



XX CONGRESSO

SOCIEDADE PORTUGUESA DE ESTATÍSTICA

PORTO 26 A 29 DE SETEMBRO 2012

MODELOS COM EQUAÇÕES ESTRUTURAIS MARIA DE FÁTIMA SALGUEIRO



CATÓLICA
PORTO

ECONOMIA E
GESTÃO



SOCIEDADE PORTUGUESA
DE ESTATÍSTICA

XX Congresso Anual da Sociedade Portuguesa de Estatística
Porto, 26 a 29 de Setembro de 2012

Modelos com Equações Estruturais

Maria de Fátima Salgueiro

Edições SPE

Título: Modelos com Equações Estruturais

Autora: Maria de Fátima Salgueiro

Editora: Sociedade Portuguesa de Estatística

Conceção Gráfica da Capa: Católica | Porto

Impressão: Instituto Nacional de Estatística

Tiragem: 500 Exemplares

ISBN: 978-972-8890-26-1

Depósito Legal: 348725/12

Prefácio

Foi com um misto de espanto e de alegria que recebi o convite que me foi dirigido pelo Prof. Dr. Pedro Duarte Silva, em nome da Direcção da Sociedade Portuguesa de Estatística (SPE) e da Comissão Organizadora do XX Congresso Anual da SPE, para realizar em 2012 o Minicurso sobre Modelos com Equações Estruturais e publicar um livro sobre o tema. Agradeço publicamente o convite que me foi endereçado, e a oportunidade que me foi dada, esperando poder corresponder ao desafio que me foi lançado!

Tratando-se de uma temática tão abrangente, alvo de inúmeros desenvolvimentos teóricos recentes, e tão atractiva do ponto de vista da aplicação a variadas problemáticas e áreas substantivas, houve que fazer escolhas em termos dos principais tópicos a abordar no dia de Minicurso, e a cobrir no presente livro que lhe serve de apoio. Optou-se por estruturar o presente livro em cinco capítulos: um primeiro capítulo introdutório, com uma breve resenha histórica, exemplos motivadores, clarificação dos principais conceitos, notação e *software* a utilizar. Um segundo capítulo faz a ponte entre a Análise Factorial Clássica e a Análise Factorial Confirmatória, também conhecida por componente de medida do modelo global. O terceiro capítulo ilustra como a Regressão Linear e a *Path Analysis* podem ser considerados casos particulares de um modelo global só com variáveis manifestas. Da junção da componente de medida com a componente

estrutural resulta um Modelo com Equações Estruturais global. O quarto capítulo apresenta uma proposta de metodologia de modelação, dando particular ênfase à estimação do modelo e à avaliação da qualidade do ajustamento modelo-dados. Estes quatro primeiros capítulos encerram uma introdução à temática em estudo. O quinto capítulo procura abordar outros tópicos considerados de particular interesse, como sejam modelos multigrupos, modelos com estruturas de médias e análise factorial de segunda ordem. Será ainda feita referência a alguns tópicos de desenvolvimento mais recente, designadamente modelos com curva de crescimento latente e modelos com estrutura multinível. Para além da listagem das referências bibliográficas referidas ao longo do texto, e demais bibliografia consultada, são apresentados três apêndices, com comandos de sintaxe e excertos anotados de outputs do LISREL.

Diz Jan de Leew na apresentação de Kaplan (2000): *Over the last 25 years Structural Equation Modeling (SEM) has become one of the most important data analysis techniques in the social sciences. In fact, it has become much more than that. It has become a language to formulate social science theories and a language to talk about the relationship between variables.* Mas, continua Jan de Leew, *SEM is not without its critics, and most researchers active in the area will admit that it can easily be misused and, in fact, it has frequently been misused!*

Várias serão as razões para o incorrecto uso da metodologia de modelação com recurso a Modelos com Equações Estruturais: porque se trata de uma técnica de modelação complicada, e muitos dos seus utilizadores não têm ainda os conhecimentos estatísticos suficientes para a compreender e conhecer, porque algum do *software* estatístico disponível é “excessivamente” *user friendly*, porque alguma da literatura disponível quase desvaloriza a componente matemática dos modelos ...

Com o Minicurso e com o presente livro pretendo ajudar a contribuir para uma cada vez mais correcta utilização dos Modelos com Equações Estruturais junto de todos quantos com eles lidam, em particular da Comunidade Científica Portuguesa.

Com todos os que assistirem ao Minicurso ou lerem este livro partilho conhecimentos e experiências. Bem Hajam! Sejam gratos a todos aqueles que nos precederam no interesse por estas temáticas e nos possibilitaram tão estimulante área de investigação e empenhemo-nos em contribuir para o legado que deixamos às gerações vindouras.

Confio este propósito ÀQUELE que é a Verdade, que sustenta e dá sentido às nossas vidas e suas canseiras, e que em cada dia eu procuro servir, também através de Modelos com Equações Estruturais.



Maria de Fátima Salgueiro

Instituto Universitário de Lisboa (ISCTE-IUL)

Lisboa, Julho de 2012

Índice

1	Introdução	1
1.1	Modelos com Equações Estruturais: Aspectos Introdutórios	1
1.2	Breve Resenha Histórica	1
1.3	Exemplos Motivadores: Problemas Tipo e Áreas de Aplicação	5
1.3.1	Exemplo de um Modelo de Regressão Linear	6
1.3.2	Exemplo de um Modelo de Análise de Caminhos	8
1.3.3	Exemplo de um Modelo de Análise Factorial Confirmatória	9
1.3.4	Exemplo de um Modelo com Equações Estruturais Global	10
1.3.5	Outros Exemplos Motivadores	11
1.4	Principais Conceitos e Notação	15
1.4.1	Conceitos Base	15
1.4.2	Notação a Utilizar	18
1.5	Software Estatístico Disponível e Páginas Web	21
1.6	Referências Bibliográficas: Breves Recomendações	23

2	Da Análise Factorial Exploratória à Análise Factorial Confirmatória	25
2.1	O Modelo Clássico de Análise Factorial Exploratória	27
2.1.1	Análise Factorial <i>versus</i> Análise em Componentes Principais	27
2.1.2	Pressupostos e Equações do Modelo Clássico de Análise Factorial	27
2.1.3	Estimação e Ajustamento do Modelo Clássico de Análise Factorial	29
2.1.4	Exemplo de um Modelo de Análise Factorial Exploratória em SPSS e em PRELIS	30
2.2	Análise Factorial Confirmatória - Modelo de Medida	34
2.2.1	Pressupostos e Equações do Modelo de Análise Factorial Confirmatória	34
2.2.2	Identificação do Modelo de Análise Factorial Confirmatória	35
2.2.3	Validade e Fiabilidade dos Indicadores e das Escalas	37
2.2.4	Exemplo de um Modelo de Análise Factorial Confirmatória em LISREL	40
3	Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global	45
3.1	Modelo Estrutural só com Variáveis Manifestas	46
3.1.1	Pressupostos e Equações do Modelo Estrutural	46
3.1.2	Identificação do Modelo Estrutural	48
3.1.3	O Modelo de Análise de Caminhos: Efeitos Directos, Indirectos e Totais	50
3.1.4	Exemplos de Modelos de Regressão Linear	52

3.1.5	Exemplo de Modelo de Análise de Caminhos	58
3.2	Modelo Equações Estruturais Global	61
3.2.1	Pressupostos e Equações do Modelo Global	61
3.2.2	Identificação do Modelo Global	64
3.2.3	Modelo MIMIC	66
3.2.4	Exemplo de Modelo com Equações Estruturais Global	67
3.2.5	Exemplo de Modelo Longitudinal, com Medidas Repetidas	72
4	Metodologia de Modelação	77
4.1	Etapas na Modelação	77
4.2	Estimação do Modelo com Equações Estruturais	83
4.2.1	Objectivo da Estimação	83
4.2.2	Funções de Ajustamento	85
4.3	Qualidade de Ajustamento Modelo-Dados	92
4.3.1	Medidas de Ajustamento Global do Modelo	93
4.3.2	Ajustamento das Componentes do Modelo	103
5	Outros Tópicos	107
5.1	Modelos Multi-Grupos	107
5.1.1	O Que Se Pode Testar	108
5.1.2	Estimação e Qualidade do Ajustamento	109
5.1.3	Soluções Estandarizadas	110
5.1.4	Um Exemplo de Análise Multi-Grupos em LISREL	111
5.2	Modelos com Estruturas de Médias	116

viii Índice

5.2.1	Extensão do Modelo com Equações Estruturais Global	116
5.2.2	Função de Ajustamento	118
5.2.3	Estimação das Médias dos Factores	119
5.3	Análise Factorial de Segunda Ordem	119
5.3.1	Equação do Modelo e Matriz de Variância-Covariância Implícita	120
5.3.2	Um Exemplo de Aplicação de um Modelo de Análise Factorial de Segunda Ordem	120
5.4	Modelos com Curva de Crescimento Latente	125
5.4.1	Aspectos Introdutórios	125
5.4.2	Modelos com Trajectórias Não Condicionadas	127
5.4.3	Modelos com Múltiplos Indicadores	129
5.4.4	Modelos com Variáveis Manifestas Ordinais	130
5.4.5	Modelos com Trajectórias Condicionadas	131
5.5	Modelos com Estrutura Multi-Nível	133
5.6	Indicadores Reflexivos versus Indicadores Formativos	134
5.7	Outros Tópicos Ainda Não Mencionados	135
	Bibliografia	137
	A Comandos e Outputs Anotados - Modelos de Análise Factorial	147
A.1	Modelo de Análise Factorial Exploratória em SPSS	147
A.2	Modelo de Análise Factorial Exploratória em PRELIS	150
A.3	Modelo de Análise Factorial Confirmatória em LISREL	154
	B Comandos e Outputs Anotados - Modelos Estruturais	163

B.1	Modelos de Regressão Linear	163
B.1.1	MRLS	163
B.1.2	MRLM	164
B.2	Modelo de Análise de Caminhos	165
B.3	Modelo Equações Estruturais Global	168
B.4	Modelo Global com Efeito Longitudinal	172
C	Comandos e Outputs Anotados - Outros Tópicos	175
C.1	Análise Multi-Grupos	175
C.1.1	Estrutura de Comandos de Sintaxe LISREL em Análise Multi-Grupos	175
C.1.2	Exemplo de AFC multi-grupos	177
C.1.3	Exemplo de modelo global multi-grupos	177
C.2	Modelo de Análise Factorial de Segunda Ordem	179

Capítulo 1

Introdução

1.1 Modelos com Equações Estruturais: Aspectos Introdutórios

Modelação estatística com recurso a Modelos com Equações Estruturais (*Structural Equation Modelling*) designa actualmente um *framework* integrador de um conjunto de modelos estatísticos, do qual podem ser considerados casos particulares os modelos de análise factorial confirmatória, de regressão linear múltipla multivariada (equações simultâneas), de equações estruturais global ou de análise factorial de segunda ordem, entre outros.

O Modelo com Equações Estruturais global permite estimar, simultaneamente, um conjunto de relações multiplas de dependência entre variáveis, algumas das quais podem ser latentes, i.e., não directamente mensuráveis mas medidas a partir de um conjunto de indicadores manifestos ou observados. São variadas e crescentes as áreas de aplicação deste tipo de modelos, incluindo a Gestão, o Marketing e a Psicologia.

1.2 Breve Resenha Histórica

Na área das Ciências Sociais muitas realidades há que não podem ser medidas directamente, mas através de outras, essas sim directamente

2 Introdução

mensuráveis. Do ponto de vista estatístico, uma variável que não pode ser medida directamente, mas antes através de um conjunto de indicadores manifestos, sujeitos a erro de medida, diz-se uma **variável latente**. Uma variável latente representa pois algo de subjacente ao que é observado. Considera-se que a variável latente explica totalmente a estrutura de associações entre indicadores manifestos - é o chamado pressuposto da independência local. O desenvolvimento de modelos com variáveis latentes, designadamente o modelo clássico de análise factorial, foi crucial para o subsequente desenvolvimento de Modelos com Equações Estruturais. O modelo de análise factorial será estudado em detalhe no Capítulo 2.

Uma outra grande linha de investigação que muito terá contribuído para a evolução dos Modelos com Equações Estruturais é a dos modelos de ***Path Analysis*** (Análise de Caminhos). Sewall Wright, um investigador nas áreas da biologia e da biometria, é o seu proponente, e já em 1934 uma tradição especializada neste campo da biometria lidava com modelos de equações simultâneas no contexto da representação e dos esquemas conhecidos por *path analysis*. Só por volta dos anos 60, porém, é que alguns metodólogos (no campo da Sociologia) mostraram o valor de combinar a simplicidade das representações através do diagrama da análise de caminhos com o rigor da especificação simultânea de equações.

No princípio dos anos 70 do século XX, a modelação causal foi-se tornando um método importante de investigação sociológica, começando a ser estudados em profundidade modelos que incluíam variáveis latentes. Faltavam, contudo, princípios matemáticos e estatísticos unificadores de uma metodologia. A combinação dos trabalhos de Jöreskog (1973), Keesling (1972) e Wiley (1973) permitiu a construção de um *framework* analítico coerente. Com efeito, o assim designado modelo de Jöreskog-Keesling-Wiley (JKW) surge como um marco, no sentido de apresentar uma plataforma de trabalho que torna possível lidar com modelos complexos de forma rotineira. O modelo de JKW apresenta duas partes distintas: uma primeira, que consiste

num modelo similar aos modelos econométricos de equações simultâneas, com a exceção de que todas as variáveis são latentes, conhecida por modelo estrutural; e uma segunda parte, conhecida por modelo de medida, que apresenta os indicadores como dependentes das variáveis latentes, à semelhança do que acontece na análise factorial. Karl Jöreskog, um estatístico da área da psicometria, é vulgarmente reconhecido como o grande promotor da **CFA - Confirmatory Factor Analysis** (Análise Factorial Confirmatória) e dos **SEM - Structural Equation Models** (Modelos com Equações Estruturais). O software estatístico LISREL (*LI*near *Str*uctural *REL*ations), de Jöreskog e Sörbom, foi o programa estatístico inequivocamente responsável pela crescente divulgação e popularidade deste tipo de modelos.

Desde cedo se foi tornando claro que modelos de análise factorial não eram apropriados para modelar dados não quantitativos, o que fez desenvolver a investigação em modelos estatísticos para dados qualitativos. Lazarsfeld e Henry propuseram em 1968 o modelo com classes latentes e, em 1974, Goodman generalizou a metodologia a variáveis nominais. Por seu turno, em 1979 Haberman propôs e desenvolveu modelos log-lineares. Os trabalhos de Vermunt e Magidson levaram ao desenvolvimento do software estatístico Latent Gold, inicialmente concebido para modelos com variáveis categoriais.

Se Karl Jöreskog teve um papel crucial e único nos anos 70 do século XX, cuja síntese implementou de forma unificada no programa LISREL, há também que reconhecer como determinantes os subsequentes contributos de investigadores tais como Albert Satorra, Peter Bentler, M.W. Browne, R.P. McDonald e, de uma forma muito especial, Bengt Muthén, contributos esses que se revelaram fundamentais para desenvolvimentos mais recentes no âmbito dos Modelos com Equações Estruturais. Um dos desenvolvimentos importante foi a extensão dos métodos de estimação por forma a possibilitar lidar com distribuições não normais. Devido ao trabalho de Browne (1984) e Muthén (1978, 1984), entre outros, passou a ser possível estimar parâmetros de Modelos com Equações Estruturais que incluem misturas de variáveis

4 Introdução

dicotômicas, categoriais ordinais e contínuas.

Os desenvolvimentos estatísticos mais recentes procuram encontrar um *framework* integrador, possibilitando assim que abordagens tradicionalmente diferentes de modelação estatística sejam consideradas casos particulares de modelação com recurso a Modelos com Equações Estruturais: modelos com variáveis latentes (categoriais e métricas), para dados com uma estrutura hierárquica (longitudinal, multinível), considerando a natureza complexa do desenho amostral. Neste âmbito se inclui o trabalho de Bengt Muthén, consubstanciado no software estatístico Mplus, a investigação de Anders Skrondal e Sophia Rabe-Hesketh, associada ao programa STATA e ao pacote GLLAMM - *Generalized Linear Latent And Mixed Models* (ver Skrondal e Rabe-Hesketh, 2004). É também no âmbito deste esforço integrador / tentativa de convergência de metodologias, que se podem entender os desenvolvimentos mais recentes propostos por Karl Jöreskog e colaboradores tais como Irini Moustaki, Kenneth Bollen ou Patrick Curran, os quais têm permitido ao LISREL acompanhar estes desenvolvimentos teóricos e estimar, na lógica dos Modelos com Equações Estruturais, modelos com variáveis ordinais, modelos de análise factorial confirmatória multinível, modelos de análise factorial exploratória para variáveis ordinais, modelos com curvas de crescimento latente (*latent growth curve models*) . . .

Paralelamente aos desenvolvimentos teóricos e científicos, acompanhados de perto pelo desenvolvimento do software estatístico que permite estimar os modelos, tem sido crescente o interesse em aplicar Modelos com Equações Estruturais a problemas substantivos em áreas científicas tão diversas como a Psicologia, a Gestão, o Marketing, a Economia, o Comportamento Organizacional, a Medicina, a Demografia, a Educação, a Sociologia ou as Finanças. Prova disso são os inúmeros artigos científicos publicados nos últimos anos em revistas de reconhecido mérito internacional, nas mais variadas áreas de conhecimento, que usam Modelos com Equações Estruturais como “ferramenta” de modelação estatística. Crescente tem sido também

o recurso da comunidade acadêmica, em diversas áreas do saber, a este tipo de modelos, consubstanciado no crescente número de Dissertações de Mestrado e de Teses de Doutorado que a utilizam, tendência à qual a comunidade científica Portuguesa não é alheia.

1.3 Exemplos Motivadores: Problemas Tipo e Áreas de Aplicação

Os Modelos com Equações Estruturais permitem estimar simultaneamente um conjunto de relações de dependência entre variáveis que podem ser latentes, entrando assim em linha de conta com a componente erro de medida. Esta sua capacidade de abarcar uma visão mais sistemática e holística dos problemas e fenómenos a modelar tem levado a uma crescente utilização deste tipo de modelos em variadas áreas de estudo.

Nesta secção são dados exemplos de diagramas de diferentes tipos de submodelos de um Modelo com Equações Estruturais global. Alguns destes exemplos serão retomados ao longo deste livro para ilustrar a metodologia de modelação com recurso a Modelos com Equações Estruturais.

Um primeiro conjunto de exemplos utiliza dados relativos a percepções de bem estar, e de satisfação com diversas dimensões da vida. Autores há que defendem que a percepção por parte dos indivíduos do seu grau de bem estar, e a sua satisfação com as condições de vida, são meios importantes de aferir o padrão de vida da sociedade em que esses indivíduos se inserem. Todavia, não há uma forma única de medir percepções de bem estar, sendo que diversos instrumentos de medida têm sido propostos na literatura. Em termos de opções metodológicas é frequente a utilização de uma medida global, ou a construção de um *score* total, obtido através da soma ou da média das respostas dos indivíduos às dimensões de interesse.

Salgueiro *et al.* (2011) propuseram uma abordagem longitudinal para

6 Introdução

modelar as trajectórias de satisfação com duas dimensões de bem estar (satisfação com aspectos materiais e satisfação com aspectos de lazer), com recurso a modelos com curva de crescimento latente (*latent growth curve models*). Foram utilizados dados do *British Household Panel Survey* (BHPS), um painel representativo dos agregados familiares no Reino Unido, da responsabilidade do *Institute for Social and Economic Research*, Universidade de Essex, Reino Unido (ver Taylor *et al.*, 2008). O BHPS existe desde 1991 e recolhe dados anualmente, tanto ao nível do agregado como do indivíduo, tendo por objectivo obter informação acerca de mudanças ocorridas aos níveis económico e social. São ainda recolhidos dados sobre um conjunto de variáveis atitudinais e comportamentais, como seja a satisfação com as condições de vida.

Neste livro, e usando dados de Salgueiro *et al.* (2011) relativos aos anos de 2002 a 2005, no que a percepções de bem estar diz respeito são consideradas seis questões de satisfação. Numa escala de 1 (nada satisfeito) a 7 (completamente satisfeito) os respondentes foram chamados a classificar o seu grau de satisfação com: i) o rendimento do agregado familiar; ii) a sua casa / habitação; iii) o seu emprego; iv) a sua vida social; v) a quantidade de tempo livre disponível e vi) a utilização feita desse tempo livre. As três primeiras questões são consideradas como indicadores de satisfação com aspectos materiais, e as três últimas como medindo a satisfação com aspectos de lazer. São ainda consideradas como possíveis variáveis explicativas da satisfação o rendimento do agregado familiar, a idade, o sexo e o número de horas de trabalho semanal do respondente.

1.3.1 Exemplo de um Modelo de Regressão Linear

Um modelo muito simples postula que a satisfação com aspectos materiais é função do rendimento do agregado familiar e pode ser representada por um modelo de regressão linear simples (MRLS):

$$Mater02 = f(\log Rend)$$

Mater02 designa um *score* de satisfação material relativo ao ano de 2002, obtido através da soma das respostas dos inquiridos às três questões de satisfação relativas a aspectos materiais (satisfação com o rendimento do agregado; com a casa / habitação e com o emprego), podendo variar de 3 (satisfação mínima) a 21 (satisfação máxima). O logaritmo do rendimento mensal do agregado familiar é designado por *logRend*.

A Figura 1.1 representa o diagrama deste MRLS: *Mater02* é a variável dependente ou explicada e *logRend* é a variável independente ou explicativa.



Figura 1.1 Exemplo de diagrama de MRLS.

Extendendo o modelo anteriormente apresentado, é possível postular que a satisfação com aspectos materiais depende não só do rendimento do agregado familiar, mas também da idade e do sexo do inquirido. Assumindo a *idade* como medida em anos e o *sexo* como uma variável binária (0 = Feminino; 1 = Masculino), é possível equacionar um modelo de regressão linear múltipla (MRLM):

$$Mater02 = f(logRend, idade, sexo)$$

A Figura 1.2 representa o diagrama deste MRLM: *Mater02* é a variável dependente ou explicada e *logRend*, *idade* e *sexo* são as variáveis independentes ou explicativas.

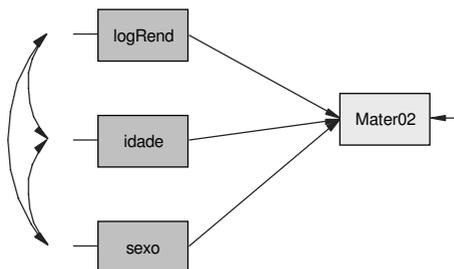


Figura 1.2 Exemplo de diagrama de MRLM.

1.3.2 Exemplo de um Modelo de Análise de Caminhos

É ainda possível pensar que a satisfação com aspectos materiais num dado momento temporal influencia a percepção de satisfação com o lazer que o indivíduo tem no momento temporal seguinte, assumindo-se assim que, por exemplo, *Mater02* influencia positivamente *Lazer03*. Pode ainda postular-se uma relação de influência negativa entre o número de horas de trabalho semanal e a percepção da satisfação com o lazer. Este conjunto de relações de dependência entre diferentes variáveis, em que variáveis independentes numa relação podem ser consideradas dependentes noutras, é o objecto de estudo da análise de caminhos e encontra-se representado no diagrama da Figura 1.3. São postuladas as duas seguintes relações estruturais:

$$\begin{cases} Mater02 = f(logRend, idade, sexo) \\ Lazer03 = f(Mater02, sexo, n_horas) \end{cases}$$

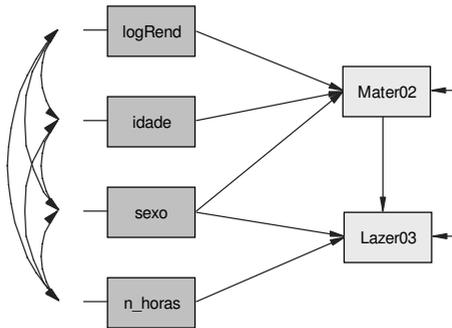


Figura 1.3 Exemplo de diagrama de modelo de análise de caminhos.

1.3.3 Exemplo de um Modelo de Análise Factorial Confirmatória

Até agora, tantos nos modelos de regressão como na análise de caminhos, as variáveis de satisfação utilizadas foram indicadores observados, já que não é possível medir de forma directa e única percepções de satisfação com aspectos materiais ou de lazer. Utilizar um só indicador de medida para cada variável latente pode ser considerado discutível, até porque o indicador utilizado pode ser imperfeito. A análise factorial confirmatória (AFC) procura assim construir um modelo de medida para as variáveis latentes.

Recorde-se que no BHPS existem seis dimensões de satisfação diferentes, com base nas quais foram calculados os dois *scores* de satisfação (material e lazer) utilizados até agora. Podemos todavia considerar que as variáveis categoriais ordinais de satisfação com o rendimento do agregado familiar (*Sat2*), satisfação com a casa / habitação (*Sat3*) e satisfação com o emprego (*Sat5*) podem ser considerados indicadores de medida de uma variável latente contínua que podemos designar por *MATER* (percepções de satisfação com aspectos materiais). As variáveis categoriais ordinais de satisfação com a vida social (*Sat6*);

10 Introdução

satisfação com a quantidade de tempo livre (*Sat7*) e satisfação com a utilização que é feita desse tempo livre (*Sat8*) podem ser consideradas indicadores de medida da variável latente contínua *LAZER* (percepções de satisfação com aspectos de lazer). Assumindo que as duas dimensões de satisfação estão correlacionadas, o diagrama que representa o modelo de medida das variáveis latentes *MATER* e *LAZER* é o que se representa na Figura 1.4, considerando como momento temporal o ano de 2003.

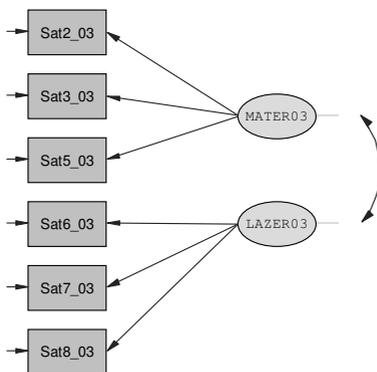


Figura 1.4 Exemplo de diagrama de AFC - modelo de medida de duas variáveis latentes.

1.3.4 Exemplo de um Modelo com Equações Estruturais Global

Se juntarmos o conjunto de relações estruturais postuladas no diagrama do modelo de análise de caminhos (Figura 1.3) com as relações postuladas no modelo de medida apresentado na Figura 1.4 obtemos um Modelo com Equações Estruturais global. A Figura 1.5 representa o diagrama do modelo global, que postula as percepções de satisfação com aspectos materiais num determinado momento temporal como

influenciadas pelo rendimento do agregado familiar e pelo sexo e idade do respondente, e a influenciar as percepções de satisfação com aspectos de lazer no momento temporal seguinte. Assume-se ainda que as percepções de satisfação com o lazer são negativamente influenciadas pelo número de horas de trabalho semanal do respondente.

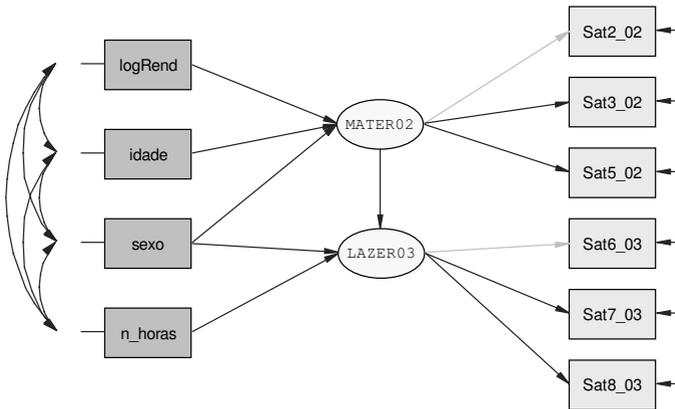


Figura 1.5 *Modelo com Equações Estruturais: exemplo de diagrama de modelo global.*

Perceber quais os pressupostos subjacentes a um modelo desta natureza, quais os métodos de estimação disponíveis, que tipo de problemas podem ocorrer em termos de estimação ou como avaliar a qualidade do ajustamento modelo-dados são objectivos deste livro.

1.3.5 Outros Exemplos Motivadores

Esta secção apresenta outros diagramas de caminhos de modelos utilizados nas áreas da Economia e Gestão e da Psicologia, com o intuito de mostrar que o grau de complexidade deste tipo de modelos pode ser elevado. Ao longo deste livro procuraremos mostrar como com-

12 Introdução

plexidade crescente do modelo postulado implica atenção redobrada do ponto de vista da construção, estimação e validação do modelo estatístico.

A Figura 1.6 representa o diagrama de um Modelo com Equações Estruturais com efeitos cruzados e desfasados entre quatro medidas repetidas de variáveis latentes de satisfação com aspectos materiais e com aspectos de lazer. Tal modelo foi apresentado por Salgueiro, M.F., Smith, P.W.F. e Vieira, M.D.T. em 2008 no Congresso Anual da SPE numa comunicação intitulada *Modelação de Trajectórias de Satisfação com dados do Bristish Household Panel Survey*. O modelo em causa será analisado em detalhe na Secção 3.2.5, em particular no que à modelação longitudinal de medidas repetidas diz respeito.

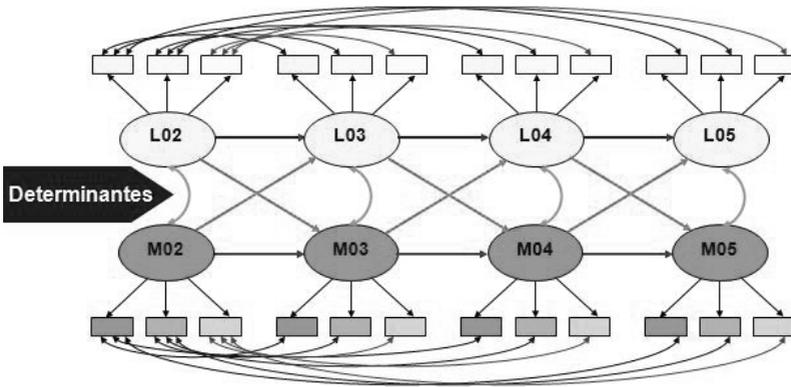


Figura 1.6 Modelo com Equações Estruturais com efeitos cruzados e desfasados entre medidas repetidas de variáveis latentes de satisfação.

A Figura 1.7 representa o diagrama de um modelo de Análise Factorial de Segunda Ordem para um constructo de Bem-Estar Psicológico (PM - *Psychological Morbidity*), medido indirectamente pelos 12 indicadores de três factores de primeira ordem: AD - *Anxiety and Depression*; SD - *Social Dysfunction* e LC - *Loss of Confidence*. Trata-se da escala do GHQ12 - *General Health Questionnaire*, analisada por

Malta (2009) com dados do *British Household Panel Survey*. Este tipo de modelos de análise factorial de segunda ordem, muito usado na Psicologia, será retomado na Secção 5.3.

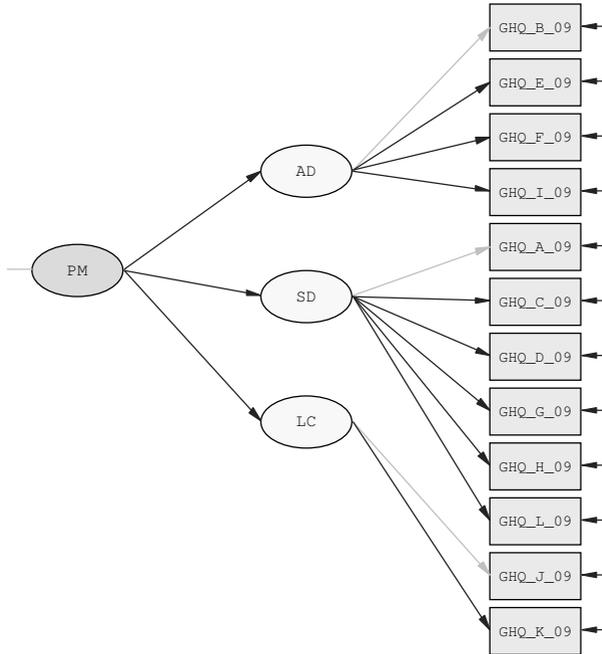


Figura 1.7 Modelo de análise factorial de segunda ordem, com 3 fatores de primeira ordem (AD, SD e LC) medidos por 12 indicadores.

A Figura 1.8 representa o diagrama de um Modelo com Equações Estruturais global com medidas repetidas e efeitos longitudinais. O modelo em causa foi proposto por Vitória et al. (2011) com o intuito de investigar o impacte da influência social na intenção de fumar e no comportamento tabágico dos jovens Portugueses, no âmbito do projecto *European Smoking prevention Framework Approach*. O modelo inclui múltiplos potenciais mediadores.

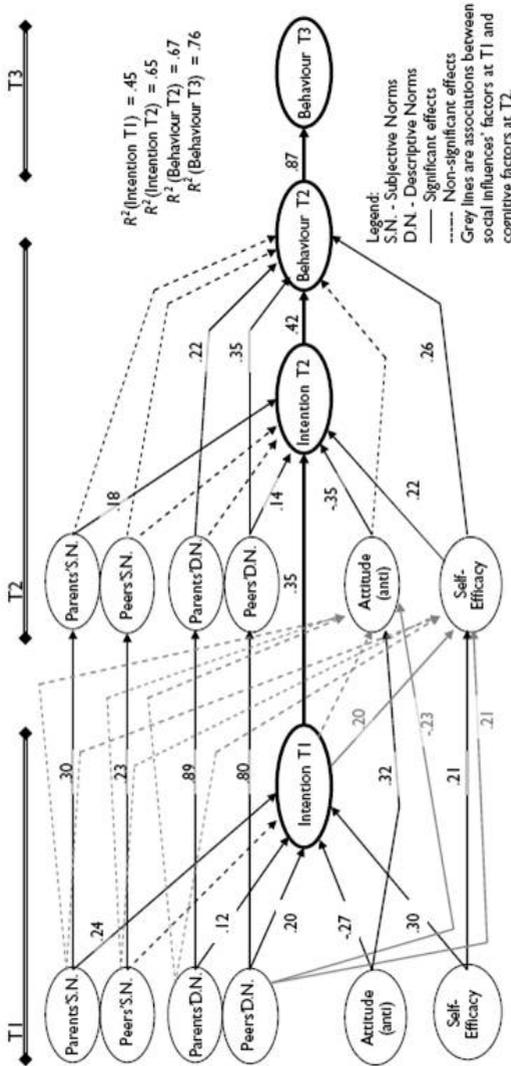


Figure 2. Path diagram of the longitudinal model proposed to investigate the impact of social influence on adolescents' intention to smoke and smoking behaviour.

Figura 1.8 Modelo com Equações Estruturais global com medidas repetidas e efeitos longitudinais.

1.4 Principais Conceitos e Notação

Por serem amplamente utilizados ao longo deste livro, apresentam-se em seguida alguns dos conceitos base (por ordem alfabética) e notação em Modelos com Equações Estruturais.

1.4.1 Conceitos Base

- **Constructo** - conceito que o investigador define em termos conceptuais e que não pode ser medido directamente.
- **Correlação Policórica** (*polychoric correlation*) é uma correlação entre duas variáveis latentes contínuas (subjacentes a duas variáveis manifestas ordinais) na sua distribuição normal bivariada (ver Olsson, 1979).
- **Correlação Polyserial** é uma correlação entre uma variável latente contínua (subjacente a uma variável manifesta ordinal) e uma variável manifesta métrica (ver Olsson et al., 1982).
- **Diagrama de Caminhos** (*Path Diagram*) - representação gráfica do conjunto de relações entre todas as variáveis do modelo. Apresenta-se na Figura 1.9 a simbologia usualmente utilizada num diagrama de caminhos.
- **Erro de Medida** - grau em que o indicador de medida observado não é explicado pela variável latente.
- **Erros de Especificação** - são erros que ocorrem aquando da especificação do modelo, como sejam o omitir variáveis importantes ou o incorrectamente pressupor que uma dada relação estrutural não existe. Na prática, porém, há mais modelos “errados” do que “certos”, pelo que o erro de especificação é muito comum, embora nem sempre reconhecido e usualmente não facilmente reconhecível.
- **Especificação do Modelo** - processo de selecção das variáveis a incluir no modelo, e do postular das relações entre elas.
- **Fiabilidade** (*Reliability*) - grau em que uma variável ou um

16 Introdução

conjunto de variáveis é consistente naquilo que é suposto medir. Difere do conceito de validade uma vez que não tem a ver com o que deve ser medido, mas sim com a forma como é medido.

