a PME, embora isso seja possível como demonstrado em VEIGA(2010). No entanto, na presença de um plano amostral informativo, uma alternativa ao uso dos pesos amostrais seria a inclusão das variáveis do desenho no modelo e a aplicação de métodos robustos de estimação. Os modelos multiníveis apresentados neste artigo incluíram algumas variáveis determinantes do desenho amostral da PME e foram estimados utilizando métodos robustos para a estimação dos erros-padrão além de controlarem pela correlação existente entre as medidas repetidas através da imposição de restrições aos parâmetros da matriz de covariância dos erros.

Referências Bibliográficas

ANTONACI, G. d. A.; SILVA, D. B. d. N. Analysis of alternative rotation patterns for the Brazilian system of integrated household surveys. In: Proceedings of the 56th Session of the International Statistical Institute (ISI). [S.l.: s.n.], 2007.

BARROS, R. P. d.; CORSEUIL, C. H.; LEITE, P. G. Labor Market and Poverty in Brazil. [S.l.], 2000.

DIGGLE, P. J.; LIANG, K.-Y.; ZEGER, S. L. The Analysis of Longitudinal Data. Second. Oxford: Oxford University Press, 2002. (Oxford Statistical Science).

FITZMAURICE, G. M.; LAIRD, N. M.; WARE, J. H. Applied Longitudinal Analysis. New York: Wiley-Interscience, 2004. (Wiley Series in Probability and Statistics).

FREES, E. W. Longitudinal and Panel Data: Analysis and Applications in the Social Sciences. New York: Cambridge University Press, 2004.

GOLDSTEIN, H. Multilevel mixed linear model analysis using iterative genera-lized least squares. Biometrika, v. 73, n. 1, p. 43-56, 1986. Disponível em: <http://biomet.oxfordjournals.org/cgi/content/abstract/73/1/43>.

GOLDSTEIN, H. Multilevel statistical models. 3rd. ed. London: Arnold, 2003. (Kendall's Library of Statistics; 3).

HECKMAN, J. J.; SINGER, B. Longitudinal Analysis of Labor Market Data. Cambridge: Cambridge University Press, 1985.

HEDEKER, D.; GIBBONS, R. D. Application of random-e ects pattern-mixture models for missing data in longitudinal studies. Psychological Methods, v. 2, n. 1, p. 64-78, 1997.

HOX, J. J. Multilevel analyses of grouped and longitudinal data. In: LITTLE, T. D.; SCHNABEL, K. U.; BAUMERT, J. (Ed.). Modelling Longitudinal Multiple-Group Data: Practical Issues, Applied Approaches, and Speci c Examples. [S.l.]: Lawrence Erlbaum associates, 2000.

IBGE. Pesquisa Mensal de Emprego. Serie Relatórios Metodológicos. [S.l.], 2002. v. 23, 74 p. http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pme_nova/srmpme_2ed.pdf.

JENKINS, S. P. Modelling household income dynamics. Journal of Population Economics, v. 13, n. 4, p. 529 - 567, 2000.

MINCER, J.; POLACHEK, S. Family investment in human capital: Earnings of women. Journal of Political Economy, v. 82, n. 2, p. S76-S108, Part II, 1974.

RABE-HESKETH, S.; SKRONDAL, A. Multilevel and Longitudinal Modeling using Stata. College Station, Tex.: Stata Press, 2005.

RABE-HESKETH, S.; SKRONDAL, A. Multilevel modelling of complex survey data. Journal of the Royal Statistical Society. Series A, v. 127, n. 4, p. 805-827, 2006.

SAS Institute Inc,Version 8. SAS OnlineDoc R. Cary, NC, 1999.

SINGER, J. D. Using SAS PROC MIXED to t multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics, v. 23, n. 4, p. 323-355, 1998.

SINGER, J. D.; WILLETT, J. B. Applied Longitudinal Data Analysis: Modeling Change and Occurrence. New York: Oxford University Press, 2003.

SNIJDERS, T. A. B.; BOSKER, R. J. Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks, CA: Sage Publications Ltd, 1999.

STEELE, F. Multilevel models for longitudinal data. Journal of the Royal Statistical Society. Series A, v. 127, n. 1, p. 5-19, 2008. Disponível em: <http://ideas.repec.org/a/bla/jorssa/v171y2008i1p5-19.html>.

TUFTE, E. R. Data analysis for politics and policy. N.J.: Prentice-Hall, Englewood Cli s, 1974.

VEIGA, A. d. C. Methods for analysing complex panel data using multilevel models with an application to the Brazilian labour force survey. Tese (Doutorado).University of Southampton, School of Social Sciences, 2010.

YANG, M.; GOLDSTEIN, H.; BROWNE, W.; WOODHOUSE, G. Multivariate multilevel analyses of examination results. Journal of the Royal Statistical Society. Series A, v. 165, n. 1, p. 137-153, 2002.

Agradecimentos

Aos professores doutores Peter W. F. Smith e James J. Brown. Este estudo foi realizado enquanto a autora cursava o doutorado pela University of Southampton (Inglaterra) e financiada pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, MEC – BEX 2392-04-6).

Abstract

There are different approaches for the analysis of longitudinal data, also known as panel or repeated measures data. The multilevel modeling approach allows the analysis longitudinal data from complex sample surveys by taking into account the main features of the data set such as: the hierarchical data structure; the complex structure of the residual correlation and the complex sampling design including rotating panels and panel attrition. This article presents an application of longitudinal multilevel models to the data from the Brazilian labour force survey (the PME). This survey has a complex sampling design which includes a non-consecutive rotating panel scheme characterised as 4-8-4. The labour income dynamics of employed heads of households is investigated by means of models that present different level of complexities and incorporate the sampling design of the PME.

Key words: Longitudinal Data; Brazilian Labour Force Survey (PME); multilevel models, correlation matrix.

Apêndice: Diagnostico dos Resíduos para o Modelo Multivariado com Estrutura Linear Geral Lag-Dependent

Figura 1: Resíduos de Nível Um: Indivíduos

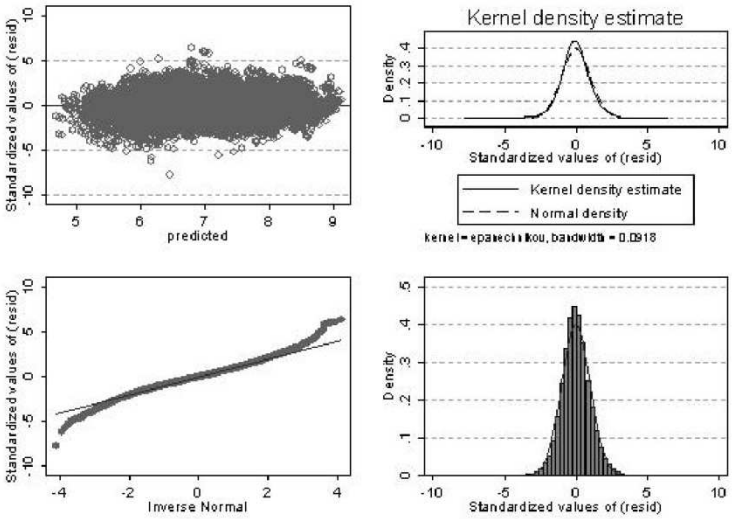
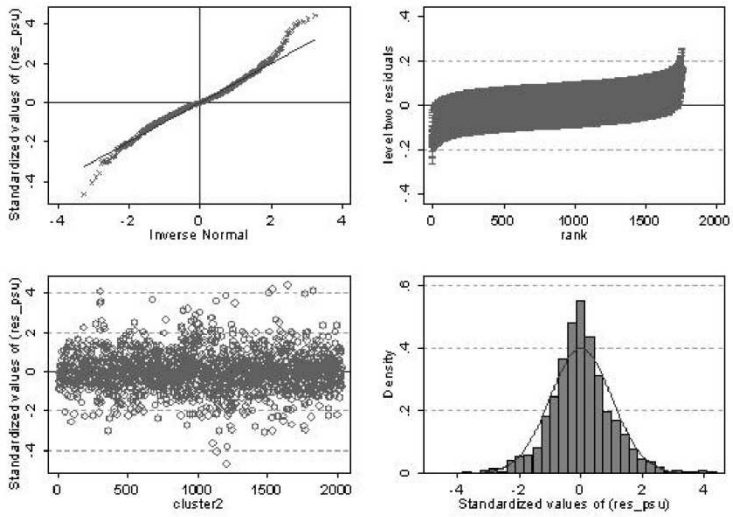


Figura 2: Resíduos de Nível Dois: UPAs



O comportamento dos índices de ações em países emergentes: uma análise com dados em painel e modelos hierárquicos

Luiz Paulo Lopes Fávero[1]
José Elias Feres de Almeida[2]

Resumo

Estudos que investigam a influência do mercado e da economia sobre índices de ações de países emergentes são escassos. Sob a hipótese de que há diferenças nos comportamentos dos índices de ações destes países, este estudo utiliza-se da modelagem de dados em painel e hierárquica para avaliar os efeitos de variáveis dos mercados acionários e investigar as influências econômicas sobre o comportamento dos índices de ações ao longo do tempo. Com a utilização de uma amostra proveniente da Compustat Global, com 60 países emergentes ao longo de 262 meses (1986-2007), totalizando 9.979 observações, verifica-se que os retornos dos dividendos, a relação preço/book value e o turnover do mercado influenciam o comportamento dos índices de ações, por meio do ajuste de um modelo de regressão linear com efeitos fixos e termos de erro AR(1). Verifica-se também que há representatividade do efeito país para a diferenciação das taxas de crescimento dos índices de ações, com destaque para o investimento em pesquisa.

Palavras-Chave: Dados em Painel; Efeito País; Medidas Repetidas; Mercado de Ações; Modelos Hierárquicos.

1 Faculdade de Economia, Administração e Contabilidade da USP – End.: Av. Prof. Luciano Gualberto, 901 - FEA 3 - Cidade Universitária – São Paulo-SP, CEP 05508-900 – email: lpfavero@usp.br
2 Universidade Federal do Espírito Santo (UFES), Avenida Fernando Ferrari, 514, CCJE, sl. 54 - Campus Universitário, Vitória – ES, CEP: 29075-910 – email: joseelias@ccje.ufes.br

1. Introdução

Muitos trabalhos com foco no mercado financeiro abordam aspectos relativos aos impactos gerados por um determinado grupo de variáveis sobre comportamentos específicos que propiciem informação aos tomadores de decisão e, neste sentido, as ferramentas econométricas proliferam em estudos importantes da área de finanças e contabilidade que têm como objetivos a criação de modelos que expliquem a realidade e a verificação empírica de fenômenos observados (Watts & Zimmerman, 1986). Este enfoque tem por objetivo descrever como se desenrola o mundo real e predizer o que irá ocorrer e, desta maneira, seu objetivo relaciona-se com a investigação das razões pelas quais as organizações tomam determinadas decisões em detrimento de outras, dentro de um contexto mercadológico e econômico (Martins & Theóphilo, 2009).

Segundo King & Levine (1993), Levine (1997), Rajan & Zingales (1998; 2003) e Roe & Siegel (2009), o desenvolvimento financeiro é atualmente visto como necessário e útil para o desenvolvimento econômico de uma nação, que retroalimenta o próprio desenvolvimento do mercado acionário. De acordo com Roe & Siegel (2009), muitos fatores podem explicar as diferenças no desenvolvimento dos mercados financeiros ao redor do mundo, em especial para os países emergentes. Além de características associadas às próprias empresas e organizações, muitas variáveis macroeconômicas e sociais podem interferir de forma diferenciada para a evolução dos indicadores do mercado financeiro ao longo do tempo. Mas o que induz uma nação a apresentar comportamentos diferenciados para a evolução de seus indicadores em relação a outras? Acredita-se, conforme apontam Dyck & Zingales (2004), que seja uma combinação de fatores relacionados a aspectos legais, econômicos e sociais, abertura comercial, condições das empresas atuantes e, de acordo com Jones & Banning (2009), a aspectos políticos. Jensen, Menezes-Filho & Sbragia (2004) ainda discutem a relação entre desenvolvimento socioeconômico dos países e inovação tecnológica. Uma discussão mais aprofundada sobre o desenvolvimento do mercado financeiro em países emergentes pode ser encontrada em Bekaert, Harvey & Lundblad (2001) e Bekaert & Harvey (2002).

De acordo com Adjasi & Biekpe (2009), o desenvolvimento do mercado acionário, em especial, torna-se, cada vez mais, um aspecto fundamental para o desenvolvimento de todo o mercado financeiro em países emergentes, já que possibilita que firmas obtenham capital por meio de mercados mais abertos, transparentes e com transações mais rápidas e eficientes, de modo a facilitar atividades de investimento e promover um crescimento mais sustentado dos meios de produção já que, segundo Tobin (1969) e Von Furstenberg (1977), a atividade do mercado acionário é positivamente correlacionada com o investimento.