- **Identificação** - um parâmetro diz-se identificado se puder ser escrito, de forma única, como função dos elementos da matriz de variância-covariância (ou de correlações) da população, ou de um outro parâmetro já identificado. O modelo diz-se identificado se todos os seus parâmetros estiverem identificados.
- **Indicador** - corresponde frequentemente a uma pergunta de um questionário e é operacionalizado numa variável manifesta (ou observada) usada para medir uma variável latente.
- **Índices de Modificação** - calculados para cada parâmetro não estimado no modelo especificado, mas passível de estimação. O seu valor indica a melhoria que se obtém no ajustamento global do modelo se o parâmetro em causa for sujeito a estimação. Podem ser usados para modificação / re-especificação do modelo. Devem ser usados com extremo cuidado, como guias orientadores, devendo apenas ser introduzidos no modelo parâmetros para os quais haja suporte teórico justificativo.
- **Limiares** (*Thresholds*) - Uma variável ordinal Y , com k categorias, pode ser encarada como uma medição imperfeita de uma variável contínua latente subjacente η , possuindo η uma distribuição normal padronizada. Os valores efectivos da variável Y podem ser arbitrários e são irrelevantes desde que a informação ordinal seja retida, isto é, *scores* baixos correspondem a categorias baixas de Y , associadas a baixos valores de η . A relação entre Y e η é tal que $Y = i$ é equivalente a $\alpha_{i-1} < \eta \leq \alpha_i$, sendo $\alpha_0 = -\infty$, $\alpha_k = +\infty$ e $\alpha_1 < \alpha_2 < \dots < \alpha_{k-1}$. Estes $k - 1$ valores α são chamados limiares ou *thresholds*. Seja n_j o número de casos da categoria j da variável ordinal. Os $k - 1$ limiares podem ser estimados a partir da distribuição marginal

como

$$\hat{\alpha}_i = \Phi^{-1} \left(\sum_{j=1}^i \frac{n_j}{N} \right), \quad \text{com } i = 1, 2, \dots, k-1$$

sendo Φ^{-1} o inverso da função de distribuição normal padronizada e N o número total de observações da variável ordinal Y . Os limiares são importantes para o cálculo das correlações policóricas e *polyserial*.

- **Modelo** - operacionalização da teoria; especificação de um conjunto de relações de dependência que podem ser testadas empiricamente. O modelo pode ser representado graficamente através de um diagrama e formalizado matematicamente através de um sistema de equações.
- **Modelo de Medida** - componente do Modelo com Equações Estruturais que especifica os indicadores de medida de cada variável latente.
- **Modelo Estrutural** - componente do Modelo com Equações Estruturais que estabelece o conjunto de relações de dependência entre as variáveis latentes do modelo. Com base na teoria, na experiência anterior e nos objetivos do estudo, o investigador define quais as variáveis independentes que vão influenciar cada uma das variáveis dependentes. O modelo estrutural expressa as relações entre variáveis independentes e dependentes, mesmo quando uma dependente se torna independente numa outra relação. As relações estruturais propostas correspondem frequentemente a hipóteses formuladas do ponto de vista teórico e substantivo.
- **Modelo Nulo** (*Null model*) - modelo usado com base para comparação e que não especifica qualquer relação entre as variáveis observadas, assumindo-as como independentes ou não correlacionadas.
- **Modelo Saturado** (*Saturated model*) - modelo com zero graus de liberdade, que inclui todas as possíveis relações entre variá-

18 Introdução

veis observadas.

- **Qualidade do Ajustamento** (*Goodness of Fit*) - grau em que o modelo especificado e estimado reproduz a estrutura de variância-covariância (ou de correlações) observada na amostra.
- **Validade** (*Validity*) - grau em que uma medida representa adequadamente o conceito em estudo. Importa pois perceber o que se pretende medir e tentar medi-lo de forma tão adequada ou correcta quanto possível. A validade tem a ver com a forma como o conceito é definido pelo indicador de medida, enquanto que a fiabilidade tem a ver com a consistência da medida.
- **Variável Latente** - é uma variável aleatória cujas realizações não observamos, por contraposição com variável manifesta em que as realizações são observadas. Uma variável latente pode ser usadas para representar um constructo hipotético. Não pode ser medida directamente, mas sim através de um ou mais indicadores (variáveis observadas ou manifestas).

1.4.2 Notação a Utilizar

O diagrama de caminhos (*path diagram*) representa o conjunto de relações entre todas as variáveis do modelo. Tal representação assenta na simbologia apresentada na Figura 1.9.

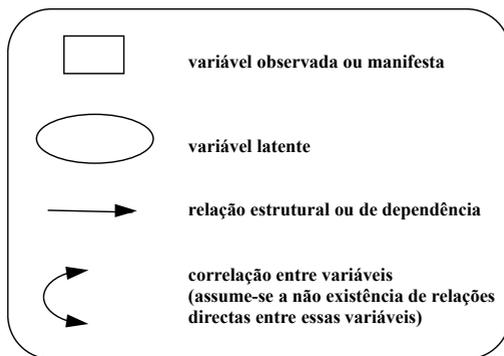


Figura 1.9 Simbologia utilizada em Diagramas de Caminhos

Dois pressupostos base estão subjacentes à construção do diagrama: i) todas as relações entre variáveis são indicadas (e o objectivo é modelar todas essas relações com o menor número possível de setas que possam ser teoricamente justificadas); ii) todas as relações são consideradas lineares. Para uma exposição mais detalhada sobre as regras propostas por Wright para a construção e interpretação de diagramas de caminhos ver Loehlin (2004, capítulo 1).

A notação que irá ser usada foi desenvolvida por Jöreskog, Keesling e Wiley e foi popularizada pelo LISREL, ao ponto de muitos autores a usarem, referindo-se a ela como sendo a notação LISREL. Trata-se de uma notação matricial, em que letras maiúsculas designam matrizes e letras minúsculas os elementos dessas matrizes. São usadas letras gregas para os parâmetros e variáveis latentes, e letras romanas para as variáveis observadas.

As variáveis do modelo são assim designadas:

- ξ - (ksi) variável latente independente
- η - (eta) variável latente dependente
- X - indicador de medida de ξ
- Y - indicador de medida de η
- δ - (delta) erro de medida de X
- ϵ - (epsilon) erro de medida de Y
- ζ - (zeta) termo residual da equação estrutural

As variáveis do modelo integram os seguintes vectores:

- no modelo de medida
 - \mathbf{X} vector das q variáveis observadas independentes X
 - \mathbf{Y} vector das p variáveis observadas dependentes Y
 - $\boldsymbol{\delta}$ vector dos erros de medida das q variáveis X
 - $\boldsymbol{\epsilon}$ vector dos erros de medida das p variáveis Y

20 Introdução

- no modelo estrutural

ξ vector das n variáveis latentes independentes ξ

η vector das m variáveis latentes dependentes η

ζ vector dos m termos residuais das relações estruturais

Para além dos vectores das variáveis, o modelo global requer ainda um conjunto de matrizes associadas ao **modelo estrutural**, constantes da Tabela 1.1, e um outro conjunto de matrizes associadas ao **modelo de medida**, constantes da Tabela 1.2.

matriz	descrição	notação elemento
B	coeficientes das relações directas entre variáveis (latentes) dependentes	β_{mm}
Γ	coeficientes das relações directas entre variáveis (latentes) independentes e dependentes	γ_{mn}
Φ	variâncias e covariâncias das variáveis (latentes) independentes	ϕ_{nn}
Ψ	variâncias e covariâncias dos termos residuais das relações estruturais	ψ_{mm}

Tabela 1.1 *Matrizes do Modelo Estrutural*

matriz	descrição	notação elemento
Λ_X	coeficientes de regressão de X sobre ξ	λ_{qn}^x
Λ_Y	coeficientes de regressão de Y sobre η	λ_{pm}^y
Θ_δ	variâncias e covariâncias dos erros de medida δ	$\theta\delta_{qq}$
Θ_ϵ	variâncias e covariâncias dos erros de medida ϵ	$\theta\epsilon_{pp}$
$\Theta_{\delta\epsilon}$	covariâncias entre os δ e os ϵ	$\theta\delta\epsilon_{qp}$

Tabela 1.2 *Matrizes do Modelo de Medida*

1.5 Software Estatístico Disponível e Páginas Web

A implementação em computador, por parte de Jöreskog e Sörbom (1993 a, b) dos modelos com relações estruturais lineares, conhecida por **LISREL** (*L*inear *S*tructural *R*ELations), tornou-se standard na área, ao ponto de ter imposto a sua notação e de se ter praticamente tornado sinónimo de Modelos com Equações Estruturais.

A versão comercial mais recente do LISREL é a versão 8.8, de Julho de 2006, existindo ainda uma *student version*, grátis que se pode descarregar da página *web* da empresa que comercializa o LISREL: www.ssicentral.com/lisrel/index.html. As principais limitações da versão de estudante face à versão comercial são o tratamento de um número máximo de 15 variáveis manifestas e a existência de algumas limitações na importação de ficheiros de outros pacotes estatísticos.

Para além do LISREL existem disponíveis:

- o **EQS** de Bentler, inicialmente considerado o principal concorrente do LISREL, actualmente na versão 6.1:
www.mvsoft.com/index.htm ;
- o **AMOS**, comercializado pela empresa que comercializa o SPSS:
www-01.ibm.com/software/analytics/spss/products/statistics/amos/. Note-se que a ênfase deste software é posta na construção do modelo a partir do diagrama conceptual. Sendo uma lógica atraente para utilizadores com menor preparação estatística ou matemática, pode tornar-se uma lógica “perigosa” em termos de utilização da ferramenta de modelação estatística;

22 Introdução

- o **Mplus**, de Bengt Muthén e Linda Muthén. Bengt Muthén foi antigo aluno e colaborador directo de Jöreskog e pedra-chave na fase inicial do LISREL. O Mplus está actualmente na versão 6.12 e pretende ser uma plataforma mais ampla de modelação estatística, da qual os Modelos com Equações Estruturais sejam um caso particular (www.statmodel.com/). Nos últimos anos tem sido o software que mais tem acompanhado, incorporado e até contribuído para os desenvolvimentos estatísticos na área dos Modelos com Equações Estruturais, tendo sido crescente o número de novos utilizadores;
- Para os utilizadores de R existem basicamente disponíveis três *packages*:
 - **sem**, desenvolvido por John Fox (ver Fox, 2006);
 - **OpenMx** (<http://openmx.psyc.virginia.edu/>), desenvolvido por Boker e colaboradores (ver Boker et al., 2012). Dizem os autores que o *OpenMx is a free open source software for fitting Structural Equation Models to observed data. OpenMx offers the features you would expect in an SEM software package. But OpenMx works in ways that will make your modeling jobs easier and will allow you to do things that other SEM packages don't*;
 - **lavaan**, desenvolvido por Yves Rosseel (ver Rosseel, 2012). Nas palavras do autor *the lavaan R package is developed to provide useRs, researchers and teachers a free open-source, but commercial-quality package for latent variable modeling. The long-term goal of lavaan is to implement all the state-of-the-art capabilities that are currently available in commercial packages. However, lavaan is still under development, and much work still needs to be done.*

As seguintes páginas web são de particular interesse e de consulta recomendada:

- SEMNET (SEM discussion list):
www.gsu.edu/~mkteer/semnet.html
- Home page de Ed. Rigdon (Prof. de Marketing, na Georgia State University) - www.gsu.edu/~mkteer/index.html
e página de perguntas frequentes (FAQ):
www2.gsu.edu/~mkteer/semfaq.html
- LISREL page - University of Texas:
<http://www.utexas.edu/cc/stat/software/lisrel/>
e também <http://ssc.utexas.edu/software>

1.6 Referências Bibliográficas: Breves Recomendações

Bollen (1989a) é a referência clássica chave, e certamente a mais citada, para a compreensão da modelação estatística com recurso a Modelos com Equações Estruturais. No dizer do autor *this book provides a comprehensive introduction to the general structural equation system, commonly known as the "LISREL model"*.

Brown (2006) é um bom ponto de partida para Modelos de Análise Factorial Confirmatória, sobretudo para investigadores com maior sensibilidade à área da Psicologia. Para investigadores com maior familiaridade com a área do Marketing, Diamantopolous e Siguaw (2000) é uma possível introdução aos Modelos com Equações Estruturais. Schumacker e Lomax (2010) é um livro que deve merecer particular atenção, apresentando muitos exemplos simples, com comandos de sintaxe em LISREL e outputs anotados. A edição de 2010 é a terceira, e os autores usam exclusivamente o LISREL. A segunda edição, de 2004, pode ser muito útil a quem estiver interessado em exemplos em AMOS e EQS.

24 Introdução

Para “dicas” de como escrever sobre Modelos com Equações Estruturais ver Hoyle (1995, capítulo 9). Para recomendações de como apresentar os resultados de um estudo com Análise Factorial Confirmatória ver Brown (2006, página 144). O capítulo 11 de Schumacker e Lomax (2010) apresenta orientações e recomendações sobre como reportar os resultados da modelação realizada com recurso a Modelos com Equações Estruturais. A literatura existente é muitíssimo vasta, espelhando assim a abrangência, as potencialidades de aplicação, a motivação para a investigação teórica, o interesse por parte dos investigadores e também a actualidade da temática que este livro procura apresentar.

Capítulo 2

Da Análise Factorial Exploratória à Análise Factorial Confirmatória

Uma síntese dos principais marcos na história da Análise Factorial pode ser encontrada em Cudeck et al. (2001, capítulo 2).

Em termos de históricos, Francis Galton (1822 - 1911) pode ser considerado o precursor directo da Análise Factorial e Charles Spearman (1863 - 1945) o “pai”. Com efeito, depois dos conceitos de correlação e regressão, nos finais dos anos 80 do século XIX Galton escreveu dois artigos que podem ser considerados o “embrião” da Análise Factorial. Spearman, um psicólogo com grande aptidão estatística, investigou um conjunto de variáveis correspondentes a testes de aptidão mental. Spearman concluiu que as pessoas com bom desempenho num teste tendiam a ter também um bom desempenho nos outros testes de aptidão mental, o que o levou a concluir que os *scores* dos indivíduos eram manifestações de uma aptidão geral subjacente, que designou por *general intelligence*. Em 1904 Spearman propõe um “modelo” de Análise Factorial com um só factor: a inteligência global. Contudo o autor não falou em “modelo”, aliás o conceito de modelo estatístico não estava ainda estabelecido nessa altura.

Louis Thurstone (1887 - 1955) levou a cabo um conjunto exaustivo de análises estatísticas, a que hoje poderíamos chamar análise factorial exploratória. Um marco importante no desenvolvimento teórico da Análise Factorial é o artigo de Anderson e Rubin (1956), considerado uma “*an unified exposition from the viewpoint of the mathematical statistician*”. Lawley e Maxwell (1971) pode ser considerado

26 Da Análise Factorial Exploratória à Análise Factorial Confirmatória

o primeiro livro sobre Análise Factorial no âmbito da tradição estatística, constituindo ainda hoje uma referência ao nível do modelo clássico de análise factorial (assumindo normalidade), em particular no que à estimação e teste de hipóteses diz respeito. Para uma exposição matemática detalhada do modelo clássico de análise factorial ver Bartholomew et. al. (2008, 2011); Everitt (1984); Kaplan (2000) ou Loehlin (2004).

Um contributo substancial para o desenvolvimento da Análise Factorial Confirmatória foi dado por Karl Joreskog (1935 -) e seus colaboradores. Para uma introdução à análise factorial confirmatória *versus* análise factorial exploratória ver Long (1983a). Para uma introdução rotulada pela autora de “não matemática” à AFC em LISREL ver Byrne (1989). Para uma apresentação detalhada do modelo de AFC escrita por um professor de um Departamento de Psicologia, com exemplos de aplicação, correspondentes comandos de sintaxe LISREL e Mplus, e apresentação de outputs comentados, ver Brown (2006). Para uma síntese das potencialidades do PRELIS em termos de Análise Factorial Exploratória ver Jöreskog et al. (2001, páginas 143–161).

A estrutura do presente capítulo é a seguinte: numa primeira secção é apresentado o modelo clássico de análise factorial exploratória, com pressupostos, equações e referência à estimação, sendo feita a ilustração com recurso a um exemplo em SPSS e em PRELIS. Numa segunda secção é apresentado o modelo de análise factorial confirmatória (pressupostos, equações, regras para ajudar à identificação do modelo), sendo depois discutida a temática da validade e fiabilidade dos indicadores e das escalas. O capítulo termina com a apresentação de um exemplo de AFC em LISREL.

2.1 O Modelo Clássico de Análise Factorial Exploratória

A Análise Factorial é o mais antigo e o mais usado modelo com variáveis latentes. O modelo clássico supõe que todas as variáveis (manifestas e latentes) são métricas. O objectivo é derivar novas variáveis (os factores ou variáveis latentes) por forma a possibilitar uma melhor compreensão da estrutura de variâncias e covariâncias das variáveis manifestas.

2.1.1 Análise Factorial *versus* Análise em Componentes Principais

Enquanto que a Análise em Componentes Principais é um *método descritivo* que produz uma transformação ortogonal das variáveis que não depende de qualquer modelo, e que visa explicar o máximo possível da variância das variáveis iniciais, a Análise Factorial é uma *técnica baseada em modelo*, que visa explicar / reproduzir a estrutura de variâncias e covariâncias das variáveis iniciais, ainda que reduzindo a dimensionalidade dos dados. Sendo uma técnica estatística que tem na sua base um modelo, a Análise Factorial tem *pressupostos* acerca da distribuição conjunta das variáveis iniciais na população, pressupostos esses que têm que ser verificados. Será então possível quantificar a qualidade do ajustamento modelo-dados e fazer inferência para a população de onde foi extraída a amostra.

2.1.2 Pressupostos e Equações do Modelo Clássico de Análise Factorial

Começemos por estudar a situação mais simples de um modelo com um só factor. Sejam X_1, X_2, \dots, X_p variáveis manifestas que se acredita medirem um só factor latente L . A forma mais simples de ex-

28 Da Análise Factorial Exploratória à Análise Factorial Confirmatória

pressar a regressão de cada variável X_i sobre L é usar um modelo linear:

$$X_i = \alpha_{i0} + \alpha_{i1}L + e_i \quad (i = 1, 2, \dots, p)$$

em que L é o factor comum a todas as variáveis manifestas X , os termos residuais e_i são os factores únicos ou específicos e α_{i1} representa os pesos factoriais.

O modelo clássico unifactorial pressupõe que:

- os e_i são independentes e $e_i \sim N(0, \sigma_i^2)$ (com $i = 1, 2, \dots, p$);
- $e_i \perp L$ (os factores específicos são independentes do factor latente);
- as variáveis manifestas X_i são condicionalmente independentes, dado o factor latente L .

Quando o factor latente L tem média zero e desvio-padrão 1, em termos teóricos,

$$\text{cov}(X_i, X_t) = \alpha_{i1}\alpha_{t1}, \quad \text{com } i \neq t, \text{ de 1 a } p.$$

Quando se procede à estimação do modelo obtêm-se as estimativas $\hat{\alpha}_{i1}$ e $\hat{\alpha}_{t1}$. O modelo terá um bom ajustamento se as estimativas $\hat{\alpha}_{i1}$ e $\hat{\alpha}_{t1}$ assumirem valores próximos dos verdadeiros valores dos parâmetros α_{i1} e α_{t1} .

Estes resultados podem ser facilmente generalizados para a situação de p variáveis manifestas e q variáveis latentes. O **modelo geral factorial linear** pode ser escrito como:

$$X_i = \alpha_{i0} + \alpha_{i1}L_1 + \alpha_{i2}L_2 + \dots + \alpha_{iq}L_q + e_i \quad (i = 1, 2, \dots, p)$$

L_1, L_2, \dots, L_q são os factores comuns ou variáveis latentes, que se assumem com média zero e desvio-padrão um, os e_i são os termos residuais e $\alpha_{i1}, \dots, \alpha_{iq}$ são os pesos factoriais. Se as variáveis X estiverem padronizadas, os $\alpha_{i1}, \dots, \alpha_{iq}$ podem ser interpretados como as correlações entre os factores latentes e as variáveis manifestas padronizadas, e desempenham papel chave na interpretação dos factores.

Os **pressupostos** do modelo clássico de Análise Factorial com dois ou mais factores são os seguintes:

- os termos residuais e_i são iid (independentes e identicamente distribuídos), sendo que $e_i \sim N(0, \sigma_i^2)$ (com $i = 1, 2, \dots, p$);
- os e_i não estão correlacionados com as variáveis latentes;
- $\mathbf{L} \sim N_q(\mathbf{0}, \mathbf{I})$ (i.e., as q variáveis latentes têm distribuição normal multivariada, com média zero, desvio padrão um e não estão correlacionadas entre si).

Em termos teóricos:

- $E(X_i) = \alpha_{i0}$;
- $var(X_i) = \alpha_{i1}^2 + \alpha_{i2}^2 + \dots + \alpha_{iq}^2 + \sigma_i^2$ ($i = 1, \dots, p$), sendo σ_i^2 a variância residual ou específica, e $\alpha_{i1}^2 + \alpha_{i2}^2 + \dots + \alpha_{iq}^2$ a **comunalidade** de X_i (variância de X_i explicada pelos factores comuns);
- $cov(X_i, X_k) = \alpha_{i1}\alpha_{k1} + \alpha_{i2}\alpha_{k2} + \dots + \alpha_{iq}\alpha_{kq}$ ($i \neq k$, de 1 a p);
- é prática comum estandardizar as variáveis manifestas X antes de fazer a Análise Factorial (em termos práticos basear a análise na matriz de correlações), uma vez que facilita a interpretação dos resultados. Note-se que, nesse caso, $E(X_i) = \alpha_{i0} = 0$.

2.1.3 Estimação e Ajustamento do Modelo Clássico de Análise Factorial

Existem vários métodos de estimação disponíveis, ainda que com pressupostos diferentes. Qualquer dos métodos começa por construir uma medida de distância entre a matriz de variância-covariância (ou de correlações) observada e a matriz implícita ao modelo postulado. O objectivo da estimação é encontrar estimativas para os parâmetros por forma a minimizar essa medida de distância.

Para ajudar a decidir quantos factores considerar e para aferir da adequação do modelo especificado e estimado aos dados, deve ser

30 Da Análise Factorial Exploratória à Análise Factorial Confirmatória

usado um **teste de qualidade de ajustamento** modelo-dados. Assumindo normalidade, a hipótese nula em teste é a de que a matriz de variância-covariância das variáveis manifestas tem a forma especificada pelo modelo factorial. Não rejeitar esta hipótese significa que o modelo se ajusta aos dados. Sob H_0 , a estatística de teste tem distribuição χ^2 com $[(p - q)^2 - (p + q)]/2$ graus de liberdade. Se um modelo com um determinado número de factores tiver fraco ajustamento devem introduzir-se mais factores, desde que haja graus de liberdade para o fazer, sem contudo perder de vista o equilíbrio entre qualidade do ajustamento e interpretabilidade do modelo.

Para modelos com mais de um factor a solução encontrada não é única. *Rotação* é o nome que se dá ao processo pelo qual se passa de uma solução para outra. Uma rotação *ortogonal* mantém os factores latentes não correlacionados. A rotação permite clarificar a estrutura subjacente aos factores e facilitar a interpretação. Obter uma *estrutura simplificada* dos pesos factoriais é obter uma solução em que cada variável tem uma elevada correlação com um só factor e correlações baixas com os restantes factores. Ao fazer a rotação os softwares estatísticos procuram, de entre o conjunto de soluções possíveis para o modelo, aquela que mais se aproxima de uma estrutura simplificada. A rotação não altera o ajustamento do modelo: apenas redistribui a variância explicada pelos factores e altera os pesos factoriais, pelo que a interpretação dos factores poderá ser diferente.

2.1.4 Exemplo de um Modelo de Análise Factorial Exploratória em SPSS e em PRELIS

Retomemos os dados do BHPS apresentados na Secção 1.3, e mais especificamente as seis percepções de satisfação com as condições de vida medidas em escalas de tipo Likert de 1 (nada satisfeito) a 7 (completamente satisfeito) que, abusivamente, trataremos como variáveis quantitativas para efeitos de análise factorial exploratória. Utilizaremos as respostas válidas dos 2008 inquiridos, relativas ao ano de

2003:

	nome da variável	designação da variável - Satisfação com:
X_1	<i>Sat2_03</i>	rendimento do agregado familiar
X_2	<i>Sat3_03</i>	casa / habitação
X_3	<i>Sat5_03</i>	emprego
X_4	<i>Sat6_03</i>	vida social
X_5	<i>Sat7_03</i>	quantidade de tempo livre
X_6	<i>Sat8_03</i>	utilização do tempo livre

Para proceder à análise exploratória prévia das seis variáveis em SPSS importa recorrer ao menu *Analyze / Descriptive Statistics / Explore* e pedir histogramas, *boxplots*, *Q-Q plots* e testes à normalidade univariada de cada uma das variáveis. Note-se que o SPSS não disponibiliza um teste à normalidade multivariada. Recorde-se que a normalidade univariada de cada uma das seis variáveis, de per si, não garante a normalidade multivariada da distribuição conjunta das seis variáveis, bastando todavia não se verificar a normalidade univariada de uma das variáveis para impedir a normalidade multivariada da distribuição conjunta das variáveis de se verificar.

A Análise Factorial a realizar tem como objectivo diminuir a dimensionalidade da estrutura de percepções de satisfação. Será primeiro testado um modelo unifactorial, e em seguida um modelo com dois factores. Os comandos de SPSS a executar para obter o modelo de análise factorial exploratória com dois factores de satisfação apresentam-se no Apêndice A.1, assim como os principais outputs obtidos.

A rejeição da hipótese nula de que o modelo unifactorial reproduz a matriz de variância-covariância das variáveis manifestas leva a considerar um modelo com 2 factores, não havendo no caso concreto graus de liberdade suficientes para estimar um modelo com três factores. O valor de significância obtido para o modelo bi-factorial leva à rejeição da hipótese nula em teste (pressupondo normalidade multivariada da distribuição conjunta das seis variáveis de satisfação, que, pela natu-

32 Da Análise Factorial Exploratória à Análise Factorial Confirmatória

reza da escala das variáveis, não se poderá verificar ...). Poder-se-á também questionar a ortogonalidade dos dois factores, até agora assumida. Do Apêndice A.1, e em termos exploratórios, é possível concluir que um dos factores está mais correlacionado com as variáveis *Sat2*, *Sat3* e *Sat5* (de satisfação com aspectos materiais), estando o outro factor mais correlacionado com as três variáveis de satisfação com aspectos de lazer (*Sat6*, *Sat7* e *Sat8*).

O PRELIS pode também ser utilizado para estimar um modelo clássico de Análise Factorial. Os comandos de PRELIS utilizados e demais outputs obtidos apresentam-se no Apêndice A.2. Comentam-se aqui os principais resultados, comparando-os com os anteriormente obtidos em SPSS.

No menu da Análise Factorial Clássica do PRELIS a escolha do número de factores é opcional. Se o utilizador não decidir quantos factores pretende no modelo, o próprio software decide quantos factores reter, com base no teste de qualidade de ajustamento apresentado.

Normal Maximum Likelihood Factor Analysis

Decision Table for Number of Factors

Factors	Chi2	df	P	DChi2	Ddf	PD	RMSEA
0	4307.06	15	0.000				0.377
1	371.22	9	0.000	3935.85	6	0.000	0.142
2	22.64	4	0.000	348.57	5	0.000	0.048
3	0.00	0	1.000	22.64	4	0.000	0.000

No exemplo em análise, na tabela obtida para ajudar a decidir quantos factores considerar no modelo, não chega a ser atingido um valor- p superior a 5% no teste à hipótese nula de que o modelo com um (ou com dois) factores reproduz a estrutura de associações na população. Não há graus de liberdade suficientes para considerar um modelo com 3 factores, escolhendo o PRELIS uma solução com dois factores (com menor valor de RMSEA do que a solução unifactorial).

Os pesos factoriais obtidos antes da rotação são idênticos aos obtidos em SPSS:

Normal Maximum Likelihood Factor Analysis with 2 Factors

Unrotated Factor Loadings

	Factor 1	Factor 2	Unique Var
	-----	-----	-----
Sat2_03	0.551	0.556	0.388
Sat3_03	0.425	0.304	0.727
Sat5_03	0.472	0.300	0.687
Sat6_03	0.824	-0.131	0.304
Sat7_03	0.774	-0.094	0.392
Sat8_03	0.845	-0.188	0.250

Em vez das communalidades, o PRELIS mostra as componentes únicas ou específicas de variância de cada uma das variáveis iniciais. Recorde-se que $\text{var}(X_i) = \sigma_i^2 + \text{comunalidade de } X_i$. Como estamos a trabalhar com uma matriz de correlações as variâncias são todas unitárias, sendo a componente específica de variância de X_i dada por $\hat{\sigma}_i^2 = 1 - (\hat{\alpha}_{i1}^2 + \hat{\alpha}_{i2}^2)$. Os pesos factoriais após rotação Varimax são também idênticos aos do SPSS.

Varimax-Rotated Factor Loadings

	Factor 1	Factor 2	Unique Var
	-----	-----	-----
Sat2_03	0.196	0.758	0.388
Sat3_03	0.213	0.477	0.727
Sat5_03	0.257	0.497	0.687
Sat6_03	0.777	0.302	0.304
Sat7_03	0.716	0.310	0.392
Sat8_03	0.825	0.264	0.250

O PRELIS mostra ainda uma solução com uma rotação oblíqua PROMAX, em que se possibilita aos dois factores estarem correlacionados, apresentando uma estimativa de correlação de 0.6.

Promax-Rotated Factor Loadings

	Factor 1	Factor 2	Unique Var
	-----	-----	-----
Sat2_03	0.823	-0.070	0.388
Sat3_03	0.482	0.062	0.727
Sat5_03	0.489	0.106	0.687
Sat6_03	0.053	0.801	0.304
Sat7_03	0.086	0.725	0.392
Sat8_03	-0.011	0.872	0.250

Factor Correlations		
	Factor 1	Factor 2
	-----	-----
Factor 1	1.000	
Factor 2	0.601	1.000

A hipótese de os dois factores estarem correlacionados, bem como a hipótese de diversos pesos factoriais serem nulos, serão testadas no âmbito da Análise Factorial Confirmatória.

2.2 Análise Factorial Confirmatória - Modelo de Medida

A Análise Factorial Confirmatória (modelo de medida) pode ser considerada um submodelo do Modelo com Equações Estruturais Global.

2.2.1 Pressupostos e Equações do Modelo de Análise Factorial Confirmatória

O sistema de equações do modelo de medida é dado por:

$$\mathbf{X} = \mathbf{\Lambda}_X \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (2.1)$$

em que $\mathbf{X}^T = (X_1, X_2, \dots, X_q)$ é o vector das q variáveis manifestas ou observadas, $\boldsymbol{\xi}^T = (\xi_1, \xi_2, \dots, \xi_n)$ é o vector das n variáveis latentes e $\boldsymbol{\delta}^T = (\delta_1, \delta_2, \dots, \delta_q)$ é o vector das q variáveis residuais (erro de medida das variáveis \mathbf{X}).

Os pressupostos do modelo são os seguintes:

- as variáveis observadas \mathbf{X} são medidas como desvios em relação à média, i.e., $E[\mathbf{X}] = \mathbf{0}$;
- $\boldsymbol{\xi}$ e $\boldsymbol{\delta}$ são vectores de variáveis aleatórias com média zero, i.e., $E[\boldsymbol{\delta}] = E[\boldsymbol{\xi}] = \mathbf{0}$;

- as variáveis residuais δ não estão correlacionadas com as variáveis latentes ξ , i.e., $E[\xi\delta^T] = E[\delta\xi^T] = \mathbf{0}$.

A matriz de variância-covariância implícita é dada por:

$$\Sigma_{\mathbf{X}\mathbf{X}} = \Lambda_{\mathbf{X}}\Phi\Lambda_{\mathbf{X}}^T + \Theta_{\delta}, \quad (2.2)$$

sendo $\Sigma_{\mathbf{X}\mathbf{X}}$ a matriz de variância-covariância das variáveis observadas \mathbf{X} , Φ a matriz de variância-covariância das variáveis latentes ξ e Θ_{δ} a matriz de variância-covariância dos termos residuais δ .

Numa solução estandardizada, as variáveis latentes ξ têm variância unitária e Φ passa a ser uma matriz de correlações. Se, para efeitos de identificação, for necessário pressupor que as variáveis latentes não estão correlacionadas, Φ é uma matriz identidade.

É fácil demonstrar que a Equação 2.2 define a matriz de variância-covariância do modelo de análise factorial confirmatória definido pela Equação 2.1:

$$\begin{aligned} \Sigma_{\mathbf{X}\mathbf{X}} &= \text{var}(\mathbf{X}) = E[\mathbf{X}\mathbf{X}^T] = E\left[(\Lambda_{\mathbf{X}}\xi + \delta)(\Lambda_{\mathbf{X}}\xi + \delta)^T\right] \\ &= E\left[(\Lambda_{\mathbf{X}}\xi + \delta)(\xi^T\Lambda_{\mathbf{X}}^T + \delta^T)\right] \\ &= E\left[\Lambda_{\mathbf{X}}\xi\xi^T\Lambda_{\mathbf{X}}^T\right] + E\left[\Lambda_{\mathbf{X}}\xi\delta^T\right] + E\left[\delta\xi^T\Lambda_{\mathbf{X}}^T\right] + E\left[\delta\delta^T\right] \\ &= \Lambda_{\mathbf{X}}E\left[\xi\xi^T\right]\Lambda_{\mathbf{X}}^T + \Lambda_{\mathbf{X}}E\left[\xi\delta^T\right] + E\left[\delta\xi^T\right]\Lambda_{\mathbf{X}}^T + E\left[\delta\delta^T\right] \\ &= \Lambda_{\mathbf{X}}\Phi\Lambda_{\mathbf{X}}^T + \mathbf{0} + \mathbf{0} + \Theta_{\delta} = \Lambda_{\mathbf{X}}\Phi\Lambda_{\mathbf{X}}^T + \Theta_{\delta} \end{aligned}$$

2.2.2 Identificação do Modelo de Análise Factorial Confirmatória

O modelo de Análise Factorial Confirmatória definido pelas Equações 2.1 e 2.2 não está identificado a não ser que se imponham restrições sobre parâmetros de $\Lambda_{\mathbf{X}}$, Φ ou Θ_{δ} . Prova-se que, se houver um conjunto de parâmetros que satisfaçam as equações do modelo, haverá um número infinito de conjuntos que as satisfazem.

36 Da Análise Factorial Exploratória à Análise Factorial Confirmatória

Seja M uma matriz $n \times n$, invertível. Seja $\ddot{\Lambda}_X = \Lambda_X M^{-1}$, $\ddot{\xi} = M\xi$ e $\ddot{\Phi} = M\Phi M^T$. Tanto as matrizes Λ_X e Φ como as matrizes $\ddot{\Lambda}_X$ e $\ddot{\Phi}$ satisfazem as Equações 2.1 e 2.2. Com efeito,

$$\ddot{\Lambda}_X \ddot{\xi} + \delta = (\Lambda_X M^{-1})(M\xi) + \delta = \Lambda_X (M^{-1}M)\xi + \delta = \Lambda_X \xi + \delta.$$

Logo, se $X = \Lambda_X \xi + \delta$, também $X = \ddot{\Lambda}_X \ddot{\xi} + \delta$. Por outro lado,

$$\begin{aligned} \ddot{\Lambda}_X \ddot{\Phi} \ddot{\Lambda}_X^T + \Theta_\delta &= (\Lambda_X M^{-1})(M\Phi M^T)(\Lambda_X M^{-1})^T + \Theta_\delta \\ &= \Lambda_X (M^{-1}M)\Phi(M^T(M^T)^{-1})\Lambda_X^T + \Theta_\delta \\ &= \Lambda_X \Phi \Lambda_X^T + \Theta_\delta. \end{aligned}$$

Logo, se $\Sigma_{XX} = \Lambda_X \Phi \Lambda_X^T + \Theta_\delta$, também $\Sigma_{XX} = \ddot{\Lambda}_X \ddot{\Phi} \ddot{\Lambda}_X^T + \Theta_\delta$. Uma vez que as matrizes "só igualam as matrizes originais se $M = I$, cada uma das matrizes M , invertível, fornece uma solução igualmente satisfatória para o modelo, isto é, o modelo não está identificado.

Impôr restrições adicionais ao modelo, para o tornar identificado, significa o seguinte:

- uma restrição sobre um elemento de Λ_X significa que a variável manifesta correspondente não depende / não é indicador de medida de um dado factor latente;
- uma restrição sobre um elemento de Φ significa que as variáveis latentes correspondentes não estão correlacionadas;
- uma restrição sobre um elemento de Θ_δ significa considerar não existir erro de medida (se é elemento da diagonal principal) ou não permitir correlação entre erros de medida (se é elemento fora da diagonal principal).

O modelo está identificado se for possível mostrar que todos os elementos do vector de parâmetros a estimar são funções únicas das variâncias e covariâncias populacionais das variáveis X . Estabelecer do ponto de vista algébrico a identificação do modelo pode tornar-se um processo extremamente difícil, embora seja o único que a garante.

Do ponto de vista prático há um conjunto de **regras de identificação do modelo de medida** que podem ser utilizadas, a saber:

- Regra- t
 - é uma condição necessária, não é uma condição suficiente;
 - requer $t \leq \frac{1}{2}q(q + 1)$, isto é, o número de parâmetros a estimar tem que ser menor ou igual que o número de elementos únicos da matriz de variância-covariância $\Sigma_{\mathbf{X}\mathbf{X}}$.
- Regra dos 3 indicadores
 - não é condição necessária, é condição suficiente;
 - requer:
 - * $n \geq 1$ (uma ou mais variáveis latentes);
 - * cada linha de $\Lambda_{\mathbf{X}}$ com um e só um elemento não nulo;
 - * três ou mais indicadores por factor;
 - * Θ_{δ} diagonal.
- Regra dos 2 indicadores
 - não é condição necessária, é condição suficiente;
 - requer:
 - * $n > 1$;
 - * Θ_{δ} diagonal;
 - * dois ou mais indicadores por factor;
 - * cada linha de $\Lambda_{\mathbf{X}}$ com um e só um elemento não nulo;
 - * não haver elementos nulos em Φ , ou cada linha de Φ ter pelo menos um elemento não nulo fora da diagonal principal.

2.2.3 Validade e Fiabilidade dos Indicadores e das Escalas

Para uma exposição detalhada sobre *Classical Test Theory* ver Blunch (2008, capítulo 2), Mueller (1996, páginas 110–121) ou Brown (2006,

páginas 337–351). Os conceitos de **validade** (*validity*) e de **fiabilidade** (*reliability*) são considerados chave na avaliação de um instrumento de medida. Um instrumento diz-se fiável se, repetindo a medição em condições idênticas, se obtiver um resultado muito similar. Um instrumento diz-se válido se medir aquilo que efectivamente é suposto medir, e nada mais.

Muitas vezes, em vez de utilizar técnicas de Análise Factorial Confirmatória, os investigadores constróiem as designadas *summated scales*, através da soma das respostas dos inquiridos aos diversos itens que constituem uma escala, cada um deles frequentemente medido numa escala de tipo Likert de 5 ou de 7 pontos. O investigador espera assim obter maior diferenciação nas medições, estando todavia a dar idêntica importância ou peso a cada um dos itens da escala. Sejam X_1, X_2, \dots, X_k os k itens medidos em escala de tipo Likert. $L = X_1 + X_2 + \dots + X_k$ designa a *summated scale* dos k itens.

Uma das propostas da literatura para medir a fiabilidade das *summated scales* é o coeficiente Alfa de Cronbach, proposto por Cronbach (1951), e dado por

$$\alpha = \frac{k}{k-1} \times \frac{\sigma_L^2 - \sum_{i=1}^k \sigma_i^2}{\sigma_L^2},$$

sendo k o número de itens da escala, σ_i^2 a variância do item X_i e σ_L^2 a variância da *summated scale*. Repare-se porém que esta abordagem não permite aferir da fiabilidade de cada um dos indicadores.

Muitos têm sido os autores que têm discutido o contributo da Análise Factorial Confirmatória para aferir da validade e fiabilidade dos indicadores e das escalas - para uma síntese ver Diamantopolous e Sigauw (2000, páginas 89–92).

Já vimos que a **validade** de um indicador tem a ver com o facto de medir aquilo que é suposto medir. Por exemplo, um teste de QI mede realmente a inteligência? Trata-se de um tipo de questão que nunca pode ser respondido com absoluta certeza. Bollen (1989a, páginas 194–206) sugere que, em Modelos com Equações Estruturais,

a validade de X_i (indicador de medida de ξ_j) seja entendida como a magnitude da relação directa entre ξ_j e X_i - dada por λ_{ij} . Por outras palavras, a validade é abordada como tendo a ver com a questão da correspondência directa entre um indicador de medida e a respectiva variável latente. Neste sentido, podem ser encaradas como medidas de validade λ_{ij} (coeficiente de validade não estandardizado), e λ_{ij}^s (coeficiente de validade estandardizado), sendo

$$\lambda_{ij}^s = \lambda_{ij} \left[\frac{\phi_{jj}}{\text{var}(X_i)} \right]^{1/2}, \text{ com } \phi_{jj} = \text{var}(\xi_j).$$

O coeficiente de validade não estandardizado deve ser significativamente não nulo, sendo que o coeficiente de validade estandardizado permite aferir qual o indicador com maior validade associado a cada variável latente.

A fiabilidade é a consistência da medição e pode ser entendida como a parte da medição que está livre do puro erro aleatório. Assim sendo, a **fiabilidade de um indicador** X_i é dada por $1 - \delta_i$, sendo δ_i o erro de medida do indicador X_i , numa solução estandardizada. Alguns autores apresentam a fiabilidade como igualando $(\lambda_{ij}^s)^2$, o quadrado do peso factorial estandardizado de X_i . Porque a fiabilidade do indicador deve exceder os 0.5, o peso factorial estandardizado do indicador (λ_{ij}^s) deverá, em valor absoluto, ser superior a 0.7.

Bollen (1989a, página 218) dá uma definição alternativa de fiabilidade de X_i , como sendo a magnitude das relações directas que todas as variáveis (excepto os δ) têm sobre X_i , e mede a fiabilidade através do coeficiente de determinação múltipla para X_i sendo que, quanto mais os valores de R^2 se aproximarem de 1, maior é a fiabilidade do indicador. Recorde-se que o R^2 de X_i é calculado como sendo igual a um menos o rácio entre a variância estimada do erro associado a X_i e a variância de X_i .

A **fiabilidade de uma variável latente** ξ_j (**CR - composite reliability**) é uma medida da consistência interna de todos os indicadores

40 Da Análise Factorial Exploratória à Análise Factorial Confirmatória de ξ_j . Pode ser calculada como:

$$\text{CR de } \xi_j = \frac{(\sum_i \lambda_{ij}^s)^2}{(\sum_i \lambda_{ij}^s)^2 + \sum_i \text{var}(\delta_i)},$$

sendo δ_i os erros de medida dos indicadores de ξ_j . Bagozzi e Yi (1988) referem como recomendáveis valores de fiabilidade compósita superiores a 0.6. Segundo a definição de Bollen, a fiabilidade da variável latente ξ_j pode ser medida pelo coeficiente de determinação R^2 a ela associado.

A **AVE - average variance extracted** (variância média extraída) indica o montante de variância capturado pelo constructo em relação ao montante de variância devida ao erro de medida - ver Fornell e Larcker (1981, página 46):

$$\text{AVE de } \xi_j = \frac{\sum_i (\lambda_{ij}^s)^2}{\sum_i (\lambda_{ij}^s)^2 + \sum_i \text{var}(\delta_i)}.$$

Valores de AVE inferiores a 0.5 indiciam que a componente erro de medida explica uma maior proporção da variância dos indicadores do que a variável latente, pondo assim em causa a validade dos indicadores e da própria variável latente. A AVE deve assumir valores superiores a 0.5, indicando assim que uma maior proporção de variância dos indicadores é capturada pela variável latente, face à variância explicada pela componente erro de medida.

2.2.4 Exemplo de um Modelo de Análise Factorial Confirmatória em LISREL

É agora ilustrada a utilização do LISREL para estimar um modelo de Análise Factorial Confirmatória, sendo considerada a natureza ordinal da escala de medida das variáveis manifestas.

O modelo de análise factorial exploratória anteriormente estimado sugere considerar dois factores de satisfação, correlacionados entre si. O factor de satisfação com aspectos materiais (*MATER*) deverá ter como indicadores de medida as variáveis *Sat2* (satisfação com o rendimento do agregado), *Sat3* (satisfação com a casa / habitação) e *Sat5* (satisfação com o emprego), ao passo que o factor de satisfação com o lazer (*LAZER*) deverá ser medido pelos indicadores *Sat6* (satisfação com a vida social), *Sat7* (satisfação com a quantidade de tempo livre) e *Sat8* (satisfação com a utilização do tempo livre). O diagrama do modelo de AFC postulado foi anteriormente apresentado na Figura 1.4, e consta agora da Figura 2.1, onde já se apresentam as estimativas obtidas em LISREL, como adiante se explicará.



Figura 2.1 AFC: modelo de medida das duas variáveis latentes de satisfação, com estimativas obtidas em LISREL considerando a natureza ordinal das seis variáveis manifestas.

42 Da Análise Factorial Exploratória à Análise Factorial Confirmatória

O modelo de medida é dado por:

$$\begin{bmatrix} X_1 & Sat2_03 \\ X_2 & Sat3_03 \\ X_3 & Sat5_03 \\ X_4 & Sat6_03 \\ X_5 & Sat7_03 \\ X_6 & Sat8_03 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^x & 0 \\ \lambda_{21}^x & 0 \\ \lambda_{31}^x & 0 \\ 0 & \lambda_{42}^x \\ 0 & \lambda_{52}^x \\ 0 & \lambda_{62}^x \end{bmatrix} \begin{bmatrix} \xi_1 & MATER03 \\ \xi_2 & LAZER03 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

Para fixar a escala de medida das variáveis latentes e evitar um problema básico de identificação, ξ_1 e ξ_2 assumem-se com variância unitária, pelo que a matriz de variância-covariância das variáveis latentes (neste caso uma matriz de correlações) é do tipo

$$\Phi = \begin{bmatrix} 1 & \\ \phi_{21} & 1 \end{bmatrix}.$$

A matriz Θ_δ é uma matriz diagonal, sendo os elementos da diagonal principal $\theta_{\delta_{11}}, \theta_{\delta_{22}}, \dots, \theta_{\delta_{66}}$ as variâncias dos erros de medida δ_i . O número total de parâmetros a estimar é de 13 ($= 6 \lambda + 6 \theta_\delta + 1 \phi$).

Modelo de Medida em LISREL

A distribuição de frequências de cada uma das seis variáveis manifestas e a matriz de correlações policóricas entre estas seis variáveis ordinais apresentam-se no Apêndice A.3. Havendo 6 variáveis manifestas e 13 parâmetros a estimar, o número de graus de liberdade do modelo é dado por $\frac{6*(6+1)}{2} - 13 = 8$. Uma explicação mais aprofundada sobre estimação de modelos com variáveis ordinais é dada na Secção 4.2.

Os comandos de sintaxe que permitem estimar o modelo de AFC representado na Figura 2.1 encontram-se no Apêndice A.3, bem como alguns dos principais outputs. É possível concluir que as estimativas obtidas têm todas valores admissíveis, e que os pesos factoriais estandardizados se encontram acima de 0.6, sendo a variável *Sat3_03*

aquela que apresenta menos valor de correlação com o factor latente do qual é indicador de medida: *MATER03* (e, conseqüentemente, menor valor de R^2).

Conforme apresentado na Tabela 2.1, os indicadores de *MATER03* têm menores valores de fiabilidade do que os indicadores de *LAZER03*. Os valores obtidos para a fiabilidade compósita e para a variância média extraída revelam que devemos ter alguma preocupação com a validade de *MATER03*, uma vez que o valor de AVE obtido é inferior a 0.5, estando o valor de fiabilidade compósita dentro do limite aceitável.

	λ^s	R^2	CR	AVE
<i>MATER03</i>			0.69	0.43
<i>Sat2_03</i>	0.73	0.54		
<i>Sat3_03</i>	0.59	0.35		
<i>Sat5_03</i>	0.63	0.40		
<i>LAZER03</i>			0.89	0.74
<i>Sat6_03</i>	0.87	0.75		
<i>Sat7_03</i>	0.82	0.67		
<i>Sat8_03</i>	0.89	0.78		

Tabela 2.1 Valores de fiabilidade, fiabilidade compósita e variância média extraída.

Em termos da qualidade de ajustamento modelo-dados, esta parece aceitável: RMSEA=0.0397; SRMR=0.023 e GFI=0.998. A questão da qualidade de ajustamento será vista com detalhe aquando do estudo do Modelo com Equações Estruturais global, no próximo capítulo.

A Figura 2.2 mostra os valores dos índices de modificação calculados pelo LISREL. O modelo de AFC em análise teria uma qualidade de ajustamento superior se sujeitássemos a estimação λ_{61} , i.e., se permitíssemos que *Sat8* fosse indicador de medida de *MATER* (o que do ponto de vista substantivo não parece fazer sentido), ou se

44 Da Análise Factorial Exploratória à Análise Factorial Confirmatória

estimassemos $\theta_{\delta_{43}}$, $\theta_{\delta_{54}}$ ou $\theta_{\delta_{63}}$, i.e., se permitissemos correlacionar erros de medida das variáveis manifestas correspondentes. Compete ao analista decidir se fará sentido permitir essas correlações, e como do ponto de vista teórico e substantivo as conseguirá justificar.

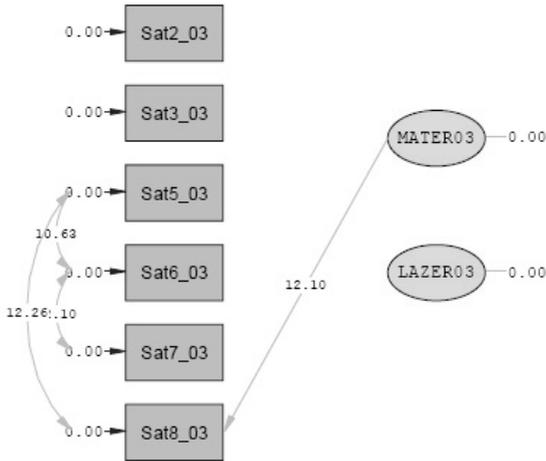


Figura 2.2 AFC: índices de modificação do modelo de medida das duas variáveis latentes de satisfação.

Capítulo 3

Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

Para uma perspectiva histórica sobre Modelos com Equações Estruturais ver Kaplan (2000, capítulo 1). Para uma introdução aos modelos de estrutura de covariância consultar Long (1983b). Para o que o autor designa pelo “essencial de um curso introdutório, de cariz não matemático, a modelos com equações estruturais nas áreas sociais e comportamentais” usando AMOS ver Blunch (2008). Para exemplos de aplicação na área da Psicologia com recurso ao LISREL, ao AMOS e ao Mplus ver, respectivamente, Byrne (1998, 2009, 2011).

Mueller (1996) e Kelloway (1998) são livros introdutórios, com os princípios base subjacentes a modelos com Equações estruturais, e com vários exemplos de aplicação, incluindo os comandos de sintaxe do LISREL utilizados para estimar os modelos e principais outputs anotados e comentados. Bollen (1989a) é a referência chave para uma exposição detalhada destes modelos.

O objectivo deste capítulo é ilustrar como os modelos de regressão linear (simples e múltipla) e de análise de caminhos podem ser encarados como casos particulares da componente estrutural de um Modelo com Equações Estruturais global, só com variáveis manifestas (Secção 3.1). Juntando a componente estrutural do modelo à componente de medida analisada em detalhe no capítulo anterior, é possível obter um Modelo Global (Secção 3.2).

3.1 Modelo Estrutural só com Variáveis Manifestas

3.1.1 Pressupostos e Equações do Modelo Estrutural

Analisaremos a componente estrutural do modelo no caso em que não há variáveis latentes mas apenas variáveis manifestas X e Y . Assume-se assim que as variáveis dependentes (Y) são explicadas pelo modelo, isto é, a estrutura de variâncias e covariâncias das variáveis dependentes é explicada pelas variáveis independentes (X). É um modelo do tipo

$$\mathbf{Y} = \mathbf{B}\mathbf{Y} + \mathbf{\Gamma}\mathbf{X} + \boldsymbol{\zeta} = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\mathbf{X} + \boldsymbol{\zeta}), \quad (3.1)$$

em que \mathbf{Y} é o vector das p variáveis dependentes, \mathbf{X} é o vector das q variáveis independentes e $\boldsymbol{\zeta}$ é o vector dos termos residuais das p equações estruturais.

Pressupõe-se que:

- $(\mathbf{I} - \mathbf{B})$ é uma matriz não singular, isto é, $(\mathbf{I} - \mathbf{B})^{-1}$ existe;
- $E[\boldsymbol{\zeta}] = \mathbf{0}$, $E[\mathbf{X}] = \mathbf{0}$, $E[\mathbf{Y}] = \mathbf{0}$;
- $\boldsymbol{\zeta}$ não está correlacionado com \mathbf{X} .

Sabe-se ainda que $\boldsymbol{\Phi}$ é a matriz de variância-covariância das variáveis manifestas independentes (X), não estando sujeita a qualquer restrição, e $\boldsymbol{\Psi}$ é a matriz de variância-covariância dos termos residuais das relações estruturais (ζ). Seja qual for o método de estimação utilizado, a estimativa de $\boldsymbol{\Phi}$ é sempre igualada a $\mathbf{S}_{\mathbf{X}\mathbf{X}}$, a matriz de variância-covariância amostral das variáveis manifestas X .

Se não se assumir que $E[\mathbf{X}] = E[\mathbf{Y}] = \mathbf{0}$, e se se dispuser do vector de médias das variáveis manifestas, o modelo apresentado pode tomar uma forma mais geral dada por

$$\mathbf{Y} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{Y} + \mathbf{\Gamma}\mathbf{X} + \boldsymbol{\zeta}, \quad (3.2)$$

em que α é o vector dos p interceptos das p equações do modelo estrutural. Podem considerar-se os seguintes **casos particulares** do modelo estrutural:

- quando $B = 0$ cada equação de 3.2 representa uma *equação de regressão linear*, simples ou múltipla, consoante o número de variáveis independentes X . Genericamente, $Y = \Gamma X + \zeta$ é uma regressão multivariada, que descreve a relação entre diversas variáveis independentes e dependentes. Trata-se de um modelo perfeitamente identificado (com zero graus de liberdade), não sendo por isso possível avaliar a qualidade do ajustamento. Este modelo permite aferir a importância explicativa das variáveis X face às Y ;
- quando B é uma matriz triangular inferior e Ψ é uma matriz diagonal (i.e., os termos residuais das equações estruturais não estão correlacionados) está-se perante um *sistema de equações recursivo*, também designado por análise de caminhos (*path analysis*);
- quando não for possível especificar B como uma matriz triangular inferior o sistema de equações é não recursivo: trata-se de um tipo de modelos muito comum na área da econometria, designado por *sistemas de equações simultâneas*. Este tipo de modelos não será abordado neste livro.

Veremos ao longo desta Secção 3.1 exemplos dos dois primeiros casos particulares mencionados: regressão linear e análise de caminhos.

A **matriz de variância-covariância implícita ao modelo estrutural** é dada por:

$$\Sigma = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} = \begin{bmatrix} A(\Gamma\Phi\Gamma^T + \Psi)A^T & A\Gamma\Phi \\ \Phi\Gamma^T A^T & \Phi \end{bmatrix}, \quad (3.3)$$

sendo $A = (I - B)^{-1}$.

Este resultado é facilmente demonstrável:

$$\begin{aligned}
 \Sigma_{YY} &= E[YY^T] = E \left[\{(I - B)^{-1}(\Gamma X + \zeta)\} \{(I - B)^{-1}(\Gamma X + \zeta)\}^T \right] \\
 &= E [A(\Gamma X + \zeta) (X^T \Gamma^T + \zeta^T) A^T] \\
 &= A \{E [\Gamma X X^T \Gamma^T] + E [\Gamma X \zeta^T] + E [\zeta X^T \Gamma^T] + E [\zeta \zeta^T]\} A^T \\
 &= A \{E [\Gamma E [X X^T] \Gamma^T + \Gamma E [X \zeta^T] + E [\zeta X^T] \Gamma^T + E [\zeta \zeta^T]\} A^T \\
 &= A (\Gamma \Phi \Gamma^T + \mathbf{0} + \mathbf{0} + \Psi) A^T = A (\Gamma \Phi \Gamma^T + \Psi) A^T
 \end{aligned}$$

$$\begin{aligned}
 \Sigma_{YX} &= E [Y X^T] = E [A(\Gamma X + \zeta) X^T] = E [A \Gamma X X^T] + E [A \zeta X^T] \\
 &= A \Gamma E [X X^T] + A E [\zeta X^T] = A \Gamma \Phi + \mathbf{0} = A \Gamma \Phi
 \end{aligned}$$

$$\Sigma_{XY} = (A \Gamma \Phi)^T = \Phi \Gamma^T A^T$$

$$\Sigma_{XX} = E [X X^T] = \Phi$$

3.1.2 Identificação do Modelo Estrutural

Se não houver qualquer restrição sobre B , Γ e Ψ , o modelo não está necessariamente identificado, pelo que a identificação pode ser obtida impondo restrições sobre elementos destas matrizes. A identificação só se pode provar algebricamente, demonstrando que os parâmetros desconhecidos são apenas função dos parâmetros identificados, e que essas funções conduzem a soluções únicas. Há contudo um conjunto de **regras** que podem ajudar na **identificação do modelo estrutural**, a saber:

- Regra- t
 - é condição necessária (ajuda a descobrir modelos não identificados), não é condição suficiente (não garante a identificação);
 - requer $t \leq \frac{1}{2}(p+q)(p+q+1)$, isto é, o número de elementos distintos da matriz de variância-covariância das $p+q$

variáveis observadas deve ser maior ou igual ao número de parâmetros a estimar;

- Regra da matriz \mathbf{B} nula
 - não é condição necessária, é condição suficiente;
 - só se aplica a modelos em que \mathbf{B} é uma matriz nula (modelos de regressão linear);
 - sendo \mathbf{B} nula, os parâmetros desconhecidos de Φ , Γ e Ψ estão identificados;
- Regra recursiva
 - não é condição necessária, é condição suficiente;
 - requer \mathbf{B} triangular inferior e Ψ diagonal (só se aplica a sistemas recursivos, i.e., modelos de análise de caminhos);
 - se Ψ é diagonal e \mathbf{B} triangular inferior, isto é, se o sistema for recursivo, então todos os parâmetros desconhecidos estão identificados;
- Condição de ordem (*Order condition*)
 - é uma condição necessária, não é uma condição suficiente;
 - avalia uma equação (e não todo o modelo);
 - requer que todos os elementos de Ψ sejam livres (sujeitos a estimação) e que o número de restrições seja maior ou igual a $p - 1$, isto é, para que uma equação esteja identificada o número de variáveis excluídas dessa equação tem que ser maior ou igual ao número de variáveis menos um;
- Condição de característica (*Rank condition*)
 - é condição necessária e suficiente;
 - avalia uma equação (e não todo o modelo);
 - requer Ψ livre e que a característica da matriz \mathbf{C}_i seja igual a $p - 1$. Para tal há que construir a matriz \mathbf{C} , dada por:

$$\mathbf{C} = \left[(\mathbf{I} - \mathbf{B}) \mid -\Gamma \right]$$

Para aferir da identificação de i -ésima equação, devem apagar-se todas as colunas da matriz C que não tenham zeros na i -ésima linha de C . As colunas que não se eliminaram formam uma nova matriz C_i . Uma condição necessária e suficiente para a identificação da i -ésima equação é que a característica de C_i seja igual ao número de variáveis menos um.

3.1.3 O Modelo de Análise de Caminhos: Efeitos Directos, Indirectos e Totais

Os modelos de análise de caminhos, introduzidos por Wright (1921, 1934) passaram a ter grande expressão na literatura na área das ciências sociais devido ao trabalho de Duncan (1966) na área da sociologia. A análise de caminhos fornece ao investigador um método multivariado (mais do que uma variável dependente) de estimar conjuntamente efeitos directos, indirectos e totais entre um determinado conjunto de variáveis, sendo para tal necessário especificar correctamente, e a partir de um modelo teórico, um diagrama. O diagrama de caminhos é assim uma representação das relações estruturais postuladas no modelo. A estimação dos parâmetros do modelo permitirá concluir acerca da estrutura de relações de dependência entre variáveis.

A análise de caminhos permite decompor as covariâncias entre pares de variáveis como função dos parâmetros do modelo, i.e., em componentes aditivas que contêm apenas elementos das matrizes B , Γ , Φ e Ψ . Tais decomposições são importantes porque mostram como é que os parâmetros se relacionam com a estrutura de covariâncias, podendo também ajudar na tarefa da identificação.

Um **efeito directo** (*direct effect*) de uma variável independente sobre uma variável dependente é medido pelo coeficiente da regressão estrutural que relaciona as duas variáveis. Um dado efeito indirecto de uma determinada variável sobre uma outra, potencialmente me-

diado por uma ou mais variáveis intermédias (ou mediadoras), é medido pelo produto dos coeficientes de regressão ao longo da cadeia de caminhos que liga as duas variáveis. A soma de todos os possíveis efeitos indirectos, para um determinado par de variáveis, designa-se por **efeito indirecto total** (*total indirect effect*). A soma dos efeitos directos com os efeitos indirectos totais designa-se por **efeito estrutural total** (*total structural effect*).

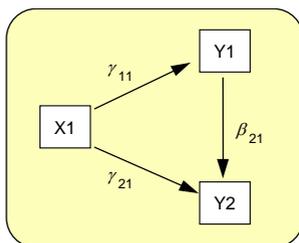


Figura 3.1 Exemplo de decomposição dos efeitos totais em directos e indirectos.

A Figura 3.1 representa os efeitos de uma variável independente (X_1) sobre duas variáveis dependentes (Y_1 e Y_2). Assim:

- β_{21} é um efeito directo de Y_1 sobre Y_2 e indica que uma unidade de variação em Y_1 resulta numa variação de β_{21} unidades em Y_2 (mantendo-se constante X). Se as variáveis estiverem padronizadas, o acréscimo de um desvio padrão em Y_1 resulta no acréscimo de β_{21} desvios-padrão em Y_2 ;
- γ_{21} tem interpretação semelhante, estando em causa um efeito directo de X_1 sobre Y_2 ;
- há um efeito indirecto de X_1 sobre Y_2 , potencialmente mediado por Y_1 , dado por $\gamma_{11}\beta_{21}$;
- o efeito total de X_1 sobre Y_2 é dado por: $\gamma_{21} + \gamma_{11}\beta_{21}$.

Os efeitos directos estão contidos nas matrizes \mathbf{B} e $\mathbf{\Gamma}$. É condição suficiente para que todos os efeitos totais existam que o valor absoluto

do maior valor próprio de BB^T seja inferior a 1. O LISREL calcula-o, designando-o por índice de estabilidade (*stability index*).

3.1.4 Exemplos de Modelos de Regressão Linear

A estimação de modelos de regressão linear (simples e múltipla) como casos particulares da componente estrutural de um Modelo com Equações Estruturais global é aqui ilustrada, com recurso ao LISREL.

Um Exemplo de Regressão Linear Simples

Retomando o exemplo da Figura 1.1, pretende-se explicar a satisfação com aspectos materiais em 2002 (variável Y) como função do (logaritmo do) rendimento do agregado familiar (variável X). Recorde-se que, num MRLS, a equação de regressão para o i -ésimo indivíduo (com $i = 1, 2, \dots, N$) é dada por:

$$Y_i = \alpha + \gamma X_i + \zeta_i,$$

sendo α o intercepto e γ o coeficiente de regressão. Uma vez estimado o modelo, a partir dos dados amostrais disponíveis, obtêm-se uma equação de regressão ajustada, dada por $\hat{Y}_i = \hat{\alpha} + \hat{\gamma}X_i$. O método de estimação tipicamente utilizado em modelos de regressão linear, OLS - *ordinary least squares*, minimiza a soma dos quadrados dos erros, isto é, das diferenças entre os valores de Y observados e os valores de Y estimados,

$$\sum_{i=1}^N \hat{\zeta}_i^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2.$$

Relembre-se que as estimativas de α e de γ , pelo método dos mínimos quadrados, são dadas por: $\hat{\gamma} = \hat{\sigma}_{YX} / \hat{\sigma}_{XX}^2$ e $\hat{\alpha} = \hat{\mu}_Y - \hat{\gamma}\hat{\mu}_X$, sendo $\hat{\mu}_X$ e $\hat{\mu}_Y$ as médias amostrais de X e de Y , $\hat{\sigma}_{YX}$ a covariância amostral entre X e Y e $\hat{\sigma}_{XX}^2$ a variância amostral de X .

Se ambas as variáveis X e Y estiverem padronizadas, o intercepto da equação de regressão é nulo ($\alpha = 0$), e a estimativa do coeficiente de

regressão iguala o coeficiente de correlação de Pearson entre as duas variáveis X e Y ($\hat{\gamma} = \hat{\rho}_{YX}$). A equação de regressão toma então a forma simplificada $\hat{Y} = \hat{\rho}_{YX}X$. Em geral é possível obter coeficientes de regressão estandardizados a partir dos não estandardizados, usando a expressão: $\hat{\gamma}_{\text{standard}} = \hat{\gamma}_{\text{não standard}}(\hat{\sigma}_X/\hat{\sigma}_Y)$, sendo $\hat{\sigma}_X$ e $\hat{\sigma}_Y$ o desvio-padrão amostral de X e de Y , respectivamente.

Para aferir da relevância da variável X no modelo importa examinar o coeficiente de regressão estimado e compará-lo com o respectivo erro padrão. Em termos práticos trata-se de ver se o valor t disponibilizado excede, em valor absoluto, $\simeq 2$. O *coeficiente de determinação* R^2 varia entre 0 e 1 e representa a proporção de variância da variável Y que pode ser atribuída à variabilidade de X .

Os comandos de sintaxe do LISREL que permitem estimar o MRLS que explica a satisfação com aspectos materiais (Y) como função do (logaritmo do) rendimento do agregado familiar (X) constam do apêndice B.1.1. A Figura 3.2 apresenta o diagrama do MRLS postulado, com as estimativas obtidas em LISREL.

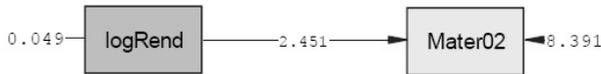


Figura 3.2 Diagrama do MRLS postulado, com estimativas obtidas.

Apresentam-se alguns excertos do output obtido. A matriz de variância-covariância e o vector de médias amostrais, a partir dos quais serão obtidos $\hat{\gamma}$ e $\hat{\alpha}$:

Covariance Matrix	
	Mater02 logRend

Mater02	8.687
logRend	0.120 0.049
Means	
	Mater02 logRend

	14.946 3.479

54 Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

Os valores de significância obtidos indiciam que os parâmetros são todos significativamente diferentes de zero. Note-se ainda que $\hat{\phi}_{11} = S_{11} = 0.049$, a variância amostral de *logRend*.

```

Number of Iterations = 0
      GAMMA
      logRend
      -----
Mater02      2.451
            (0.292)
            8.402

      PHI
      logRend
      -----
            0.049
            (0.002)
            31.670

      PSI
      Mater02
      -----
            8.391
            (0.265)
            31.670
    
```

O coeficiente de determinação R^2 é de 0.034, indicando que apenas 3.4% da variabilidade de *Mater02* é explicada pela variabilidade de *logRend*, o que é muito baixo... O modelo tem zero graus de liberdade, e por isso não é possível aferir a qualidade de ajustamento modelo-dados.

```

Squared Multiple Correlations for Structural Equations
      Mater02
      -----
            0.034

                        Goodness of Fit Statistics
                        Degrees of Freedom = 0
                        The Model is Saturated, the Fit is Perfect !
    
```

Está-se pois perante um modelo muito fraco e demasiado simplista. O próximo passo será utilizar um modelo de regressão linear múltipla, procurando explicar Y a partir de várias variáveis independentes.

Um Exemplo de Regressão Linear Múltipla

Retomemos o MRLM apresentado no diagrama da Figura 1.2, em que está em causa explicar a satisfação com aspectos materiais em 2002 (variável Y) como função do (logaritmo do) rendimento do agregado familiar (variável X_1), da idade (variável X_2) e do sexo do respondente (variável X_3 , com 0 = Feminino e 1 = Masculino). A Figura 3.3 apresenta o diagrama do MRLM postulado, com as estimativas obtidas em LISREL.

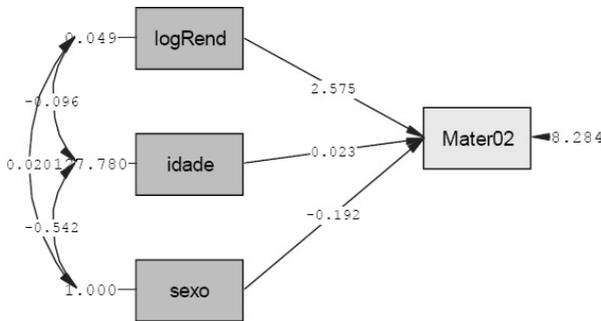


Figura 3.3 Diagrama do MRLM postulado, com estimativas obtidas.

A equação de regressão para o i -ésimo indivíduo (com $i = 1, 2, \dots, N$) é dada por:

$$Y_i = \alpha + \gamma_{11}X_{1i} + \gamma_{12}X_{2i} + \gamma_{13}X_{3i} + \zeta_i,$$

sendo Y a variável dependente, α o intercepto e γ_{1j} o coeficiente de regressão associado à variável independente X_j (com $j = 1, 2, 3$).

Note-se que a estimativa do intercepto $\hat{\alpha}$ só tem interpretação substantiva quando todas as variáveis X no modelo podem assumir o valor zero, o que não é o caso no nosso exemplo devido à forma como as variáveis foram medidas e codificadas. Assim, o intercepto inclui-se aqui meramente para efeitos de ilustração de como fazê-lo.

56 Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

O valor não padronizado do coeficiente de regressão $\hat{\gamma}_{1j}$ representa a variação unitária esperada em Y quando há uma variação unitária em X_j e todas as restantes variáveis independentes permanecem inalteradas. Tal como vimos no caso da regressão linear simples, numa solução estandardizada $\alpha = 0$, e a estimativa do coeficiente de regressão padronizado representa o acréscimo, em desvios-padrão, de Y , quando X_j tem um acréscimo de um desvio-padrão e as restantes variáveis independentes permanecem inalteradas. Os coeficientes de regressão padronizados são úteis para mostrar a “importância” ou contributo relativo das variáveis independentes para explicar a variável dependente.

O ficheiro de sintaxe do LISREL que permite estimar o modelo de regressão linear múltipla que explica *Mater02* como função de *logRend*, *idade* e *sexo* apresenta-se no apêndice B.1.2.

Os valores de significância associados às estimativas (não padronizadas) dos coeficientes de regressão e do intercepto indiciam que os parâmetros do MRLM são todos significativos. Note-se ainda que os elementos da matriz Φ igualam as variâncias e covariâncias amostrais das variáveis independentes X .

```

Number of Iterations = 0
      GAMMA
      logRend      idade      sexo
-----
Mater02      2.575      0.023      -0.192
             (0.291)      (0.006)      (0.065)
             8.837      3.991      -2.977

      PHI
      logRend      idade      sexo
-----
logRend      0.049
             (0.002)
             31.654
idade      -0.096      127.780
             (0.056)      (4.037)
             -1.722      31.654
sexo      0.020      -0.542      1.000
             (0.005)      (0.253)      (0.032)
             4.067      -2.143      31.654
    
```

```

PSI
  Mater02
-----
      8.284
    (0.262)
    31.654
ALPHA
  Mater02
-----
      5.058
    (1.050)
    4.819
    
```

O coeficiente de determinação R^2 , que representa a proporção de variância da variável dependente explicada pelas variáveis independentes no modelo, assume um valor de 0.046, i.e., apenas 4.6% da variabilidade de *Mater02* é explicada pelas três variáveis independentes.

```

Squared Multiple Correlations for Structural Equations
  Mater02
-----
      0.046
    
```

A partir das estimativas padronizadas (ver Apêndice B.1.2) é possível concluir que, em termos relativos, a variável *logRend* parece ser, das três variáveis independentes, a que mais contribui para explicar a variável dependente *Mater02*; todavia estamos perante um MRLM com um ajustamento muito fraco.

Breve Comentário sobre Multicolinearidade

Um problema potencialmente sério em MRLM é o da multicolinearidade, isto é, uma situação em que uma ou mais variáveis independentes estão altamente correlacionadas com outra ou outras variáveis independentes, também incluídas no modelo. Sinais que podem indicar a existência de multicolinearidade incluem:

- coeficientes de correlação entre variáveis independentes elevados (acima de 0.7, em valor absoluto);

- uma ou mais estimativas dos coeficientes de regressão com sinais contraditórios face ao que seria esperado;
- uma ou mais estimativas dos coeficientes de regressão padronizados com valores muito elevados;
- estimativas elevadas dos erros padrão dos coeficientes de regressão;
- valor elevado de R^2 , mas poucas (ou nenhuma) variáveis independentes significativamente explicativas.

No caso de alguma destas situações ocorrer, compete ao analista decidir qual ou quais as variáveis explicativas a excluir do modelo.

3.1.5 Exemplo de Modelo de Análise de Caminhos

Retomemos agora o diagrama do modelo de análise de caminhos anteriormente apresentado na Figura 1.3, e agora na Figura 3.4 já com as estimativas obtidas em LISREL (numa solução standardizada).

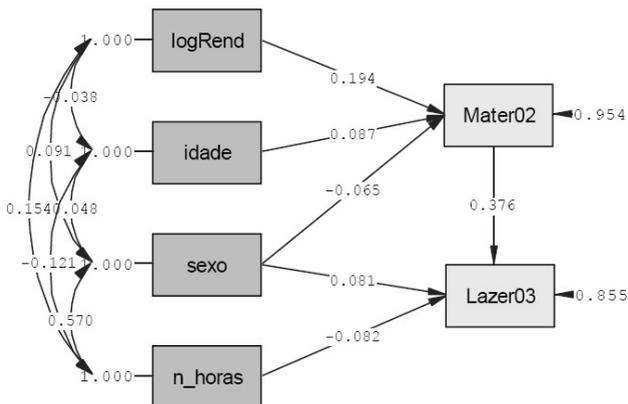


Figura 3.4 Diagrama de modelo de análise de caminhos, com estimativas obtidas (solução standardizada).

Neste modelo estão postuladas as seguintes relações estruturais:

- a percepção de satisfação com aspectos materiais em 2002 influencia a percepção de satisfação com aspectos de lazer em 2003;
- a percepção de satisfação com aspectos materiais é influenciada pelo rendimento do agregado familiar, pela idade e pelo sexo do respondente;
- a percepção de satisfação com aspectos de lazer é influenciada pelo sexo do respondente e pelo número de horas de trabalho semanal (sendo que maior número de horas de trabalho implicará menor satisfação com lazer).