Um dos fundamentos de finanças preconiza que o comportamento do mercado acionário reflete o valor presente dos dividendos futuros esperados. Como as firmas pagam dividendos a partir de seus ganhos, os quais dependem da atividade econômica real, os preços das ações e, consequentemente, os índices de bolsas de valores, refletem as expectativas da atividade econômica real atual e futura no mercado. Desde que a teoria macroeconômica aponta uma significativa relação entre política macroeconômica e atividade econômica futura esperada, deve existir, segundo Lee (1997), uma forte relação intertemporal entre os índices de ações, as características do mercado e as variáveis macroeconômicas.

A relação entre os índices de bolsas de valores e as características do mercado acionário e da economia dos países desenvolvidos é bem documentada. Porém estudos que investigam a influência de variáveis mercadológicas e econômicas sobre os índices de ações de países em desenvolvimento, sob uma perspectiva temporal, ainda são escassos (Kwon, Shin & Bacon, 1997).

O presente estudo tem o intuito de investigar os efeitos das características dos mercados acionários e de variáveis econômicas no desempenho mensal dos índices de ações de países emergentes e, para tanto, são utilizados modelos de dados em painel e de níveis hierárquicos com componentes de variância. Assim sendo, são apresentadas as questões que se pretende responder por meio da aplicação destas modelagens: Como os índices de ações dos países reagem ao efeito país? Existem diferenças nos comportamentos dos índices entre os países emergentes ao longo do tempo causadas pelas variáveis explicativas? E mais especificamente, é possível medir estas relações? Esta perspectiva permitirá verificar como o principal índice de ações de cada país reage ao chamado "efeito país".

Por meio desta abordagem, exploram-se os efeitos do mercado acionário e do país de origem sobre o índice de ações em uma escala temporal, com a utilização de uma amostra proveniente da Compustat Global com dados econômicos e do mercado de ações de 60 países emergentes, em um período de 262 meses (1986-2007).

Este trabalho procura investigar se existem diferenças nos índices de ações, tanto em termos médios quanto em termos de taxas de crescimento, entre países ao longo do tempo, e quais as razões econômicas para a existência destes fenômenos. Também traz contribuições ao estudo das fontes de heterogeneidade da evolução dos índices de ações de países emergentes ao longo do tempo, especificamente no que se refere aos efeitos do mercado e do país, por meio das técnicas de dados em painel e de modelos multinível.

Como os índices de ações sofrem influências e reagem por meio da consideração de uma estrutura semi-forte de eficiência dos mercados de capitais, mesmo que a velocidades diferentes, justifica-se analisar as variáveis do mercado de capitais (nível firma – dividendos, price-to-book, volume de negócios) e também aquelas relacionadas aos aspectos macroeconômicos (nível país – PIB per capita, população economicamente ativa, investimentos em educação e pesquisa, importações e exportações), conforme discutem Valadkhani, Chancharat & Havie (2009). Um maior incentivo à educação, um desenvolvimento do comércio internacional ou um crescimento do investimento em pesquisa podem representar um eventual aumento do índice de ações de um país, pela geração de novas tecnologias, pelo desenvolvimento de novos setores produtivos na economia e pela entrada de capital estrangeiro, que propiciam um incremento de renda, uma diversificação da carteira de investimentos e um crescimento da oferta pública de novas ações.

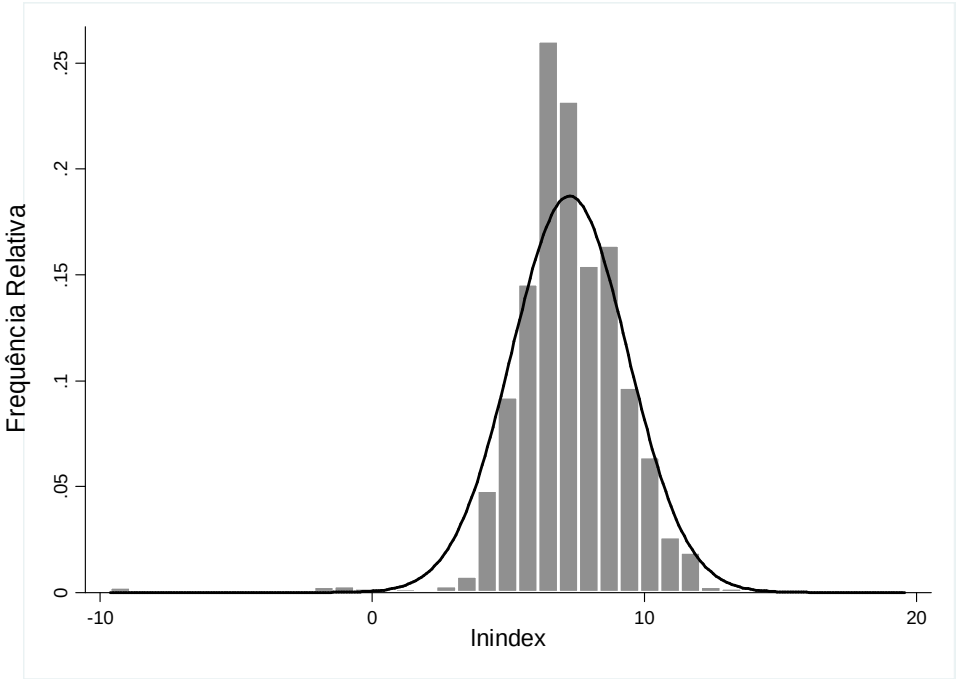
Primeiramente, são avaliadas as influências das características dos mercados acionários sobre a evolução dos índices, por meio do ajuste de diversos modelos de dados em painel. Na sequência, são analisadas as variações dos índices de ações provenientes de países diferentes ao longo do tempo e, posteriormente, investigadas as variáveis econômicas que eventualmente explicam a variação destes índices entre os próprios países. Desta forma, será utilizada também uma abordagem hierárquica em dois níveis, sendo o primeiro nível relativo à variação temporal (medida repetida) e o segundo nível às características econômicas dos países.

A seção 1 apresenta um diagnóstico da evolução dos índices de ações destes 60 países emergentes, com enfoque para as variações ao longo do tempo ou para um dado país (within variation) e para as variações dos índices que ocorrem entre países (between variation). Nas duas seções seguintes, é apresentada uma revisão da literatura sobre a modelagem de dados em painel e a modelagem hierárquica de dois níveis e com medidas repetidas. Na seção 4, são apresentados a amostragem, a definição das variáveis, as hipóteses e os modelos propostos no trabalho. A seção 5 destina-se aos resultados e às respectivas discussões. As principais conclusões são elaboradas na última seção.

2. Evolução mensal dos índices de ações de países emergentes

Inicialmente, explicita-se que a variável de desempenho adotada no presente estudo refere-se ao logaritmo natural do principal índice mensal de ações de cada um dos 60 países emergentes considerados (lnindex). Como o índice original pode variar consideravelmente de país para país, em função do número total de ações e de seus respectivos preços naquele mercado, faz-se necessária a aplicação do logaritmo natural, que propicia uma distribuição normal dos dados ao longo do período em análise, conforme pode ser observado na figura 1.

Figura 1: Distribuição do Logaritmo Natural dos Índices de Ações de Países Emergentes (1986-2007)

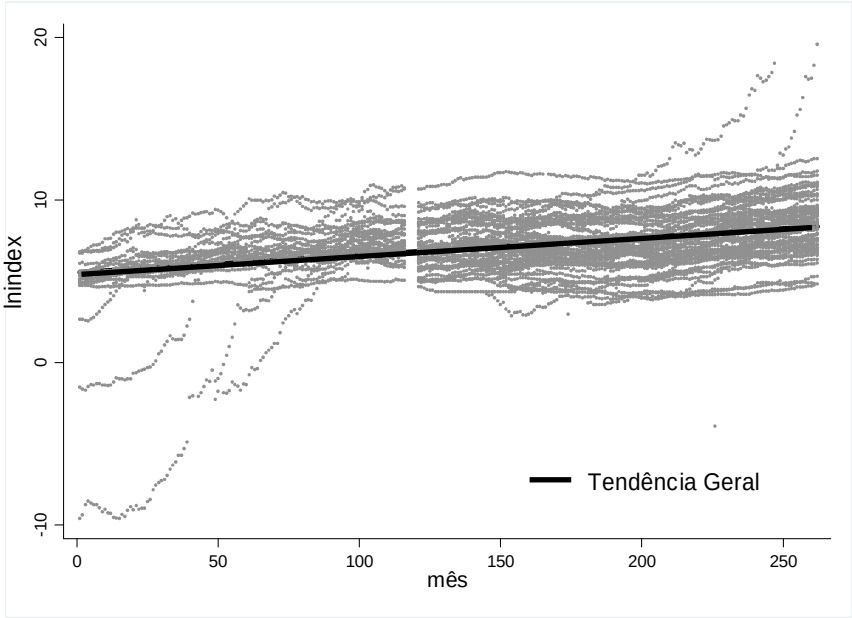


Cada ponto na figura 2 representa o valor de lnindex de um país em um determinado mês. Este comportamento sugere a elaboração de modelos longitudinais. Segundo Cameron & Trivedi (2009), muitos métodos microeconométricos enfatizam a existência de correlação, ao longo do tempo, entre as observações referentes a um dado indivíduo, no caso país, com independência entre indivíduos distintos em qualquer tempo. Porém, em diversos casos, quando se pretende modelar uma relação entre uma variável resposta e variáveis explicativas a partir de dados em painel provenientes de diferentes países, é aconselhável admitir no modelo inicial a existência de correlação entre as observações referentes a um mesmo país em tempos distintos, bem como entre diferentes países em um mesmo instante de tempo.

Ao longo do período 1986-2007, muitos países emergentes apresentaram crescimentos expressivos nos seus principais índices de ações em um ou mais meses, como pode ser observado na figura 2 para os 60 países considerados no presente estudo. Por meio da mesma figura, porém, é possível notar a existência de uma tendência geral média, linear, para o crescimento do logaritmo destes índices ao longo do tempo. Ademais, poucos países, e em um número relativamente baixo de meses, apresentaram valores dos logaritmos de seus índices de ações com padrões

discrepantes, o que sugere a existência de comportamentos similares, mesmo que, por vezes, ocorram diferenças nas médias e nas inclinações entre cada série individual.

Figura 2: Evolução do Logaritmo dos Índices de Ações de Países Emergentes



A justificativa para a adoção dos modelos de dados em painel no presente trabalho também se relaciona com o fato de que alguns regressores, como certas características econômicas dos países, são invariantes no tempo, enquanto outros, como uma tendência geral, são invariantes entre países. A maioria deles neste estudo, porém, acaba variando entre países e ao longo do tempo, conforme será apresentado e discutido na seção 4. Enquanto a figura 3 apresenta a variação do logaritmo de cada índice de ações ao longo do tempo para cada país, ou seja, mostra os desvios do logaritmo do índice de ações em relação à média individual de cada país (within variance), a figura 4 apresenta a variação dos índices de ações entre os países, ou seja, mostra os desvios dos logaritmos dos índices de ações dos países em relação à média geral para cada instante de tempo (between variation).

Figura 3: Desvios do Logaritmo do Índice de Ações em Relação à Média de Cada País ao Longo do Tempo

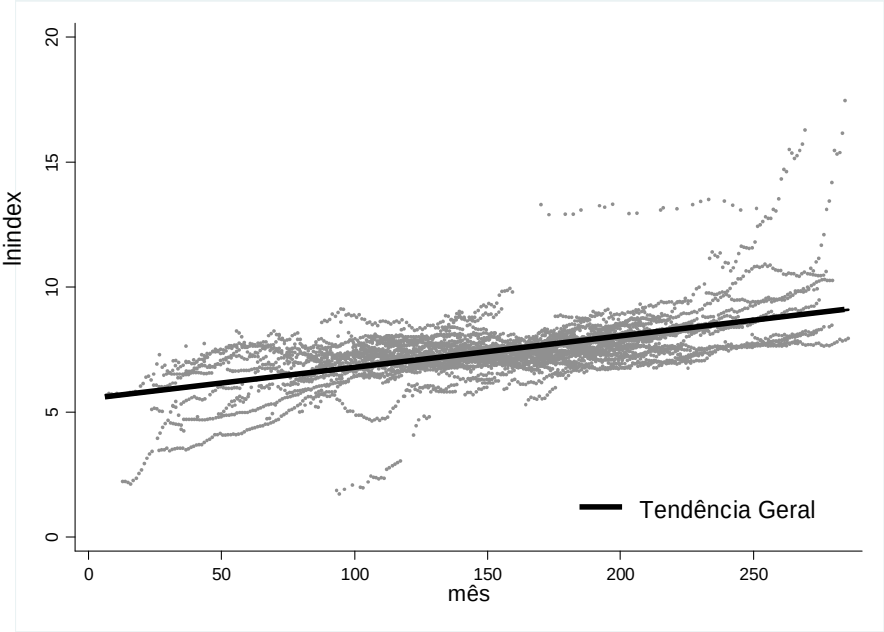
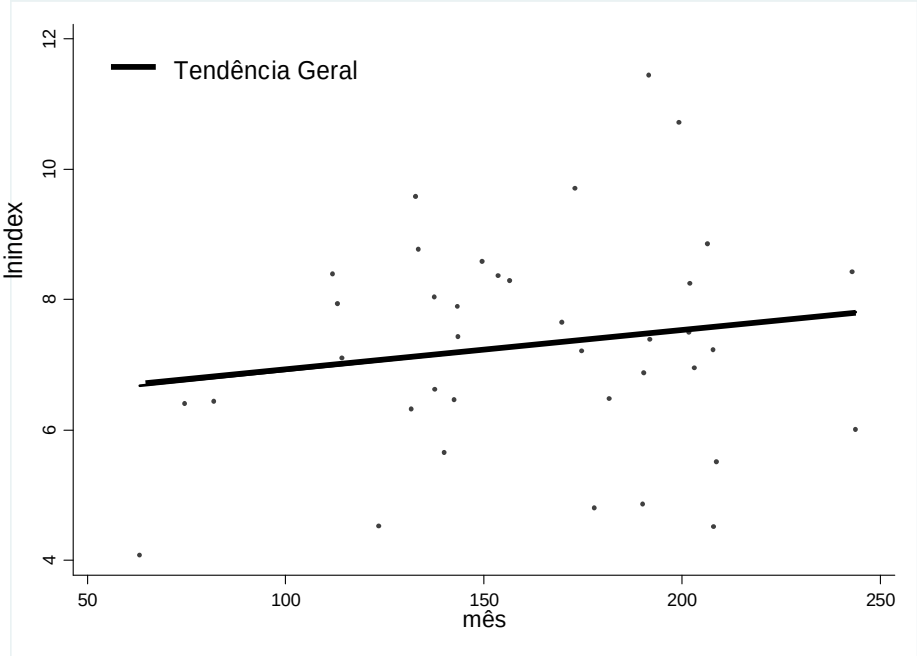