As relações estruturais entre variáveis são representadas no diagrama da esquerda para a direita, e de cima para baixo. Por outras palavras, cada variável dependente é representada à direita e / ou abaixo de todas as variáveis X ou Y que a antecedem, permitindo assim que o modelo seja recursivo. Designando *Mater02* por Y_1 , *Lazer03* por Y_2 , *logRend* por X_1 , *idade* por X_2 , *sexo* por X_3 e *n_horas* por X_4 , ao modelo recursivo de análise de caminhos representado na Figura 3.4 correspondem as duas seguintes equações estruturais:

$$\begin{cases} Y_1 = \alpha_1 & +\gamma_{11}X_1 & +\gamma_{12}X_2 & +\gamma_{13}X_3 & & +\zeta_1 \\ Y_2 = \alpha_2 & +\beta_{21}Y_1 & & +\gamma_{23}X_3 & +\gamma_{24}X_4 & +\zeta_2 \end{cases}$$

O modelo pode ainda ser escrito na forma matricial (recorde-se a forma geral de um modelo estrutural só com variáveis manifestas dada pela Equação 3.2) como:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ \gamma_{12} & 0 \\ \gamma_{13} & \gamma_{23} \\ 0 & \gamma_{24} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}.$$

O ficheiro de sintaxe que permite estimar este modelo em LISREL, bem como os restantes outputs obtidos, apresentam-se no Apêndice B.2. Foram as seguintes as estimativas obtidas para os parâmetros das matrizes \mathbf{B} , $\mathbf{\Gamma}$, e $\mathbf{\Psi}$, numa solução estandardizada (em que as variáveis têm variância unitária):

60 Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

Standardized Solution

BETA					
	Mater02	Lazer03			
	-----	-----			
Mater02	- -	- -			
Lazer03	0.376	- -			
GAMMA					
	logRend	idade	sexo	n_horas	
	-----	-----	-----	-----	
Mater02	0.194	0.087	-0.065	- -	
Lazer03	- -	- -	0.081	-0.082	
PSI					
Note: This matrix is diagonal.					
	Mater02	Lazer03			
	-----	-----			
	0.954	0.855			

Seguem-se os valores dos coeficientes de determinação R^2 . Apenas 4.6% da variabilidade de *Mater02* pode ser atribuída à variabilidade de *logRend*, *idade* e *sexo*, ao passo que 14.5% da variância de *Lazer03* é explicada pelas restantes variáveis no modelo, das quais é função.

Squared Multiple Correlations for Structural Equations

Mater02	Lazer03
-----	-----
0.046	0.145

Goodness of Fit Statistics

Degrees of Freedom = 3

Minimum Fit Function Chi-Square = 7.284 (P = 0.0634)

Normal Theory Weighted Least Squares Chi-Square = 7.270 (P = 0.0638)

Apresentam-se no Apêndice B.2 os valores obtidos para os efeitos indirectos e totais. Por exemplo, a variável *sexo* tem um efeito directo sobre *LAZER03* de $\hat{\gamma}_{23} = 0.081$ (numa solução estandardizada); e tem também um efeito indirecto, potencialmente mediado por *MATER02*, de $-0.065 \times 0.376 = -0.024$. Assim, o efeito total da variável *sexo* sobre *LAZER03* é de $0.081 - 0.024 = 0.057$. No caso da variável *idade*, o efeito total sobre *LAZER03* é igual ao efeito indirecto (já que o modelo não postula qualquer efeito directo) e iguala, numa solução estandardizada, $0.087 \times 0.0376 = 0.033$. O LISREL calcula todos estes efeitos, pelo que o utilizador apenas terá de os

perceber e interpretar. Note-se que, para além de testes de hipóteses à nulidade de cada um dos efeitos directos postulados no modelo, o LISREL disponibiliza testes de hipóteses à nulidade dos efeitos totais e à nulidade do conjunto dos efeitos indirectos, não sendo todavia possível aferir da significância de cada um dos efeitos indirectos, de per si. Vitória et al. (2011) ilustram como calcular os diferentes efeitos indirectos e aferir da sua significância com base na abordagem de *bootstrap* proposta por Preacher e Hayes (2008).

3.2 Modelo Equações Estruturais Global

3.2.1 Pressupostos e Equações do Modelo Global

Pressupostos

Para além de assumir amostras aleatórias, observações independentes e relações estruturais lineares nos parâmetros e nas variáveis, o Modelo com Equações Estruturais global tem ainda os seguintes pressupostos:

- as variáveis são medidas como desvios face aos seus valores médios, pelo que :

$$E[\boldsymbol{\eta}] = E[\boldsymbol{\xi}] = E[\boldsymbol{\zeta}] = E[\mathbf{Y}] = E[\mathbf{X}] = E[\boldsymbol{\epsilon}] = E[\boldsymbol{\delta}] = \mathbf{0}$$

- os erros de medida não estão correlacionados com as variáveis latentes, isto é, $\boldsymbol{\epsilon}$ não está correlacionado com $\boldsymbol{\eta}$, nem $\boldsymbol{\delta}$ está correlacionado com $\boldsymbol{\xi}$, pelo que:

$$\begin{aligned} E[\boldsymbol{\xi}\boldsymbol{\delta}^T] &= E[\boldsymbol{\delta}\boldsymbol{\xi}^T] = E[\boldsymbol{\eta}\boldsymbol{\epsilon}^T] = E[\boldsymbol{\epsilon}\boldsymbol{\eta}^T] = \mathbf{0} \\ E[\boldsymbol{\xi}\boldsymbol{\epsilon}^T] &= E[\boldsymbol{\epsilon}\boldsymbol{\xi}^T] = E[\boldsymbol{\eta}\boldsymbol{\delta}^T] = E[\boldsymbol{\delta}\boldsymbol{\eta}^T] = \mathbf{0} \end{aligned}$$

- os erros de medida não estão correlacionados entre si, nem com os termos residuais das equações estruturais, isto é:

$$\begin{aligned} E[\boldsymbol{\delta}\boldsymbol{\epsilon}^T] &= E[\boldsymbol{\epsilon}\boldsymbol{\delta}^T] = \mathbf{0} \\ E[\boldsymbol{\zeta}\boldsymbol{\epsilon}^T] &= E[\boldsymbol{\epsilon}\boldsymbol{\zeta}^T] = E[\boldsymbol{\zeta}\boldsymbol{\delta}^T] = E[\boldsymbol{\delta}\boldsymbol{\zeta}^T] = \mathbf{0} \end{aligned}$$

62 Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

- ξ e ζ não estão correlacionados: $E[\xi\zeta^T] = E[\zeta\xi^T] = \mathbf{0}$
- B tem zeros na diagonal e $(I - B)$ é uma matriz não singular, pelo que $(I - B)^{-1}$ existe.

Equações

O Modelo com Equações Estruturais global é definido pelas três seguintes equações (as duas primeiras constituem o modelo de medida e a última constitui a componente estrutural do modelo):

$$\begin{aligned} X &= \Lambda_X \xi + \delta \\ Y &= \Lambda_Y \eta + \epsilon \\ \eta &= B\eta + \Gamma\xi + \zeta \end{aligned} \quad (3.4)$$

Sabe-se ainda que $E[\xi\xi^T] = \Phi$ e $E[\zeta\zeta^T] = \Psi$, sendo Φ e Ψ , respectivamente, as matrizes de variância-covariância das variáveis latentes independentes ξ e dos termos residuais das relações estruturais ζ .

Matriz de Variância-Covariância Implícita

Os pressupostos e as equações anteriormente apresentadas implicam a seguinte matriz de variância-covariância para as variáveis manifestas:

$$\begin{aligned} \Sigma &= \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} \\ &= \begin{bmatrix} \Lambda_Y A (\Gamma\Phi\Gamma^T + \Psi) A^T \Lambda_Y^T + \Theta_\epsilon & \Lambda_Y A \Gamma\Phi\Lambda_X^T \\ \Lambda_X \Phi\Gamma^T A^T \Lambda_Y^T & \Lambda_X \Phi\Lambda_X^T + \Theta_\delta \end{bmatrix}, \end{aligned} \quad (3.5)$$

com $A = (I - B)^{-1}$.

Este resultado é facilmente demonstrável. Note-se que Σ_{XX} corresponde à matriz de variância-covariância implícita ao modelo de medida para X e iguala $\Lambda_X \Phi\Lambda_X^T + \Theta_\delta$, como demonstrado aquando da apresentação da Equação 2.2. Por seu turno:

$$\begin{aligned}
 \Sigma_{YY} &= E[\mathbf{Y}\mathbf{Y}^T] = E [(\Lambda_Y \boldsymbol{\eta} + \boldsymbol{\epsilon})(\Lambda_Y \boldsymbol{\eta} + \boldsymbol{\epsilon})^T] \\
 &= E [(\Lambda_Y \boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\eta}^T \Lambda_Y^T + \boldsymbol{\epsilon}^T)] \\
 &= E [\Lambda_Y \boldsymbol{\eta} \boldsymbol{\eta}^T \Lambda_Y^T] + E [\Lambda_Y \boldsymbol{\eta} \boldsymbol{\epsilon}^T] + E [\boldsymbol{\epsilon} \boldsymbol{\eta}^T \Lambda_Y^T] + E [\boldsymbol{\epsilon} \boldsymbol{\epsilon}^T] \\
 &= \Lambda_Y E [\boldsymbol{\eta} \boldsymbol{\eta}^T] \Lambda_Y^T + \Lambda_Y E [\boldsymbol{\eta} \boldsymbol{\epsilon}^T] + E [\boldsymbol{\epsilon} \boldsymbol{\eta}^T] \Lambda_Y^T + E [\boldsymbol{\epsilon} \boldsymbol{\epsilon}^T] \\
 &= \Lambda_Y [\mathbf{A} (\boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Gamma}^T + \boldsymbol{\Psi}) \mathbf{A}^T] \Lambda_Y^T + \mathbf{0} + \mathbf{0} + \boldsymbol{\Theta}_\epsilon \\
 &= \Lambda_Y \mathbf{A} (\boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Gamma}^T + \boldsymbol{\Psi}) \mathbf{A}^T \Lambda_Y^T + \boldsymbol{\Theta}_\epsilon
 \end{aligned}$$

Note-se que o valor que se obtém para $E [\boldsymbol{\eta} \boldsymbol{\eta}^T]$ é idêntico ao que se obtém para $E [\mathbf{Y}\mathbf{Y}^T]$ na matriz de variância-covariância implícita ao modelo estrutural só com variáveis manifestas (ver Equação 3.3).

$$\begin{aligned}
 \Sigma_{YX} &= E[\mathbf{Y}\mathbf{X}^T] = E [(\Lambda_Y \boldsymbol{\eta} + \boldsymbol{\epsilon})(\Lambda_X \boldsymbol{\xi} + \boldsymbol{\delta})^T] \\
 &= E [(\Lambda_Y \boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\xi}^T \Lambda_X^T + \boldsymbol{\delta}^T)] \\
 &= E [\Lambda_Y \boldsymbol{\eta} \boldsymbol{\xi}^T \Lambda_X^T] + E [\Lambda_Y \boldsymbol{\eta} \boldsymbol{\delta}^T] + E [\boldsymbol{\epsilon} \boldsymbol{\xi}^T \Lambda_X^T] + E [\boldsymbol{\epsilon} \boldsymbol{\delta}^T] \\
 &= \Lambda_Y E [\boldsymbol{\eta} \boldsymbol{\xi}^T] \Lambda_X^T + \Lambda_Y E [\boldsymbol{\eta} \boldsymbol{\delta}^T] + E [\boldsymbol{\epsilon} \boldsymbol{\xi}^T] \Lambda_X^T + E [\boldsymbol{\epsilon} \boldsymbol{\delta}^T] \\
 &= \Lambda_Y (\mathbf{A} \boldsymbol{\Gamma} \boldsymbol{\Phi}) \Lambda_X^T + \mathbf{0} + \mathbf{0} + \mathbf{0} = \Lambda_Y \mathbf{A} \boldsymbol{\Gamma} \boldsymbol{\Phi} \Lambda_X^T
 \end{aligned}$$

Note-se que o valor de $E [\boldsymbol{\eta} \boldsymbol{\xi}^T]$ é idêntico ao valor de $E [\mathbf{Y}\mathbf{X}^T]$ na matriz de variância-covariância implícita ao modelo estrutural só com variáveis manifestas (ver Equação 3.3). Por último:

$$\Sigma_{XY} = E[\mathbf{X}\mathbf{Y}^T] = (\Lambda_Y \mathbf{A} \boldsymbol{\Gamma} \boldsymbol{\Phi} \Lambda_X^T)^T = \Lambda_X \boldsymbol{\Phi} \boldsymbol{\Gamma}^T \mathbf{A}^T \Lambda_Y^T$$

Efeitos Directos, Indirectos e Totais

É possível decompor os parâmetros do modelo global e analisar os efeitos directos, indirectos e totais. Num Modelo global com Equações

Estruturais os efeitos directos constam das matrizes \mathbf{B} , $\mathbf{\Gamma}$, $\mathbf{\Lambda}_X$ e $\mathbf{\Lambda}_Y$. A Tabela 3.1 ilustra a decomposição dos efeitos totais em efeitos directos e efeitos indirectos, bem como a relação destes efeitos com os parâmetros do modelo.

	Directos	Indirectos	Totais
$\xi \rightarrow \eta$	$\mathbf{\Gamma}$	$(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma} - \mathbf{\Gamma}$	$(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}$
$\eta \rightarrow \eta$	\mathbf{B}	$(\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} - \mathbf{B}$	$(\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I}$
$\xi \rightarrow Y$	$\mathbf{0}$	$\mathbf{\Lambda}_Y(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}$	$\mathbf{\Lambda}_Y(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}$
$\xi \rightarrow X$	$\mathbf{\Lambda}_X$	$\mathbf{0}$	$\mathbf{\Lambda}_X$
$\eta \rightarrow Y$	$\mathbf{\Lambda}_Y$	$\mathbf{\Lambda}_Y(\mathbf{I} - \mathbf{B})^{-1} - \mathbf{\Lambda}_Y$	$\mathbf{\Lambda}_Y(\mathbf{I} - \mathbf{B})^{-1}$
$\eta \rightarrow X$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$

Tabela 3.1 *Decomposição dos efeitos totais em efeitos directos e efeitos indirectos.*

3.2.2 Identificação do Modelo Global

Especificado que esteja o modelo, e antes de proceder à estimação do mesmo, importa procurar garantir a sua identificação: se todos os parâmetros do modelo estiverem identificados, o modelo está identificado.

A forma mais segura de estabelecer a identificação (e a única que efectivamente a prova) consiste em explicitar cada um dos parâmetros do modelo como função das variâncias e covariâncias amostrais, e de outros parâmetros já identificados. Tal tarefa será mais fácil se se demonstrar primeiro que os parâmetros do modelo de medida estão identificados, e só depois que o estão os parâmetros do modelo estrutural (tentando explicitá-los em função dos parâmetros do modelo de medida que já se provaram identificados, e não directamente em termos de covariância das variáveis manifestas). O analista deve ter a preocupação de garantir a identificação local: o modelo, globalmente

considerado, pode ter graus de liberdade positivos mas existirem componentes do modelo que estão sub-identificadas. Um exemplo: se o modelo estrutural tiver 10 graus de liberdade e o modelo de medida estiver sub-identificado ($df = -1$), então o modelo global terá 9 graus de liberdade mas não estará identificado, o que põe em causa as estimativas obtidas.

Para modelos com elevado grau de complexidade os procedimentos algébricos tornam-se difíceis, fastidiosos e muito sujeitos a erro. Existem algumas **regras** que ajudam a detectar a **identificação do modelo global**, embora não exista nenhuma condição necessária e suficiente de identificação:

- Regra- t

- é uma condição necessária, não é uma condição suficiente;
- requer que o número de parâmetros independentes a serem estimados, t , seja inferior ao número de elementos únicos de Σ , i.e.,

$$t \leq \frac{1}{2}(p+q)(p+q+1).$$

- Regra dos dois passos

- não é condição necessária, é condição suficiente;
- inclui 2 passos:
 - 1º tratar o modelo como se de análise factorial confirmatória se tratasse, ignorando \mathbf{B} , $\mathbf{\Gamma}$ e $\mathbf{\Psi}$, e aferir a identificação do modelo de medida. Se este modelo estiver identificado, passar ao 2º passo;
 - 2º tratar as variáveis latentes como se fossem manifestas e estabelecer a identificação do modelo estrutural, i.e., ver se os parâmetros de \mathbf{B} , $\mathbf{\Gamma}$ e $\mathbf{\Psi}$ estão identificados;
- se ambas as condições se verificarem esta é uma regra suficiente de identificação; mas porque não é uma condição necessária, pode acontecer que um modelo não verifique estas condições e ainda assim esteja identificado.

Em termos práticos, para além de verificar se o número de graus de liberdade do modelo é não negativo, convém verificar se foi atribuída uma unidade de medida a cada uma das variáveis latentes do modelo (bastando para tal igualar a variância das variáveis latentes à unidade, ou então, para cada variável latente, igualar à unidade o peso factorial associado à variável manifesta de referência). No caso de haver um só indicador para uma variável latente importa garantir que se assumiu que essa variável foi medida sem erro, pelo que a correspondente variância do erro de medida na matriz Θ_{δ} (ou Θ_{ϵ}) é igual a zero.

3.2.3 Modelo MIMIC

Um caso particular do Modelo com Equações Estruturais global, muito utilizado, é o chamado *Multiple Indicators and Multiple Causes model* - modelo MIMIC. O modelo contém uma só variável latente (é um modelo unifactorial) η_1 , medida por um ou mais indicadores, as variáveis Y . A variável latente η_1 é directamente influenciada por uma ou mais variáveis explicativas X (para mais detalhes ver Bollen, 1989a, página 331 ou Brown, 2006, páginas 304–316).

No caso concreto dos dados que temos vindo a analisar, poderíamos considerar o factor latente de satisfação com aspectos materiais em 2003 (*MATER03*) medido pelos três indicadores de satisfação com o rendimento do agregado familiar (*Sat2_03*), com a casa / habitação (*Sat3_03*) e com o emprego (*Sat5_03*) - como vimos no modelo de AFC da Figura 2.1 - e influenciado pelas variáveis explicativas (logaritmo do) rendimento do agregado familiar (*logRend*), idade do respondente (*idade*) e sexo do respondente (*sexo*), assumindo-se as três variáveis explicativas como correlacionadas entre si.

3.2.4 Exemplo de Modelo com Equações Estruturais Global

Vamos retomar o modelo global da Figura 1.5 do capítulo introdutório, agora apresentado na Figura 3.5 já com as estimativas obtidas em LISREL (numa solução estandarizada, com estimação WLS). No diagrama do modelo estão postuladas as seguintes relações estruturais e de medida:

- a percepção de satisfação com aspectos materiais em 2002 (*MATER02*) influencia a percepção de satisfação com aspectos de lazer no ano seguinte (*LAZER03*);
- a satisfação com aspectos materiais é influenciada pelo (logaritmo do) rendimento do agregado familiar (X_1), pela idade (X_2) e pelo sexo (X_3) do respondente (recorde-se que 0 = Feminino e 1 = Masculino);
- a satisfação com aspectos de lazer é influenciada pelo sexo do respondente (X_3) e pelo n° de horas de trabalho semanal (X_4);
- a variável latente *MATER02* (η_1) é medida pelos indicadores de satisfação com o rendimento do agregado familiar (*Sat2_02*), com a casa / habitação (*Sat3_02*) e com o emprego (*Sat5_02*), respectivamente variáveis Y_1, Y_2 e Y_3 ;
- a variável latente *LAZER03* (η_2) é medida pelos indicadores de satisfação com a vida social (*Sat6_03*), com a quantidade de tempo livre (*Sat7_03*) e com a utilização feita do tempo livre (*Sat8_03*), respectivamente variáveis Y_4, Y_5 e Y_6 .

Vamos assumir que a escala de medida de *MATER02* é determinada pela escala de medida de *Sat2_02*, e que a escala de *LAZER03* é determinada pela de *Sat6_03*, o que implicará especificar como iguais à unidade os dois pesos factoriais correspondentes ($\lambda_{11}^y = \lambda_{42}^y = 1$).

Assim, o modelo postulado na Figura 3.5 é definido por um conjunto de duas equações para a componente estrutural e seis equações para a componente de medida. As duas equações do modelo estrutural são

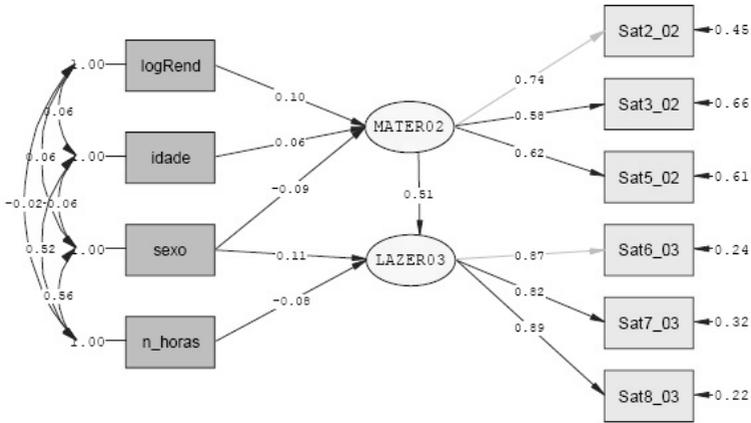


Figura 3.5 Diagrama de modelo de equações estruturais global, com estimativas obtidas (solução estandardizada).

dadas por:

$$\begin{cases} \eta_1 = \gamma_{11}X_1 + \gamma_{12}X_2 + \gamma_{13}X_3 + \zeta_1 \\ \eta_2 = \beta_{21}\eta_1 + \gamma_{23}X_3 + \gamma_{24}X_4 + \zeta_2 \end{cases}$$

Alternativamente, em termos matriciais:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ \gamma_{12} & 0 \\ \gamma_{13} & \gamma_{23} \\ 0 & \gamma_{24} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}.$$

As seis equações do modelo de medida das duas variáveis latentes dependentes podem ser escritas, usando notação matricial, como:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21}^y & 0 \\ \lambda_{31}^y & 0 \\ 0 & 1 \\ 0 & \lambda_{52}^y \\ 0 & \lambda_{62}^y \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}.$$

As restantes matrizes do modelo são Φ (a matriz de associações das 4 variáveis explicativas X), Ψ (a matriz de variância-covariância

70 Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

LAMBDA-Y		
	MATER02	LAZER03
	-----	-----
Sat2_02	0.740	- -
Sat3_02	0.580	- -
Sat5_02	0.621	- -
Sat6_03	- -	0.870
Sat7_03	- -	0.824
Sat8_03	- -	0.886

Tal como se pode ver no Apêndice B.3 os parâmetros β e γ sujeitos a estimação são todos significativos, pelo que se validam as relações estruturais postuladas no modelo. A variável que mais contribui para explicar a satisfação com aspectos materiais em 2002 é o (logaritmo do) rendimento do agregado familiar, seguida do sexo do respondente, sendo que em 2002 os homens têm valores médios de satisfação material inferiores às mulheres (a estimativa de γ_{13} é negativa). Todavia, em 2003 os homens têm valores médios de satisfação com aspectos de lazer superiores às mulheres (a estimativa de γ_{23} é positiva). A satisfação com o lazer em 2003 é muito influenciada pela satisfação com o material em 2002, sendo que um acréscimo no número de horas de trabalho semanal leva a menores valores de satisfação com o lazer. As correlações estimadas são as seguintes:

Correlation Matrix of ETA and KSI						
	MATER02	LAZER03	logRend	idade	sexo	n_horas
	-----	-----	-----	-----	-----	-----
MATER02	1.000					
LAZER03	0.497	1.000				
logRend	0.101	0.059	1.000			
idade	0.072	-0.010	0.060	1.000		
sexo	-0.089	0.026	0.059	-0.058	1.000	
n_horas	-0.022	-0.025	-0.023	0.520	0.557	1.000

Seguem-se os valores obtidos para os R^2 das equações estruturais e das variáveis Y :

Squared Multiple Correlations for Structural Equations	
MATER02	LAZER03
-----	-----
0.023	0.256

Squared Multiple Correlations for Y - Variables

Sat2_02	Sat3_02	Sat5_02	Sat6_03	Sat7_03	Sat8_03
0.547	0.336	0.386	0.756	0.680	0.784

Os coeficientes de determinação R^2 associados às duas equações estruturais indicam que:

- 2.3% da variância da variável latente *MATER02* é explicada pelas variáveis manifestas *logRend*, *idade* e *sexo*;
- 25.6% da variância de *LAZER03* é explicada pela variáveis latente *MATER02* e pelas variáveis manifestas *sexo* e *n_horas*.

Recorde-se que os coeficientes de determinação R^2 associados a cada uma das variáveis manifestas podem ser encarados como uma medida da fiabilidade dos indicadores e, idealmente, deverão aproximar-se do valor 1. Seguem-se algumas medidas da qualidade de ajustamento:

Goodness of Fit Statistics

Degrees of Freedom = 27

Minimum Fit Function Chi-Square = 152.506 (P = 0.0)

Root Mean Square Error of Approximation (RMSEA) = 0.0481

90 Percent Confidence Interval for RMSEA = (0.0408 ; 0.0557)

P-Value for Test of Close Fit (RMSEA < 0.05) = 0.646

Normed Fit Index (NFI) = 0.972

Non-Normed Fit Index (NNFI) = 0.961

Parsimony Normed Fit Index (PNFI) = 0.583

Comparative Fit Index (CFI) = 0.977

Incremental Fit Index (IFI) = 0.977

Relative Fit Index (RFI) = 0.953

Standardized RMR = 0.101

Goodness of Fit Index (GFI) = 0.994

Adjusted Goodness of Fit Index (AGFI) = 0.988

Parsimony Goodness of Fit Index (PGFI) = 0.488

São também disponibilizados índices de modificação, i.e., para cada parâmetro do modelo não sujeito a estimação é calculada a melhoria no ajustamento global do modelo (a redução na estatística de χ^2) caso esse parâmetro fosse sujeito a estimação. O índice de modificação mais elevado está associado ao parâmetro que, uma vez estimado, mais contribui para melhorar o ajustamento global do modelo.

72 Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

Da análise dos índices de modificação apresentados na Figura 3.6 é possível concluir que o ajustamento global do modelo melhoraria se se sujeitassem a estimação algumas correlações entre erros de medida, nomeadamente $\theta_{\epsilon_{51}}$ ou $\theta_{\epsilon_{63}}$, ou se se considerasse a variável *Sat8_03* como indicador de medida de *MATER02* (o que não faz qualquer sentido). Como vimos anteriormente, a decisão de modificar, ou não, o modelo deve ser conduzida por um suporte teórico substantivo, e não pelos dados da amostra.



Figura 3.6 Diagrama de modelo de equações estruturais global: índices de modificação.

3.2.5 Exemplo de Modelo Longitudinal, com Medidas Repetidas

Rogosa (1979) discute Modelos com Equações Estruturais para dados longitudinais (em painel), designadamente modelos com efeitos cruzados e desfasados e correlações entre erros de medida de medidas repetidas. Vamos agora retomar o modelo da Figura 1.6, com cariz longitudinal, e que incorpora medidas repetidas em quatro momentos temporais distintos, de 2002 a 2005, dos indicadores de satisfação com aspectos materiais e com aspectos de lazer.

Na Secção 3.2.4 foi analisado um Modelo com Equações Estruturais global que postula que a satisfação com aspectos materiais em 2002 influencia a satisfação com aspectos de lazer em 2003 (ver Figura 3.5). No modelo representado na Figura 1.6 considera-se que cada uma das dimensões de satisfação num determinado momento temporal influencia a referida dimensão no momento temporal seguinte, i.e., são postulados efeitos desfasados (com um momento temporal de desfasamento). Postulam-se ainda efeitos cruzados, também com um ano de desfasamento, i.e., considera-se que a satisfação com cada uma das dimensões num determinado momento temporal influencia a satisfação com a outra dimensão no momento temporal seguinte.

Em cada momento temporal as variáveis latentes de satisfação são medidas pelos três indicadores que têm vindo a ser utilizados nos exemplos anteriores. Uma vez que para cada indicador de satisfação existem quatro medidas repetidas, o modelo postulado permite correlacionar os erros de medida dessas medidas repetidas (para assim espelhar o efeito longitudinal) - a matriz de covariância dos erros de medida deixa de ser diagonal, representando as setas entre indicadores na Figura 1.6 essas correlações. Por outro lado, e porque em cada momento temporal estamos a modelar dois factores de satisfação que se assumem correlacionados, os termos residuais das duas relações estruturais, em cada momento temporal, também se assumem correlacionados - a matriz Ψ deixa de ser diagonal. Foram considerados como determinantes ou variáveis explicativas das duas dimensões de satisfação as variáveis que têm sido usadas nos exemplos anteriores: (logaritmo do) rendimento do agregado familiar, idade, sexo e o número de horas de trabalho semanal do respondente. A sequência de comandos de sintaxe que permite estimar por *Robust* ML o modelo postulado consta do Anexo B.4. Apresentam-se as estimativas obtidas para os parâmetros de especial interesse no modelo: os efeitos cruzados e desfasados entre dimensões de satisfação e os efeitos das variáveis explicativas (todos numa solução estandardizada):

74 Do Modelo Estrutural com Variáveis Manifestas ao Modelo Global

BETA						
	MATER02	LAZER02	MATER03	LAZER03	MATER04	LAZER04
MATER02	--	--	--	--	--	--
LAZER02	--	--	--	--	--	--
MATER03	0.726	0.047	--	--	--	--
LAZER03	0.010	0.704	--	--	--	--
MATER04	--	--	0.805	-0.002	--	--
LAZER04	--	--	0.105	0.661	--	--
MATER05	--	--	--	--	0.798	0.002
LAZER05	--	--	--	--	0.027	0.688

GAMMA				
	logRend	idade	sexo	n_horas
MATER02	0.248	0.102	-0.098	--
LAZER02	--	--	0.058	-0.089

O modelo aqui apresentado pode ser considerado como um ponto de partida: conforme se pode ver através dos valores-*t* apresentados no Anexo B.4 há parâmetros no modelo cujas estimativas não se revelaram significativas, tanto ao nível dos efeitos das variáveis explicativas da satisfação, como ao nível de efeitos cruzados. Note-se porém que todos os efeitos desfasados se revelaram significativos. Poder-se-ia pois estimar um novo modelo sem incluir todos ou alguns dos parâmetros associados a estimativas não significativas no modelo actual.

Poder-se-ia ainda testar a hipótese de invariância ao longo do tempo, quer do modelo factorial, quer por exemplo dos efeitos desfasados (já que a maioria dos efeitos cruzados não se revelou significativa). Por exemplo, testar a invariância dos efeitos desfasados significa comparar o modelo agora apresentado com um outro modelo que postula os três coeficientes de regressão entre as medidas repetidas de cada uma das dimensões de satisfação como sendo iguais. O novo modelo teria menos 4 parâmetros a estimar, e conseqüentemente mais 4 graus de liberdade, e poder-se-ia aferir da invariância dos coeficientes de regressão através do teste da diferença de qui-quadrado para modelos encaixados (*nested models*). Procedimento análogo vai ser seguido para testar a invariância do modelo factorial.

Testar a invariância factorial ao longo do tempo

No caso de estudos longitudinais em que há medidas repetidas de variáveis latentes é importante questionar e testar a invariância dos factores, i.e., a estabilidade do modelo de medida ao longo do tempo. Está em causa perceber se o constructo medido em diferentes ocasiões permanece o mesmo, ou se a sua estrutura muda. É possível testar a invariância da estrutura factorial impondo restrições de igualdade na estrutura dos pesos factoriais e testando depois a significância da diferença obtida nos valores de χ^2 dos dois modelos (com e sem restrições de invariância).

O modelo com restrições terá menos parâmetros, mais graus de liberdade e maior valor de teste da estatística de χ^2 . Se $\Delta\chi^2$ for maior que o valor crítico de uma distribuição de qui-quadrado com graus de liberdade dados por Δdf é porque a diferença entre modelos é significativa, sendo de rejeitar a hipótese nula de invariância. Se $\Delta\chi^2$ for inferior ao valor crítico é porque a diferença entre modelos não é significativa e portanto as restrições de invariância impostas nos pesos factoriais são plausíveis, devendo optar-se pelo modelo com restrições de invariância.

O procedimento descrito pode ser adoptado para modelos encaixados, no caso de amostras grandes. Para mais detalhes sobre o teste de diferenças de qui-quadrado ver Bollen (1989a, páginas 292–295).

Capítulo 4

Metodologia de Modelação

O objectivo deste capítulo é apresentar as etapas de uma metodologia de modelação com recurso a Modelos com Equações Estruturais (Secção 4.1), em particular no que à estimação (Secção 4.2) e à avaliação da qualidade de ajustamento modelo-dados (Secção 4.3) diz respeito.

Para mais detalhes sobre abordagens à metodologia de modelação ver, por exemplo, Hoyle (1995, capítulos 1 a 5) ou Hair et al. (2010, capítulos 11 e 12). Para uma exposição detalhada sobre aspectos práticos a ter em linha de conta na modelação ver Bentler e Chou (1987) e Breckler (1990).

4.1 Etapas na Modelação

Em termos de estratégia de modelação, desenvolvido que esteja o modelo substantivo de base teórica que se deseja estimar, e uma vez especificadas as hipóteses a testar, importa construir o diagrama do modelo, deixando claro quais as variáveis latentes e as relações estruturais (de dependência) entre elas, e quais os indicadores a usar para medir cada uma das variáveis latentes do modelo.

Note-se que, no que ao **desenvolvimento de um modelo de base teórica** diz respeito, a convicção com que o investigador formula as hipóteses e postula relações de dependência entre construtos deve basear-se na revisão de literatura efectuada e no conhecimento teórico-substantivo da temática. Tratando-se de uma abordagem confirmató-

ria, o investigador deve ser guiado por um suporte teórico que forneça uma explicação consistente e global do fenómeno em estudo, e não pelos dados empíricos de que dispõe. Não será excessivo enfatizar que é crucial ao longo de todas as etapas do processo de modelação o conhecimento substantivo teórico da problemática em estudo por parte do analista. O modelo postulado deve ser conciso, procurando-se garantir o equilíbrio entre incluir no modelo todas as variáveis tidas por relevantes e a capacidade de interpretação dos resultados quando o número de variáveis é muito elevado (e o que tal pode exigir em termos de dimensão da amostra).

Importa depois procurar especificar as equações do modelo de medida e do modelo estrutural, procurando identificar cada um dos parâmetros do modelo a estimar. Definida que esteja a escala de medida de cada uma das variáveis latentes no modelo, importa aferir da **identificação do modelo** global, começando pela componente de medida e utilizando as regras de identificação apresentadas no Capítulo 3.

Face à estrutura de parâmetros especificada no modelo, é gerada uma matriz de variância-covariância implícita (dada pelas Equações 2.2, 3.3 ou 3.5 consoante o tipo de modelo). Pode porém acontecer que haja várias estruturas ou conjuntos de valores dos parâmetros a gerar a mesma matriz de variância-covariância implícita, caso em que se está perante estruturas equivalentes. Se um dado parâmetro assume um mesmo valor em todas as estruturas equivalentes diz-se que esse parâmetro está identificado. Se todos os parâmetros a estimar no modelo estiverem identificados e puderem ser explicitados de forma única em termos das variâncias e covariâncias (ou correlações) das variáveis manifestas, diz-se que o modelo está identificado.

Podem ser sintomas de problemas de identificação a incapacidade do software em inverter a matriz de variâncias-covariâncias (ou de correlações) amostrais; erros padrão elevados para um ou mais coeficientes estimados; estimativas não razoáveis ou mesmo inadmissíveis (tais como variâncias negativas); e correlações elevadas (0.9 ou mais) entre os coeficientes estimados. Se a matriz de informação que o pro-

grama utiliza durante a estimação for definida positiva, será quase certo que o modelo está identificado. Esta não é porém uma condição necessária e suficiente de identificação; a única garantia é a de que se a matriz for singular o modelo não está identificado. Restando ainda dúvidas, o analista pode estimar o modelo por diversas vezes, usando valores iniciais diferentes para o algoritmo de estimação, e ver se o modelo converge sempre para as mesmas estimativas dos parâmetros. Se tal não acontecer é porque há problemas de identificação. A solução para problemas de identificação passa por impôr mais restrições ao modelo, isto é, por eliminar, de forma gradual, alguns dos parâmetros a estimar.

Face ao tipo de dados recolhidos e à dimensão da amostra, e após a análise exploratória dos dados, importa escolher o programa estatístico e o método de estimação a usar e proceder à estimação do modelo. Uma cuidada **análise exploratória dos dados** possibilita ao investigador uma familiarização com as suas características e problemas: As variáveis manifestas são quantitativas, categoriais ordinais ou meramente nominais? Existem valores omissos e em que proporção? Como se vai lidar com eles (irá ser usada alguma forma de imputação)? Existem valores atípicos de resposta (*outliers*)? Qual a distribuição observada das respostas? Qual a dimensão da amostra que vai efectivamente ser usada? Face à escala das variáveis faz sentido calcular variâncias e covariâncias (ou correlações de Pearson) e testar se existe normalidade multivariada da sua distribuição conjunta? Ou está-se perante variáveis ordinais, devendo ser calculadas correlações policóricas? Note-se que os Modelos com Equações Estruturais foram inicialmente concebidos para usar matrizes de variância-covariância (daí a designação “análise de estruturas de covariância”). Todavia, em muitas aplicações nas áreas das Ciências Sociais as unidades de medida das variáveis observadas são, muitas vezes, arbitrárias e irrelevantes, caso em que deve ser usada uma matriz de correlações. Procedimento análogo deve ser adoptado quando, num mesmo modelo, há simultaneamente variáveis quantitativas e variáveis ordinais.

O software a utilizar permite fazer a distinção do tipo de escala das variáveis? O LISREL e o Mplus sim, o AMOS não! Face às respostas a todas estas questões, que método de estimação pode e deve ser escolhido? A questão da **estimação do modelo** será abordada com mais detalhe na Secção 4.2.

Uma vez estimado o modelo há que aferir a qualidade do ajustamento modelo-dados, usando para tal medidas de ajustamento global do modelo e das suas diferentes componentes. Importa ainda verificar se todas as estimativas obtidas são admissíveis do ponto de vista estatístico e se fazem sentido do ponto de vista substantivo, face ao modelo postulado. A questão da **qualidade do ajustamento** será analisada com mais detalhe na Secção 4.3.

Compete depois ao analista interpretar as estimativas e o modelo obtido e, eventualmente, modificá-lo. Uma análise cuidadosa da solução obtida implica uma **análise dos resíduos**, bem como dos valores dos índices de modificação e das mudanças esperadas nos parâmetros em causa (se forem sujeitos a estimação). Os resíduos ajustados representam a diferença entre a matriz de variância-covariância (ou de correlações) amostral, S , e a estimada ou reproduzida pelo modelo, $\hat{\Sigma}$. O LISREL calcula ainda resíduos padronizados, de fácil interpretação, uma vez que não estão dependentes das unidades de medida das variáveis. Um resíduo positivo padronizado elevado significa que o modelo sub-estima a covariância (ou a correlação) entre as 2 variáveis em questão, ao passo que um resíduo negativo padronizado elevado indica que o modelo a sobre-estima. No primeiro caso, importará melhorar o modelo introduzindo relações que expliquem a covariância (ou a correlação) entre as duas variáveis, enquanto que no segundo caso não deve ser sujeito a estimação o parâmetro correspondente.

O LISREL fornece informação acerca da distribuição amostral dos resíduos padronizados, sob a forma gráfica, através de dois tipos de gráficos: o diagrama de caule-e-folhas e o gráfico quantil-quantil. Um “bom” ajustamento do modelo é caracterizado por um diagrama

de caule-e-folha em que os resíduos se distribuem aleatoriamente em torno do zero, estando a maioria no centro e poucos nas abas, e por um gráfico Q-Q em que os valores se distribuem aleatoriamente em torno da linha dos 45.

Os programas estatísticos que possibilitam a estimação de Modelos com Equações Estruturais calculam e disponibilizam **índices de modificação** e as mudanças esperadas nos valores das estimativas caso os parâmetros fossem sujeitos a estimação. O índice de modificação mede o decréscimo esperado na estatística de χ^2 (ou seja, a redução na discrepância entre o modelo e os dados) se o parâmetro em causa for sujeito a estimação. Em estimação ML, GLS e WLS o índice de modificação tem uma distribuição χ^2 com um grau de liberdade, desde que o modelo inicial tenha ajustamento razoável e a amostra seja suficientemente grande. O maior índice de modificação corresponderá ao parâmetro que, uma vez estimado, mais contribuirá para a melhoria do ajustamento do modelo, medida pela redução na estatística χ^2 , redução essa aproximadamente igual ao valor do índice de modificação (Sörbom, 1989).

Os autores dos programas defendem a seguinte metodologia de utilização dos índices de modificação:

- se o valor da estatística de χ^2 for grande face aos graus de liberdade, devem ser examinados os índices de modificação e deve ser sujeito a estimação o parâmetro que tiver maior índice associado, desde que tal faça sentido, do ponto de vista teórico e substantivo;
- se teoricamente não fizer sentido sujeitar a estimação esse parâmetro, deve procurar-se o segundo maior índice de modificação, e assim sucessivamente, até já não haver índices de modificação elevados ou parâmetros que faça sentido sujeitar a estimação. Se os sinais de determinados parâmetros forem, à priori, especificados como sendo positivos ou negativos, as mudanças esperadas nas estimativas dos parâmetros calculadas podem ser

82 Metodologia de Modelação

usadas como guia para excluir alterações de modelos em que o sinal da mudança seja oposto ao pretendido.

Todavia, ao longo dos anos a literatura tem sido muito crítica em relação à utilização dos índices de modificação (ver, por exemplo, MacCallum et al., 1992 e Saris et al., 1987), com os autores dos programas a argumentar que estes índices podem ser usados para detectar modelos equivalentes e para dar um contributo importante para a detecção de erros de especificação do modelo (ver Jöreskog e Sörbom, 1990), e com os críticos a chamar a atenção para o problema da *capitalization on chance* como um processo *data driven* (conduzido pelas características dos dados em análise) e também para o facto de o cálculo da estatística de χ^2 ser diferentemente sensível a diferentes parâmetros do modelo devido à potência do teste.

Em caso de modificação do modelo inicialmente postulado será conveniente que se procure proceder à **validação do modelo** final escolhido. A literatura sugere três possíveis abordagens:

- Embora não seja um procedimento frequentemente utilizado, pode usar-se uma amostra para estudar o modelo teórico e uma outra para validar o modelo. Pode depois usar-se uma abordagem multi-grupos (ver Secção 5.1), com diferentes amostras, para testar a invariância do modelo proposto.
- Porque replicar o estudo com um outro conjunto de dados é muitas vezes proibitivo por restrições de tempo, custo, ou outras, a alternativa pode passar por um procedimento de **validação cruzada**: dividir aleatoriamente a amostra inicial em estudo em duas, e usar uma subamostra para estimar o modelo (amostra de estimação) e a outra subamostra para comparar os resultados (amostra de validação), naturalmente desde que a dimensão da amostra inicial o permita. Cudeck e Browne (1983) propõem o cálculo de um **CVI - split sample Cross-Validation Index**. Para mais detalhes ver Scumacker e Lomax (2010, páginas 229–234).

- Quando os dois procedimentos já referidos não são exequíveis, pode tentar-se uma **abordagem *bootstrap***. Uma subamostra aleatória dos dados é considerada um substituto da população (uma pseudo-população) e a partir dela é feita reamostragem, sendo depois estimados os modelos e obtidas as estimativas e os erros padrão para as diferentes subamostras geradas. O estimador *bootstrap* é obtido como uma média das estimativas, sendo ainda calculado um “intervalo de confiança” em torno da média das estimativas obtidas por reamostragem. Este estimador, e o “intervalo de confiança” que lhe está associado, são utilizados para aferir a estabilidade das estimativas obtidas para o modelo proposto, face à amostra em estudo. Schumacker e Lomax (2010, páginas 234–240) ilustram como usar este procedimento com recurso ao PRELIS.

Note-se ainda que o facto de se ter chegado a um bom ajustamento modelo - dados, e a um modelo em que todos os parâmetros são significativos e as suas estimativas são interpretáveis do ponto de vista substantivo, não garante que se tenha conseguido “o melhor modelo”, e muito menos o “modelo verdadeiro” já que modelos são sempre representações mais ou menos aproximadas da realidade. Os Modelos com Equações Estruturais têm pois as mesmas **limitações** que outras metodologias empíricas: só se pode refutar um modelo, nunca se pode provar que um modelo é válido; um bom ajustamento modelo-dados não significa que se tenha encontrado o “modelo verdadeiro”.

4.2 Estimação do Modelo com Equações Estruturais

4.2.1 Objectivo da Estimação

O objectivo geral da estimação é encontrar estimativas para os parâmetros do modelo especificado que reproduzam, tanto quanto pos-

sível, a matriz de variância-covariância amostral das variáveis manifestas, \mathbf{S} . A matriz de variância-covariância da população, $\mathbf{\Sigma}$, está relacionada com os parâmetros da população através da equação de covariância (que, como já vimos, varia segundo o tipo de modelo especificado - Equações 2.2, 3.3 ou 3.5).

Trata-se pois de encontrar estimativas para os parâmetros sujeitos a estimação das matrizes $\mathbf{\Lambda}_X$, $\mathbf{\Lambda}_Y$, \mathbf{B} , $\mathbf{\Gamma}$, $\mathbf{\Phi}$, $\mathbf{\Psi}$, $\mathbf{\Theta}_\delta$ e $\mathbf{\Theta}_\epsilon$ que, face ao modelo especificado, gerem uma matriz de variância-covariância estimada, $\hat{\mathbf{\Sigma}}$, que esteja tão próximo quanto possível da matriz \mathbf{S} . Para definir a proximidade de $\mathbf{\Sigma}$ a \mathbf{S} é requerida uma função de ajustamento (*fit function*), designada por $F(\mathbf{S}, \mathbf{\Sigma})$. Quando a matriz de variância-covariância implícita ao modelo, $\mathbf{\Sigma}$, é substituída, na função de ajustamento, pela sua estimativa, $\hat{\mathbf{\Sigma}}$, obtém-se o valor da função de ajustamento $F(\mathbf{S}, \hat{\mathbf{\Sigma}})$ para esse conjunto de estimativas dos parâmetros.

As funções de ajustamento que se apresentam possuem as seguintes propriedades: $F(\mathbf{S}, \mathbf{\Sigma})$ é um escalar; $F(\mathbf{S}, \mathbf{\Sigma}) \geq 0$; $F(\mathbf{S}, \mathbf{\Sigma}) = 0$ se e só se $\mathbf{\Sigma} = \mathbf{S}$ (o valor da função de ajustamento é nulo se e só se o ajustamento for perfeito); e $F(\mathbf{S}, \mathbf{\Sigma})$ é contínua em \mathbf{S} e $\mathbf{\Sigma}$. Dependendo da forma como é definida a relação entre distância de matrizes, parâmetros no modelo e variâncias e covariâncias amostrais, podem ser obtidas diferentes funções de ajustamento, associadas a diferentes métodos de estimação de modelos.

Os métodos de estimação que se apresentam nesta Secção são iterativos, necessitando de valores de inicialização para desencadear os cálculos. As estimativas iniciais podem ser obtidas através dos métodos IV (*Instrumental Variables*) e TSLS (*Two-Stage Least Squares*). Embora as estimativas resultantes não sejam assintoticamente eficientes, são consistentes e muito fáceis de calcular. Se o modelo em análise se ajustar bem aos dados, os valores iniciais calculados estão frequentemente tão próximo da solução final que basta um pequeno número de iterações para que a convergência e a solução final sejam atingidas.

A cada um dos procedimentos iterativos de estimação, que minimizam a função de ajustamento, está associada uma **matriz de informação** para os parâmetros, cuja ordem é igual ao número de parâmetros a estimar no modelo. Os elementos da matriz de informação são os valores esperados das segundas derivadas da função de ajustamento no ponto de solução. A inversa da matriz de informação contém na diagonal principal as variâncias assintóticas, sendo os restantes elementos as covariâncias assintóticas entre todos os pares possíveis de estimativas dos parâmetros. Os erros padrão das estimativas dos parâmetros igualam a raiz quadrada dos elementos da diagonal principal da inversa da matriz de informação. As correlações entre as estimativas são obtidas dividindo os elementos fora da diagonal principal da inversa da matriz de informação pelos correspondentes pares de erros padrão. Como foi dito anteriormente, se a matriz de informação não for definida positiva, o modelo não está identificado. Para uma exposição detalhada sobre não convergência, soluções impróprias e escolha de estimativas iniciais ver Boomsma (1985).

4.2.2 Funções de Ajustamento

A maioria dos programas estatísticos para estimar Modelos com Equações Estruturais incorpora quatro métodos de estimação principais: ULS (*Unweighted Least Squares*), ML (*Maximum Likelihood*), GLS (*Generalised Least Squares*) e métodos ADF (*Asymptotically Distribution Free*), designadamente o WLS (*Weighted Least Squares*).

O estimador de mínimos quadrados não ponderados **ULS** (*Unweighted Least Squares*) minimiza a função

$$F_{ULS} = \frac{1}{2} tr \left[(\mathbf{S} - \mathbf{\Sigma})^2 \right],$$

em que $tr \left[(\mathbf{S} - \mathbf{\Sigma})^2 \right]$ representa o traço da matriz dos termos residuais $\mathbf{S} - \mathbf{\Sigma}$, de ordem $p + q$, sendo \mathbf{S} a matriz das variâncias e covariâncias amostrais e $\mathbf{\Sigma}$ a matriz de variância-covariância implícita ao modelo.

Trata-se de um estimador consistente e relativamente fácil de calcular, sem ser preciso estabelecer nenhum pressuposto acerca da distribuição das variáveis observadas. Não é porém o estimador assintoticamente mais eficiente (o estimador de máxima verosimilhança tem maior eficiência, isto é, gera variâncias menores em grandes amostras). Também não é um estimador independente da escala de medida das variáveis: os valores mínimos de F_{ULS} diferem com mudanças de escala, e até pelo simples facto de se analisar uma matriz de variância-covariância em vez de uma matriz de correlações.

O estimador de mínimos quadrados generalizados **GLS** (*Generalized Least Squares*) minimiza a função de ajustamento

$$F_{GLS} = \frac{1}{2} \text{tr} \left[(\mathbf{I} - \mathbf{\Sigma} \mathbf{S}^{-1})^2 \right].$$

Trata-se de minimizar a soma dos quadrados dos resíduos ponderados pela inversa da matriz \mathbf{S} . O estimador ULS implicitamente pondera todos os elementos de $\mathbf{S} - \mathbf{\Sigma}$ como se eles tivessem iguais variâncias e covariâncias com todos os outros elementos. Tem-se então um problema similar ao que ocorre quando, em análise de regressão, se aplica o estimador OLS a uma equação em que há heterocedasticidade ou autocorrelação dos resíduos. Nesses casos de análise de regressão deve usar-se o GLS, que pondera as observações por forma a corrigir as desiguais variâncias ou as covariâncias não nulas entre os resíduos. Analogamente se aplica aqui uma função de ajustamento GLS, que pondera os elementos de $\mathbf{S} - \mathbf{\Sigma}$ de acordo com as suas variâncias e covariâncias com todos os outros elementos, usando para tal como matriz de ponderação \mathbf{S}^{-1} . O estimador GLS é independente da escala de medida das variáveis e é um estimador consistente e assintoticamente eficiente (embora pouco se saiba do seu comportamento quando as amostras são pequenas). Para que estas propriedades assintóticas se verifiquem é pressuposto que a distribuição conjunta das variáveis manifestas seja normal multivariada.

O estimador de máxima verosimilhança **ML** (*Maximum Likelihood*) minimiza a função:

$$F_{ML} = \log |\Sigma| + tr(\mathbf{S}\Sigma^{-1}) - \log |\mathbf{S}| - (p + q),$$

desde que \mathbf{S} e Σ sejam matrizes definidas positivas (não singulares). É pressuposto que as variáveis manifestas \mathbf{X} e \mathbf{Y} tenham distribuição conjunta normal multivariada para que F_{ML} tenha as desejadas propriedades assintóticas.

Destacam-se como principais **propriedades** do estimador F_{ML} o ser consistente, assintoticamente não enviesado e assintoticamente eficiente (de entre os consistentes é o que tem menor variância assintótica). F_{ML} é um estimador independente da escala de medida das variáveis, pelo que não há consequências pelo facto de as variáveis manifestas terem diferentes unidades de medida, ou pelo facto de se mudarem as unidades de medida de uma ou mais variáveis manifestas. Estas propriedades são, como vimos, assintóticas, i.e., verificam-se para dimensões de amostra a tender para infinito. A questão que se põe é a de saber qual deverá ser a dimensão (mínima) da amostra para que seja possível tirar o desejado partido destas propriedades. Não há, infelizmente, resposta inequívoca e definitiva a esta questão. Autores há que defendem um mínimo recomendável de 5 a 10 observações por cada parâmetro a estimar no modelo. Todavia, o analista deve estar ciente de que quanto mais complexos forem os modelos e mais parâmetros tiverem, maior deverá ser a dimensão da amostra. Com efeito, pode acontecer que 200 observações sejam suficientes para estimar por máxima verosimilhança um modelo muito simples, com poucos parâmetros e poucas variáveis, sendo os indicadores manifestos quantitativos, com distribuição conjunta normal multivariada. Mas tal situação raramente ocorre com dados reais, pelo que os analista terá de recorrer a um método de estimação com pressupostos mínimos acerca da distribuição conjunta das variáveis, o que implica necessariamente uma amostra de grande dimensão.

O estimador ML (à semelhança do GLS) fornece um teste de quali-

dade do ajustamento, para modelos com graus de liberdade positivos. A distribuição assintótica de $(N - 1) \times F_{ML}$ é uma distribuição de χ^2 com $\frac{1}{2}(p + q)(p + q - 1) - t$ graus de liberdade (sendo t o número total de parâmetros a estimar). Em amostras grandes os valores da estatística de χ^2 estimados por ML e por GLS devem estar próximos.

Os estimadores ULS, GLS e ML podem ser considerados casos particulares de uma função mais geral - a de mínimos quadrados ponderados - **WLS (Generally Weighted Least Squares)**, com $F_{WLS} =$

$$[\mathbf{s} - \boldsymbol{\sigma}]^T \mathbf{W}^{-1} [\mathbf{s} - \boldsymbol{\sigma}] = \sum_{g=1}^k \sum_{h=1}^g \sum_{i=1}^k \sum_{j=1}^i w^{gh,ij} (s_{gh} - \sigma_{gh}) (s_{ij} - \sigma_{ij}),$$

sendo:

- $\mathbf{s}^T = [s_{11}, s_{21}, s_{22}, s_{31}, \dots, s_{kk}]$ o vector dos elementos da parte inferior da matriz de variância-covariância amostral \mathbf{S} , de ordem $k \times k$ (incluindo os elementos da diagonal principal), usados para ajustar o modelo (com $k = p + q$);
- $\boldsymbol{\sigma}^T = [\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \dots, \sigma_{kk}]$ o vector dos elementos correspondentes na matriz de variância-covariância implícita ao modelo $\boldsymbol{\Sigma}$;
- $w^{gh,ij}$ é o elemento tipo de uma matriz de pesos \mathbf{W}^{-1} , definida positiva, de ordem $\frac{k(k+1)}{2} \times \frac{k(k+1)}{2}$. A forma mais usual de escolher \mathbf{W} , de elemento tipo $w_{gh,ij}$, é deixar $w_{gh,ij}$ ser uma estimativa consistente da covariância assintótica entre s_{gh} e s_{ij} . Browne (1984) demonstrou que, se for esse o caso, F_{WLS} é assintoticamente eficiente;
- a covariância assintótica entre s_{ij} e s_{gh} é dada por

$$\text{Acov}(s_{ij}, s_{gh}) = N^{-1}(\sigma_{ijgh} - \sigma_{ij}\sigma_{gh}),$$

com $\sigma_{ijgh} = E(X_i - \mu_i)(X_j - \mu_j)(X_g - \mu_g)(X_h - \mu_h)$, o momento de quarta ordem em torno da média, sendo σ_{ij} e σ_{gh} as covariâncias populacionais de X_i com X_j e de X_g com X_h , respectivamente. Para mais detalhes ver Bollen (1989a, páginas 425-428).

Principais vantagens da utilização do estimador WLS:

- são mínimos os pressupostos acerca da distribuição das variáveis: estamos perante um estimador ADF - *Asymptotically Distribution Free*;
- desde que a matriz de pesos seja escolhida por forma a que os seus elementos sejam estimadores consistentes das covariâncias assintóticas, o estimador F_{WLS} é eficiente e fornece um teste estatístico com distribuição de χ^2 à qualidade de ajustamento do modelo.

Principais desvantagens do estimador WLS:

- para que se dê a convergência, aquando da estimação, é requerida uma amostra de dimensão superior à que seria necessária com F_{ML} , F_{ULS} ou F_{GLS} . Com efeito, a estimação WLS requer amostras muito grandes;
- dificuldades práticas do ponto de vista do cálculo: requer a inversão de uma matriz quadrada de ordem $\frac{1}{2}(p+q)(p+q+1)$, o que se pode tornar quase impraticável se o número de variáveis manifestas for muito elevado (porque requer muito tempo e especialmente memória de computador) - note-se que o número de elementos da matriz aumenta rapidamente com o número de variáveis. Porque está em causa estimar momentos de quarta ordem, para que a precisão seja aceitável são requeridas amostras muito grandes, o que por vezes se torna uma dificuldade prática;
- não é claro o quanto severa deve ser a não normalidade para exigir estimação WLS, isto é, não é claro que o estimador F_{WLS} seja efectivamente melhor que o F_{ML} ou o F_{GLS} quando a não normalidade das variáveis observadas é apenas moderada, sobretudo se as amostras não forem muito grandes.

DWLS - *Diagonally Weighted Least Squares*

O cálculo da matriz de pesos \mathbf{W}^{-1} usada em WLS requer, como vimos, elevado tempo de processamento e grande disponibilidade de memória de computador se o número de variáveis for elevado. Em LISREL é proposta uma aproximação alternativa que consiste em calcular somente as variâncias assintóticas dos coeficientes estimados.

Seja w_{gh} uma estimativa da variância assintótica de s_{gh} . Estas estimativas podem ser usadas com uma função de ajustamento do tipo:

$$F_{DWLS} = \sum_{g=1}^k \sum_{h=1}^g \left(\frac{1}{w_{gh}} \right) (s_{gh} - \sigma_{gh})^2$$

Trata-se afinal de usar uma matriz de pesos, \mathbf{W}^{-1} , tal como em F_{WLS} . Contudo, no caso de F_{DWLS} , a matriz \mathbf{W}^{-1} é uma matriz diagonal com elementos $\frac{1}{w_{gh}}$.

Embora esta função não conduza a estimativas assintoticamente eficientes dos parâmetros do modelo, estamos perante uma solução de compromisso entre o ULS e o WLS. O DWLS pode também ser usado quando são analisadas matrizes de correlação.

Estimação WLS *versus* Robust ML

Quando as variáveis manifestas quantitativas têm uma distribuição conjunta que se afasta da normal multivariada põe-se a questão de que método de estimação usar: WLS ou um método “robusto” de máxima verosimilhança (designado por *Robust Maximum Likelihood*). Para uma síntese sobre estimação de modelos no caso de não normalidade em variáveis contínuas, ou na presença de variáveis ordinais, ver Kaplan (2000, páginas 80–87).

Quando a distribuição das variáveis manifestas se afasta da normal, por assimetria elevada ou curtose excessiva, as estimativas obtidas por ML, os erros padrão e a estatística de teste de χ^2 poderão não ser robustos. No caso de grandes desvios da normalidade deve recorrer-se a

estimação WLS o que, como já vimos, requer uma amostra de elevada dimensão para que a matriz das variâncias e covariâncias assintóticas das variâncias e covariâncias (ou correlações) amostrais, a ser usada como matriz de pesos na estimação WLS, possa ser estimada com rigor. Mas, e se a amostra não for suficientemente grande? E se as variáveis só se desviarem ligeiramente da normalidade? Enquanto que a violação da assimetria não parece afectar grandemente a consistência do estimador F_{ML} , uma curtose excessiva usualmente elimina a eficiência assintótica, tornando a matriz de variância-covariância estimada e o valor de χ^2 potencialmente não rigorosos.

No caso de a dimensão da amostra não ser suficientemente grande para permitir usar WLS, é recomendado usar o método *Robust ML*, que pode ser considerado um tipo de método ADF. O método *Robust ML* foi inicialmente proposto por Browne (1987) para análise factorial e modelos afins, e foi extendido por Satorra e Bentler (1988, 1994) para modelos de estrutura de covariância. Tal método encontra-se implementado na versão 8.8 do LISREL. Trata-se afinal de utilizar o método de máxima verosimilhança para estimar os parâmetros do modelo, sendo os erros padrão, os valores das estatísticas de teste t e o valor da estatística de χ^2 corrigidos, usando para tal a matriz de variâncias e covariâncias assintóticas das variâncias e covariâncias (ou das correlações) amostrais. Estudos de simulação revelam que, para amostras não muito grandes, a estimação *Robust ML* parece mais indicada do que a estimação WLS.

Quando as variáveis manifestas são ordinais assume-se que um respondente que seleccionou uma determinada categoria de resposta tem mais dessa característica do que se tivesse seleccionado uma categoria inferior, mas não é verdadeiramente possível quantificar o quanto mais. Jöreskog (2005) defende ser errada a prática, muitas vezes corrente de tratar variáveis ordinais como se de quantitativas se tratasse (com propriedades métricas). O cálculo de correlações de Pearson e a estimação por ML de modelos em que haja uma ou mais variáveis categoriais pode levar a estimativas dos parâmetros altamente dis-

torcidas e a valores incorrectos da estatística χ^2 e dos erros padrão. Devem antes calcular-se correlações policóricas (entre duas variáveis latentes contínuas, subjacentes às duas variáveis manifestas ordinais) e correlações *polyserial* (entre uma variável manifesta contínua e a variável latente contínua subjacente à manifesta ordinal) Para efeitos de estimação deve recorrer-se à estimação WLS, com uma correcta matriz de pesos ou, se a dimensão da amostra para tal não for suficiente, recorrer à estimação *Robust ML* .

No LISREL, quando é utilizado o método de estimação *Robust ML* (bastando para tal indicar ML como método de estimação e indicar o ficheiro que contém a matriz de covariância assintótica), é disponibilizado no output o valor da *Satorra-Bentler scaled chi-square statistic*, devendo ser esse o valor da medida de χ^2 de ajustamento global a ser usado. Para mais detalhes sobre cálculo dos erros padrão e da estatística de χ^2 em estimação *Robust ML* ver Jöreskog et al. (2001, páginas 179–182 e Apêndice A).

4.3 Qualidade de Ajustamento Modelos-Dados

Uma vez obtidas as estimativas dos parâmetros, importa analisar o ajustamento entre o modelo estimado e os dados disponíveis. Trata-se de um assunto que nem sempre reúne consenso, sendo por isso vasta a literatura publicada nesta área. Ao longo dos anos muitos foram os que se dedicaram a investigar nesta área, tendo desenvolvido e proposto um grande número de medidas de avaliação da qualidade de ajustamento, basicamente em três perspectivas diferentes: medidas de ajustamento absoluto (*absolute fit indices*), medidas de ajustamento incremental (*incremental fit indices*), e medidas de parcimónia (*parsimony fit indices*) e de comparação de modelos. Para além do ajustamento global do modelo (Secção 4.3.1) importa ainda analisar o ajustamento das diferentes componentes do modelo (Secção 4.3.2).

4.3.1 Medidas de Ajustamento Global do Modelo

Designaremos o modelo em análise por *modelo proposto* e o modelo de independência (em que as covariâncias entre variáveis são todas nulas) por *modelo nulo*. O modelo nulo vai servir como base de comparação no caso das medidas de ajustamento incremental. Para uma apresentação síntese das medidas de ajustamento e valores considerados aceitáveis ver Schumacker e Lomax (2010, capítulo 5).

i) Medidas de Ajustamento Absoluto

Está em causa a hipótese nula de que $\Sigma = \Sigma(\theta)$, isto é, que a matriz de variância-covariância da população é igualada à matriz implícita ao modelo (já vimos que o modelo estatístico e os seus pressupostos implicam uma determinada estrutura de covariância $\Sigma(\theta)$ para as variáveis observadas). Como os parâmetros da população são desconhecidos, na prática usa-se S , a matriz de variância-covariância amostral, em vez de Σ . O ajustamento do modelo será perfeito quando $S = \hat{\Sigma}(\hat{\theta})$, i.e., quando a matriz de variância-covariância amostral for igualada pela matriz de variância-covariância estimada / reproduzida pelo modelo em teste.

Estatística de χ^2

A estatística de χ^2 iguala $(N - 1) \times F$, sendo N a dimensão de amostra e F o valor mínimo da função de ajustamento para o modelo especificado. Como vimos anteriormente, F , e a função de ajustamento, variam consoante o método de estimação utilizado.

Pressupõe-se que, se o modelo estiver correctamente especificado, se a amostra for suficientemente grande, se for analisada uma matriz de variância-covariância e se as variáveis observadas seguirem distribuição normal multivariada, $(N - 1) \times F$ tem distribuição assintótica de χ^2 . Verificando-se os pressupostos enumerados, a estatística de χ^2 pode ser usada como um teste estatístico, sendo $H_0 : \Sigma = \Sigma(\theta)$, isto

94 Metodologia de Modelação

é, testa a hipótese de que o modelo especificado reproduz a estrutura da variâncias e covariâncias na população. Os graus de liberdade são dados por $df = \frac{1}{2}(p + q)(p + q + 1) - t$, sendo $p + q$ o número total de variáveis observadas, e t o número total de parâmetros a estimar. Para não se rejeitar a hipótese nula de que o modelo especificado reproduz a estrutura de associações na população, o valor- p tem que ser maior ou igual ao nível de significância α desejado.

GFI - Goodness-of-Fit Index

Trata-se de uma medida que varia entre zero (fraco ajustamento) e um (ajustamento perfeito). Mede a quantidade relativa de variância e covariância conjuntamente explicadas pelo modelo, comparando para tal o valor da estatística χ^2 do modelo proposto com o valor da estatística χ^2 do modelo nulo:

$$GFI = 1 - \frac{\chi_{proposto}^2}{\chi_{nulo}^2}.$$

RMR - Root Mean Square Residual

Um resíduo individual é dado por $(s_{ij} - \hat{\sigma}_{ij})$, sendo s_{ij} um elemento de \mathbf{S} e $\hat{\sigma}_{ij}$ um elemento de $\hat{\Sigma}(\hat{\theta})$. Se o resíduo for positivo é porque o modelo subestima a covariância entre duas variáveis, se for negativo é porque a covariância estimada é demasiado elevada. Idealmente todos os resíduos deviam estar próximos de zero. O *RMSR* é uma medida de média dos resíduos ajustados, dada por:

$$RMR = \left[2 \sum_{i=1}^{p+q} \sum_{j=1}^i (s_{ij} - \hat{\sigma}_{ij})^2 / (p + q)(p + q + 1) \right]^{1/2}.$$

Se for usada uma matriz de correlações, o RMR varia entre zero e um e idealmente deve assumir valores inferiores a 0.05; se for usada uma matriz de variância-covariância, o RMR é sempre positivo, não tendo

limite superior. Só pode pois ser interpretado em relação à dimensão das variâncias e covariâncias observadas em \mathbf{S} (o que se torna mais fácil se todas as variáveis observadas estiverem padronizadas) e não pode ser utilizado para comparação de modelos baseados em variáveis diferentes.

CN - *Critical N*

Medida proposta por Hoelter (1983), põe o ênfase na dimensão da amostra: trata-se de estimar a dimensão da amostra necessária para a “aceitação” de um dado modelo. Foi apresentada pelo autor como

$$CN = \frac{Z_{critical} + \sqrt{2df - 1}}{4F} + G,$$

em que G é o número de grupos analisado simultaneamente, F é o valor mínimo da função de ajustamento e $Z_{critical}$ é o valor crítico da distribuição normal para um dado nível de significância α escolhido.

A medida pode também ser apresentada como

$$CN = \frac{\chi_{critical}^2}{F} + 1,$$

em que o numerador representa o valor crítico da distribuição χ^2 , com df graus de liberdade, para um dado nível de significância α escolhido. A medida CN dá-nos a dimensão da amostra com a qual o valor F levaria a rejeitar a hipótese nula de que $\Sigma = \Sigma(\theta)$, face ao valor de α escolhido.

Hoelter sugere $CN \geq 200 \times G$, ou seja, que para dimensões da amostra superiores a 200 vezes o número de grupos analisado, o modelo reproduz adequadamente a estrutura de associações observada. Bollen e Liang (1988) discutem as propriedades do CN.

RMSEA - *Root Mean Square Error of Approximation*

Browne e Cudeck (1993) propõem um conjunto de medidas de ajustamento que tomam em consideração o erro de aproximação à popu-

lação e a precisão de medida em si mesma. Definem uma estimativa da função de discrepância da população (PDF) como

$$\widehat{PDF} = \text{Max} \left\{ F - \frac{df}{N-1}, 0 \right\},$$

sendo F o valor mínimo da função de ajustamento. Os autores sugerem a utilização de um intervalo de confiança a 90% para o erro de aproximação à população. Uma vez que \widehat{PDF} tende a decrescer à medida que são introduzidos parâmetros adicionais no modelo, os autores sugerem a utilização do *Root Mean Square Error of Approximation* (RMSEA), de Steiger (1990), como medida da discrepância por grau de liberdade. Assim:

$$RMSEA = \sqrt{\widehat{PDF}/df}.$$

Valores de $RMSEA < 0.05$ indicam bom ajustamento, sendo aceitáveis valores até 0.08. Para uma exposição mais detalhada ver Kaplan (2000, páginas 110–114).

ii) Medidas de Ajustamento Incremental

São medidas que comparam o modelo proposto com um outro modelo de referência, em geral o modelo nulo ou de independência. O modelo nulo fornece assim um padrão para um “mau” ajustamento, com o qual o modelo em análise é comparado.

NFI - *Normed Fit Index*

Medida proposta por Bentler e Bonett (1980), é uma das medidas de ajustamento mais usadas e pode ser definida como

$$NFI = \frac{\chi_{nulo}^2 - \chi_{proposto}^2}{\chi_{nulo}^2} = \frac{F_{nulo} - F_{proposto}}{F_{nulo}},$$

sendo χ_{nulo}^2 e F_{nulo} os valores da estatística de χ^2 e do mínimo da função de ajustamento para o modelo nulo, e $\chi_{proposto}^2$ e $F_{proposto}$

os correspondentes valores para o modelo proposto. O NFI mede a melhoria que se obtém no ajustamento ao passar do modelo nulo, de independência, para o modelo em análise.

O NFI varia entre zero e um, e quanto mais perto estiver de 1 melhor é o ajustamento (o valor 1 ocorre quando $\chi_{proposto}^2 = 0$, o que não é realista esperar. Esta medida é criticada por não entrar em linha de conta com o número de graus de liberdade do modelo e por depender da dimensão da amostra.

IFI - Incremented Fit Index

Bollen (1989b) propôs este índice que ajusta o NFI pelos graus de liberdade e pela dimensão da amostra. Assim,

$$IFI = \frac{\chi_{nulo}^2 - \chi_{proposto}^2}{\chi_{nulo}^2 - df_{proposto}} = \frac{F_{nulo} - F_{proposto}}{F_{nulo} - \frac{df_{proposto}}{N-1}}.$$

Valores de *IFI* muito abaixo de 1 indicam fraco ajustamento, valores muito acima de 1 podem indicar sobre-identificação; idealmente $IFI \approx 1$.

RFI - Relative Fit Index

É similar ao NFI, mas entra em linha de conta com os graus de liberdade:

$$\rho_1 = RFI = \frac{\frac{\chi_{nulo}^2}{df_{nulo}} - \frac{\chi_{proposto}^2}{df_{proposto}}}{\frac{\chi_{nulo}^2}{df_{nulo}}} = \frac{\frac{F_{nulo}}{df_{nulo}} - \frac{F_{proposto}}{df_{proposto}}}{\frac{F_{nulo}}{df_{nulo}}}.$$

O RFI varia entre zero e um, e quanto maior o seu valor melhor é o ajustamento do modelo.

TLI - Tucker-Lewis Index

Tucker e Lewis (1973) desenvolveram um índice para modelos de análise factorial, que posteriormente foi extendido ao contexto de Modelos

com Equações Estruturais, que corrige IFI pelos graus de liberdade:

$$TLI = \frac{\frac{\chi_{nulo}^2}{df_{nulo}} - \frac{\chi_{proposto}^2}{df_{proposto}}}{\frac{\chi_{nulo}^2}{df_{nulo}} - 1} = \frac{\frac{F_{nulo}}{df_{nulo}} - \frac{F_{proposto}}{df_{proposto}}}{\frac{F_{nulo}}{df_{nulo}} - \frac{1}{N-1}}.$$

O TLI varia entre zero e um e idealmente deve aproximar-se de 1.

CFI - Comparative Fit Index

$$CFI = 1 - \frac{\chi_{proposto}^2 - df_{proposto}}{\chi_{nulo}^2 - df_{nulo}}.$$

Trata-se de um dos índices mais utilizados, até por não ser muito sensível à complexidade do modelo. Varia entre zero e um, sendo recomendáveis valores superiores a 0.9.

iii) Medidas de Parcimónia e de Comparação de Modelos

As medidas de parsimónia entram em linha de conta com o grau de complexidade do modelo e o número de parâmetros a estimar. O objectivo é ver se o ajustamento foi conseguido à custa de sobre-identificação, i.e., com um número excessivo de parâmetros. Alguns exemplos:

AGFI - Adjusted Goodness-of-Fit Index

Trata-se de ajustar o *GFI* pelos graus de liberdade do modelo:

$$AGFI = 1 - \frac{(p+q)(p+q+1)}{2df} \times (1 - GFI).$$

O AGFI varia entre zero e um, sendo recomendáveis valores superiores a 0.9.

χ^2/df

Trata-se de corrigir o valor de χ^2 pelos graus de liberdade (df). Com efeito, se o valor da medida χ^2 é grande face aos graus de liberdade do modelo, é porque será possível obter informação adicional a partir dos dados. Não é claro o critério de magnitude deste rácio, necessário para não refutar que o modelo em análise reproduz a matriz de variâncias-covariâncias da população. Alguns autores falam em valores inferiores a 2 ou 3. Trata-se, porém, de uma simples recomendação, sem consenso na literatura. Por outro lado é um rácio muito sensível à dimensão da amostra, tendendo a assumir valores elevados para amostras grandes.

PNFI - Parsimonious Normed Fit Index

O PNFI ajusta o NFI pela perda de graus de liberdade:

$$PNFI = \frac{df_{proposto}}{df_{nulo}} \times NFI_{proposto},$$

sendo $df_{nulo} = (p + q)(p + q - 1)/2$. Varia entre 0 e 1, sendo usualmente inferior ao NFI.

PGFI - Parsimonious Goodness-of-Fit Index

$$PGFI = \frac{df_{proposto}}{df_{nulo}} \times GFI_{proposto},$$

sendo $df_{nulo} = (p + q)(p + q + 1)/2$. O PGFI varia entre 0 e 1.

AIC - Akaike Information Criterion

O AIC é um critério de comparação de modelos, especialmente para modelos com diferente número de parâmetros (Akaike, 1987). É dado por:

$$AIC = (N - 1) \times F + 2t = \chi^2 + 2t,$$

sendo t o número total de parâmetros a estimar. Assim, *Model AIC* = $\chi^2_{proposto} + 2t$ e *Independence AIC* = $\chi^2_{nulo} + 2t$. Terá melhor ajustamento o modelo com o menor AIC.

tamento o modelo que tiver menor valor de AIC.