Figura 4: Desvios dos Logaritmos dos Índices de Ações dos Países em Relação à Média Geral para Cada Instante de Tempo



A seção 4 apresenta as variações within e between de cada um dos regressores que serão levados em consideração no presente estudo. Como será discutido mais adiante e de acordo com Cameron & Trivedi (2009), os modelos de dados em painel podem permitir que os regressores sejam endógenos pela existência de correlação com um componente do erro que seja invariante com o tempo (efeitos fixos), ou assumir que os regressores sejam completamente exógenos (efeitos aleatórios). Ambos estimadores serão considerados neste artigo.

Como a amostra oferece dados de 60 países em 262 meses, o painel pode ser considerado longo (T > N). Desta forma, como a influência temporal é bastante importante em séries longas, modelos de efeitos fixos e aleatórios serão também aplicados com a consideração de componentes auto-regressivos (AR(1)) para os resíduos, o que pode resultar em parâmetros com estimativas mais eficientes para painéis longos. De acordo com Kothari (2001), o entendimento dos determinantes dos processos auto-regressivos pode gerar evidências empíricas adequadas quando do estudo das características de firmas, de setores ou macroeconômicas, o que acaba por motivar a utilização de tais modelos.

Elaborado o diagnóstico inicial que embasará muitos dos critérios adotados na modelagem ao longo do estudo, espera-se que não haja uma propensão à existência de tendências idiossincráticas na amostra, que poderiam ser resultantes de comportamentos sistemáticos discrepantes dos índices de ações entre países ao longo do tempo. Em outras palavras, espera-se que existam variáveis econômicas relevantes para a explicação do comportamento dos índices de ações ao longo do tempo, o que justifica a utilização de modelos hierárquicos.

Parte-se para uma apresentação dos modelos de dados em painel e dos modelos hierárquicos de dois níveis em uma perspectiva temporal.

3. Modelagem de dados em painel

Existem muitos modelos diferentes que podem ser utilizados para dados em painel. A distinção básica entre eles, segundo Greene (2007), é a existência de efeitos fixos ou aleatórios. O termo "efeitos fixos" dá uma ideia equivocada da modelagem, pois para

ambos os casos, os efeitos no nível do indivíduo (neste caso, país) são aleatórios. Assim, segundo Cameron & Trivedi (2009), os modelos de efeitos fixos apresentam a complicação adicional de que os regressores sejam correlacionados com os efeitos do nível do indivíduo e, portanto, uma estimação consistente dos parâmetros do modelo requer uma eliminação ou controle dos efeitos fixos. Assim, um modelo que leva em conta os efeitos específicos do indivíduo para uma variável dependente y_{it} especifica que:

$$y_{it} = \beta_{0i} + x'_{it}\,\beta_1 + \varepsilon_{it} \qquad (1)$$

em que x_{it} são regressores, β_{0i} são os efeitos aleatórios específicos de indivíduo e ε_{it} representa o erro idiossincrático.

Fazendo o termo do erro ser $\mu_{it} = \beta_{0i} + \varepsilon_{it}$ e permitindo que x'_{it} seja correlacionado com o termo de erro invariante no tempo (β_{0i}), assume-se que x'_{it} não é correlacionado com o erro idiossincrático ε_{it}. O modelo de efeitos fixos implica que $E(y_{it}|\beta_{0i}, x_{it}) = \beta_{0i} + x'_{it}\beta_1$, assumindo que $E(\varepsilon_{it}|\beta_{0i}, x_{it}) = 0$, de modo que $\beta_j = \partial E(y_{it}|\beta_{0i}, x_{it})/\partial x_{j,it}$. A vantagem do modelo de efeitos fixos é que pode ser obtido um estimador consistente do efeito marginal do j-ésimo regressor de $E(y_{it}|\beta_{0i}, x_{it})$, dado que $x_{j,it}$ varia no tempo.

No modelo de efeitos aleatórios, por outro lado, assume-se que β_{0i} é puramente aleatório, ou seja, que não é correlacionado com os regressores. A estimação, portanto, é elaborada com um estimador FGLS (*feasible generalized least squares*). A vantagem do modelo de efeitos aleatórios é que este estima todos os coeficientes, mesmo dos regressores invariantes no tempo, e, portanto, os efeitos marginais. Ademais, $E(y_{it}|x_{it})$ pode ser estimado. Porém a grande desvantagem é que estes estimadores são inconsistentes se o modelo de efeitos fixos for mais apropriado.

Conforme já discutido, a variável dependente e os regressores podem potencialmente variar simultaneamente ao longo do tempo e entre indivíduos. Enquanto a variação, ao longo do tempo ou para um dado indivíduo, é conhecida por *within variance*, a variação entre indivíduos é chamada de *between variance*. De acordo com Wooldridge (2002), no modelo de efeitos fixos, o coeficiente de um regressor com baixa variação *within* será imprecisamente estimado e não será identificado se não houver

qualquer *within variance*. Assim, é de fundamental importância a distinção entre estas variações para a definição do melhor modelo de dados em painel.

A variação total das observações de um regressor x em torno da média geral $\bar{x} = 1/\sum_i T_i \sum_i \sum_t x_{it}$ no conjunto de dados pode ser decomposta na soma da variação *within* ao longo do tempo para cada indivíduo em torno de $\bar{x}_i = 1/T \sum_t x_{it}$ e na variação *between* entre indivíduos (para $\bar{x}_i$ em torno de $\bar{x}$). De acordo com Cameron & Trivedi (2009):

Variância *Within*: $s_{xW}^2 = \frac{1}{\sum_i T_i - 1} \sum_i \sum_t (x_{it} - \bar{x}_i + \bar{x})^2$

Variância *Between*: $s_{xB}^2 = \frac{1}{N-1} \sum_i (\bar{x}_i - \bar{x})^2$

Variância Geral (*Overall*): $s_{xO}^2 = \frac{1}{\sum_i T_i - 1} \sum_i \sum_t (x_{it} - \bar{x})^2$

As notações N e $\Sigma_i T_i$ correspondem, respectivamente, ao número de indivíduos e ao número total de observações ao longo do tempo. Na seção 4, são apresentadas e discutidas as variâncias de cada um dos regressores a serem considerados nos modelos no presente trabalho.

Ainda de acordo com Cameron & Trivedi (2009), os estimadores dos parâmetros β_1 de um modelo de efeitos fixos para a equação (1) eliminam os efeitos fixos β_{0i}, ou seja, é elaborada uma transformação *within* pela diferenciação de médias. Desta forma, uma estimação *within* elabora uma modelagem com os dados diferenciados em torno da média e não se pode estimar um coeficiente de um regressor sem variação ao longo do tempo. Assim, os efeitos fixos β_{0i} na equação (1) podem ser eliminados pela subtração das médias de cada indivíduo $\bar{y}_i = \bar{x}'_i \beta_1 + \bar{\varepsilon}_i$ no modelo correspondente, resultando o modelo *within*, ou modelo de diferenças de média:

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)'\beta_1 + (\varepsilon_{it} - \bar{\varepsilon}_i) \qquad (2)$$

em que $\bar{x}_i = T_i^{-1} \sum_{t=1}^{T_i} x_{it}$ e o estimador *within* é o estimador OLS deste modelo. De acordo com Cameron & Trivedi (2009), pelo fato de β_{0i} ter sido eliminado, o estimador

OLS oferece estimativas consistentes de β_1, mesmo se β_{0i} for correlacionado com x_{it}, como é o caso do modelo de efeitos fixos.

O estimador *between* utiliza somente a variação entre indivíduos (*cross-sections*) e é o estimador OLS de uma regressão de $\bar{y}_i$ em função de $\bar{x}_i$, apresentada a seguir (equação (3)). Por levar em consideração apenas as variações *cross-section* nos dados, o coeficiente de qualquer regressor que seja invariante entre indivíduos não pode ser identificado.

$$\bar{y}_i = \beta_0 + \bar{x}'_i \beta_1 + (\beta_{0i} - \beta_0 + \bar{\varepsilon}_i) \qquad (3)$$

A consistência deste estimador requer que o termo do erro $(\beta_{0i} - \beta_0 + \bar{\varepsilon}_i)$ não seja correlacionado com x_{it}, o que ocorre quando β_{0i} for um efeito aleatório, mas não quando for um efeito fixo. Segundo Hsiao (2003), este estimador é raramente utilizado, pois os estimadores de efeitos aleatórios acabam sendo mais consistentes.

O estimador de efeitos aleatórios, por outro lado, é um estimador FGLS na equação (1). Assim, o modelo de efeitos aleatórios é o modelo de efeitos individuais:

$$y_{it} = x'_{it} \beta_1 + (\beta_{0i} + \varepsilon_{it}) \qquad (4)$$

com $\beta_{0i} \sim (\beta_0, \sigma^2_\alpha)$ e $\varepsilon_{it} \sim (0, \sigma^2_\varepsilon)$. Desta forma, o termo de erro $\mu_{it} = \beta_{0i} + \varepsilon_{it}$ é correlacionado ao longo do tempo t, para uma dada observação i, com correlação:

$$\mathrm{corr}(\mu_{it}, \mu_{is}) = \sigma^2_\alpha / (\sigma^2_\alpha + \sigma^2_\varepsilon), \text{ para todo } s \neq t \qquad (5)$$

O estimador de efeitos aleatórios é o estimador FGLS de β_1 da equação (4), dadas as correlações dos erros da expressão (5).

Segundo Cameron & Trivedi (2009), em modelos com erros heterocedásticos e autocorrelacionados, o estimador GLS (*general least squares*) pode ser calculado como um estimador OLS em um modelo que tenha erros não-correlacionados homocedásticos obtido de (4) por uma transformação linear apropriada. No caso do modelo de efeitos aleatórios da equação (4), tal modelo transformado é dado pela equação (6).

$$(y_{it} - \theta_i \bar{y}_i) = (1 - \theta_i)\beta_0 + (x_{it} - \theta_i \bar{x}_i)'\beta_1 + \{(1 - \theta_i)\beta_{0i} + (\varepsilon_{it} - \theta_i \bar{\varepsilon}_i)\} \qquad (6)$$

Um estimador FGLS (*feasible* GLS) é obtido substituindo-se θ_i, que é dado por uma estimativa consistente indicada por:

$$\theta_i = 1 - \sqrt{\sigma_\varepsilon^2 / (T_i \sigma_\alpha^2 + \sigma_\varepsilon^2)} \qquad (7)$$

O estimador de efeitos aleatórios será consistente e completamente eficiente se o modelo de efeitos aleatórios for apropriado, porém será inconsistente se o modelo de efeitos fixos for apropriado, uma vez que a correlação entre x_{it} e β_{0i} resulta numa correlação entre os regressores e o termo do erro na expressão (6). Da mesma forma, ainda segundo Cameron & Trivedi (2009), se não ocorrerem efeitos fixos, então o estimador de efeitos aleatórios será consistente, mas ineficiente e, portanto, uma estimação com erros-padrão robustos 'clusterizados' deveria ser obtida.

A expressão da estimativa por mínimos quadrados generalizados factíveis de um coeficiente de regressão do modelo (1), supondo efeitos aleatórios, torna-se igual à da estimativa do mesmo coeficiente em um modelo de efeitos fixos (estimação *within*) se $\hat{\theta}_i = 1$.