CAIC

Medida proposta por Bozdogan (1987) e conhecida pela versão consistente do AIC, ajusta o AIC pelo efeito dimensão da amostra e é dada por:

$$CAIC = (N - 1) \times F \times (1 + \ln N) \times t,$$

sendo t o número total de parâmetros a estimar. Deve escolher-se o modelo com menor CAIC.

ECVI - *single sample Cross Validation Index*

Medida proposta por Browne e Cudeck (1993), é dada por:

$$ECVI = \frac{\chi^2}{N - 1} + \frac{2t}{N - 1}.$$

O ECVI é uma medida do grau em que se espera que o modelo proposto se replique noutra amostra da mesma população. Deve escolher-se o modelo com menor valor de *ECVI*.

iv) Ajustamento do Modelo: Controvérsia e Consenso

A controvérsia acerca dos índices de ajustamento em Modelos com Equações Estruturais pode ser encarada como parte integrante de uma controvérsia maior, que tem a ver com a melhor forma de medir o ajustamento de um modelo estatístico.

Uma característica chave na literatura sobre Modelos com Equações Estruturais é o uso de um teste estatístico para a hipótese nula de que $\Sigma = \Sigma(\theta)$, sendo θ o vector dos parâmetros (livres ou a estimar) do modelo. Se o modelo estiver correctamente especificado e se os pressupostos de distribuição dos dados forem verificados, o analista pode usar o teste estatístico com uma distribuição assintótica de χ^2 ,

para testar a hipótese nula de que o modelo especificado reproduz exactamente a matriz de variância-covariância da população.

Cudeck e Browne (1983) criticaram esta hipótese nula, dizendo que os modelos são sempre aproximações da realidade, pelo que não será muito realista considerar $\Sigma = \Sigma(\theta)$, mas antes $\Sigma \approx \Sigma(\theta)$: em vez de se perguntar se o modelo em causa reproduz exactamente a matriz de variância-covariância da população deve antes perguntar-se se, de entre um conjunto de alternativas, o modelo em causa é o que mais se aproxima da realidade, uma vez que não é possível encontrar o “modelo correcto”, mas apenas a melhor aproximação.

Também Jöreskog chamou a atenção para a interpretação do teste estatístico de χ^2 dizendo que, na maior parte do trabalho empírico, um modelo deve ser encarado como uma aproximação da realidade, pelo que, do ponto de vista estatístico, o problema não deve ser o de testar uma dada hipótese (que à priori pode ser considerada falsa...), antes sim o de ajustar o modelo aos dados e decidir se o ajustamento é ou não adequado. A estatística de χ^2 deve então ser encarada, não como um teste estatístico, mas como uma medida da qualidade do ajustamento, sendo que valores elevados desta medida correspondem a fraco ajustamento e valores baixos correspondem a bom ajustamento. Os graus de liberdade servem como um padrão para se julgar se a estatística de χ^2 é grande ou pequena.

Isto encorajou a prática de dividir a estatística de χ^2 pelos graus de liberdade. Infelizmente, porém, o critério recomendado para um bom ajustamento varia largamente - até 2, ou até 5..., além de que se mantêm as outras críticas a fazer à estatística de χ^2 :

- é muito sensível à dimensão da amostra sendo que, para grandes amostras é grande a aptidão do teste para detectar desvios, ainda que pequenos, do modelo aos dados e indicar a rejeição da hipótese nula;
- o seu valor tende a decrescer quando novos parâmetros são adicionados ao modelo, o que pode levar o analista a, desneces-

sariamente, introduzir parâmetros no modelo com o objectivo de ver reduzida a estatística de χ^2 e, por essa via, melhorar o ajustamento.

A partir do trabalho de Tucker e Lewis (1973), Bentler e Bonett (1980) tomaram uma orientação diferente no desenvolvimento de medidas de qualidade de ajustamento, tendo surgido as medidas de ajustamento incremental. A ideia chave foi gerar um modelo base, com o qual o modelo em estudo e o seu teste estatístico pudessem ser comparados. Todavia, muitos destes índices têm a propriedade de aumentar pela incorporação de parâmetros adicionais (ou seja, consegue obter-se um melhor ajustamento do modelo pelo simples facto de se sujeitarem a estimação mais parâmetros, sendo o valor máximo do ajustamento atingido para o modelo saturado).

Surgem então as medidas de parcimónia, em que se combinam dois tipos de informações acerca do modelo: a qualidade do ajustamento e a parsimónia, uma vez que se entra em linha de conta com a perda de graus de liberdade. Se um dado índice indicia bom ajustamento e, ao tomar-se em consideração a parsimónia, se obtém um ajustamento moderado, é porque muito do bom ajustamento é devido à existência de muitos parâmetros sujeitos a estimação, não sendo explicado pelo modelo em questão. Deve ser objectivo do analista alcançar parsimónia, isto é, possuir o maior número possível de graus de liberdade, o que significa procurar um bom ajustamento à custa do menor número possível de parâmetros a estimar. Os índices de parsimónia são especialmente úteis para comparação de modelos, particularmente quando o número de parâmetros e de graus de liberdade é diferente.

Se é verdade que os autores nem sempre estão de acordo quanto à melhor forma de avaliar o ajustamento do modelo, pontos há que parecem merecer **consenso**:

- o melhor guia na avaliação do ajustamento do modelo é um forte suporte teórico para o modelo. Se o modelo tiver pouco

sentido do ponto de vista teórico, será difícil justificá-lo, mesmo que o ajustamento estatístico seja excelente;

- o teste de χ^2 não deve ser a base única para determinar o ajustamento do modelo, até porque a hipótese nula subjacente ao teste estatístico é demasiado rígida na maioria dos casos (a natureza dos modelos em Ciências Sociais é muito mais aproximativa);
- não se deve confiar exclusivamente em nenhuma medida de qualidade do ajustamento, seja ela qual for, devem antes ser usadas várias medidas, de diferentes tipos;
- é preferível considerar um conjunto de modelos alternativos do que examinar um só modelo. Frequentemente, o conhecimento em dada área não é suficientemente detalhado para fornecer uma especificação única de um modelo. Estimar vários modelos permite ao investigador explorar estruturas plausíveis;
- não deve ser ignorado o ajustamento das componentes do modelo: um modelo com excelente ajustamento global pode não ser aceitável devido ao ajustamento das componentes (os R^2 , as soluções impróprias, a magnitude das estimativas ...).

4.3.2 Ajustamento das Componentes do Modelo

O ajustamento global do modelo nada permite concluir acerca da qualidade do ajustamento das diversas componentes. É possível obter **coeficientes de determinação múltiplos**, R^2 , separadamente para cada variável observada, e para cada equação estrutural.

O coeficiente de determinação múltiplo para a i -ésima variável observável é dado por:

$$R^2_{\text{variável } i} = 1 - \hat{\Theta}_{ii}/s_{ii},$$

em que $\hat{\Theta}_{ii}$ é a variância estimada do erro de medida (δ_i ou ϵ_i) e s_{ii} é a variância da variável manifesta i . Para que a variável i seja um bom indicador de medida da variável latente, o valor de R^2 deve ser elevado (idealmente perto de 1).

104 Metodologia de Modelação

O coeficiente de determinação múltiplo para a j -ésima equação estrutural é dado por:

$$R^2_{\text{equação } j} = 1 - \frac{\widehat{\text{var}}(\zeta_j)}{\widehat{\text{var}}(\eta_j)},$$

em que $\widehat{\text{var}}(\zeta_j)$ e $\widehat{\text{var}}(\eta_j)$ são, respectivamente, as variâncias estimadas de ζ_j e de η_j . Quanto mais perto este valor estiver de 1, melhor.

Examinar a solução obtida implica prestar particular atenção às estimativas dos parâmetros (seus sinais e magnitude), às correlações entre as estimativas e aos erros padrão. Há ainda que verificar a existência de **estimativas inadmissíveis**, i.e., valores tidos por não razoáveis para determinados parâmetros, como sejam:

- variâncias negativas dos erros de medida e dos termos residuais nas relações estruturais (elementos na diagonal principal das matrizes Θ_δ , Θ_ϵ e Ψ);
- coeficientes padronizados muito próximo ou acima do valor 1;
- todos os coeficientes estimados com erros padrão associados muito elevados;
- estimativas de parâmetros altamente correlacionadas;
- correlações maiores que um (em magnitude);
- matrizes de variância-covariância e de correlação que não são definidas positivas.

Se estes problemas existirem o analista deve tentar resolvê-los, sendo que a correcção passa muitas vezes por uma análise detalhada do modelo para detectar erros de especificação e, conseqüentemente, pela re-especificação do modelo.

O facto de não existirem valores não razoáveis para as estimativas dos parâmetros ou o facto de o ajustamento global do modelo ter sido considerado aceitável não garantem, por si só, um bom ajustamento das componentes estrutural e de medida do modelo.

O **modelo de medida** encontra-se ajustado se:

- a significância associada aos pesos factoriais indicar que estes são todos estatisticamente significativos. Adicionalmente, o valor de R^2 de cada variável manifesta deve ser superior a 0.5. Em termos práticos isso significa que o correspondente peso factorial padronizado deve ser superior a 0.7. A literatura recomenda que, se tal não acontecer para um dado indicador, o mesmo seja eliminado ou substituído por um outro que seja melhor indicador de medida da variável latente em causa;
- a fiabilidade de cada variável latente, medida pelo R^2 ou, alternativamente, pela *composite reliability*, exceder 0.6.

Ter-se-á um bom ajustamento para o **modelo estrutural** quando:

- os coeficientes estruturais (β e γ) forem estatisticamente significativos;
- as correlações entre as estimativas das variáveis latentes não forem muito elevadas. Se tal acontecer, devem adoptar-se medidas correctivas no sentido de eliminar uma variável latente ou de reformular as relações estruturais;
- os R^2 de cada equação estrutural estiverem tão próximos de 1 quanto possível, indicando uma boa proporção de variância explicada.

Capítulo 5

Outros Tópicos

Este capítulo encontra-se estruturado em sete secções e refere, de forma breve, modelos com estrutura multi-grupos, modelos com estrutura de médias, modelos de análise factorial de segunda ordem, modelos com curva de crescimento latente e modelos com estrutura multinível. A penúltima secção faz referência a indicadores formativos e o capítulo termina com uma curta alusão a outros tópicos anteriormente não mencionados neste livro.

5.1 Modelos Multi-Grupos

Tudo o que vimos até agora tem a ver com dados que se assumem provenientes de uma única população. Situações há em que importa analisar, em simultâneo, dados de diversas populações, forçando a igualdade de alguns parâmetros do modelo (ou mesmo de todos) para diversos grupos populacionais - está-se perante um Modelo com Equações Estruturais multi-grupos. Será então possível determinar até que ponto o modelo de medida e o modelo estrutural se ajustam nas várias populações e examinar diferenças entre grupos.

Consideremos um conjunto de G populações (grupos de indivíduos mutuamente exclusivos e claramente definidos), sendo que é possível medir um conjunto de variáveis para um determinado número de indivíduos de cada população. O modelo global para o grupo g , com $g = 1, 2, \dots, G$, é definido pelas seguintes matrizes de parâmetros:

$$\Lambda_Y^{(g)}, \Lambda_X^{(g)}, B^{(g)}, \Gamma^{(g)}, \Phi^{(g)}, \Psi^{(g)}, \Theta_\delta^{(g)}, \Theta_\epsilon^{(g)}.$$

Se não houver restrições entre grupos (por exemplo determinados parâmetros serem iguais em 2 ou mais grupos), cada grupo pode ser analisado separadamente. Todavia, se houver restrições entre grupos, os dados dos diferentes grupos devem ser analisados simultaneamente, para se obterem estimativas eficientes e comparáveis dos parâmetros. Para uma exposição sobre modelação com múltiplos grupos ver Kaplan (2000, páginas 63–77).

5.1.1 O Que Se Pode Testar

A análise multi-grupos pode ser usada para testar diferentes restrições de igualdade entre parâmetros, como sejam, por exemplo:

- que as propriedades de medida das variáveis manifestas são as mesmas em todos os grupos, designadamente:

- pesos factoriais iguais nos vários grupos:

$$\Lambda_Y^{(1)} = \Lambda_Y^{(2)} = \dots = \Lambda_Y^{(G)} \text{ e / ou}$$

$$\Lambda_X^{(1)} = \Lambda_X^{(2)} = \dots = \Lambda_X^{(G)};$$

- variâncias e covariâncias dos factores latentes iguais nos vários grupos:

$$\Phi^{(1)} = \Phi^{(2)} = \dots = \Phi^{(G)};$$

- variâncias e covariâncias dos erros de medida iguais nos vários grupos:

$$\Theta_\epsilon^{(1)} = \Theta_\epsilon^{(2)} = \dots = \Theta_\epsilon^{(G)} \text{ e / ou}$$

$$\Theta_\delta^{(1)} = \Theta_\delta^{(2)} = \dots = \Theta_\delta^{(G)};$$

- que as relações estruturais são invariantes nos vários grupos:

$$B^{(1)} = B^{(2)} = \dots = B^{(G)} \text{ e / ou } \Gamma^{(1)} = \Gamma^{(2)} = \dots = \Gamma^{(G)}$$

Em termos gerais, qualquer grau de invariância pode ser testado, desde um extremo em que se pressupõem todos os parâmetros invariantes em todos os grupos, ao outro extremo, em que não há restrições

entre grupos. Compete ao investigador especificar as hipóteses a testar.

5.1.2 Estimação e Qualidade do Ajustamento

Para estimar simultaneamente todos os G modelos, o LISREL minimiza a seguinte função de ajustamento (uma combinação ponderada do ajustamento em cada um dos grupos):

$$F = \sum_{g=1}^G \frac{N_g}{N} F_g \left(\mathbf{S}^{(g)}, \mathbf{\Sigma}^{(g)}, \mathbf{W}^{(g)} \right),$$

sendo:

- F_g a função de ajustamento (ULS, GLS, ML, WLS ou DWLS) para o grupo g ;
- N_g a dimensão da amostra para o grupo g ;
- N a dimensão total da amostra ($= N_1 + \dots + N_G$);
- $\mathbf{S}^{(g)}$ e $\mathbf{\Sigma}^{(g)}$ as matrizes de variância-covariância amostral e populacional, para o grupo g ;
- $\mathbf{W}^{(g)}$ a matriz de pesos para o grupo g .

São primeiro calculadas estimativas iniciais, pelos métodos TSLS ou IV, para cada grupo separadamente, ignorando as restrições de igualdade entre grupos. Os parâmetros que foram especificados como sendo iguais entre grupos são depois substituídos pelo valor médio das estimativas iniciais encontradas. Em seguida, e de forma iterativa, é minimizada a função de ajustamento F . Quanto mais próximo a matriz de variância-covariância amostral de cada grupo, $\mathbf{S}^{(g)}$, estiver da matriz de variância-covariância implícita ao modelo, para cada grupo, $\mathbf{\Sigma}^{(g)}$, melhor é o ajustamento do modelo. A medida de qualidade do ajustamento, χ^2 , continua a ser definida como N vezes o mínimo de F . É uma medida do ajustamento global conjunto, isto é, de todos os modelos em todos os grupos, incluindo todas as restrições. O número

de graus de liberdade é dado por $df = \frac{1}{2}G(p+q)(p+q+1) - t$, sendo t o número total de parâmetros independentes estimados em todos os grupos. Para além do valor global da estatística χ^2 são disponibilizados, separadamente para cada grupo, o contributo relativo de cada grupo para o valor global da estatística χ^2 e os valores das medidas RMR e GFI.

5.1.3 Soluções Estandarizadas

Em LISREL é possível obter dois tipos de soluções estandarizadas, SS e SC, sendo que numa solução SS apenas as variáveis latentes são padronizadas, enquanto que numa solução SC também as variáveis manifestas o são. Na análise multi-grupos, cada uma dessas estandarizações pode ser feita dentro de cada grupo (*within group standardized solution*) ou entre grupos (*solution standardized to a common metric*), o que resulta em quatro tipos de estandarização:

- **solução estandarizada a uma métrica comum** - se na linha de Output pedirmos a opção SS, o LISREL estandariza as variáveis latentes por forma a que uma média ponderada da sua matriz de variância-covariância seja uma matriz de correlações, obtendo assim uma métrica / escala comum a todos os grupos. Esta estandarização não faz com que a matriz de variância-covariância das variáveis latentes seja uma matriz de correlações em cada grupo. Antes, uma média ponderada destas matrizes de variância-covariância é que é uma matriz de correlações. Esta solução estandarizada a uma métrica comum é disponibilizada na parte do output designada por *Solution Standardized to a Common Metric*;
- **solução estandarizada dentro de cada grupo** - a estandarização é feita por forma a que as variáveis latentes tenham, para cada grupo, variância unitária. Esta estandarização obtém-se pedindo a opção SS na linha de Output, e ana-

lisando a parte do output designada por *Within Group Standardized Solution*. Não é porém uma estandardização aconselhável para proceder a comparações entre grupos;

- **solução completamente estandardizada a uma métrica comum** - se na linha OU pedirmos a opção SC tanto as variáveis latentes como as manifestas são padronizadas, mas numa métrica comum a todos os grupos. Tal estandardização é disponibilizada na parte do output designada por *Common Metric Completely Standardized Solution*. É a estandardização recomendada para fazer comparações das estimativas entre grupos;
- **solução completamente estandardizada dentro de cada grupo** - para cada grupo são padronizadas tanto as variáveis latentes como as manifestas. Para tal, pedir a opção SC na linha OU e analisar a parte do output designada por *Within Group Completely Standardized Solution*.

Se na análise multi-grupos houver restrições de igualdade entre grupos não se devem estandardizar as variáveis dentro de cada grupo: **para possibilitar a comparação** de parâmetros pertencentes a grupos diferentes, **as variáveis devem ser padronizadas a uma métrica comum para todos os grupos**. As escalas das variáveis latentes devem ser fixadas com recurso a **variáveis manifestas de referência**, atribuindo valores não nulos a elementos em Λ_X e Λ_Y , de tal forma que as variáveis latentes fiquem numa escala comum.

5.1.4 Um Exemplo de Análise Multi-Grupos em LISREL

Vamos retomar o exemplo do modelo global, representado no diagrama da Figura 1.5 e considerar dois grupos de respondentes: os que pertencem a agregados familiares em que há crianças, e aqueles

que vivem em agregados familiares onde não existem crianças. Com efeito há literatura que postula que a existência de crianças no agregado familiar conduz a valores mais baixos de satisfação com aspectos de lazer.

AFC multi-grupos e invariância do modelo de medida

Começaremos por uma AFC multi-grupos, primeiro sem restrições de igualdade entre os dois grupos, e depois com restrições, por forma a testar a invariância do modelo de medida nos dois grupos. Tal como no exemplo da Figura 1.5 tanto *MATER02* como *LAZER03* são medidas por 3 indicadores cada, assumindo-se que os dois factores de satisfação estão correlacionados.

Escolhemos *Sat2_02* e *Sat6_03* como variáveis de referência, respectivamente para as variáveis latentes *MATER02* e *LAZER03*. A estrutura de comandos de sintaxe LISREL que permite estimar o modelo por *Robust ML* consta do Apêndice C.1.2. Apresentam-se alguns excertos dos outputs obtidos. Primeiro as estimativas dos parâmetros do modelo de medida de cada um dos grupos, numa solução completamente standardizada a uma métrica comum, por forma a possibilitar a comparação das estimativas nos dois grupos:

Common Metric Completely Standardized Solution

```
GRUPO: SEM CRIANCAS
      LAMBDA-X
            MATER02    LAZER03
            -----    -
Sat2_02    0.721      - -
Sat3_02    0.579      - -
Sat5_02    0.597      - -
Sat6_03    - -        0.856
Sat7_03    - -        0.768
Sat8_03    - -        0.870
      PHI
            MATER02    LAZER03
            -----    -
MATER02    0.997
LAZER03    0.462      1.018
```

GRUPO: COM CRIANCAS		
LAMBDA-X		
	MATERO2	LAZER03
	-----	-----
Sat2_02	0.721	- -
Sat3_02	0.544	- -
Sat5_02	0.602	- -
Sat6_03	- -	0.856
Sat7_03	- -	0.866
Sat8_03	- -	0.886
PHI		
	MATERO2	LAZER03
	-----	-----
MATERO2	1.004	
LAZER03	0.563	0.972

Note-se como as estimativas obtidas para os pesos factoriais têm valores similares nos dois grupos. As medidas de qualidade do ajustamento modelo-dados revelam um bom ajustamento, apesar de ser de rejeitar a hipótese nula de que, para ambos os grupos, a matriz de correlações populacional é reproduzida pela matriz de correlações implícita ao modelo, como aliás quase sempre acontece em Modelos com Equações Estruturais

Global Goodness of Fit Statistics
 Degrees of Freedom = 16
 Minimum Fit Function Chi-Square = 66.153 (P = 0.000)
 Satorra-Bentler Scaled Chi-Square = 41.428 (P = 0.000480)
 Root Mean Square Error of Approximation (RMSEA) = 0.0398
 90 Percent Confidence Interval for RMSEA = (0.0251 ; 0.0549)
 Normed Fit Index (NFI) = 0.992
 Non-Normed Fit Index (NNFI) = 0.991
 Comparative Fit Index (CFI) = 0.995

Testaremos em seguida a hipótese de igualdade de pesos factoriais nos 2 grupos, ou seja, a hipótese de que a matriz $\Lambda_{\mathbf{X}}$ é invariante nos 2 grupos. Para tal, há que declarar a matriz $\Lambda_{\mathbf{X}}$ como invariante na linha MO, de definição do modelo, do grupo 2, mantendo-se os restantes comandos do ficheiro de sintaxe mantém-se. Obtém-se então as seguintes estimativas para os parâmetros do modelo de medida, numa solução completamente estandardizada a uma métrica comum:

114 Outros Tópicos

Common Metric Completely Standardized Solution

LAMBDA-X						
	MATERO2	LAZER03				
	-----	-----				
Sat2_02	0.720	- -				
Sat3_02	0.566	- -				
Sat5_02	0.600	- -				
Sat6_03	- -	0.856				
Sat7_03	- -	0.807				
Sat8_03	- -	0.876				
GRUPO: SEM CRIANCAS						
PHI						
	MATERO2	LAZER03				
	-----	-----				
MATERO2	1.008					
LAZER03	0.458	0.988				
THETA-DELTA						
	Sat2_02	Sat3_02	Sat5_02	Sat6_03	Sat7_03	Sat8_03
	-----	-----	-----	-----	-----	-----
	0.479	0.671	0.641	0.260	0.389	0.233
GRUPO: COM CRIANCAS						
PHI						
	MATERO2	LAZER03				
	-----	-----				
MATERO2	0.988					
LAZER03	0.573	1.019				
THETA-DELTA						
	Sat2_02	Sat3_02	Sat5_02	Sat6_03	Sat7_03	Sat8_03
	-----	-----	-----	-----	-----	-----
	0.484	0.695	0.639	0.281	0.284	0.233

O valor da estatística de χ^2 é agora de 49.089, com 20 graus de liberdade. A diferença na estatística de χ^2 deste modelo face ao modelo anterior pode ser encarada como o valor da estatística de χ^2 para um teste de hipóteses à hipótese nula de que a matriz $\Lambda_{\mathbf{X}}$ é invariante nos dois grupos. $\Delta\chi^2 = 49.089 - 41.428 = 7.661$ e $\Delta df = 20 - 16 = 4$. Considerando um nível de significância $\alpha = 5\%$, o valor crítico de uma distribuição de χ^2 com quatro graus de liberdade é de 9.488, pelo que não será de rejeitar a hipótese de que a matriz $\Lambda_{\mathbf{X}}$ é invariante nos dois grupos, i.e., de que os pesos factoriais são iguais nos 2 grupos. Os valores das estimativas dos pesos factoriais, idênticos nos dois grupos, são os seguintes:

	LAMBDA-X	
	MATER02	LAZER03
	-----	-----
Sat2_02	0.720	- -
Sat3_02	0.566	- -
Sat5_02	0.600	- -
Sat6_03	- -	0.856
Sat7_03	- -	0.807
Sat8_03	- -	0.876

Analogamente, poder-se-ia também testar a invariância da matriz Φ nos dois grupos, e também a invariância da matriz Θ_{δ} .

Invariância das relações estruturais

Por último ilustrar-se-á como testar a hipótese de invariância das relações estruturais postuladas no modelo global apresentado na Figura 1.5 nos dois grupos de agregados familiares: com e sem crianças. Com efeito, se a magnitude dessas relações não variar com o grupo, tal indicia que a variável existência de crianças no agregado não tem efeito moderador nas relações estruturais do modelo.

A estrutura de comandos que permite estimar um modelo multi-grupos, sem e com restrições de invariância dos parâmetros β e γ nos dois grupos, é apresentada no Apêndice C.1.3.

O modelo sem restrições tem $df = 54$ e $\chi^2 = 29.83$, ao passo que o modelo com restrições tem $df = 60$ e $\chi^2 = 51.06$. Assim, $\Delta df = 6$ e $\Delta\chi^2 = 21.23$. O valor crítico de uma distribuição de χ^2 com 6 graus de liberdade é de 12.592, pelo que é de rejeitar a hipótese nula de invariância dos parâmetros nos dois grupos. Por outras palavras, as relações estruturais têm magnitudes diferentes quando há, ou não, crianças no agregado familiar, pelo que esta variável funciona como moderadora no modelo em estudo.

Assim, apresentam-se as estimativas obtidas para os parâmetros que definem as relações estruturais em cada um dos dois grupos, numa solução completamente estandardizada a uma métrica comum e que, por isso têm valores que são comparáveis:

116 Outros Tópicos

Common Metric Completely Standardized Solution

GRUPO: SEM CRIANCAS

BETA				
	MATERO2	LAZER03		
	-----	-----		
MATERO2	- -	- -		
LAZER03	0.428	- -		
GAMMA				
	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
MATERO2	0.287	0.113	-0.057	- -
LAZER03	- -	- -	0.100	-0.073

GRUPO: COM CRIANCAS

BETA				
	MATERO2	LAZER03		
	-----	-----		
MATERO2	- -	- -		
LAZER03	0.511	- -		
GAMMA				
	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
MATERO2	0.339	-0.058	-0.123	- -
LAZER03	- -	- -	0.187	-0.231

Repare-se, por exemplo, como o número de horas de trabalho semanal tem um maior impacto (negativo) na satisfação com aspectos de lazer nos agregados familiares com crianças, e como a idade do respondente tem maior impacto sobre a satisfação material em agregados sem crianças.

5.2 Modelos com Estruturas de Médias

5.2.1 Extensão do Modelo com Equações Estruturais Global

O modelo de equações estruturais global foi definido no Capítulo 3, pelas Equações 3.4, pressupondo que as variáveis aleatórias têm média nula. Este pressuposto será agora relaxado e far-se-á uma extensão do modelo por forma a incluir quatro vectores adicionais de parâmetros:

α , τ_Y , τ_X e κ . Estes novos vectores contêm os interceptos das equações (estruturais e de medida) e os valores das médias das variáveis latentes independentes. O modelo global com estrutura de médias é definido pelas seguintes equações (a primeira constitui a componente estrutural do modelo e as restantes constituem os modelos de medida):

$$\begin{aligned}\eta &= \alpha + B\eta + \Gamma\xi + \zeta \\ Y &= \tau_Y + \Lambda_Y\eta + \epsilon \\ X &= \tau_X + \Lambda_X\xi + \delta.\end{aligned}\tag{5.1}$$

Os vectores de interceptos são α (de ordem número de variáveis η por 1), τ_Y (de ordem número de variáveis Y por 1), e τ_X (de ordem número de variáveis X por 1).

Todos os pressupostos do modelo global definidos na Secção 3.2.1 se mantêm, com 2 excepções:

- $E[\xi]$ deixa de ser nulo para ser igual a κ , sendo κ um vector de ordem n° de variáveis ξ por 1;
- $E[\eta]$ deixa de ser nulo para ser igual a $(I - B)^{-1}(\alpha + \Gamma\kappa)$.

Os vectores de médias das variáveis manifestas são:

$$\begin{aligned}\mu_Y &= \tau_Y + \Lambda_Y(I - B)^{-1}(\alpha + \Gamma\kappa); \\ \mu_X &= \tau_X + \Lambda_X\kappa.\end{aligned}\tag{5.2}$$

Por *default* no LISREL cada uma dos quatro novos vectores de parâmetros está definido como sendo um vector nulo, o que conduz ao modelo definido pelas Equações 3.4. Para serem incluídos no modelo, os novos vectores têm que ser especificados na linha MO, de especificação do modelo, no ficheiro de comandos. A designação para estes quatro vectores é a seguinte: TY (τ_Y), TX (τ_X), AL (α) e KA (κ). Cada um destes vectores pode ser declarado como *Fixed*, *Free* e, em análise multi-grupos *SP* - *same pattern* ou *IN* - *invariant*.

Em geral, numa só população, os parâmetros em α , τ_Y , τ_X e κ não estarão identificados se não forem impostas restrições adicionais. No caso de uma análise multi-grupos facilmente se podem impôr restrições simples, por forma a tornar identificados todos os parâmetros do modelo. Um modo expedito de o fazer é declarar como nulo o vector ou os vectores de parâmetros da estrutura de médias de um determinado grupo (por exemplo do grupo 1: $\tau_X^{(1)}$, $\tau_Y^{(1)}$, $\alpha^{(1)}$ e / ou $\kappa^{(1)}$ declarados como *Fixed* na linha MO do grupo 1) e sujeitar a estimação o vector ou os vectores de parâmetros dos restantes grupos. Em termos de interpretação das estimativas obtidas, esta deve ser feita por comparação com os valores nulos das estimativas do grupo considerado de referência.

5.2.2 Função de Ajustamento

A função de ajustamento do modelo global de equações estruturais com estrutura de médias é a seguinte:

$$F = \sum_{g=1}^G \frac{N_g}{N} F_g, \quad \text{sendo } F_g =$$

$$\frac{1}{2} \left[\mathbf{s}^{(g)} - \boldsymbol{\sigma}^{(g)} \right]^T \mathbf{W}_{(g)}^{-1} \left[\mathbf{s}^{(g)} - \boldsymbol{\sigma}^{(g)} \right] + \frac{1}{2} \left[\bar{\mathbf{z}}^{(g)} - \boldsymbol{\mu}^{(g)} \right]^T \mathbf{V}_{(g)}^{-1} \left[\bar{\mathbf{z}}^{(g)} - \boldsymbol{\mu}^{(g)} \right],$$

com $\boldsymbol{\mu}^{(g)} = [\boldsymbol{\mu}_Y^{(g)} \quad \boldsymbol{\mu}_X^{(g)}]^T$. O primeiro termo de F_g é semelhante a F_{WLS} , apresentado na Secção 4.2. O segundo termo envolve:

- $\bar{\mathbf{z}}^{(g)}$ o vector de médias amostrais;
- $\boldsymbol{\mu}^{(g)}$ o vector de médias populacionais (função de $\boldsymbol{\mu}_X$ e $\boldsymbol{\mu}_Y$, definidos pelas Equações 5.2);
- $\mathbf{V}_{(g)}$ uma matriz de pesos que iguala

$$\begin{aligned} & \mathbf{S}^{(g)}, \text{ no caso de estimação ULS, GLS, WLS e DWLS} \\ & \hat{\boldsymbol{\Sigma}}^{(g)}, \text{ no caso de estimação ML.} \end{aligned}$$

Se α , τ_Y , τ_X e κ forem todos vectores nulos, o segundo termo de F_g é constante, pelo que o problema se reduz a uma análise multi-grupos, como descrita no Capítulo 5.1, sem qualquer estrutura de médias.

5.2.3 Estimação das Médias dos Factores

Embora a média de uma variável latente não esteja identificada numa análise com um só grupo, podem ser estimadas diferenças nas médias das variáveis latentes em vários grupos desde que, em todos os grupos, as variáveis estejam todas na mesma escala. Em 1974, Sörbom estendeu o modelo clássico de invariância factorial, por forma a ser possível estimar médias de factores. O modelo de Sörbom tem as estruturas de médias assim definidas:

$$\mathbf{X}^{(g)} = \tau_X + \Lambda_X \xi^{(g)} + \delta^{(g)}, \quad \text{com } g = 1, 2, \dots, G,$$

sendo a média de $\xi^{(g)}$ igual a $\kappa^{(g)}$. Para definir as origens e unidades de medida dos factores ξ pode fazer-se $\kappa^{(1)} = \mathbf{0}$ e fixar um valor não nulo em cada coluna de Λ_X . Os parâmetros a serem estimados são:

- τ_X e Λ_X , que se assumem invariantes para todos os grupos;
- $\kappa^{(2)}, \kappa^{(3)}, \dots, \kappa^{(G)}$, vectores de médias de ξ , nos vários grupos;
- $\Phi^{(1)}, \Phi^{(2)}, \dots, \Phi^{(G)}$, matrizes de variância-covariância de ξ , nos vários grupos;
- $\Theta_\delta^{(1)}, \Theta_\delta^{(2)}, \dots, \Theta_\delta^{(G)}$, matrizes de variância-covariância dos erros nos vários grupos. Podem ser declaradas como invariantes entre grupos, sendo na maioria dos casos matrizes diagonais.

5.3 Análise Factorial de Segunda Ordem

Estamos perante um modelo de análise factorial de segunda ordem quando os factores de primeira ordem são explicados por uma deter-

minada estrutura factorial de ordem superior. Trata-se de um tipo de modelo muito utilizado na área da psicometria.

5.3.1 Equação do Modelo e Matriz de Variância-Covariância Implícita

A equação $\mathbf{Y} = \mathbf{\Lambda}_Y \boldsymbol{\eta} + \boldsymbol{\epsilon}$ representa um modelo de análise factorial para \mathbf{Y} , sendo $\boldsymbol{\epsilon}$ o vector dos erros de medida e $\boldsymbol{\eta}$ o vector dos factores de primeira ordem. Por seu turno, as variáveis $\boldsymbol{\eta}$ podem ser explicadas por um conjunto de factores $\boldsymbol{\xi}$, chamados factores de segunda ordem, por forma a que $\boldsymbol{\eta} = \mathbf{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta}$, sendo $\mathbf{\Gamma}$ a matriz dos pesos factoriais de segunda ordem e $\boldsymbol{\zeta}$ o vector das componentes únicas ou específicas das variáveis $\boldsymbol{\eta}$.

Se combinarmos a primeira equação com esta última, obtemos um modelo de análise factorial de segunda ordem, em que não há variáveis X e em que \mathbf{B} é uma matriz nula, modelo esse do tipo

$$\mathbf{Y} = \mathbf{\Lambda}_Y (\mathbf{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta}) + \boldsymbol{\epsilon}, \quad (5.3)$$

com matriz de variância-covariância implícita dada por

$$\mathbf{\Lambda}_Y (\mathbf{\Gamma} \boldsymbol{\Phi} \mathbf{\Gamma}^T + \boldsymbol{\Psi}) \mathbf{\Lambda}_Y^T + \boldsymbol{\Theta}_\epsilon \quad (5.4)$$

(matriz que se obtém da Equação 3.5 fazendo $\mathbf{B} = \mathbf{0}$ e, consequentemente, $\mathbf{A} = \mathbf{I}$).

5.3.2 Um Exemplo de Aplicação de um Modelo de Análise Factorial de Segunda Ordem

O exemplo que se segue, retirado de Malta (2009), ilustra a aplicação de um modelo de análise factorial de segunda ordem ao estudo das percepções de bem-estar, no que à debilidade psicológica diz respeito, medida por 12 itens da escala *General Health Questionnaire* - GHQ12.

De acordo com vários autores, a escala de debilidade psicológica GHQ12 incorpora três dimensões distintas: ansiedade e depressão

(medida por 4 itens); disfunção social (medida por 6 itens); e perda de confiança (medida por 2 itens). Os 12 itens da escala GHQ12 apresentam-se na Tabela 5.1.

Ansiedade e Depressão (AD)
B - perda de sono por preocupações
E - constantemente sobre pressão
F - problemas em superar dificuldades
I - sentimento de tristeza ou depressão
Disfunção Social (SD)
A - facilidade de concentração
C - sentimento de utilidade
D - capacidade de tomar decisões
G - capacidade de apreciar actividades diárias
H - capacidade de encarar problemas
L - sentimento de felicidade
Perda de Confiança (LC)
J - perda de confiança em si próprio
K - perda de auto-estima

Tabela 5.1 *Itens da escala GHQ12.*

Os itens em estudo foram retirados do *British Household Panel Survey*. Estão todos medidos numa escala de tipo Likert de quatro pontos. No caso das escalas de AD e de LC valores mais baixos de resposta significam ausência do sintoma, ao passo que valores mais elevados correspondem a aumento do sintoma e enunciam estados psicológicos negativos. Inversamente, valores mais baixos de resposta aos itens da escala de SD estão associados a maior bem-estar psicológico.

Os dados em análise são relativos ao ano de 2009. A amostra inclui um total de 4562 respondentes, com respostas válidas em todos os itens. O diagrama do modelo de análise factorial de segunda ordem a testar foi apresentado na Figura 1.7. Antes de estimar o modelo de

122 Outros Tópicos

análise factorial de segunda ordem optou-se por estimar dois modelos de AFC: um modelo unifactorial com um só factor de debilidade psicológica e um modelo com três factores, AD, SD e LC, que se assumem correlacionados entre si (uma vez que se assumem medir um só factor de debilidade psicológica).