Para dados em painel longo, ou seja, com muitos períodos para um número relativamente menor de indivíduos, como é o caso deste trabalho (T = 262 meses e N = 60 países), os efeitos individuais β_{0i} (neste artigo, os efeitos de país) podem ser incorporados em x_{it} como variáveis *dummy* para cada período, de acordo com o seguinte modelo:

$$y_{it} = \beta_{0i} + \gamma_t + x'_{it}\beta_1 + \varepsilon_{it} \qquad (8)$$

de modo a haver muitos efeitos de tempo γ_t (no presente artigo, efeitos mensais). Um modelo *pooled*, para T > N, em que os regressores x_{it} contemplam o intercepto, o efeito temporal e possivelmente um vetor de variáveis de indivíduo, pode ser escrito como:

$$y_{it} = x'_{it}\beta_1 + \mu_{it} \qquad (9)$$

Como T é relativamente maior do que N, passa a ser necessária a especificação de um modelo que considere a existência de correlação serial do erro. Desta forma, para dados em painel longo, os modelos *pooled* com métodos de estimação OLS (POLS) e FGLS passam a ser mais adequados, já que permitem a utilização de um modelo AR(1) para μ_{it} ao longo do tempo que μ_{it} seja heterocedástico. Assim:

$$\mu_{it} = \rho_i \mu_{i,t-1} + \varepsilon_{it} \qquad (10)$$

em que os termos ε_{it} são não-correlacionados no tempo, porém com correlação entre indivíduos diferente de zero (corr(ε_{it}, ε_{is}) = σ_{ts}).

Alternativamente à inclusão de um vetor de variáveis *dummy* para cada período, estima-se, finalmente, um modelo de efeitos individuais com termos de erro AR(1), que representa um modelo melhor do que aquele que considera os termos de erro i.i.d. Assim:

$$y_{it} = \beta_{0i} + x'_{it}\beta_1 + \mu_{it} \qquad (11)$$

Logo, segundo Cameron & Trivedi (2009), este modelo potencialmente gerará estimativas dos parâmetros mais eficientes. Neste caso, dada a estimativa de $\hat{\rho}$ por meio da equação (9), primeiramente elimina-se o efeito do erro AR(1) e, na sequência, elimina-se o efeito individual por meio da aplicação da diferença de médias. Assim, a modelagem pode considerar β_{0i} como sendo um efeito fixo ou um efeito aleatório.

Após a apresentação dos modelos de dados em painel, explicita-se que este trabalho aplicará seis diferentes tipos de modelagens, a fim de propiciar um melhor entendimento dos diversos tipos de estimadores, bem como apresentar um modelo adequado para a explicação do comportamento dos índices de ações de países emergentes em uma perspectiva longitudinal, em função das características de cada um dos respectivos mercados acionários. O quadro 1 apresenta uma síntese dos modelos a serem estimados.

Quadro 1: Modelos de Dados em Painel a Serem Estimados

Modelo	Descrição
Efeitos Fixos	$y_{it} = \beta_{0i} + x'_{it}\beta_1 + \varepsilon_{it}$ Os parâmetros β_{0i} podem ser correlacionados com os regressores x_{it}, o que permite uma forma limitada de endogeneidade. Assume-se que x_{it} não é correlacionado com o erro idiossincrático ε_{it}.
Efeitos Aleatórios	$y_{it} = x'_{it}\beta_1 + (\beta_{0i} + \varepsilon_{it})$ Os parâmetros β_{0i} e os termos de erro idiossincrático ε_{it} são independentes e identicamente distribuídos. O estimador de efeitos aleatórios é o FGLS de β_1, dado que $\mathrm{corr}(\mu_{it}, \mu_{is}) = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$.
Efeitos Fixos com Termos de Erro AR(1)	$y_{it} = \beta_{0i} + x'_{it}\beta_1 + \mu_{it}$ Com $\mu_{it} = \rho_i\mu_{i,t-1} + \varepsilon_{it}$. Considera-se β_{0i} como sendo um efeito fixo.
Efeitos Aleatórios com Termos de Erro AR(1)	$y_{it} = \beta_{0i} + x'_{it}\beta_1 + \mu_{it}$ Com $\mu_{it} = \rho_i\mu_{i,t-1} + \varepsilon_{it}$. Considera-se β_{0i} como sendo um efeito aleatório.
Pooled com Método de Estimação OLS e com Termos de Erro AR(1)	$y_{it} = \beta_{0i} + \gamma_t + x'_{it}\beta_1 + \varepsilon_{it}$ Com $\mu_{it} = \rho_i\mu_{i,t-1} + \varepsilon_{it}$, em que os ε_{it} são serialmente não-correlacionados, mas com correlação entre países igual a corr(ε_{it}, ε_{is}) = $\sigma_{ts} \neq 0$.
Com Método de Estimação FGLS e com Termos de Erro AR(1)	$y_{it} = \beta_{0i} + \gamma_t + x'_{it}\beta_1 + \varepsilon_{it}$ Similar ao modelo *pooled* com método de estimação OLS, mas com estimador FGLS.

4. Modelos hierárquicos de dois níveis com medidas repetidas

Segundo Gelman (2006), os modelos hierárquicos são uma generalização dos métodos de regressão e, portanto, podem ser utilizados para uma variedade de propósitos, incluindo predição, redução dos dados e inferência causal a partir de experimentos e estudos observacionais. Contribuições significativas sobre o assunto são encontradas em Hofmann (1997), Kreft & de Leeuw (1998), Snijders & Bosker (1999),

Raudenbush & Bryk (2002), Hox (2002), Goldstein (2003), de Leeuw & Meijer (2007), Heck & Thomas (2008) e Rabe-Hesketh & Skrondal (2008).

Em comparação com os modelos clássicos de regressão ou com os modelos de análise de covariância, os modelos multinível apresentam a vantagem de levar em consideração a análise de dados hierarquicamente estruturados. Estes modelos propõem uma estrutura de análise dentro da qual podem ser reconhecidos os distintos níveis em que se articulam os dados, estando cada sub-nível representado pelo seu próprio modelo (Draper, 1995). Cada um dos sub-modelos, de acordo com Soto & Morera (2005) e Fávero et al. (2009), expressa a relação entre as variáveis dentro de um determinado nível e especifica como as variáveis deste nível influenciam as relações que se estabelecem em outros níveis.

Este trabalho pretende estudar também o efeito dos países sobre os índices de ações sob uma perspectiva temporal, ou seja, por meio da utilização de modelos hierárquicos com medidas repetidas. Portanto, o nível 1 não levará em consideração aspectos relacionados a firmas ou ações, como exposto até o presente momento, mas considerará os efeitos temporais para a evolução dos índices. Portanto, o nível 1 é conhecido, nestes casos, por medida repetida e, após a introdução deste tipo de modelagem, faz-se necessária a apresentação de um modelo mais adequado que leva em consideração os aspectos temporais no nível 1.

Segundo Hofmann, Jacobs & Baratta (1993) e Short, Ketchen Jr., Bennett & du Toit (2006), os modelos hierárquicos de dois níveis com medidas repetidas são capazes de identificar padrões de mudança sistemática de forma individual da variável de desempenho entre as observações em estudo, além de oferecerem vantagens adicionais por permitirem aos pesquisadores a modelagem com a inclusão de variáveis preditoras específicas no segundo nível de análise, oferecendo respostas de quão exatas são as influências dos países ao longo do tempo. Assim, este método de análise considera uma trajetória de desempenho única para cada país, levando em consideração a presença de efeitos aleatórios entre as observações para a explicação da evolução dos respectivos índices de ações.

Os modelos de 2 níveis consistem de 2 sub-modelos, em que há $t = 1, \ldots, T_i$ meses no nível 1, os quais são aninhados em cada $i = 1, \ldots, n$ países. Assim, tem-se, para o nível 1, que:

$$y_{ti} = \beta_{0i} + \beta_{1i}.\text{mês}_{ti} + r_{ti} \qquad (11)$$

Em que:

- mês: variável explicativa do nível 1 (medida repetida) correspondente a cada mês da análise (262 meses, de 1986 a 2007);

- $t = 1,2, \ldots, T_i$ (meses), e $i = 1,2, \ldots, n$ (países);

- β_{0i}: valor esperado (média) da variável de desempenho (índice de ações) do país i no mês 1;

- β_{1i} é a taxa de crescimento da variável de desempenho do país i; e

– σ^2 é a variância de r_i (variância de um determinado país ao longo do tempo). Assume-se que o termo aleatório $r_{ti} \sim N(0, \sigma^2)$.

Cada coeficiente do nível 1 torna-se uma variável dependente no modelo do nível 2. Assim, este pode ser escrito como:

$$\beta_{pi} = \gamma_{p0} + \sum_{q=1}^{Q_p} \gamma_{pq}.x_{qi} + \mu_{pi} \qquad (12)$$

Em que:

- γ_{pq} ($q = 0, 1, \ldots, Q_p$) são os coeficientes do nível 2;

- x_{qi} é o vetor de regressores do nível 2; e

- μ_{pi} é o efeito aleatório do nível 2. Assume-se que, para cada unidade i, o vetor $(\mu_{0i}, \mu_{1i}, \ldots, \mu_{pi})'$ apresenta uma distribuição normal multivariada, em que cada elemento μ_{pi} tenha média zero e variância $Var(\mu_{pi}) = \tau_{pp}$.

Em função da natureza não balanceada dos dados na maioria dos modelos hierárquicos, ou seja, em função do fato de n variar entre as unidades, os métodos tradicionais de estimação dos componentes de variância-covariância falham nas estimações. Assim, são utilizadas técnicas computacionais interativas, a fim de que seja possível a estimação de σ^2 e τ por máxima verossimilhança.

De maneira geral, os modelos de dois níveis são estimados por máxima verossimilhança no conceito restrito (REML, ou *restricted maximum likelihood*). Com esta abordagem, os componentes de variância-covariância são estimados por meio de máxima verossimilhança, ponderando todos os possíveis valores de efeitos fixos (Lee & Nelder, 2001), que são estimados via métodos dos mínimos quadrados generalizados. Já pela estimação por máxima verossimilhança no conceito integral (ML, ou *full maximum likelihood*), os parâmetros de variância-covariância e os coeficientes de efeitos fixos do nível 2 são estimados pela maximização de sua probabilidade conjunta.

Inferências estatísticas sobre os coeficientes fixos do nível 2, γ's, baseiam-se na premissa de que os efeitos aleatórios em cada nível são normalmente distribuídos. Com a utilização de grandes amostras, é possível elaborar inferências estatísticas sobre γ's que não são baseadas nesta premissa por meio da aplicação do método de estimação generalizada de equações (GEE, ou *generalized estimating equations*). A comparação dos resultados provenientes do GEE com aqueles resultantes do HLM (*hierarchical linear modeling*) oferece uma possibilidade de verificar se as inferências oriundas do HLM sobre os γ's são sensíveis à violação desta premissa. O modelo mais simples GEE assume que y_{ti} para o caso i é independente de $y_{ti'}$ para outro caso i' e que y_{ti} e $y_{ti'}$ apresentam variância constante. Sob estas simples premissas, a estimação dos coeficientes γ's, pelo método dos mínimos quadrados ordinários, seria justificada. Por outro lado, se estas premissas não foram corretas e a adoção do método OLS for feita, as estimações dos erros-padrão seriam viesadas e inconsistentes, conforme apontam Raudenbush & Bryk (2002), Goldstein (2003) e Brown & Uyar (2004).

A seguir, no quadro 2, são apresentados os testes de hipótese para os efeitos fixos e para os componentes de variância dos modelos hierárquicos de dois níveis que serão elaborados no presente trabalho.

Quadro 2: Testes de Hipótese para os Efeitos Fixos e para os Componentes de Variância em Modelos Hierárquicos de Dois Níveis

Hipótese	Teste Estatístico
Efeitos Fixos Parâmetros: H_0: γ's = 0 H_1: γ's $\neq$ 0	Teste t
Efeitos Aleatórios (Componentes de Variância-Covariância) Parâmetros: H_0: τ's = 0 H_1: τ's > 0	Teste Qui-quadrado (χ^2)

Modelagens similares foram extensivamente utilizadas na literatura para se investigar os efeitos dos países sobre algum comportamento específico, merecendo destaque os trabalhos de Brouthers (1998), Christmann, Day & Yip (1999), Hawawini, Subramanian & Verdin (2004), Makino, Isobe & Chan (2004) e Makino, Beamish & Zhao (2004).

Esta linha de pesquisa remonta aos trabalhos de Schmalensee (1985) e Rumelt (1991), sendo seguidos por uma série de outros estudos que analisaram os efeitos da firma, do setor ou do tempo sobre variáveis de desempenho, como os de Bergh (1993), Bergh (1995), Roquebert, Phillips & Westfall (1996), Deadrick, Bennett & Russell (1997), McGahan & Porter (1997), Brush & Bromiley (1997), Brush, Bromiley & Hendrickx (1999), Bowman & Helfat (2001), Adner & Helfat (2003), Hough (2006), Misangyi, Lepine, Algina & Goeddeke (2006), Short, Ketchen Jr., Bennett & du Toit (2006) e Short, Ketchen Jr., Palmer & Hult (2007).

De forma complementar aos estudos destacados, a literatura também apresenta pesquisas que analisaram a influência dos efeitos institucionais e dos países por meio da utilização de variáveis contábeis e financeiras que possam influenciar o valor de mercado ou provocar reações nos preços das ações das firmas. Neste caso, merecem destaque os trabalhos de Ball, Kothari e Robin (2000), Bushman e Piotroski (2006), Grambovas, Giner e Christodoulou (2006) e Ball, Robin e Sadka (2008).

5. Amostragem, definição das variáveis, hipóteses e método

A utilização de modelos de dados em painel e hierárquicos na avaliação do mercado financeiro tem sido cada vez mais frequente e representa novos desafios quando da formulação de problemas que têm como objetivo a mensuração de desempenho. A aplicação destes modelos permite estimar a relação do desempenho com diversas variáveis ambientais simultaneamente e, por conseguinte, acredita-se que a utilização deste tipo de modelagem torna possível o desenvolvimento de novas pesquisas científicas que têm por objetivo a determinação das influências de aspectos temporais, mercadológicas e até mesmo regionais sobre os índices de ações. A seguir, apresentam-se a amostragem, a definição das variáveis, as hipóteses a serem testadas no estudo e o método propriamente dito.

5.1. Amostragem e definição das variáveis

A evolução dos índices de ações de cada país emergente, bem como as características de cada mercado acionário e as variáveis econômicas dos países foram coletadas por meio da base da Compustat Global. Inicialmente, os dados foram extraídos de todos os conjuntos de países disponíveis na base no período compreendido entre 1986 e 2007. A base original contava com 60 países empresas em um período de 262 meses, totalizando 9.979 observações.

A base de dados oferece, pela própria característica de cada mercado acionário, um painel desbalanceado. Além disso, a transformação da variável dependente para o logaritmo natural fez com que algumas observações se perdessem quando da aplicação dos modelos de dados em painel e hierárquicos. Ademais, seguindo a mesma lógica proposta por McGahan & Porter (1997) e Goldszmidt, Brito & Vasconcelos (2007) em relação aos critérios de exclusão de observações, foram eliminados, quando da modelagem hierárquica, países que apresentaram algum dado faltante em relação aos regressores econômicos ou à variável de desempenho (logaritmo natural do índice de ações). Além disto, os critérios buscaram evitar a indeterminação na alocação dos

diferentes efeitos pela presença de menos de dois casos em cada fator. Mantiveram-se, portanto, apenas países que apresentaram dois ou mais períodos da variável dependente relatados na base.

Finalmente, 5.491 observações compuseram a amostra final para a modelagem de dados em painel e 2.968 observações para a modelagem hierárquica, com apenas 23 países neste último caso. Por meio deste tratamento, adquire-se vantagem na utilização de modelos multinível, uma vez que estes permitem verificar quais efeitos de país mais bem explicam a variância no índice de ações ao longo do tempo.

A variável de desempenho adotada para análise, em todos os modelos considerados neste artigo, é o logaritmo do principal índice de ações de um país (lnindex). As variáveis referentes aos mercados de capitais, ou seja, aquelas correspondentes aos regressores dos modelos de dados em painel, e as variáveis econômicas dos países, a serem incluídas no nível 2 da modelagem multinível, também foram obtidas por meio deste mesmo banco de dados e encontram-se no quadro 3 a seguir:

Quadro 3: Definição das Variáveis para os Modelos de Dados em Painel e Hierárquicos

Variáveis de Mercado Acionário (Regressores nos Modelos de Dados em Painel)	
lnyield	Logaritmo natural dos dividendos no mercado
lnp_bv	Logaritmo natural da relação preço/*book value* do mercado
lnvtrdus	Logaritmo natural do valor transacionado no mercado, em milhões de dólares
lntover	Logaritmo natural da taxa de *turnover* do mercado
lnnush	Logaritmo natural do número de ações transacionadas no mercado, em milhões
days	Número de dias com operações no mercado
Variáveis Econômicas dos Países (Nível 2 dos Modelos Hierárquicos)	
pibcap	Produto interno bruto *per capita*
pea	População economicamente ativa, em percentual da população residente
educpib	Investimento em educação, em percentual do PIB
pesqpib	Investimento em pesquisa, em percentual do PIB
import	Importações, em milhões de dólares
export	Exportações, em milhões de dólares

A consideração inicial das variáveis referentes aos indicadores do mercado acionário de cada país para a explicação da evolução dos índices de ações é baseada nos trabalhos de Brouthers (1998) e de Andrezo & Lima (2007). Já a consideração de

variáveis econômicas no nível 2 da modelagem hierárquica baseia-se nos estudos de Collins (1990), de Hawawini, Subramanian & Verdin (2004) e de Goldszmidt, Brito & Vasconcelos (2007), em que foram testadas e verificadas as significâncias estatísticas de variáveis como PIB e competitividade de países.

As variáveis apresentadas serão utilizadas para sustentar, ou não, as hipóteses de pesquisa, formuladas a seguir.

5.2. Hipóteses a serem testadas

Este trabalho procura investigar quais as principais variáveis do mercado acionário que influenciam os índices de ações em uma perspectiva longitudinal. Além disso, procura verificar se existem diferenças, ao longo do tempo, nos índices de ações de países emergentes e se estas diferenças ocorrem entre países por conta das variações existentes nas variáveis econômicas.

Assim sendo, as hipóteses a serem testadas podem ser descritas da seguinte maneira:

Hipótese 1: Há características dos mercados acionários dos países emergentes que explicam a variação no índice de ações ao longo do tempo.

Hipótese 2: Há variabilidade significativa nos índices de ações de países emergentes ao longo do tempo (período de 1986 a 2007, 262 meses).

Hipótese 3: Há variabilidade significativa nos índices de ações, ao longo do tempo, entre países.

Hipótese 4: Os índices mensais de ações de países emergentes seguem uma tendência linear ao longo do período compreendido entre os anos de 1986 e 2007, e há diferenças nesta tendência entre países.

Hipótese 5: Há características econômicas dos países que explicam a variação no índice de ações ao longo do tempo.

Para a verificação de cada um das hipóteses apresentadas, necessita-se apresentar o método a ser utilizado, bem como os modelos propostos.

5.3. Método

Elaboradas as considerações sobre as variáveis de cada modelo, parte-se para a apresentação dos modelos propriamente ditos.

A modelagem de dados em painel é baseada pela existência de variâncias geral, within e between em cada regressor, conforme discutido anteriormente. A tabela 1 apresenta esta decomposição de variância para cada um dos regressores apresentados no quadro 3.

Tabela 1: Estatísticas do Painel e Decomposição da Soma de Quadrados Within e Between

Variável	Decomposição	Média	Desvio-Padrão	Mínimo	Máximo	Observações
país	geral					N.T = 9.979
	between					N = 60
	within		0,000			
mês	geral		69,073	1,000	262,00	N.T = 9.979
	between		38,807	79,006	256,91	N = 60
	within		60,389	18,645	290,61	
lnyield	geral		1,379	-4,605	7,19	N.T = 6.544
	between	0,726	1,141	-3,212	2,06	N = 46
	within		0,953	-4,786	5,85	
lnp_bv	geral		0,765	-4,605	6,51	N.T = 6.942
	between	0,622	0,518	-0,706	1,84	N = 52
	within		0,618	-4,263	6,11	
lnvtrdus	geral		3,022	-4,605	13,71	N.T = 9.938
	between	5,273	2,735	-0,952	10,52	N = 60
	within		1,437	-2,729	10,96	
lntover	geral		1,479	-4,605	4,36	N.T = 9.785
	between	0,426	1,276	-2,552	2,95	N = 60
	within		0,829	-4,612	3,67	
lnnush	geral		3,770	-7,684	16,38	N.T = 9.889
	between	5,074	3,265	-2,626	12,91	N = 60
	within		1,919	-6,588	14,12	
days	geral		10,127	0,000	986,00	N.T = 9.933
	between	20,411	1,511	13,993	24,19	N = 60
	within		10,040	-0,529	987,02	

De acordo com a tabela 1, nota-se que o país é obviamente invariante no tempo e, portanto, apresenta variação *within* igual a zero. Por outro lado, a variável referente ao tempo (mês) não é invariante entre indivíduos, já que se trata de um painel desbalanceado e, portanto, a sua variação *between*, mesmo sendo relativamente menor do que a *within*, não é igual a zero. Das demais variáveis, a maioria apresenta maior variação entre indivíduos (*between*) do que ao longo do tempo (*within*), porém ainda não é possível afirmar que a estimação *within* resultará numa perda de eficiência, já que a proporção entre as variâncias *within* e *between* de cada variável é diferente e ainda não se conhecem as significâncias estatísticas de cada um delas nos modelos. A tabela 1, todavia, oferece um maior embasamento para a adoção dos modelos de dados em painel e a aplicação de diversos estimadores. As colunas *"Mínimo" e "Máximo"* apresentam, respectivamente, os valores mínimos e máximos de x_{it} para a linha "geral", $\overline{x}_i$ para a linha *"between"* e $\left(x_{it} - \overline{x}_i + \overline{x}\right)$ para a linha *"within".*

Conforme discutido na seção 2 e resumido no quadro 1, serão elaborados 6 modelos de dados em painel, com diferentes considerações sobre os estimadores e os termos de erro. O modelo geral a ser adotado, derivado da equação (1), obedece ao que segue:

$$\begin{aligned} \text{lnindex}_{it} &= \beta_{0i} + \beta_1.(\text{lnyield}) + \beta_2.(\text{lnp_bv}) + \beta_3.(\text{lnvtrdus}) + \\ &+ \beta_4.(\text{lntover}) + \beta_5.(\text{lnnush}) + \beta_6.(\text{days}) + \varepsilon_{it} \end{aligned} \qquad (13)$$

Com as devidas considerações para cada um dos modelos específicos, esta etapa auxilia na verificação da hipótese 1 anteriormente proposta.

Para a verificação das hipóteses 2 e 3 apresentadas, propõe-se um modelo com ausência de variáveis preditoras (modelo nulo), que oferece estimações dos componentes de variância ao longo do tempo e entre países, já que a modelagem proporciona um teste χ^2 para o componente de variância entre países (Raudenbush, Bryk, Cheong, Congdon & du Toit, 2004). Assim, o modelo nulo pode ser escrito como:

Modelo Nulo

Nível 1 (Medida Repetida):

$\text{lnindex}_{ti} = \beta_{0i} + r_{ti}$, $\qquad r_{ti} \sim \text{NID}(0, \sigma^2)$ $\qquad$ (14)

lnindex: variável de desempenho representada pelo logaritmo do índice de ações;

$t=1,2, \ldots, T_i$ (meses) e $i=1,2, \ldots, n$ (países);

β_{0i}: lnindex esperado (médio) do índice de ações do país i no mês 1; e

σ^2: variância "dentro" do país.

Nível 2 (País):

$\beta_{0i} = \gamma_{00} + \mu_{0i}, \qquad \mu_{0i} \sim NID(0, \tau_{\beta_0})$ (15)

γ_{00}: média geral dos lnindex's; e

τ_{β_0}: variância entre os lnindex's esperados.

Para a verificação da quarta hipótese, são propostos dois modelos que incluem um componente de tendência (variação ao longo do tempo) no nível 1. O primeiro modelo não inclui efeitos aleatórios e testa apenas se os índices de ações dos países seguem uma tendência temporal linear. Já o modelo seguinte apresenta a inclusão de efeitos aleatórios e testa se há variância significativa da tendência dos índices entre países ao longo do tempo.

Modelo de Tendência **Linear sem Efeitos Aleatórios**

Nível 1 (Medida Repetida):

$lnindex_{ti} = \beta_{0i} + \beta_{1i}.mês_{ti} + r_{ti}, \qquad r_{ti} \sim NID(0, \sigma^2)$ (16)

β_{1i}: taxa de crescimento de lnindex do país i.

Nível 2 (País):

$\beta_{0i} = \gamma_{00} + \mu_{0i}, \qquad \mu_{0i} \sim NID(0, \tau_{\beta_0})$ (17)

$\beta_{1i} = \gamma_{10}$

γ_{10}: média geral das taxas de crescimento dos lnindex's esperados.

Modelo de Tendência Linear com Efeitos Aleatórios

Nível 1 (Medida Repetida):

$lnindex_{ti} = \beta_{0i} + \beta_{1i}.mês_{ti} + r_{ti}, \qquad r_{ti} \sim NID(0, \sigma^2)$ (18)

Nível 2 (País):

$\beta_{0i} = \gamma_{00} + \mu_{0i}, \qquad \mu_{0i} \sim NID(0, \tau_{\beta_0})$ (19)

$\beta_{1i} = \gamma_{10} + \mu_{1i}, \qquad \mu_{1i} \sim NID(0, \tau_{\beta_1})$

τ_{β_1}: variância entre as taxas de crescimento esperadas entre países.

A significância das mudanças individuais no índice de ações é testada de duas formas. A primeira consiste num teste χ^2 que compara as estatísticas de desvios entre o

modelo com o efeito do mês (modelo de tendência linear) e o modelo nulo. A segunda oferece um teste t para os efeitos fixos e χ^2 para os componentes de variância. A significância do efeito fixo para os períodos sugere que o efeito do tempo é constante para todos os países. Porém a inclusão de efeitos aleatórios auxilia na verificação da existência de variabilidade significativa nos índices de ações, ao longo do tempo, entre os países emergentes.