A Tabela 5.2 apresenta as estimativas obtidas para os pesos factoriais, numa solução estandardizada. Foi utilizado o método de estimação *Robust ML*, tendo as variáveis manifestas sido modeladas como ordinais.

Itens	Modelo Unifactorial	Modelo Trifactorial		
		AD	SD	LC
B	0.69	0.75	—	—
E	0.71	0.80	—	—
F	0.80	0.85	—	—
I	0.84	0.88	—	—
A	0.70	—	0.75	—
C	0.61	—	0.72	—
D	0.67	—	0.80	—
G	0.71	—	0.77	—
H	0.72	—	0.82	—
L	0.77	—	0.81	—
J	0.83	—	—	0.93
K	0.80	—	—	0.89

Tabela 5.2 *Estimativas dos pesos factoriais num modelo unifactorial, e num modelo com três factores: AD, SD e LC.*

Obtiveram-se os seguintes valores estimados para as correlações entre os três factores no modelo trifactorial: 0.72 para a correlação entre AD e SD; 0.73 para a correlação entre SD e LC; e 0.84 para a correlação entre AD e LC. Apresentam-se na Tabela 5.3 os valores obtidos para algumas medidas de qualidade do ajustamento dos dois modelos em análise.

	Modelo Unifactorial	Modelo Trifactorial
AIC	2623.1	677.6
RMSEA	0.101	0.049
SRMR	0.09	0.052
χ^2	2575.1	623.57
g.l.	54	51

Tabela 5.3 *Medidas de qualidade do ajustamento dos modelos unifactorial e trifactorial.*

É possível concluir que a qualidade do ajustamento do modelo trifactorial é superior, uma vez que está associada a um menor valor de AIC. Assim, o modelo unifactorial de debilidade psicológica deve ser preterido face a um modelo que considera três dimensões. Uma vez que as correlações estimadas entre os três factores são relativamente elevadas, fará sentido estimar o modelo de análise factorial de segunda ordem que se apresenta na Figura 5.1, já com as estimativas obtidas em LISREL, numa solução estandardizada.

Os comandos que permitem estimar o modelo em estudo apresenta-se no Apêndice C.2. Obtiveram-se os seguintes valores de R^2 :

Squared Multiple Correlations for Structural Equations		
AD	SD	LC
----- 0.836	----- 0.625	----- 0.851

Note-se que não são apresentadas as medidas de qualidade do ajustamento do modelo de segunda ordem uma vez que, havendo três factores de primeira ordem, elas são idênticas às do modelo trifactorial. Os dois modelos têm igual número de parâmetros a estimar, de graus de liberdade e de valores das medidas da qualidade do ajustamento. Na prática, em vez de três correlações entre factores de primeira ordem, estão a estimar-se três pesos factoriais. Todavia, tal igualdade deixará de ocorrer se o número de factores de primeira ordem for superior a três.

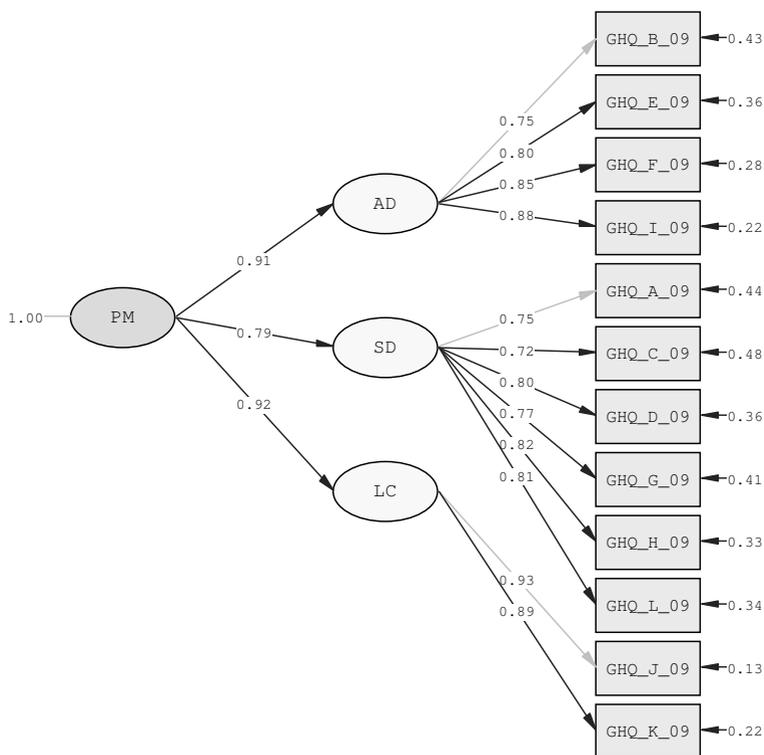


Figura 5.1 Diagrama de modelo de análise factorial de segunda ordem da debilidade psicológica.

Porque a preocupação aqui é sobretudo metodológica no sentido de ilustrar como estimar um modelo de análise factorial de segunda ordem, não se irá proceder a qualquer re-especificação do modelo com o intuito de melhorar a qualidade do ajustamento modelo-dados. É porém de referir que o índice de modificação mais elevado sugere correlacionar os erros de medida dos itens I e J.

5.4 Modelos com Curva de Crescimento Latente

5.4.1 Aspectos Introdutórios

Também designados por Modelos com Trajectória Latente, *Latent Growth Curve Models* (LGCM), *Latent Change Analysis*, *Growth Modelling* ou *Latent Trajectory Models*, são modelos estatísticos para dados longitudinais, que permitem que cada indivíduo em análise tenha uma trajectória distinta de mudança (ou curva de crescimento) ao longo do tempo. Estes modelos permitem a investigação das diferenças entre indivíduos nas trajectórias de mudança ao longo do tempo, bem como das variáveis explicativas dessas trajectórias.

Trata-se de um tipo de modelação que tem sido alvo de grandes desenvolvimentos nas últimas décadas, tendo vindo a ser crescente a sua aplicação. Para uma introdução a LGCM ver Preacher *et al.* (2008) e Kaplan (2000, capítulo 8). Bollen and Curran (2006) pode ser considerada a referência-chave à abordagem dos LGCM no âmbito dos Modelos com Equações Estruturais. Raykov e Marcoulides (2006) apresentam exemplos de comandos de sintaxe e outputs anotados de modelação em LISREL e em Mplus.

Duncan *et al.* (2006) mostram como, integrando o *framework* dos Modelos com Equações Estruturais, os LGCM permitem tirar partido das suas potencialidades, designadamente a existência de medidas de qualidade do ajustamento global; a modelação de trajectórias de variáveis latentes medidas por múltiplos indicadores; a inclusão no modelo de covariáveis latentes; ou a modelação simultânea de dois ou mais processos, permitindo a estimação das relações de dependência entre os parâmetros das trajectórias latentes. Todavia, continuam a ser necessárias amostras de grande dimensão, sendo a capacidade de lidar com variáveis manifestas categoriais ordinais é ainda um problema em alguns *software*.

Chave para os desenvolvimentos mais recentes dos LGCM tal como os conhecemos hoje foi também o trabalho desenvolvido por McArdle (1988) e por Meredith and Tisak (1990). Para uma apresentação de desenvolvimentos recentes ver Collins e Sayer (2001), Muthén (2004) e McArdle (2009).

Podem encontrar-se aplicações de LGCM em variadas áreas substantivas. Dias (2009) e Serra (2010) discutem a utilização de LGCM para modelar trajectórias de satisfação laboral em Portugal, usando Mplus e LISREL. Note-se que o Mplus permite considerar a natureza complexa do desenho amostral para efeitos de estimação, mesmo quando as variáveis manifestas são categoriais ordinais, enquanto que o LISREL só o permite para variáveis manifestas contínuas. Malta (2009) propõe um LGCM para modelar a trajectória de um factor de segunda ordem de debilidade psicológica, medido indirectamente por 12 indicadores de três factores latentes de primeira ordem, usando a escala GHQ12 e dados do BHPS. Salgueiro (2009) propõe a utilização de LGCM para modelar trajectórias de satisfação com o emprego com dados do BHPS, e Salgueiro (2010) usa modelos não condicionais para descrever trajectórias de satisfação com as condições de vida, com recurso a dados Portugueses do *European Community Household Panel*. Salgueiro et al. (2010) discutem a introdução de covariáveis para explicar as trajectórias de satisfação, e Salgueiro et al. (2011) propõem um LGCM bivariado para modelar conjuntamente percepções de bem-estar medidas por dois factores de satisfação (com aspectos materiais e com aspectos de lazer), tendo em conta a natureza complexa do desenho amostral usado na recolha dos dados. Vicente e Salgueiro (2012) contrastam a utilização de um Modelo com Equações Estruturais com efeitos cruzados e desfasados com a utilização de um LGCM para estudar, longitudinalmente, duas dimensões de privação material, com dados Portugueses do ICOR (Inquérito às Condições de Vida e Rendimento).

5.4.2 Modelos com Trajectórias Não Condicionadas

Em LGCM o efeito temporal é incorporado no modelo sob a forma de restrições nos pesos factoriais de um modelo com variáveis latentes, em que os factores latentes representam os parâmetros que definem a curva ou trajectória de crescimento: o intercepto e o declive. O tipo de trajectória (linear, quadrática, ...) depende do número de variáveis latentes especificado e da forma como os respectivos pesos factoriais variam em função do tempo. As medidas repetidas são vistas como medições imperfeitas da trajectória latente subjacente, trajectória essa que pode ser estimada a partir da estrutura de médias e de covariâncias das medidas repetidas observadas.

A Figura 5.2 representa o diagrama de um LGCM com trajectória não condicionada, i.e., sem variáveis explicativas que a expliquem. As quatro medidas repetidas (y_1, y_2, y_3, y_4) representam as variáveis contínuas manifestas (indicadores) das duas variáveis latentes: o intercepto e o declive da trajectória linear. Todas as observações se assumem medidas nos mesmos momentos temporais $t = 1, \dots, 4$. Para especificar uma trajectória linear há que impôr a restrição de que os pesos factoriais do intercepto (α) são todos iguais a 1, sendo os pesos factoriais do declive (β) iguais ao ranking da ocasião de medida: 0, 1, 2, 3, no exemplo em análise. γ denota o coeficiente de regressão do declive sobre o intercepto.

Mais genericamente, seja

$$y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it} \quad (5.5)$$

o valor da variável y para o indivíduo $i = 1, \dots, N$ no momento temporal $t = 1, \dots, T$. O vector de dimensão T dos termos residuais para o i -ésimo indivíduo, ε_i , segue distribuição $N(\mathbf{0}, \Theta)$. No caso de trajectória linear é usual assumir $\lambda_t = t - 1$. O intercepto e o declive

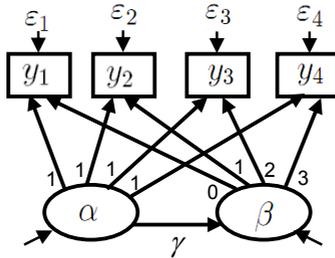


Figura 5.2 Diagrama de um LGCN não condicional, com quatro medidas repetidas de uma variável contínua y_t .

aleatórios são dados por:

$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i} \tag{5.6}$$

$$\beta_i = \mu_\beta + \gamma\alpha_i + \zeta_{\beta_i}, \tag{5.7}$$

com $\zeta_{\alpha_i} \sim N(0, \Psi_\alpha)$; $\zeta_{\beta_i} \sim N(0, \Psi_\beta)$; e ε_{it} , ζ_{α_i} e ζ_{β_i} mutuamente independentes, para $i = 1, \dots, N$ e $t = 1, \dots, T$.

Os parâmetros de especial interesse neste modelo são as médias e as variâncias dos efeitos aleatórios, e o coeficiente de regressão do declive sobre o intercepto. A média do intercepto iguala a média da variável manifesta no primeiro momento temporal (\bar{y}_1). A média do declive representa o crescimento médio da variável y_t por unidade de tempo. A variância do intercepto mede a variabilidade dos indivíduos em torno da média do intercepto; a variância do declive indica a variabilidade dos indivíduos em torno da trajetória média de crescimento. No modelo representado na Figura 5.2 assume-se que o nível inicial influencia a trajetória de mudança (e daí o coeficiente de regressão γ). Uma outra possível parameterização seria considerar a correlação entre o intercepto e o declive aleatórios. Caso se pretendesse considerar uma trajetória não linear, havia que incorporar uma outra variável latente, associada ao termo quadrático, com pesos factoriais iguais a 0, 1, 4 e 9 (o quadrado dos pesos factoriais do declive).

5.4.3 Modelos com Múltiplos Indicadores

Se em vez da trajectória de uma só variável observada se pretender modelar a trajectória de uma variável latente, medida por vários indicadores, há que considerar um LGCM com múltiplos indicadores (também designado na literatura por LGCM de segunda ordem). A Figura 5.3 representa o diagrama de um LGCM com trajectória linear para a variável latente η_t , medida por três variáveis manifestas contínuas (y_{1t}, y_{2t}, y_{3t}), em $t = 1, \dots, 4$ momentos temporais distintos. A estrutura dos pesos factoriais pode ser considerada invariante ao longo do tempo, impondo-se para tal restrições de igualdade dos pesos factoriais das medidas repetidas. É também recomendável correlacionar os erros de medida das medidas repetidas (por simplicidade na Figura apenas se representam as correlações que envolvem as medidas dos três indicadores no primeiro momento temporal).

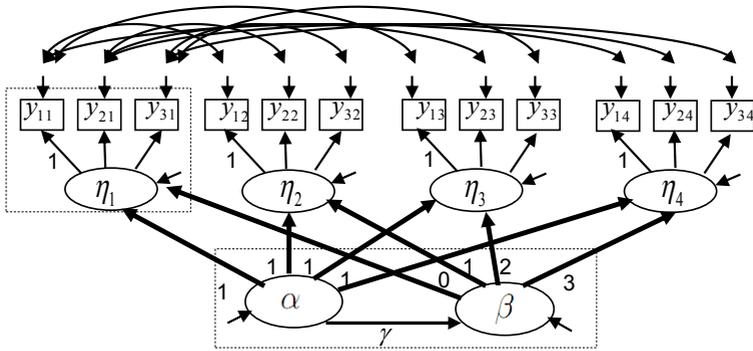


Figura 5.3 Diagrama de um LGCM de segunda ordem, com quatro medidas repetidas de uma variável latente η_t .

Mais genericamente, o valor do indicador $j = 1, \dots, J$ para o indivíduo i no momento temporal t é dado por

$$y_{ijt} = \nu_{jt} + \Lambda_{jt}\eta_{it} + v_{ijt}, \tag{5.8}$$

sendo Λ_{jt} o peso factorial do indicador j no momento temporal t .

130 Outros Tópicos

A matriz de dimensão $J \times T$ dos termos residuais para o i -ésimo indivíduo, \mathbf{v}_i , segue distribuição $N(\mathbf{0}, \Theta)$. A estrutura da matriz Θ depende de que termos residuais estão correlacionados. O modelo com trajetória latente para a variável η_t é dado por

$$\eta_{it} = \alpha_i + \lambda_t \beta_i + \zeta_{it}, \quad (5.9)$$

com efeitos aleatórios dados pelas equações 5.6 e 5.7. O modelo pressupõe que $\zeta_{\alpha_i} \sim N(0, \Psi_\alpha)$; $\zeta_{\beta_i} \sim N(0, \Psi_\beta)$; e $\zeta_{it} \sim N(0, \Psi_t)$. Por outro lado, \mathbf{v}_i , ζ_{it} , ζ_{α_i} e ζ_{β_i} são mutuamente independentes para $i = 1, \dots, N$ e $t = 1, \dots, T$.

Para além da média do declive, das variâncias dos efeitos aleatórios, e do coeficiente de regressão do declive sobre o intercepto, são agora parâmetros de especial interesse os pesos factoriais do modelo de medida. Porque η_t é uma variável latente com média zero, a média do intercepto deixou de ser um parâmetro de interesse no modelo, sendo também igual a zero.

5.4.4 Modelos com Variáveis Manifestas Ordinais

A Figura 5.4 pode ser obtida por extensão do rectângulo a tracejado no canto superior esquerdo da Figura 5.3. Se a variável latente η_t representada na Figura 5.3 tivesse sido medida por indicadores categoriais ordinais, o LGCM teria um nível extra de complexidade, correspondente à inclusão de y_{jt}^* , uma variável latente contínua que determina a variável manifesta ordinal y_{jt} .

Substituindo y por y^* na Equação 5.8, o modelo passa a ser definido como

$$y_{ijt}^* = \nu_{jt} + \Lambda_{jt} \eta_{it} + v_{ijt}. \quad (5.10)$$

A ligação entre a variável categorial ordinal y_{jt} e a variável contínua subjacente y_{jt}^* é dada pela equação

$$y_{ijt} = c \quad \text{se} \quad \tau_{c-1} < y_{ijt}^* \leq \tau_c, \quad (5.11)$$

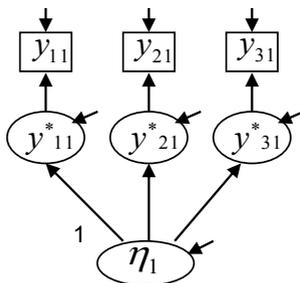


Figura 5.4 Diagrama de um modelo com indicadores ordinais.

sendo $c = 1, 2, \dots, C$ o número total de categorias da variável y_{jt} ; e τ_{c-1} e τ_c o limiar (*threshold*) inferior e o limiar superior da categoria c (com $\tau_0 = -\infty$ e $\tau_C = \infty$). Os $C - 1$ valores limiares finitos encontram-se ordenados de forma crescente (para mais detalhes ver Bollen e Curran, 2006, capítulo 8). Note-se que é prática corrente em LGCM com medidas ordinais repetidas assumir a invariância dos valores limiares das medidas repetidas.

5.4.5 Modelos com Trajectórias Condicionadas

Os modelos com trajectórias não condicionadas descrevem as curvas de crescimento, enquanto que os modelos com trajectórias condicionadas procuram explicar as curvas de crescimento. As covariáveis explicativas a introduzir no modelo podem ser consideradas e modeladas como invariantes ao longo do tempo (TIC - *Time Invariant Covariate*) ou como variantes no tempo (TVC - *Time-varying covariate*). A Figura 5.5 representa o diagrama de um LGCM com trajectória condicionada, que inclui simultaneamente uma TIC (x) e uma TVC (w_t). É assumida uma trajectória linear para a variável latente η_t , medida por três variáveis manifestas ordinais (y_{1t}, y_{2t}, y_{3t}) em $t = 1, 2, 3$ momentos temporais distintos, sendo que y_{jt}^* denota a variável contínua latente subjacente à variável manifesta ordinal y_{jt} .

132 Outros Tópicos

A TIC x pode ser uma variável *dummy* ou uma variável contínua, e é uma variável explicativa do intercepto e do declive aleatórios. A TVC w_t é também uma variável *dummy* ou uma variável contínua, e as suas medidas repetidas influenciam directamente a variável η_t .

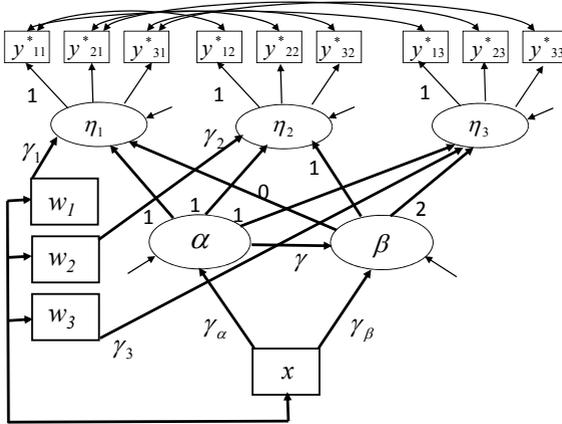


Figura 5.5 Diagrama de um LGC para uma variável latente η_t medida por múltiplos indicadores, com covariáveis variantes (w_t) e invariantes (x) ao longo do tempo.

Genericamente,

$$y_{ijt}^* = \nu_{jt} + \Lambda_{jt}\eta_{it} + v_{ijt}, \quad (5.12)$$

e

$$\eta_{it} = \alpha_i + \lambda_t\beta_i + \gamma_t w_{it} + \zeta_{it}, \quad (5.13)$$

com efeitos aleatórios dados por

$$\alpha_i = \mu_\alpha + \gamma_\alpha x_i + \zeta_{\alpha_i} \quad (5.14)$$

$$\beta_i = \mu_\beta + \gamma_\beta x_i + \zeta_{\beta_i}. \quad (5.15)$$

O modelo assume que $\zeta_{\alpha_i} \sim N(0, \Psi_\alpha)$; $\zeta_{\beta_i} \sim N(0, \Psi_\beta)$; e $\zeta_{it} \sim N(0, \Psi_t)$. Adicionalmente, $(w_{i1}, \dots, w_{it}, x_i)$, \mathbf{v}_i , ζ_{it} , ζ_{α_i} e ζ_{β_i} são mutuamente independentes com $i = 1, \dots, N$ e $t = 1, \dots, T$.

Os novos parâmetros de especial interesse no modelo são os coeficientes de regressão dos efeitos aleatórios sobre a TIC x (γ_α e γ_β), bem como os coeficientes de regressão das medidas repetidas de η_t sobre as medidas repetidas da TVC w_t (γ_t). Pode obter-se um modelo com maior parsimônia restringindo os efeitos γ_t a ser invariantes ao longo do tempo, i.e., $\gamma_1 = \gamma_2 = \dots = \gamma_T$.

5.5 Modelos com Estrutura Multi-Nível

Nas áreas das ciências sociais e comportamentais as problemáticas a estudar e os dados disponíveis têm muitas vezes uma estrutura hierárquica, com indivíduos e grupos em níveis distintos. Nesse caso não é apropriado utilizar procedimentos estatísticos de análise multivariada que não entrem em linha de conta com a estrutura hierárquica, mesmo que as análises apenas incluam variáveis ao nível do indivíduo, uma vez que o pressuposto de observações independentes e idênticamente distribuídas pode não ser válido. As consequências de usar métodos com um só nível de análise para dados com estrutura multinível são a obtenção de estimativas não enviesadas (mas não eficientes), erros padrão enviesados e, conseqüentemente, efeitos espúrios significativos - ver Hox (2010) e Snijders e Bosker (2012). No caso de modelos de regressão linear múltipla as técnicas de análise multinível estão desenvolvidas e consolidadas, existindo *software* específico amplamente divulgado (ver Raudenbush e Bryk, 2002, e Goldstein, 2011).

Muthén (1989, 1994) propõe uma abordagem para estimar Modelos com Equações Estruturais para dados com estrutura multinível usando programas como o LISREL, o EQS, o AMOS ou o Mplus. Heck e Thomas (2008) apresentam um exemplo detalhado da abordagem proposta por Muthén e discutem a sua implementação em LISREL e Mplus.

Uma abordagem alternativa à proposta por Muthén, designada por abordagem multinível multivariada, sugere a estimação das matrizes de covariância separadamente para cada nível, com recurso a programas específicos para estimar modelos multinível, usando depois essas matrizes de covariância como input da análise subsequente a realizar com recurso a um *software* para Modelos com Equações Estruturais (ver Hox, 2010, e Snijders e Bosker, 2012).

Hox e Maas (2004) descrevem sumariamente as duas abordagens referidas e discutem as vantagens de cada uma delas. Schumacker e Lomax (2010, páginas 307–320) ilustram a utilização do LISREL para estimar um modelo de regressão linear com estrutura multinível. Kaplan (2000) dedica o capítulo 7 a Modelos de Equações Estruturais com estrutura multinível e Jöreskog et al. (2001, capítulo 2) sintetiza as potencialidades do LISREL em termos de modelação para dados com estrutura multinível.

5.6 Indicadores Reflexivos versus Indicadores Formativos

Até agora temos vindo a usar indicadores reflexivos, que reflectem a variável latente subjacente que estão a medir. No diagrama do modelo tal é indicado por setas direccionadas, da variável latente para cada uma das variáveis manifestas. É o princípio subjacente à Análise Factorial Exploratória e Confirmatória: assume-se que os indicadores de uma mesma variável latente estão necessariamente correlacionados, sendo que os constructos latentes influenciam / explicam os indicadores observados, resultando os erros de medida da não capacidade dessa explicação das variáveis manifestas ser total. A *classical test theory* e os conceitos de validade e fiabilidade assumem indicadores reflexivos.

Alguns autores advogam o uso de indicadores formativos. Para uma introdução à temática ver Diamantopolous e Sigauw (2000, páginas 14, 16 e Apêndice 2A); para uma referência mais detalhada a modelos com indicadores formativos ver Brown (2006, páginas 351–361). Como o próprio nome indica, os indicadores formativos formam ou definem a variável “latente”, que na prática não é efectivamente uma variável latente mas sim uma variável manifesta que corresponde a uma média (ou soma) ponderada dos indicadores observados. Em termos de diagrama, as setas direccionadas teriam origem em cada um dos indicadores observados e apontariam para a tal variável soma ou média ponderada. Neste caso assume-se que as correlações entre os indicadores sevem ser mínimas. Na prática, em vez de modelos com indicadores formativos podem usar-se *summated scales* ou componentes principais. Os indicadores formativos são muito utilizados na abordagem PLS (*Partial Least Squares*). Para uma introdução à abordagem PLS ver Hair et al. (2010, páginas 775–778). Para um exemplo de modelação com recurso a PLS *path modelling* ver, por exemplo, Afonso (2008).

5.7 Outros Tópicos Ainda Não Mencionados

Para uma síntese sobre estimação de Modelos com Equações Estruturais com dados omissos ver Kaplan (2000, páginas 87–96). Para modelos com relações não lineares e efeitos de interacção ver Schumacker e Marcoulides (1998) ou Schumacker e Lomax (2010, páginas 327–341). Jöreskog et al. (2001, páginas 121–177) ilustram a estimação de modelos com interacção e efeitos não lineares em LISREL.

Para testar a validade de constructo usando o modelo *Multi-Trait Multi-Method* ver Byrne (1998, capítulo 6), ou Schumacker e Lomax (2010, páginas 277–286). Baron e Kenny (1986) é um artigo clássico

136 Outros Tópicos

sobre efeitos mediadores e moderadores. Para uma breve introdução ao estudo da mediação e da moderação em Modelos com Equações Estruturais ver Hair et al. (2010, páginas 766–773).

Por restrições de tempo e de espaço não é possível abordar neste curso e neste livro todos os tópicos, comprovando-se assim como os Modelos com Equações Estruturais são uma vasta e estimulante área de investigação e de aplicação.

Bibliografia

1. Afonso, J.A. (2008). *O Capital Humano, a Empregabilidade e a Participação dos Empregados na Propriedade da Empresa em que Trabalham*. Tese de Doutoramento. Instituto Universitário de Lisboa (ISCTE-IUL).
2. Akaike, H. (1987). Factor Analysis and AIC. *Psychometrika*, 52, 317–332.
3. Anderson, T.W. e Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, V, 111–150.
4. Bagozzi, R.P. e Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16, 74–94.
5. Baron, R.M. e Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research. *Journal of Personality and Social Psychology*, 51, 1173–1182.
6. Bartholomew, D., Knott, M. e Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. 3ª edição. John Wiley & Sons.
7. Bartholomew, D., Steele, F., Galbraith, J. e Moustaki, I. (2008). *Analysis of Multivariate Social Science Data*. 2ª edição. Chapman & Hall.
8. Bentler, P. (1986). Structural modeling and Psychometrika: an historical perspective on growth and achievements. *Psychometrika*, 51 (1), 35–51.
9. Bentler, P. e Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88 (3),

138 Bibliografia

- 588–606.
10. Bentler, P. e Chou, C-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16 (1), 78–113.
 11. Blunch, N.J. (2010). *Introduction to Structural Equation Modelling Using SPSS and AMOS*. SAGE Publications.
 12. Bollen, K.A. (1987). Outliers and improper solutions: confirmatory factor analysis example. *Sociological Methods & Research*, 15 (4), 375–384.
 13. Bollen, K.A. (1989a). *Structural Equations with Latent Variables*. John Wiley & Sons.
 14. Bollen, K.A. (1989b). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, 17 (3), 303–316.
 15. Bollen, K.A. (1990). Overall fit in covariance structure models: two types of sample size effects. *Psychological Bulletin*, 107 (2), 256–259.
 16. Bollen, K.A. e Curran, P.J. (2006). *Latent Curve Models: a Structural Equation Perspective*. John Wiley & Sons.
 17. Bollen, K. e Liang, J. (1988). Some properties of Hoelster's CN. *Sociological Methods & Research*, 16 (4), 492–503.
 18. Boker, S.M., Neale, M.C., Maes, H.H., Wilde, M.J., Spiegel, M., Brick, T.R., Estabrook, R., Bates, T.C., Mehta, P., von Oertzen, T., Gore, R.J., Hunetr, M.D., Hackett, D.C., Karch, J. e Brandmaier, A. (2012). *OpenMx - Reference Manual e OpenMx User Guide*.
 19. Boomsma, A. (1985). Nonconvergence, improper solutions and starting values in LISREL ML Estimation. *Psychometrika*, 50 (2), 229–242.
 20. Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, 52, 345–370.
 21. Breckler, S. (1990). Applications of covariance structure modeling in Psychology: cause for concern?. *Psychological Bulletin*, 107 (2), 260–273.
 22. Brown, T.A. (2006). *Confirmatory Factor Analysis for Applied Rese-*

- arch.* The Guilford Press.
23. Browne, M.W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.
 24. Browne, M.W. (1987). Robustness in statistical inference in factor analysis and related models. *Biometrika*, 74, 375–384.
 25. Browne, M.W. e Cudeck, R. (1993). Alternative ways of assessing model fit. Em Bollen, K.A e Long, J.S. (editores). *Testing Structural Equation Models*. SAGE Publications, 136–162.
 26. Byrne, B.M. (1989). *A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models*. Springer-Verlag.
 27. Byrne, B.M. (1998). *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Lawrence Erlbaum Associates.
 28. Byrne, B.M. (2009). *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming*. 2ª edição. Taylor & Francis / Routledge.
 29. Byrne, B.M. (2011). *Structural Equation Modeling with Mplus: Basic Concepts, Applications and Programming*. Taylor & Francis / Routledge.
 30. Collins, L.M. e Sayer, A.G., editores (2001). *New Methods for the Analysis of Change*. American Psychological Association.
 31. Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
 32. Cudeck, R. e Browne, M.W. (1983). Cross-validation of Covariance Structures. *Multivariate Behavioral Research*, 18 (2), 147–167.
 33. Cudeck, R., du Toit, S. e Sörbom, D., editores (2001). *Structural Equation Modeling: Present and Future. A Festschrift in honor of Karl Jöreskog*. Scientific Software International.
 34. Diamantopoulos, A. e Siguaw, J.A. (2000). *Introducing LISREL: a Guide for the Uninitiated*. SAGE Publications.

140 Bibliografia

35. Dias, A.L. (2009). *Satisfação no Trabalho em Portugal: uma análise longitudinal com recurso a Latent Growth Curve Models*. Dissertação de Mestrado. Instituto Universitário de Lisboa (ISCTE-IUL);
36. Duncan, O.D. (1966). Path analysis: Sociological examples. *American Journal of Sociology*, 72, 1–16.
37. Duncan, T.E, Duncan, S.C. e Strycker, L.A. (2006). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues and Applications*. 2ª edição. Lawrence Erlbaum Associates.
38. Everitt, B.S. (1984). *An Introduction to Latent Variable Models*. Chapman & Hall.
39. Fornell, C. e Larcker, D.F. (1981). Structural equation models with unobserved variables and measurement error. *Journal of Marketing Research*, 18, 39–50.
40. Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modelling: A Multidisciplinary Journal*, 13 (3), 465–486.
41. Golstein, H. (2011). *Multilevel Statistical Models*. 4ª edição. John Wiley & Sons.
42. Hair, J.F., Black, W.C., Babin, B.J. e Anderson, R.E. (2010). *Multivariate Data Analysis: A Global Perspective*. 7ª edição. Pearson Prentice Hall.
43. Heck, R.H. e Thomas, S.L. (2008). *An Introduction to Multilevel Modeling Techniques*. 2ª edição. Taylor & Francis / Routledge.
44. Hoyle, R.H., editor (1995). *Structural Equation Modeling - Concepts, Issues and Applications*. SAGE Publications.
45. Hoelter, J. (1983). The analysis of covariance structures: goodness-of-fit indices. *Sociological Methods & Research*, 11 (3), 325–344.
46. Hox, J.J. e Maas, C. (2004). Multilevel structural equation models: the limited information approach and the multivariate multilevel approach. Em van Monfort, K., Oud, J. e Satorra, A., (editores). *Recent Developments on Structural Equation Models: Theory and Applications*. Kluwer

- Academic Publishers, 135–149.
47. Hox, J.J. (2010). *Multilevel Analysis: Techniques and Applications*. 2ª edição. Taylor & Francis / Routledge.
 48. Jöreskog (1973). A general method for estimating a linear structural equation system. Em Goldberger, A.S. e Duncan, O.D. (editores), *Structural Equation Models in the Social Sciences*, Academic Press, 85–112.
 49. Jöreskog (2005). Structural Equation Modeling with Ordinal Variables using LISREL. *Online technical documentation*, Scientific Software International.
 50. Jöreskog, K. e Sörbom, D. (1990). Model search with Tetrad II and LISREL. *Sociological Methods & Research*, 19 (1), 93–106.
 51. Jöreskog, K. e Sörbom, D. (1993a). *LISREL 8 - User's Reference Guide*. Scientific Software International.
 52. Jöreskog, K. e Sörbom, D. (1993b). *PRELIS 2 - User's Reference Guide*. Scientific Software International.
 53. Jöreskog, K., Sörbom, D., du Toit, S. e du Toit, M. (2001). *LISREL 8: New Statistical Features*. Scientific Software International.
 54. Kaplan, D. (2000). *Structural Equation Modeling - Foundations and Extensions*. SAGE Publications.
 55. Kaplan, D., editor (2004). *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. SAGE Publications.
 56. Kelloway, E.K. (1998). *Using LISREL for Structural Equation Modeling - a Researcher's Guide*. SAGE Publications.
 57. Keesling, J.W. (1972). *Maximum Likelihood Approaches to Causal Analysis*. Dissertação de Doutorado não publicada, Universidade de Chicago.
 58. Loehlin, J.C. (2004). *Latent Variable Models: an Introduction to Factor, Path and Structural Equation Analysis*. 4ª edição. Lawrence Erlbaum Associates.
 59. Long, J.S. (1983a). *Confirmatory Factor Analysis - A Preface to LIS-*

142 Bibliografia

- REL*. Sage Publications.
60. Long, J.S. (1983b). *Covariance Structure Models - An Introduction to LISREL*. Sage Publications.
 61. Lawley, D.N. e Maxwell, A.E. (1971). *Factor analysis as a statistical method*. 2ª edição. Butterworths.
 62. MacCallum, R.C., Roznowski, M. e Necowitz, L.B. (1992). Model modifications in Covariance Structure Analysis: the problem of capitalization on chance. *Psychological Bulletin*, 111 (3), 490–504.
 63. Malta, J. (2009). *Modelação Longitudinal do Bem-Estar Psicológico com Modelos de Trajectória Latente*. Dissertação de Mestrado. Instituto Universitário de Lisboa (ISCTE-IUL);
 64. Marcoulides, G.A. e Moustaki, I., editores (2002). *Latent Variable and Latent Structure Models*. Lawrence Erlbaum Associates.
 65. McArdle, J.J. (1988). Dynamic but structural equation modeling of repeated measures. Em Nesselroade, J.R. e Cattell, R.B. (editores). *The Handbook of Multivariate Experimental Psychology*, vol.2, 561–614.
 66. McArdle, J.J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605.
 67. Meredith, W. e Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107–122.
 68. Mueller, R.O. (1996). *Basic Principles of Structural Equation Modeling - An Introduction to LISREL and EQS*. Springer-Verlag.
 69. Muthén, B.O. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551–560.
 70. Muthén, B.O. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, 49, 115–132.
 71. Muthén, B.O. (1989). Latent variable modelling in heterogeneous populations. *Psychometrika*, 54, 557–585.
 72. Muthén, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376–398.