Caso as hipóteses anteriores sejam verificadas, parte-se para a inclusão das variáveis econômicas no nível 2, com o intuito de se verificar a hipótese 5. O modelo passa a ser:

Modelo Completo

Nível 1:

$$\ln index_{ti} = \beta_{0i} + \beta_{1i}.mês_{ti} + r_{ti} \qquad (20)$$

Nível 2:

$$\beta_{ki} = \gamma_{k0} + \gamma_{k1}.(pibcap) + \gamma_{k2}.(pea) + \gamma_{k3}.(educpib) + \gamma_{k4}.(pesqpib) + \gamma_{k5}.(import) + \gamma_{k6}.(export) + \mu_{ki} \qquad (21)$$

Em que k = 0 quando o parâmetro se referir ao intercepto e k = 1 quando se referir à inclinação.

Após a apresentação de cada um dos modelos, o quadro 4 oferece a relação entre cada hipótese a ser testada e o respectivo modelo.

Quadro 4: Hipóteses a Serem Testadas e Respectivos Modelos

Hipótese	Modelo
Hipótese 1: Há características dos mercados acionários dos países emergentes que explicam a variação no índice de ações ao longo do tempo.	Modelos de Dados em Painel com Diversos Estimadores
Hipótese 2: Há variabilidade significativa nos índices de ações de países emergentes ao longo do tempo (período de 1986 a 2007, 262 meses).	Modelagem Hierárquica (Modelo Nulo)
Hipótese 3: Há variabilidade significativa nos índices de ações, ao longo do tempo, entre países.	
Hipótese 4: Os índices mensais de ações de países emergentes seguem uma tendência linear ao longo do período compreendido entre os anos de 1986 e 2007, e há diferenças nesta tendência entre países.	Modelagem Hierárquica (Modelo de Tendência Linear sem e com Efeitos Aleatórios)
Hipótese 5: Há características econômicas dos países que explicam a variação no índice de ações ao longo do tempo.	Modelagem Hierárquica (Modelo Completo)

6. Resultados

Inicialmente são apresentados, por meio da tabela 2, os resultados de cada um dos 6 modelos propostos de dados em painel.

Tabela 2: Modelos de Dados em Painel

Variável	Efeitos Fixos	Efeitos Aleatórios	Efeitos Fixos com Erros AR(1)	Efeitos Aleatórios com Erros AR(1)	*Pooled* com OLS e Erros AR(1)	Estimador FGLS com Erros AR(1)
lnyield	-0,05667**	-0,05641**	-0,01489**	-0,0153**	-0,01405**	-0,07781**
	(0,01168)	(0,01169)	(0,00343)	(0,00345)	(0,00477)	(0,01830)
lnp_bv	0,69881**	0,70081**	0,07049**	0,07226**	0,26191**	1,04583**
	(0,02169)	(0,02170)	(0,00705)	(0,00709)	(0,01178)	(0,03530)
lnvtrdus	0,27373**	0,27347**	-0,00039	0,00231	0,1885**	0,07608**
	(0,01812)	(0,01811)	(0,00977)	(0,00979)	(0,01193)	(0,02093)
lntover	-0,41884**	-0,41901**	0,05051**	0,04731**	-0,24614**	-0,34874**
	(0,02162)	(0,02164)	(0,00881)	(0,00883)	(0,00990)	(0,03414)
lnnush	0,28239**	0,28134**	-0,00339	-0,00197	0,08216**	0,13542**
	(0,01124)	(0,01123)	(0,00471)	(0,00472)	(0,00574)	(0,01081)
days	-0,00229	-0,00225	-0,00078	-0,00094	-0,00011	0,01079
	(0,00483)	(0,00483)	(0,00088)	(0,00088)	(0,00124)	(0,00932)
constante	4,36695**	4,17368**	9,15036**	6,95966**	5,01411**	5,62987**
	(0,11274)	(0,45059)	(0,00229)	(0,40521)	(0,04832)	(0,20229)
N	5491	5491	5451	5491	5491	5491
R^2	0,544				0,939	
R^2 (geral)	0,158	0,158	0,071	0,079		
R^2 *(between)*	0,029	0,029	0,012	0,054		
R^2 *(within)*	0,544	0,544	0,052	0,293		
F	1083,51		49,37			
P. F	0,000		0,000			
Wald χ^2		6483,22		300,78	1770,85	1642,98
P. χ^2		0,000		0,000	0,000	0,000

Obs.: Erros-padrão entre parênteses.

** $P < 0,01$.

De acordo com a tabela 2, é possível verificar que os coeficientes estimados variam entre os modelos. Primeiramente, nota-se a existência de erros-padrão bem superiores nos modelos de efeitos fixos e aleatórios (mais de 100%) em comparação com aqueles reportados pelos respectivos modelos com efeitos AR(1) nos termos de erro, mesmo que os primeiros tenham sido elaborados com erros-padrão robustos 'clusterizados'. Este fato talvez tenha ocorrido pela natureza do painel em análise, ou seja, pelo fato de ser longo, já que a obtenção de erros-padrão robustos 'clusterizados' é feita sob as hipóteses de que os erros são independentes entre indivíduos e que $N \to \infty$, que ocorre para dados em painel curto.

Desta forma, para dados em painel longo, os modelos *pooled* com métodos de estimação OLS e FGLS passam a ser mais adequados. Permitindo que os termos de erro sejam correlacionados entre países, nota-se uma redução da ordem de 50% para os erros-padrão do modelo *pooled* com estimador OLS em comparação com aqueles obtidos anteriormente por meio dos modelos de efeitos fixos e aleatórios, porém as estimativas obtidas pelo modelo com estimador FGLS oferecem erros-padrão bem mais elevados.

Conforme já discutido, os modelos de efeitos fixos e aleatórios oferecem uma alternativa, para dados em painel longo, em que são considerados os efeitos individuais com termos de erro AR(1), e representam melhores modelos do que aqueles que consideram os termos de erro i.i.d., o que poderá gerar estimativas dos parâmetros mais eficientes. De fato os modelos de efeitos fixos e aleatórios com termos de erro AR(1) apresentam erros-padrão da ordem de 50% a 70% menores do que aqueles obtidos pelos respectivos modelos sem a consideração de termos de erro AR(1). Além disso, estes modelos apresentam apenas as variáveis referentes aos retornos dos dividendos, à relação preço/*book value* e à taxa de *turnover* do mercado como sendo significativas, com 99% de nível de confiança, para explicar o comportamento dos índices de ações ao longo do tempo.

Em relação à adequação dos modelos propriamente ditos, nota-se a significância estatística do conjunto de variáveis em todos os casos (P. F para os modelos de efeitos fixos e P. Wald χ^2 para os modelos de efeitos aleatórios). Embora haja relativa importância das estatísticas R^2 para efeitos de predição, seus valores não são consideravelmente elevados nos modelos em análise. Um resultado importante, porém,

relaciona-se com a existência de maiores valores para os R^2 *within* em todos os modelos em que esta estatística é calculada. Tendo por base as expressões de cada uma das estatísticas R^2

R^2 geral: $\rho^2\{(y_{it} - \overline{y}_i),(x'_{it}\hat{\beta} - \overline{x}'_i\hat{\beta})\}$

R^2 *between*: $\rho^2(\overline{y}_i, \overline{x}'_i\hat{\beta})$

R^2 *within*: $\rho^2(y_{it}, x'_{it}\hat{\beta})$

em que $\rho^2(x,y)$ representa correlação quadrática entre x e y, nota-se que os estimadores *within* explicam melhor a variação *within* em todos os modelos, mesmo aqueles com efeitos aleatórios (Wooldridge, 2002). O teste de Hausman aplicado aos modelos de efeitos fixos e aleatórios com termos de erro AR(1) auxilia na rejeição da hipótese nula de que o modelo com efeitos aleatórios oferece estimativas consistentes dos parâmetros, já que, neste caso, $\chi^2 = 597{,}93$ (P. $\chi^2 = 0.000$).

Enquanto o coeficiente da variável que se refere ao retorno dos dividendos apresenta sinal negativo, os sinais dos coeficientes das variáveis referentes à relação preço/*book value* e à taxa de *turnover* são positivos. Desta forma, a elasticidade estimada dos dividendos do mercado no índice de ações é de -0,01489, já que o modelo é logarítmico. Seguindo o mesmo raciocínio, as elasticidades da relação preço/*book value* e da taxa de *turnover* são, respectivamente, 0,07049 e 0,05051 sobre o índice de ações. Em outras palavras, a cada cem unidades de variação de cada um destes regressores, *ceteris paribus*, o índice de ações se altera em -1,5, +7,0 e +5.0 pontos, respectivamente. A hipótese 1 é, portanto, sustentada.

Segundo Islam (1995), a principal utilidade da modelagem de dados em painel é a sua habilidade em permitir que diferenças ocorram entre países, o que faz com que os resultados sejam significativamente diferentes daqueles obtidos por meio de regressões isoladas para cada país. Assim sendo, apresenta-se a seguir, na tabela 3, as estimativas dos coeficientes dos regressores obtidas por meio de regressões separadas estimadas por OLS e referentes aos retornos dos dividendos, à relação preço/*book value* e à taxa de *turnover* do mercado para cada um dos países emergentes considerados no estudo.

Tabela 3: Estimativas dos Coeficientes dos Regressores e da Constante para Cada País

país	lnyield	lnp_bv	lntover	cons.	país	lnyield	lnp_bv	lntover	cons.
Arábia Saudita	0,355	1,125	0,276	6,023	Jordânia	-0,646	1,701	0,564	5,188
Argentina	1,140	1,062	3,313	3,889	Letônia	-0,624	0,224	-0,173	5,831
Bangladesh	-0,113	0,811	-0,016	6,598	Malásia	-0,415	0,178	0,318	6,243
Botsuana	0,506	2,04	-0,023	3,171	Marrocos	0,038	0,325	0,337	11,186
Brasil	-1,088	0,055	0,352	-7,745	México	-0,015	2,464	-1,762	8,261
Bulgária	-12,989	-10,597	0,161	-9,848	Nigéria	0,276	0,528	0,846	8,300
Chile	-0,021	1,640	0,043	7,663	Peru	-0,046	0,000	0,185	7,107
Colômbia	-0,101	0,073	-0,171	6,212	Polônia	0,245	1,794	-0,883	9,780
Coréia do Sul	0,002	0,350	0,019	6,290	Portugal	0,333	0,488	0,073	7,157
Costa do Marfim	-0,997	0,945	0,090	6,445	Quênia	0,574	0,991	0,019	6,169
Croácia	-0,280	-1,341	0,158	8,287	Rep. Checa	0,063	1,064	0,389	5,394
Egito	-0,295	1,320	0,031	7,633	Rússia	-0,733	0,805	0,339	7,212
Equador	-0,144	-0,408	0,248	5,864	Sérvia	-0,072	0,010	-0,046	6,021
Eslováquia	0,046	0,105	-0,096	5,040	Sri Lanka	-1,369	-0,964	0,194	9,099
Filipinas	-0,141	0,133	0,190	7,258	Tailândia	-0,074	0,788	-0,139	6,076
Gana	-1,764	0,063	0,037	10,101	Taiwan	0,237	0,593	0,137	7,526
Grécia	-0,022	0,876	0,358	5,624	Tunísia	-0,613	0,114	0,093	7,758
Hungria	0,471	0,445	0,693	6,665	Turquia	-0,090	0,727	0,461	6,574
Índia	0,874	1,710	0,664	3,166	Venezuela	0,006	0,271	0,195	8,649
Jamaica	-0,432	0,575	-0,037	10,791	Zimbábue	0,297	2,030	0,775	7,420

Obs.: Variável dependente lnindex.

Alguns países não apresentaram coeficientes pelas razões mencionadas na seção 4. Nota-se também a existência de diferentes elasticidades de cada regressor em cada país. A existência de coeficientes diferentes para os regressores e para os valores da constante explicitam a importância da consideração de dados em painel e suscita a investigação das razões econômicas pelas quais cada país apresenta comportamentos diferenciados do índice de ações ao longo do tempo. Logo, parte-se para a modelagem hierárquica de dois níveis com medidas repetidas.

Conforme apresentado na seção 4, é primeiramente aplicado o modelo nulo, seguindo os procedimentos propostos por Snijders & Bosker (1999) e Raudenbush & Bryk (2002).

A tabela 4 apresenta a decomposição de variância entre os níveis. Assim, um percentual considerável (75,03%) da variabilidade dos índices de ações ocorre entre países (χ^2 = 6.790,536, P. = 0,000), o que corrobora o diagnóstico inicial e a perspectiva teórica elaborados na seção 1, e um percentual menos relevante de variância (24,97%) deve-se à evolução temporal em cada país.

Tabela 4: Decomposição de Variância - Modelo Nulo

Efeito Fixo	Coeficiente	Erro-padrão	t
Média geral de lnindex (γ_{00})	7,186**	0,367	19,569
Efeito Aleatório	**Componente de Variância**	**gl**	χ^2
Variação Temporal (r_{ti})	1,072		
Variação entre Países (μ_{0i})	3.221**	22	6.790,536
Decomposição da Variância	**% por Nível**		
Nível 1 (Tempo)	24,97		
Nível 2 (País)	75,03		

** **P < 0,01.**

O modelo nulo propicia a afirmação de que há variabilidade significativa nos índices de ações de países emergentes ao longo do tempo (período de 1986 a 2007, 262 meses), porém a variabilidade nos índices de ações é mais significativa entre países para o período em análise. Logo, as hipóteses 2 e 3 também são suportadas.

As tabelas 5 e 6 oferecem os resultados dos modelos com a inclusão da tendência no nível 1, sem e com os efeitos aleatórios, respectivamente. O modelo sem efeitos aleatórios mostra que a variável correspondente ao mês (tendência linear) com efeito fixo é significante (t = 3,669, P. = 0,000).

Tabela 5: Decomposição de Variância - Modelo de Tendência Linear sem Efeitos Aleatórios

Efeito Fixo	Coeficiente	Erro-padrão	t
Média Geral de lnindex (γ_{00})	5,667**	0,534	10,609
Média Geral das Taxas de Crescimento de lnindex (γ_{10})	0,009**	0,003	3,669
Efeito Aleatório	**Componente de Variância**	**gl**	χ^2
Nível 1			
Variação Temporal (r_{ti})	0,767		
Nível 2			
Média de lnindex dos Países (μ_{0i})	2,988**	22	8.662,823

** **P < 0,01.**

Por meio da análise da tabela 6, que apresenta os resultados do modelo de tendência linear com efeitos aleatórios, é possível verificar que o componente de variância para a tendência linear também é significante (χ^2 = 4.456,537, P. = 0,000), o que demonstra que os índices mensais de ações de países emergentes seguem uma tendência linear ao longo do período compreendido entre os anos de 1986 e 2007, porém há diferenças nesta tendência entre os países emergentes. Portanto, a hipótese 4 também pode ser suportada.

Tabela 6: Decomposição de Variância - Modelo de Tendência Linear com Efeitos Aleatórios

Efeito Fixo	Coeficiente	Erro-padrão	t
Média Geral de lnindex (γ_{00})	5,467**	0,517	10,565
Média Geral das Taxas de Crescimento de lnindex (γ_{10})	0,013**	0,004	3,210
Efeito Aleatório	**Componente de Variância**	**gl**	χ^2
Nível 1			
Variação Temporal (r_{ti})	0,314		
Nível 2			
Média de lnindex dos Países (μ_{0i})	6,055**	15	4.587,343
Taxa de Crescimento entre Países (μ_{1i})	$3,7 \times 10^{-4}$ **	15	4.456,537

** **P < 0,01.**

Com as hipóteses 1 a 4 sendo suportadas pelos modelos apresentados até o presente momento, parte-se para a elaboração do modelo completo com a inclusão das variáveis econômicas no nível 2. Das variáveis econômicas consideradas, apenas a variável referente ao percentual de investimento em pesquisa em relação ao PIB apresentou significância estatística para a diferenciação das taxas de crescimento dos índices de ações dos países. O modelo completo final passa a ser:

$$\begin{aligned} lnindex_{ti} &= \beta_{0i} + \beta_{1i}.mês_{ti} + r_{ti} \\ \beta_{0i} &= \gamma_{00} + \mu_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}.(pesqpib) + \mu_{1i} \end{aligned} \qquad (22)$$

A tabela 7 apresenta os resultados deste modelo.

Tabela 7: Decomposição de Variância - Modelo Completo

Efeito Fixo	Coeficiente	Erro-padrão	t
Média Geral de lnindex (γ_{00})	5,461**	0,518	10,536
Média Geral das Taxas de Crescimento de lnindex (γ_{10})	0,016*	0,006	2,757
pesqpib (γ_{11})	0,004*	0,002	2,246
Efeito Aleatório	**Componente de Variância**	**gl**	χ^2
Nível 1			
Variação Temporal (r_{ti})	0,314		
Nível 2			
Média de lnindex dos Países (μ_{0i})	6,070**	15	4.590,35
Taxa de Crescimento entre Países (μ_{1i})	$3,6 \times 10^{-4}$ **	14	4.143,47

** **P < 0,01.**
* **P < 0,05.**

Por meio da análise da tabela 7, nota-se que a variável referente ao percentual de investimento em pesquisa em relação ao PIB de um determinado país é significante (P. = 0,035 < 0,05) e com coeficiente com sinal positivo, para a diferenciação da taxa de crescimento do índice de ações entre os países emergentes. Em outras palavras, um aumento de um ponto percentual do investimento em pesquisa em relação ao PIB faz com que a taxa de crescimento do índice de ações de um determinado país suba, em média, 0,4% ao mês, *ceteris paribus*. Este coeficiente representa uma semi-elasticidade mensal, já que a variável dependente é dada pelo logaritmo natural do índice de ações.

Desta maneira, todas as cinco hipóteses apresentadas inicialmente foram suportadas pelos resultados dos modelos de dados em painel e hierárquicos. Os resultados obtidos com estes modelos longitudinais e de análise de componentes de variância não se restringem, como apresentado, apenas à especificação da composição da variância do índice de ações, mas propiciam a análise da influência de determinadas características dos próprios mercados acionários e da economia dos países sobre este desempenho ao longo dos períodos analisados.

7. Conclusões

Muitos trabalhos que adotaram a modelagem longitudinal e hierárquica buscam explicar as variabilidades significativas existentes em indicadores entre países e ao longo do tempo. No presente estudo, a representatividade de variáveis relativas ao mercado acionário e à economia de países em desenvolvimento suscita diversas questões discutidas por Andrezo & Lima (2007) sobre a relação entre o desenvolvimento econômico e do mercado de capitais dos países. O fato de o investimento em pesquisa ser significativo para o incremento de um índice de ações de um país pode ser explicado pela geração de novas tecnologias, pelo desenvolvimento de novos setores produtivos na economia e pela entrada de capital estrangeiro, que propiciam um incremento de renda, uma diversificação da carteira de investimentos e um crescimento da oferta pública de novas ações. E este fenômeno afetou heterogeneamente a evolução dos índices de ações dos países nas duas últimas décadas, com o desenvolvimento mais forte do mercado acionário de alguns países emergentes em relação a outros, possivelmente pela internacionalização dos mercados financeiros, de créditos disponíveis e maior sofisticação dos investidores principalmente institucionais.

A contribuição do trabalho relaciona-se com a possibilidade de avaliação de diversos estimadores quando da modelagem de dados em painel, com vistas à determinação do modelo mais adequado, em função da caracterização de um painel longo. A análise das elasticidades dos regressores correspondentes às características dos mercados acionários também suscita uma discussão sobre o comportamento de índices de ações em países com mercados acionários ainda não tão desenvolvidos.

Outra contribuição refere-se à estimação do impacto da variação temporal no índice de ações de países emergentes, bem como entre os próprios países. Se o efeito país for mais forte, como é o caso deste estudo, maior deve ser o foco dispensado aos impactos relativos às diferenças econômicas e entre os mercados acionários para o entendimento da evolução de índices, ou seja, prevalecem os fundamentos econômicos de cada país. Se, por outro lado, as diferenças ao longo do tempo explicassem a maior parte da variância, menor deveria ser a preocupação, por parte de um investidor

estrangeiro, na decisão de alocação de seus recursos em determinados países emergentes, em detrimento de outros.

A terceira contribuição do presente estudo é a tentativa de adição de variáveis preditoras no nível dos países. Muitos trabalhos estudaram a decomposição de variância, porém sem a avaliação dos impactos decorrentes da presença de determinadas variáveis, objetivando apenas a determinação de onde a maior parte da variabilidade ocorre (Rumelt, 1991). As aplicações com modelos hierárquicos lineares oferecem aos pesquisadores novas possibilidades de testar hipóteses mais complexas, sem que haja o risco de violação das premissas inerentes a outras técnicas como, por exemplo, a regressão pelo método dos mínimos quadrados ordinários.

De acordo com Barth, Beaver & Landsman (2001), as técnicas econométricas são utilizadas em muitos trabalhos relevantes que, se não as possuíssem, poderiam apresentar limitação da validade de suas inferências. Porém, é fundamental que a escolha apropriada de cada técnica estatística seja baseada em uma teoria subjacente e esteja de acordo com o que se deseja investigar. Segundo Ramos- Rodriguez & Ruiz-Navarro (2004), as origens do desempenho representam um tema central em pesquisas que buscam compreender os componentes de variância de comportamentos específicos ao longo do tempo e abordagens alternativas merecem ser estudadas em relação aos índices de ações. A inclusão de outros regressores nos modelos de dados em painel e no nível de país dos modelos hierárquicos pode ser elaborada para a determinação de novas estratégias e para a criação de diferentes modelos. A própria alteração do período de análise pode propiciar um entendimento mais profundo dos mecanismos que regem os fatores de desempenho dos mercados acionários de países emergentes.

Referências bibliográficas

Adjasi, C.K.D. & Biekpe, N. (2009). Do stock markets matter in investment growth in Africa? The Journal of Developing Areas, 43(1): 109-120.

Adner, R. & Helfat, C.E. (2003). Corporate effects and dynamic managerial capabilities. Strategic Management Journal, 24(10): 1011-1025.

Andrezo, A.F. & Lima, I.S. (2007). Mercado financeiro: aspectos históricos e conceituais. Atlas, São Paulo, 3. ed., 368 p.

Ball, R.; Kothari, S. & Robin, A. (2000). The effect of international institutional factors on properties of accounting earnings. Journal of Accounting and Economics, 29(1): 1-51.

Ball, R.; Robin, A. & Sadka, G. (2008). Is financial reporting shaped by equity markets or by debt markets? An international study of timeliness and conservatism. Review of Accounting Studies, 13(2-3): 168–205.

Barth, M.E.; Beaver, W.H. & Landsman, W. (2001). The relevance of the value-relevance literature for financial accounting standard setting: another view. Journal of Accounting and Economics, 31(1-3): 77-104.

Bekaert, G. & Harvey, C.R. (2002). Research in emerging markets finance: looking to the future. Emerging Markets Review, 3(4), 429-448.

Bekaert, G.; Harvey, C.R & Lundblad, C. (2001). Emerging equity markets and economic growth. Journal of Development Economics, 66, 465-504.

Bergh, D.D. (1993). Watch the time carefully: the use and misuse of time effects in management research. Journal of Management, 19(3): 683-705.

_____. (1995). Problems with repeated measures analysis: demonstration with a study of the diversification and performance relationship. Academy of Management Journal, 38(6): 1692-1708.

Bowman, E.H. & Helfat, C.E. (2001). Does corporate strategy matter? Strategic Management Journal, 22(1): 1-23.

Brouthers, L.E. (1998). Explaining MNC profitability: country-specific, industry-specific and country-industry interactive influences. Management International Review, 38(4): 345-361.

Brush, T.H. & Bromiley, P. (1997). What does a small corporate effect mean? A variance components simulation of corporate and business effects. Strategic Management Journal, 18(10): 825-835.

Brush, T.H.; Bromiley, P. & Hendrickx, M. (1999). The relative influence of industry and corporation on business segment performance: an alternative estimate. Strategic Management Journal, 20(6): 519-547.

Bushman, R.M. & Piotroski, J.D. (2006). Financial reporting incentives for conservative accounting: the influence of legal and political institutions. Journal of Accounting and Economics, 42(1-2): 107-148.

Cameron, A.C. & Trivedi, P.K. (2009). Microeconometrics using Stata. Stata Press, College Station, Texas, 692 p.

Christmann, P.; Day, D.L. & Yip, G.S. (1999). The relative influence of country conditions, industry structure and business strategy on MNC subsidiary performance. Journal of International Management, 5(4): 241-265.

Collins, J. M. (1990). A market performance comparison of U.S. firms active in domestic, developed, and developing countries. Journal of International Business Studies, 21(2): 271-288.

De Leeuw, J. & Meijer, E. (Ed.) (2007). Handbook of multilevel analysis. Springer, New York, New York, 487 p.

Deadrick, D.L.; Bennett, N. & Russell, C.J. (1997). Using hierarchical linear modeling to examine dynamic performance criteria over time. Journal of Management, 23(6): 745-757.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. Journal of Educational and Behavioral Statistics, 20(2): 115-147.

Dyck, A. & Zingales, L. (2004). Private benefits of control: an international comparison. Journal of Finance, 59(2): 537-600.

Fávero, L.P.; Belfiore, P.; Silva, F.L. & Chan, B.L. (2009). Análise de dados: modelagem multivariada para tomada de decisões. Campus Elsevier, Rio de Janeiro, 646 p.

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can and cannot do. Technometrics, 48(3): 432-435.

Goldstein, H. (2003). Multilevel statistical models. Arnold, London, 3. ed., 272 p.

Goldszmidt, R.G.B.; Brito, L.A.L. & Vasconcelos, F.C. (2007). O efeito país sobre o desempenho da firma: uma abordagem multinível. Anais do III Encontro de Estudos em Estratégia, São Paulo.

Grambovas, C.A.; Giner, B. & Christodoulou, D. (2006). Earnings conservatism: panel data evidence from the European Union and the United States. Abacus, 42(3-4): 354-378.

Greene, W.H. (2007). Econometric analysis. Prentice Hall, Upper Saddle River, New Jersey, 6. ed., 1216 p.

Hawawini, G.; Subramanian, V. & Verdin, P. (2004). The home country in the age of globalization: how much does it matter for firm performance? Journal of World Business, 39(2): 121-135.

Heck, R.H. & Thomas, S.L. (2008). An introduction to multilevel modeling techniques. Psychology Press, New York, New York, 2. ed., 280 p.

Hofmann, D.A. (1997). An overview of the logic and rationale of hierarchical linear models. Journal of Management, 23(6): 723-744.

Hofmann, D.A.; Jacobs, R. & Baratta, J.E. (1993). Dynamic criteria and the measurement of change. Journal of Applied Psychology, 78(2): 194-204.

Hough, J.R. (2006). Business segment performance redux: a multilevel approach. Strategic Management Journal, 27(1): 45-61.

Hox, J. (2002). Multilevel analysis: techniques and applications. Lawrence Erlbaum, Mahwah, New Jersey, 304 p.

Hsiao, C. (2003). Analysis of panel data. Cambridge University Press, Cambridge, 2. ed., 384 p.

Islam, N. (1995). Growth empirics: a panel data approach. The Quarterly Journal of Economics, 110(4): 1127-1170.

Jensen, J.; Menezes-Filho, N. & Sbragia, R. (2004). Os determinantes dos gastos em P&D no Brasil: uma análise com dados em painel. Estudos Econômicos, 34(4): 661-691.

Jones, S.T.; Banning, K. (2009). US elections and monthly stock market returns. Journal of Economics and Finance, 33(3), 273-287.

King, R.G. & Levine, R. (1993). Finance and growth: Schumpeter might be right. The Quarterly Journal of Economics, 108(3): 717-737.

Kothari, S. (2001). Capital markets research in accounting. Journal of Accounting and Economics, 31(1-3), 105-231.

Kreft, I. & de Leeuw, J. (1998). Introducing multilevel modeling. Sage, London, 160 p.

Kwon, C.S.; Shin, T.S. & Bacon, F.W. (1997). The effect of macroeconomic variables on stock market returns in developing markets. Multinational Business Review, 5(2): 63-70.

Lee, U. (1997). Stock market and macroeconomic policies: new evidence from Pacific Basin countries. Multinational Finance Journal, 1(4): 273-289.

Lee, Y. & Nelder, J.A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika, 88(4): 987-1006.

Levine, R. (1997). Financial development and economic growth: views and agenda. Journal of Economic Literature, 35(2): 688-726.

Makino, S.; Beamish, P.W. & Zhao, N.B. (2004). The characteristics and performance of Japanese FDI in less developed and developed countries. Journal of World Business, 39(4): 377-392.

Makino, S.; Isobe, T. & Chan, C.M. (2004). Does country matter? Strategic Management Journal, 25(10): 1027-1043.

Martins, G.A. & Theóphilo, C.R. (2009). Metodologia da investigação científica para ciências sociais aplicadas. Atlas, São Paulo, 2. ed., 252 p.

McGahan, A.M. & Porter, M.E. (1997). How much does industry matter, really? Strategic Management Journal, 18(S1): 15-30.

Misangyi, V.; Lepine, J.A.; Algina, J. & Goeddeke Jr., F. (2006). The adequacy of repeated-measures regression for multilevel research. Organizational Research Methods, 9(1): 5-28.

Natis, L. (2000). Modelos lineares hierárquicos. Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. Dissertação de Mestrado.

Rabe-Hesketh, S. & Skrondal, A. (2008). Multilevel and longitudinal modeling using Stata. Stata Press, College Station, Texas, 2. ed., 562 p.

Rajan, R.G. & Zingales, L. (1998). Financial dependence and growth. American Economic Review, 88(3): 559-586.

_____. (2003). The great reversals: the politics of financial development in the twentieth century. Journal of Financial Economics, 69(1): 5-50.

Ramos-Rodriguez, A.R. & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: a bibliometric study of the Strategic Management Journal, 1980-2000. Strategic Management Journal, 25(10): 981-1004.

Raudenbush, S. & Bryk, A. (2002). Hierarchical linear models: applications and data analysis methods. Sage, Thousand Oaks, California, 2. ed., 512 p.

Raudenbush, S.; Bryk, A.; Cheong, Y.F.; Congdon, R. & du Toit, M. (2004). HLM 6: hierarchical linear and nonlinear modeling. Scientific Software International, Inc., Lincolnwood, Illinois, 297 p.

Roe, M.J. & Siegel, J.I. (2009). Political instability: its effects on financial development, its roots in the severity of economic inequality. SSRN. Disponível em: http://ssrn.com/abstract=963214. Acesso em: 18 set 2009.

Roquebert, J.A.; Phillips, R.L. & Westfall, P.A. (1996). Market vs. management: what ´drives´ profitability? Strategic Management Journal, 17(8): 653-664.

Rumelt, R.P. (1991). How much does industry matter? Strategic Management Journal, 12(3): 167-185.

Schmalensee, R. (1985). Do markets differ much? American Economic Review, v. 75(3): p. 341-351.

Short, J.C.; Ketchen Jr., D.J.; Bennett, N. & du Toit, M. (2006). An examination of firm, industry, and time effects on performance using random coefficients modeling. Organizational Research Methods, 9(3): 259-284.

Short, J.C.; Ketchen Jr., D.J.; Palmer, T.B. & Hult, G.T.M. (2007). Firm, strategic group, and industry influences on performance. Strategic Management Journal, 28(2): 147-167.

Snijders, T.A.B. & Bosker, R.J. (1999). Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage, London, 272 p.

Soto, J.L.G. & Morera, M.C. (2005). Modelos jerárquicos lineales. La Muralla, Madrid, 136 p.

Steenbergen, M.R. & Jones, B.S. (2002). Modeling multilevel data structures. American Journal of Political Science, 46(1): 218-237.

Tobin, J. (1969). A general equilibrium approach to monetary theory. Journal of Money, Credit and Banking, 1(1): 15-29.

Valadkhani, A.; Chancharat, S. & Havie, C. (2009). Analysing the impact of international stock markets and macroeconomic variables on the Thai Stock Market. The Business Review, 12(2): 50-56.

Von Furstenberg, G.M. (1977). Corporate investment: does market valuation matter in the aggregate? Brookings Papers on Economic Activity, 8(2): 347-408.

Watts, R.L. & Zimmerman, J.L. (1986). Positive accounting theory. Prentice Hall, Englewood Cliffs, 400 p.

Wooldridge, J.M. (2002). Econometric analysis of cross section and panel data. MIT Press, Cambridge, Massachusetts, 776 p.

Abstract

Studies investigating the influence of market and economic conditions on stock indexes in emerging countries are still scarce. Under the assumption that there are differences on the evolution of stock indexes among these countries, this study applies panel data and hierarchical modeling to assess the effects of stock markets' variables and to investigate the economic influences on the evolution of stock indexes over time. Through a sample of 60 emerging countries over 262 months (1986-2007) obtained from the Compustat Global data basis, totaling 9,979 observations, one can verify that the market dividend yield, the price to book value and turnover influence the evolution of stock indexes, after adjusting a linear regression model with fixed effects and AR(1) error terms. It's possible to verify that the country effect is also representative for the differentiation of stock indexes growth rates, especially due to the research investment.

Keywords: Country Effect; Hierarchical Modeling; Panel Data; Repeated Measures; Stock Market.

REVISTA BRASILEIRA DE ESTATÍSTICA - RBEs

POLÍTICA EDITORIAL

A Revista Brasileira de Estatística - RBEs publica trabalhos relevantes em Estatística Aplicada, não havendo limitação no assunto ou matéria em questão. Como exemplos de áreas de aplicação, citamos as áreas de advocacia, ciências físicas e biomédicas, criminologia, demografia, economia, educação, estatísticas governamentais, finanças, indústria, medicina, meio ambiente, negócios, políticas públicas, psicologia e sociologia, entre outras. A RBEs publicará, também, artigos abordando os diversos aspectos de metodologias relevantes para usuários e produtores de estatísticas públicas, incluindo planejamento, avaliação e mensuração de erros em censos e pesquisas, novos desenvolvimentos em metodologia de pesquisa, amostragem e estimação, imputação de dados, disseminação e confiabilidade de dados, uso e combinação de fontes alternativas de informação e integração de dados, métodos e modelos demográfico e econométrico.

Os artigos submetidos devem ser inéditos e não devem ter sido submetidos simultaneamente a qualquer outro periódico.

O periódico tem como objetivo a apresentação de artigos que permitam fácil assimilação por membros da comunidade em geral. Os artigos devem incluir aplicações práticas como assunto central, com análises estatísticas exaustivas e apresentadas de forma didática. Entretanto, o emprego de métodos inovadores, apesar de ser incentivado, não é essencial para a publicação.

Artigos contendo exposição metodológica são também incentivados, desde que sejam relevantes para a área de aplicação pela qual os mesmos foram motivados, auxiliem na compreensão do problema e contenham interpretação clara das expressões algébricas apresentadas.

A RBEs tem periodicidade semestral e também publica artigos convidados e resenhas de livros, bem como incentiva a submissão de artigos voltados para a educação estatística.

Artigos em espanhol ou inglês só serão publicados caso nenhum dos autores seja brasileiro e nem resida no País.

Todos os artigos submetidos são avaliados quanto à qualidade e à relevância por dois especialistas indicados pelo Comitê Editorial da RBEs.

O processo de avaliação dos artigos submetidos é do tipo 'duplo cego', isto é, os artigos são avaliados sem a identificação de autoria e os comentários dos avaliadores também são repassados aos autores sem identificação.

INSTRUÇÃO PARA SUBMISSÃO DE ARTIGOS À RBEs

O processo editorial da RBEs é eletrônico. Os artigos devem ser submetidos para o site http://rbes.submitcentral.com.br/login.php

Secretaria da RBEs

Revista Brasileira de Estatística – RBEs

ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS - IBGE

Rua André Cavalcanti, 106, sala 503A

Centro, Rio de Janeiro – RJ

CEP: 20231-050

Tels.: 55 21 2142-0123 ramal 3956 (Marilene Pereira Piau Câmara – Secretária)

55 21 2142-4957 (Pedro Luis do Nascimento Silva – Editor–Executivo)

Fax: 55 21 2142-0501

INSTRUÇÕES PARA PREPARO DOS ORIGINAIS

Os originais enviados para publicação devem obedecer às normas seguintes:

1. Podem ser submetidos originais processados pelo editor de texto *Word for Windows* ou originais processados em LaTeX (ou equivalente) desde que estes últimos sejam encaminhados e acompanhados de versões em pdf, conforme descrito no item 3, a seguir;

2. A primeira página do original (folha de rosto) deve conter o título do artigo, seguido do(s) nome(s) completo(s) do(s) autor(es), indicando-se, para cada um, a afiliação e endereço para correspondência. Agradecimentos a colaboradores e instituições, e auxílios recebidos, se for o caso de constarem no documento, também devem figurar nesta página;

3. No caso de a submissão não ser em *Word for Windows*, três arquivos do original devem ser enviados. O primeiro deve conter os originais no processador de texto utilizado (por exemplo, LaTeX). O segundo e terceiro devem ser no formato pdf, sendo um com a primeira página, como descrito no item 2, e outro contendo apenas o título, sem a identificação do(s) autor(es) ou outros elementos que possam permitir a identificação da autoria;

4. A segunda página do original deve conter resumos em português e inglês (*abstract*), destacando os pontos relevantes do artigo. Cada resumo deve ser digitado seguindo o mesmo padrão do restante do texto, em um único parágrafo, sem fórmulas, com, no máximo, 150 palavras;

5. O artigo deve ser dividido em seções, numeradas progressivamente, com títulos concisos e apropriados. Todas as seções e subseções devem ser numeradas e receber título apropriado;

6. Tratamentos algébricos exaustivos devem ser evitados ou alocados em apêndices;

7. A citação de referências no texto e a listagem final de referências devem ser feitas de acordo com as normas da ABNT;

8. As tabelas e gráficos devem ser precedidos de títulos que permitam perfeita identificação do conteúdo. Devem ser numeradas sequencialmente (Tabela 1, Figura 3, etc.) e referidas nos locais de inserção pelos respectivos números. Quando houver tabelas e demonstrações extensas ou outros elementos de suporte, podem ser empregados apêndices. Os apêndices devem ter título e numeração, tais como as demais seções de trabalho;

9. Gráficos e diagramas para publicação devem ser incluídos nos arquivos com os originais do artigo. Caso tenham que ser enviados em separado, devem ter nomes que facilitem a sua identificação e posicionamento correto no artigo (ex.: Gráfico 1; Figura 3; etc.). É fundamental que não existam erros, quer no desenho, quer nas legendas ou títulos;

10. Não serão permitidos itens que identifiquem os autores do artigo dentro do texto, tais como: número de projetos de órgãos de fomento, endereço, *e-mail*, etc. Caso ocorra, a responsabilidade será inteiramente dos autores; e

11. No caso de o artigo ser aceito para a publicação após a avaliação dos pareceristas, serão encaminhadas as sugestões/comentários aos autores sem a sua identificação. Uma vez nesta condição, é de responsabilidade única dos autores fazer o *download* da formatação padrão da revista (em doc ou em LaTeX) para o envio da versão corrigida.