73. Muthén, B.O. (2004). Latent variable analysis: growth mixture modeling and related techniques for longitudinal data. Em Kaplan, D. (editor). *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. SAGE Publications, 345–368.
74. Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44 (4), 443–459.
75. Olsson, U., Drasgow, F. e Dorans, N. (1982). The polyserial correlation coefficient. *Psychometrika*, 47 (3), 337–347.
76. Pereira, H.G. (2009). *Determinantes da Fidelização na Compra de Produtos Turísticos Online*. Tese de Doutoramento. Instituto Universitário de Lisboa (ISCTE-IUL).
77. Preacher, K.J. e Hayes, A.F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879–891.
78. Preacher, K.J., Wichman, A.L., MacCallum, R.C e Briggs, N.E. (2008). *Latent Growth Curve Modeling*. Sage Publications.
79. Raudenbush, S.W. e Bryk, A.S. (2002). *Hierarchical Linear Models*. 2ª edição. Sage Publications.
80. Raykov, T. e Marcoulides, G.A. (2006). *A First Course in Structural Equation Modeling*. 2ª edição. Lawrence Erlbaum Associates.
81. Rogosa, D. (1979). Causal models in longitudinal research: rationale, formulation and interpretation. Em *Longitudinal Research in the Study of Behavior and Development*. Academic Press, 263–302.
82. Rosseel, Y. (2012). lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software*, 48 (2).
83. Salgueiro, M.F. (1995). *Modelos com Equações Estruturais - Aplicação do LISREL a um Estudo de Socialização no Trabalho*. Dissertação de Mestrado. ISCTE.
84. Salgueiro, M.F. (2009). Modelling job satisfaction trajectories in the British Household Panel Survey using latent growth curve models. Em: Manso, J.P. e Monteiro, J.D. (editores), *Anais de Economia Aplicada*

144 Bibliografia

- ASEPELT'09 - XXIII International Congress of Applied Economics*, 23, 581–593.
85. Salgueiro, M.F. (2010). Latent Growth Curve Modelling of Portuguese Life Satisfaction Trajectories. *International Journal of Academic Research*, 2 (2), 188–194.
86. Salgueiro, M.F., Smith, P.W.F. e McDonald, J.W. (2008). The manifest association structure of the single-factor model: insights from partial correlations. *Psychometrika*, 73 (4), 665–670.
87. Salgueiro, M.F., Smith, P.W.F. e McDonald, J.W. (2010). Connections between single-factor analysis and graphical Gaussian models. *Multivariate Behavioral Research*, 45 (1), 135–152.
88. Salgueiro, M.F., Smith, P.W.F. e Vieira, M.D.T. (2011). A Multi-Process Second-Order Latent Growth Curve Model for Subjective Well-Being. *Quality & Quantity* (online first).
89. Salgueiro, M.F., Vieira, M.D.T. e Smith, P.W.F. (2010). Latent growth curve modelling of life satisfaction trajectories in the British Household Panel Survey. Em *Proceedings of the Statistics Canada Symposium 2009 - Longitudinal Surveys: from Design to Analysis*.
90. Saris, W., Satorra, A. e Sörbom, D. (1987). The Detection and Correction of Specification Errors in Structural Equation Models. *Sociological Methodology*, 17, 105–129.
91. Satorra, A. e Bentler E.M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. Em *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 308–313.
92. Satorra, A. e Bentler E.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. Em von Eye, A. e Clogg, C.C. (editores). *Latent Variable Analysis: Applications for Developmental Research*. SAGE Publications, 399–419.
93. Schumacker, R.E. e Marcoulides, G.A. (1998). *Interaction and Nonlinear Effects in Structural Equation Modeling*. Lawrence Erlbaum Asso-

- ciates.
94. Schumacker, R.E. e Lomax, R.G. (2010). *A Beginner's Guide to Structural Equation Modeling*. 3ª edição. Lawrence Erlbaum Associates.
 95. Serra, P.G. (2010). *Modelação Longitudinal da Satisfação Laboral em Portugal*. Dissertação de Mestrado. Instituto Universitário de Lisboa (ISCTE-IUL);
 96. Skrondal, A. e Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling - Multilevel, Longitudinal and Structural Equation Models*. Chapman & Hall.
 97. Snijders, T.A.B. e Bosker, R.J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. 2ª edição. Sage Publications.
 98. Sörbom, D. (1989). Model modification. *Psychometrika*, 54 (3), 371–384.
 99. Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.
 100. Taylor, M.F., Brice, J., Buck, N. e Prentice-Lane, E., editores (2008). *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. University of Essex.
 101. Tucker, L.R. e Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.
 102. van Monfort, K., Oud, J. e Satorra, A., editores (2004). *Recent Developments on Structural Equation Models: Theory and Applications*. Kluwer Academic Publishers.
 103. Vicente, P.C.R. e Salgueiro, M.F. (2012). Cross-lagged structural equation models vs latent growth curve models: a study of material deprivation in Portugal with ICOR. Em *Proceedings do XIX Congresso Anual da Sociedade Portuguesa de Estatística*. Aceite para publicação.
 104. Vitória, P.D., Salgueiro, M.F., Silva, S.A. e de Vries, H. (2009). The impact of social influence on adolescent intention to smoke: combining

146 Bibliografia

- types and referents of influence. *British Journal of Health Psychology*, 14, 681–699.
105. Vitória, P.D., Salgueiro, M.F., Silva, S.A. e de Vries, H. (2011). Social influence, intention to smoke, and adolescent smoking behaviour longitudinal relations. *British Journal of Health Psychology*, 16, 779–798.
106. Wiley, D.E. (1973). The identification problem in structural equation models with unmeasured variables. Em Goldberger, A.S. e Duncan, O.D. (editores). *Structural Equation Models in the Social Sciences*, Academic Press, 69–83.
107. Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.
108. Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161–215.

Apêndice A

Comandos e Outputs

Anotados - Modelos de

Análise Factorial

A.1 Modelo de Análise Factorial Exploratória em SPSS

Para proceder à Análise Factorial em SPSS, com estimação por máxima verossimilhança, obtenção de dois factores e uma solução com rotação varimax, importa realizar a seguinte sequência de comandos:

```
ANALYZE / DIMENSION REDUCTION / FACTOR
  VARIABLES: Sat2_03; Sat3_03; Sat5; Sat6; Sat7; Sat8_03
  DESCRIPTIVES
    STATISTICS: Initial solution
    CORRELATION MATRIX: Coefficients
  EXTRACTION
    METHOD: Maximum Likelihood
    DISPLAY: Unrotated factor solution;
    EXTRACT: fixed number of factors 2
  ROTATION
    METHOD: Varimax
    DISPLAY: Rotated solution
```

A Figura A.1 apresenta a matriz de correlações de Pearson das seis variáveis manifestas (abusivamente assumidas como quantitativas),

sobre as quais se irá aplicar análise factorial clássica (exploratória) com recurso ao SPSS. Testes de hipóteses (não apresentados) indiciam que várias dessas correlações são significativamente diferentes de zero; a própria matriz de correlações é significativamente diferente de uma matriz identidade (pelo teste de Bartlett), fazendo todo o sentido realizar uma Análise Factorial.

Correlation Matrix						
	Sat2_03	Sat3_03	Sat5_03	Sat6_03	Sat7_03	Sat8_03
Sat2_03	1,000	,407	,425	,377	,372	,366
Sat3_03	,407	1,000	,285	,296	,294	,316
Sat5_03	,425	,285	1,000	,377	,351	,315
Sat6_03	,377	,296	,377	1,000	,647	,721
Sat7_03	,372	,294	,351	,647	1,000	,673
Sat8_03	,366	,316	,315	,721	,673	1,000

Figura A.1 *Correlações entre as seis percepções de satisfação.*

Optou-se pelo método de estimação por máxima verosimilhança e, numa primeira fase, pediu-se ao SPSS que estimasse um modelo unifactorial. Alguns dos pesos factoriais obtidos foram muito baixos, tendo-se então optado por pedir um modelo com dois factores, que aliás parece estar mais de acordo com a estrutura de correlações entre variáveis manifestas que foi obtida. A Figura A.2 mostra os pesos factoriais, isto é, as correlações entre cada uma das seis variáveis manifestas e os dois factores latentes no modelo, antes e depois de uma rotação ortogonal *Varimax*. A solução rodada é de mais fácil interpretação: o Factor 1 está mais correlacionado com as variáveis *Sat8*, *Sat6* e *Sat7* (satisfação com aspectos de lazer), enquanto que o Factor 2 está mais correlacionado com as variáveis *Sat2*, *Sat5* e *Sat3* (satisfação com aspectos materiais). Note-se porém que, porque de um modelo de análise factorial clássica se trata, ambos os factores influenciam as seis variáveis manifestas, sendo os dois factores ortogonais.

Factor Matrix ^a			Rotated Factor Matrix ^a		
	Factor		Factor		
	1	2	1	2	
Sat8_03	,845	-,188	,825	,264	
Sat6_03	,824	-,131	,777	,302	
Sat7_03	,774	-,094	,716	,310	
Sat5_03	,472	,300	,196	,758	
Sat3_03	,425	,304	,257	,497	
Sat2_03	,551	,556	,213	,477	

Extraction Method: Maximum Likelihood.
a. 2 factors extracted. 5 iterations required.

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Figura A.2 Factor loadings *antes e depois da rotação* Varimax

A Figura A.3 mostra as comunialidades de cada uma das seis variáveis manifestas, isto é, a variância de cada variável X_i explicada pelos dois factores comuns no modelo. *Sat2*, *Sat3* e *Sat5* são as variáveis pior explicadas pelo modelo de dois factores ajustado, com valores de comunialidade inferiores a 0.5.

Communalities			Goodness-of-fit Test		
	Initial	Extraction	Chi-Square	df	Sig.
Sat2_03	,311	,612	22,644	4	,000
Sat3_03	,208	,273			
Sat5_03	,249	,313			
Sat6_03	,584	,696			
Sat7_03	,522	,608			
Sat8_03	,600	,750			

Extraction Method: Maximum Likelihood.

Figura A.3 *Comunalidades e ajustamento - solução com dois factores comuns*

A.2 Modelo de Análise Factorial Exploratória em PRELIS

Ilustra-se agora como utilizar o PRELIS (o pré-processador do LISREL) para estimar modelos de análise factorial exploratória.

A partir do LISREL, há que importar o ficheiro de dados do SPSS (designado por SPE2012_AFE.SAV) e gravá-lo como um ficheiro de dados do PRELIS (com extensão *.PSF*). Para tal há que executar os seguintes comandos:

```
FILE
      IMPORT DATA IN FREE FORMAT
      FILES OF TYPE: SPSS FOR WINDOWS
      SPE2012_AFE.sav
      SAVE AS: SPE2012_AFE.psf
```

Numa fase de exploração prévia dos dados em PRELIS há que ter o cuidado de definir a escala de medida das variáveis a utilizar na análise factorial exploratória. O comando

```
DATA
      DEFINE VARIABLES
```

permite, uma vez seleccionadas as variáveis, declará-las como contínuas, usando as opções

```
VARIABLE TYPE
      CONTINUOUS
      APPLY TO ALL - OK - OK
```

Note-se que, se este passo não for efectuado, e porque têm menos de 15 categorias de resposta, as variáveis serão tratadas pelo LISREL como ordinais, deixando o modelo e as estimativas obtidas de serem comparáveis com as do SPSS anteriormente apresentadas.

O comando

```
STATISTICS
DATA SCREENING
```

gera um ficheiro de output que contém estatísticas univariadas para cada uma das seis variáveis em análise, bem como testes univariados à normalidade de cada uma das variáveis. Apresentam-se alguns excertos:

```

P R E L I S  2.80 (STUDENT)
BY
Karl G. Jöreskog & Dag Sörbom
The following lines were read from file C:\LISREL_SPE2012\SPE2012_AFE.PR2:
!PRELIS SYNTAX: Can be edited
!Contents of PSFFILE:
!-----
!DA NI=6 NO=2008 MI= -999999 TR=LI
!LA
!Sat2_03 Sat3_03 Sat5_03 Sat6_03 Sat7_03 Sat8_03
!CO Sat2_03
!CO Sat3_03
!CO Sat5_03
!CO Sat6_03
!CO Sat7_03
!CO Sat8_03
!End of Contents
!-----
SY='C:\LISREL_SPE2012\SPE2012_AFE.PSF'
OU

```

São declaradas 6 variáveis de input, 2008 observações, o valor -999999 para os valores omissos, e um tratamento *listwise* das não respostas. São definidos os nomes das 6 variáveis, sendo todas elas declaradas como contínuas. É gerado um ficheiro de dados SPE2012_AFE.psf, que futuramente poderá ser lido directamente pelo PRELIS com a opção FILE OPEN.

Total Sample Size = 2008

152 Comandos e Outputs Anotados - Modelos de Análise Fatorial

Univariate Summary Statistics for Continuous Variables									
Variable	Mean	St.Dev.	T-Value	Skewness	Kurtosis	Min.	Freq.	Max.	Freq.
Sat2_03	4.816	1.293	166.836	-0.632	0.202	1.00	30	7.00	130
Sat3_03	5.427	1.212	200.620	-0.955	1.011	1.00	17	7.00	337
Sat5_03	4.980	1.343	166.136	-0.813	0.519	1.00	45	7.00	183
Sat6_03	4.895	1.256	174.674	-0.502	0.141	1.00	20	7.00	169
Sat7_03	4.449	1.390	143.399	-0.328	-0.288	1.00	53	7.00	113
Sat8_03	4.725	1.315	161.040	-0.441	-0.106	1.00	29	7.00	139

Para cada uma das variáveis são calculadas as seguintes medidas descritivas: a média, o desvio-padrão, o mínimo e o máximo (e a frequência com que ocorrem), uma medida de assimetria (*skewness*) e uma medida de achatamento (*kurtosis*). Recorde-se que numa distribuição normal o 3º momento estandardizado é igual a zero, e o 4º momento é igual a 3. Se o 3º momento estandardizado for maior que zero, a medida de *skewness* é positiva e a cauda direita da distribuição é mais extensa; se for menor que zero a medida de *skewness* é negativa. Se o 4º momento estandardizado for maior que 3, a medida de *kurtosis* é positiva e há menor concentração nas abas do que numa distribuição normal; se for menor que 3 a medida de *kurtosis* é negativa.

Test of Univariate Normality for Continuous Variables						
Variable	Skewness		Kurtosis		Skewness and Kurtosis	
	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
Sat2_03	-10.684	0.000	1.853	0.064	117.571	0.000
Sat3_03	-14.944	0.000	9.262	0.000	309.094	0.000
Sat5_03	-13.177	0.000	4.755	0.000	196.253	0.000
Sat6_03	-8.724	0.000	1.291	0.197	77.783	0.000
Sat7_03	-5.865	0.000	-2.639	0.008	41.360	0.000
Sat8_03	-7.740	0.000	-0.966	0.334	60.836	0.000

Os testes apresentados são testes às hipóteses nulas de medida de *skewness* nula, de medida de *kurtosis* nula e de normalidade univariada (simultaneamente medida de *skewness* nula e medida de *kurtosis* nula) - ver Bollen (1989a, páginas 418–421). No exemplo em análise é possível concluir que a variável *Sat3* é aquela cuja distribuição mais se afasta de uma distribuição normal.

Face aos resultados obtidos está comprometida a normalidade multi-

variada da distribuição conjunta das seis variáveis em análise. Todavia, para efeitos de ilustração da utilização do software, prosseguiremos com este exemplo.

O modelo clássico de Análise Factorial obtem-se em PRELIS com a seguinte sequência de comandos:

```

STATISTICS
  CLASSICAL FACTOR ANALYSIS
    VARIABLE LIST
      Sat2_03, Sat3_03, Sat5, Sat6, Sat7_03, Sat8_03
    SELECT
  ML FACTOR ANALYSIS
  OUTPUT OPTIONS
    PERFORM TESTS MULTIVARIATE NORMALITY
  OK
  RUN
    
```

É então gerado um ficheiro de output. Apresentam-se alguns excertos:

```

!PRELIS SYNTAX: Can be edited
SY='C:\LISREL_SPE2012\SPE2012_AFE.PSF'
SE 1 2 3 4 5 6
FA
OU MA=KM
    
```

O ficheiro de sintaxe criado selecciona as seis variáveis, executa Análise Factorial Exploratória e trabalha com uma matriz de correlações.

Test of Multivariate Normality for Continuous Variables							
Skewness			Kurtosis			Skewness and Kurtosis	
Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Square	P-Value
3.708	28.775	0.000	64.070	21.515	0.000	1290.908	0.000

São realizados testes de hipóteses multivariados de medida de *skewness* nula, de medida de *kurtosis* nula e de normalidade multivariada

154 Comandos e Outputs Anotados - Modelos de Análise Factorial

(simultaneamente medida de *skewness* multivariada nula e medida de *kurtosis* multivariada nula) - ver Bollen (1989a, páginas 422–424). Como era de esperar, face aos resultados dos testes univariados, é de rejeitar a hipótese nula de que as seis variáveis provêm de uma população com distribuição normal multivariada ($\text{valor-}p = 0.000 < 5\%$).

Correlation Matrix						
	Sat2_03	Sat3_03	Sat5_03	Sat6_03	Sat7_03	Sat8_03
Sat2_03	1.000					
Sat3_03	0.407	1.000				
Sat5_03	0.425	0.285	1.000			
Sat6_03	0.377	0.296	0.377	1.000		
Sat7_03	0.372	0.294	0.351	0.647	1.000	
Sat8_03	0.366	0.316	0.315	0.721	0.673	1.000
Means						
	Sat2_03	Sat3_03	Sat5_03	Sat6_03	Sat7_03	Sat8_03
	4.816	5.427	4.980	4.895	4.449	4.725
Standard Deviations						
	Sat2_03	Sat3_03	Sat5_03	Sat6_03	Sat7_03	Sat8_03
	1.293	1.212	1.343	1.256	1.390	1.315

É disponibilizada a matriz de correlações amostrais de Pearson das 6 variáveis iniciais, bem como o vector de médias amostrais e o vector de desvios-padrão amostrais.

A.3 Modelo de Análise Factorial Confirmatória em LISREL

Utilizando o PRELIS, uma breve exploração dos dados com o comando

```
STATISTICS
```

```
DATA SCREENING
```

permite concluir ser a seguinte a distribuição das respostas dos 2008 respondentes, para as seis categorias de resposta das seis variáveis manifestas ordinais:

Sat2_03 Frequency Percentage			Sat6_03 Frequency Percentage		
1	30	1.5	1	20	1.0
2	85	4.2	2	62	3.1
3	197	9.8	3	185	9.2
4	367	18.3	4	403	20.1
5	693	34.5	5	679	33.8
6	506	25.2	6	490	24.4
7	130	6.5	7	169	8.4
Sat3_03 Frequency Percentage			Sat7_03 Frequency Percentage		
1	17	0.8	1	53	2.6
2	30	1.5	2	128	6.4
3	102	5.1	3	306	15.2
4	236	11.8	4	469	23.4
5	504	25.1	5	595	29.6
6	782	38.9	6	344	17.1
7	337	16.8	7	113	5.6
Sat5_03 Frequency Percentage			Sat8_03 Frequency Percentage		
1	45	2.2	1	29	1.4
2	67	3.3	2	82	4.1
3	153	7.6	3	248	12.4
4	337	16.8	4	422	21.0
5	605	30.1	5	639	31.8
6	618	30.8	6	449	22.4
7	183	9.1	7	139	6.9

Recorde-se que se assume subjacente a cada variável ordinal uma variável latente contínua, com distribuição normal, com cujos valores é possível calcular correlações policóricas. Uma correlação policórica é uma correlação entre duas variáveis latentes contínuas (subjacentes a duas variáveis manifestas ordinais) na sua distribuição normal bivariada. O PRELIS calcula as correlações policóricas, bem como a matriz das variâncias e covariâncias assintóticas destas correlações amostrais, matrizes essas que serão gravadas em ficheiros externos, que serão posteriormente utilizados pelo LISREL para efeitos de estimação do modelo em estudo. Para obter tais matrizes há que efectuar a seguinte sequência de comandos, recorrendo aos menus do PRELIS:

STATISTICS

OUTPUT OPTIONS

MOMENT MATRIX: CORRELATIONS

SAVE TO FILE: GLOBAL.COR

```

ASYMPTOTIC COVARIANCE MATRIX
SAVE TO FILE: GLOBAL.ACM
PRINT BIVARIATE FREQUENCY TABLES
PRINT TESTS UNDERLYING BIVARIATE NORMALITY
OK
    
```

No ficheiro de output do PRELIS é disponibilizada informação referente às variáveis latentes contínuas subjacentes às manifestas ordinais. Trata-se de variáveis estandardizadas com média zero e desvio-padrão unitário. Como cada variável ordinal tem $m = 7$ categorias são calculados $m - 1 = 6$ limiares (*thresholds*), π , para cada uma das variáveis latentes subjacentes. Cada π_i está associado à probabilidade de resposta na categoria i da correspondente variável manifesta ordinal:

Univariate Marginal Parameters

Variable	Mean	St. Dev.	Thresholds					
-----	----	-----	-----	-----	-----	-----	-----	-----
Sat2_03	0.000	1.000	-2.172	-1.578	-1.014	-0.418	0.477	1.516
Sat3_03	0.000	1.000	-2.388	-1.988	-1.445	-0.872	-0.144	0.963
Sat5_03	0.000	1.000	-2.006	-1.591	-1.117	-0.525	0.256	1.334
Sat6_03	0.000	1.000	-2.328	-1.741	-1.112	-0.430	0.445	1.378
Sat7_03	0.000	1.000	-1.937	-1.340	-0.698	-0.060	0.747	1.587
Sat8_03	0.000	1.000	-2.185	-1.596	-0.920	-0.282	0.545	1.482

As correlações policóricas são estimadas a partir das distribuições marginais das variáveis contínuas subjacentes às manifestas ordinais, face aos limiares calculados. No exemplo em estudo foram obtidos os seguintes valores para as correlações policóricas:

Correlation Matrix

	Sat2_03	Sat3_03	Sat5_03	Sat6_03	Sat7_03	Sat8_03
-----	-----	-----	-----	-----	-----	-----
Sat2_03	1.000					
Sat3_03	0.452	1.000				
Sat5_03	0.456	0.326	1.000			
Sat6_03	0.400	0.333	0.414	1.000		
Sat7_03	0.400	0.326	0.385	0.687	1.000	
Sat8_03	0.391	0.349	0.348	0.754	0.711	1.000

O cálculo de correlações policóricas pressupõe a normalidade bivariada dos pares de variáveis contínuas latentes subjacentes às variáveis ordinais manifestas. Associado a cada coeficiente de correlação estimado, o PRELIS disponibiliza um teste à hipótese nula de normalidade bivariada (a estatística de teste tem distribuição de χ^2). Jöreskog diz tratar-se de um teste muito sensível a desvios de normalidade, razão pela qual desenvolveu uma medida tipo RMSEA, de discrepância face à população. Defende o autor que rejeitar a hipótese de normalidade bivariada das variáveis latentes subjacentes às ordinais manifestas não será grave desde que a medida RMSEA proposta seja inferior a 0.1. O autor desenvolveu ainda um teste de hipóteses à hipótese nula de que o valor de RMSEA na população é inferior a 0.1. O PRELIS disponibiliza os valores-*p* associados a cada um destes testes de hipóteses:

Correlations and Test Statistics (PE=Pearson Product Moment, PC=Polychoric, PS=Polyserial)								
Variable vs.	Variable	Correlation	Test of Model			Test of Close Fit		
			Chi-Squ.	D.F.	P-Value	RMSEA	P-Value	
Sat3_03 vs.	Sat2_03	0.452 (PC)	154.834	35	0.000	0.041	1.000	
Sat5_03 vs.	Sat2_03	0.456 (PC)	149.085	35	0.000	0.040	1.000	
Sat5_03 vs.	Sat3_03	0.326 (PC)	125.723	35	0.000	0.036	1.000	
Sat6_03 vs.	Sat2_03	0.400 (PC)	90.049	35	0.000	0.028	1.000	
Sat6_03 vs.	Sat3_03	0.333 (PC)	117.428	35	0.000	0.034	1.000	
Sat6_03 vs.	Sat5_03	0.414 (PC)	116.943	35	0.000	0.034	1.000	
Sat7_03 vs.	Sat2_03	0.400 (PC)	105.494	35	0.000	0.032	1.000	
Sat7_03 vs.	Sat3_03	0.326 (PC)	85.975	35	0.000	0.027	1.000	
Sat7_03 vs.	Sat5_03	0.385 (PC)	76.959	35	0.000	0.024	1.000	
Sat7_03 vs.	Sat6_03	0.687 (PC)	217.444	35	0.000	0.051	1.000	
Sat8_03 vs.	Sat2_03	0.391 (PC)	106.398	35	0.000	0.032	1.000	
Sat8_03 vs.	Sat3_03	0.349 (PC)	120.931	35	0.000	0.035	1.000	
Sat8_03 vs.	Sat5_03	0.348 (PC)	112.902	35	0.000	0.033	1.000	
Sat8_03 vs.	Sat6_03	0.754 (PC)	201.774	35	0.000	0.049	1.000	
Sat8_03 vs.	Sat7_03	0.711 (PC)	463.084	35	0.000	0.078	1.000	

Neste caso é possível concluir ser de rejeitar a hipótese de normalidade bivariada para todos os pares de variáveis latentes subjacentes às manifestas ordinais, não sendo o efeito da não-normalidade grave em nenhum dos casos, pois nunca se rejeita a hipótese de que $RMSEA < 0.1$. No caso de a normalidade bivariada das variáveis la-

158 Comandos e Outputs Anotados - Modelos de Análise Fatorial

tentes subjacentes às manifestas ordinais não se verificar pode optar-se por tentar agrupar categorias de resposta de uma das variáveis (especialmente se houver células com muito poucas observações), ou mesmo por retirar da análise a variável que parecer mais problemática.

Estimação WLS do modelo de AFC

O LISREL permite estimar o modelo de AFC especificado na Secção 2.2.4 utilizando a seguinte sequência de comandos de sintaxe:

```
TI Analise Factorial Confirmatoria: Material e Lazer
DA NI=30 NO=2008 MA=KM
LA
pid idade sexo criancas n_horas logRend
Sat2_02 Sat3_02 Sat5_02 Sat6_02 Sat7_02 Sat8_02
Sat2_03 Sat3_03 Sat5_03 Sat6_03 Sat7_03 Sat8_03
Sat2_04 Sat3_04 Sat5_04 Sat6_04 Sat7_04 Sat8_04
Sat2_05 Sat3_05 Sat5_05 Sat6_05 Sat7_05 Sat8_05
KM FI='C:\LISREL_SPE2012\global.cor' SY
AC FI=C:\LISREL_SPE2012\global.acm
SE
Sat2_03 Sat3_03 Sat5_03 Sat6_03 Sat7_03 Sat8_03 /
MO NX=6 NK=2 LX=FU,FI PH=SY,FI TD=DI,FR
LK
MATER03 LAZER03
FR LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2)
VA 1 PH(1,1) PH(2,2)
FR PH(2,1)
PD
OU ME=WL ND=3 MI
```

Após especificar o título (comando TI), o número de variáveis no ficheiro de input (comando NI) e o número de observações na amostra (comando NO), importa especificar as etiquetas das variáveis (comando LA) e os ficheiros que contém as matrizes de correlação a serem utilizadas (KM designa a matrix de correlações e AC designa a matriz das variâncias e covariâncias assintóticas das correlações amostrais). Seleccionadas as variáveis manifestas a incluir no modelo (comando SE) importa especificar o modelo, isto é, declarar as variáveis manifestas como indicadores independentes, declarar as variáveis latentes

também como independentes, especificar os parâmetros do modelo e, se as houver, especificar as restrições sobre parâmetros do modelo. Por último, há que escolher o método de estimação (neste caso optou-se por WLS) e que output se pretende visualizar: pediram-se 3 casas decimais (comando ND) e a disponibilização dos índices de modificação (comando MI). Note-se que declarar WLS como método de estimação a usar implica definir qual a matriz de pesos cuja inversa irá ser usada na função de ajustamento para efeitos de estimação - a matriz de variâncias e covariâncias assintóticas das correlações amostrais (linha de comando AC).

Os 13 parâmetros a estimar são os seguintes (sequencialmente numerados pelo LISREL de 1 a 13):

```

Parameter Specifications
      LAMBDA-X
            MATER03    LAZER03
            -----    -----
Sat2_03           1           0
Sat3_03           2           0
Sat5_03           3           0
Sat6_03           0           4
Sat7_03           0           5
Sat8_03           0           6
      PHI
            MATER03    LAZER03
            -----    -----
MATER03           0
LAZER03           7           0
      THETA-DELTA
            Sat2_03    Sat3_03    Sat5_03    Sat6_03    Sat7_03    Sat8_03
            -----    -----    -----    -----    -----    -----
                        8           9           10          11          12          13
    
```

As estimativas obtidas para Λ_X foram as seguintes:

LAMBDA-X		
	MATER03	LAZER03
	-----	-----
Sat2_03	0.732 (0.021) 34.988	- -
Sat3_03	0.593 (0.023) 25.819	- -
Sat5_03	0.631 (0.024) 26.385	- -
Sat6_03	- -	0.867 (0.012) 73.726
Sat7_03	- -	0.820 (0.013) 63.998
Sat8_03	- -	0.885 (0.013) 66.014

As estimativas obtidas são todas admissíveis. Porque estamos a trabalhar com uma matriz de correlações e as variáveis latentes têm variância unitária (solução estandardizada), as estimativas obtidas têm todas de estar entre -1 e 1 . As estimativas obtidas para Φ e Θ_{δ} têm todas valores admissíveis:

PHI						
	MATER03	LAZER03				
	-----	-----				
MATER03	1.000					
LAZER03	0.681 (0.022) 30.507	1.000				
THETA-DELTA						
	Sat2_03	Sat3_03	Sat5_03	Sat6_03	Sat7_03	Sat8_03
	-----	-----	-----	-----	-----	-----
	0.464 (0.038) 12.249	0.648 (0.035) 18.386	0.602 (0.038) 16.043	0.248 (0.030) 8.213	0.328 (0.031) 10.697	0.216 (0.033) 6.642

Os valores de R^2 são elevados para os indicadores de medida de LAZER03. Os indicadores de MATER03 Sat3_03 e Sat5_03 são os

que têm os valores de R^2 mais baixos, sendo também os que possuem os menores valores de pesos factoriais estandardizados.

Squared Multiple Correlations for X - Variables

Sat2_03	Sat3_03	Sat5_03	Sat6_03	Sat7_03	Sat8_03
0.536	0.352	0.398	0.752	0.672	0.78

Apresentam-se algumas medidas da qualidade de ajustamento modelodados:

```

Goodness of Fit Statistics
Degrees of Freedom = 8
Minimum Fit Function Chi-Square = 33.242 (P = 0.000)
Root Mean Square Error of Approximation (RMSEA) = 0.0397
90 Percent Confidence Interval for RMSEA = (0.0263 ; 0.0541)
Comparative Fit Index (CFI) = 0.993
Standardized RMR = 0.0233
Goodness of Fit Index (GFI) = 0.998
    
```

Estimação *Robust ML* do modelo de AFC

Para estimar este mesmo modelo de AFC por *Robust ML* importa substituir no ficheiro de sintaxe a linha de comandos de output por

```
OU ME=ML ND=3 MI
```

Neste exemplo concreto as estimativas obtidas são muito similares, o mesmo se passando com as medidas de qualidade de ajustamento. Apresentam-se os valores de algumas das medidas obtidas:

```

Goodness of Fit Statistics
Degrees of Freedom = 8
Minimum Fit Function Chi-Square = 59.827 (P = 0.00)
Satorra-Bentler Scaled Chi-Square = 33.152 (P = 0.000)
Root Mean Square Error of Approximation (RMSEA) = 0.0396
90 Percent Confidence Interval for RMSEA = (0.0262 ; 0.0540)
Comparative Fit Index (CFI) = 0.996
Standardized RMR = 0.0216
Goodness of Fit Index (GFI) = 0.990
    
```

162 Comandos e Outputs Anotados - Modelos de Análise Factorial

Seguem-se os **resíduos** estandardizados obtidos:

```
Largest Negative Standardized Residuals
Residual for Sat5_03 and Sat3_03 -3.175
Residual for Sat6_03 and Sat2_03 -3.827
Residual for Sat8_03 and Sat2_03 -4.165
Largest Positive Standardized Residuals
Residual for Sat6_03 and Sat5_03 5.817
Residual for Sat7_03 and Sat5_03 4.473
```

Note-se que o maior resíduo positivo standardizado está associado a *Sat5_03* e *Sat6_03*, seguido de um outro associado a *Sat5_03* e *Sat7_03*, o que indicia que o modelo sub-estima a relação entre *Sat5_03* e as outras duas variáveis *Sat6_03* e *Sat7_03*. Uma vez que não há razão substantiva para deixar correlacionar erros de medida entre estas variáveis, a opção a tomar passa por, ou deixar o modelo de medida como está, ou considerar um outro modelo de medida que não inclua o indicador *Sat5_03*, que já anteriormente revelou ter baixa fiabilidade. No âmbito deste livro optámos pela primeira solução.

Apêndice B

Comandos e Outputs

Anotados - Modelos

Estruturais

B.1 Modelos de Regressão Linear

B.1.1 MRLS

Comandos de sintaxe do LISREL que permitem estimar o modelo de regressão linear simples da Secção 3.1.4:

```
TI Exemplo de Regressao Linear Simples
DA NI=14 NO=2008 MA=CM
LA
pid idade sexo criancas n_horas logRend Lazer02 Lazer03 Lazer04
Lazer05 Mater02 Mater03 Mater04 Mater05
CM FI='C:\LISREL_SPE2012\estrutural.cov' SY
ME FI='C:\LISREL_SPE2012\estrutural.meas'
SE
Mater02 logRend /
MO NX=1 NY=1 GA=FU,FR PS=SY,FR AL=FR
PD
OU SC ND=3
```

Para efeitos de ilustração optou-se por introduzir o intercepto no modelo: o parâmetro α é Free , como se pode ver na linha que define o Modelo. Contudo, no caso do exemplo em análise, o intercepto não tem uma interpretação substantiva relevante.

Numa solução estandardizada de um MRLS a estimativa do coeficiente de regressão iguala o coeficiente de correlação entre X e Y :

164 Comandos e Outputs Anotados - Modelos Estruturais

```
Standardized Solution
      GAMMA
      logRend
      -----
Mater02      0.184
      Correlation Matrix of Y and X
      Mater02      logRend
      -----      -----
Mater02      1.000
logRend      0.184      1.000
      PSI
      Mater02
      -----
      0.966
```

B.1.2 MRLM

Comandos de sintaxe do LISREL que permitem estimar o modelo de regressão linear múltipla da Secção 3.1.4:

```
TI Exemplo de Regressao Linear Multipla
DA NI=14 NO=2008 MA=CM
LA
pid idade sexo criancas n_horas logRend Lazer02 Lazer03 Lazer04
Lazer05 Mater02 Mater03 Mater04 Mater05
CM FI='C:\LISREL_SPE2012\estrutural.cov' SY
ME FI='C:\LISREL_SPE2012\estrutural.mea'
SE
Mater02 logRend idade sexo /
MO NX=3 NY=1 GA=FU,FR PS=SY,FR AL=FR
PD
OU SC ND=3
```

Continua a incluir-se o vector de médias das três variáveis independentes já que se pretende estimar o intercepto da equação de regressão. Apresentam-se alguns excertos do ficheiro de output obtido. A matriz de variância-covariância e o vector de médias amostrais:

```
Covariance Matrix
      Mater02      logRend      idade
      -----      -----      -----
Mater02      8.687
logRend      0.120      0.049
idade      2.761      -0.096      127.780
```

Means		
Mater02	logRend	idade
-----	-----	-----
14.946	3.479	40.617

As estimativas obtidas numa solução estandardizada:

Standardized Solution			
GAMMA			
	logRend	idade	sexo
-----	-----	-----	-----
Mater02	0.194	0.087	-0.065
Correlation Matrix of Y and X			
	Mater02	logRend	idade
-----	-----	-----	-----
Mater02	1.000		
logRend	0.184	1.000	
idade	0.083	-0.038	1.000
PSI			
	Mater02		
-----	-----		
	0.954		

B.2 Modelo de Análise de Caminhos

Para efeitos de estimação do modelo usando o LISREL a matriz B vai ser declarada como SD - *subdiagonal* (triangular inferior), a matriz Φ como SY - simétrica, a matriz Γ como FU - *full* (já que se trata de um vector em que todos os coeficientes devem ser estimados) e a matriz Ψ como DI - *diagonal* (já que apenas as variâncias dos termos residuais das relações estruturais são estimadas; recorde-de que num modelo recursivo de análise de caminhos se assume que os termos residuais das relações estruturais não estão correlacionados).

Uma vez que não estamos interessados nos interceptos das equações estruturais, não vamos fornecer os vectores de médias das variáveis manifestas nem especificar o vector α na linha de definição do modelo. Os comandos de sintaxe do LISREL que permitem estimar o modelo recursivo de análise de caminhos em estudo são os seguintes:

166 Comandos e Outputs Anotados - Modelos Estruturais

```

TI Exemplo de Path Analysis
DA NI=14 NO=2008 MA=CM
LA
pid idade sexo criancas n_horas logRend Lazer02 Lazer03 Lazer04
Lazer05 Mater02 Mater03 Mater04 Mater05
CM FI='C:\LISREL_SPE2012\estrutural.cov' SY
SE
Mater02 Lazer03 logRend idade sexo n_horas /
MO NX=4 NY=2 GA=FU,FI BE=SD PS=DI,FR PH=SY,FR
FR GA(1,1) GA(1,2) GA(1,3) GA(2,3) GA(2,4)
PD
OU SC EF ND=3

```

O comando de selecção de variáveis SE permite declarar primeiro as variáveis Y (por ordem: Y_1 e Y_2) e só depois as variáveis X (por ordem: X_1 , X_2 , X_3 e X_4). Se nem todas as variáveis declaradas em NI forem incluídas no modelo, a linha correspondente ao comando SE tem que acabar com um traço /. O comando SC na linha de Output solicita uma solução completamente estandardizada (as variáveis manifestas passam a ter variância unitária). O comando EF permite solicitar o cálculo dos efeitos indirectos e totais. Foram obtidos os seguintes valores para as estimativas, valores dos testes t e erros-padrão:

Number of Iterations = 5

BETA					
	Mater02	Lazer03			
	-----	-----			
Mater02	- -	- -			
Lazer03	0.448	- -			
	(0.025)				
	18.175				
GAMMA					
	logRend	idade	sexo	n_horas	
	-----	-----	-----	-----	
Mater02	2.575	0.023	-0.192	- -	
	(0.291)	(0.006)	(0.065)		
	8.835	3.990	-2.977		
Lazer03	- -	- -	0.285	-0.027	
			(0.088)	(0.008)	
			3.227	-3.267	

PHI

	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
logRend	0.049 (0.002) 31.646			
idade	-0.096 (0.056) -1.721	127.780 (4.038)		
sexo	0.020 (0.005) 4.066	-0.542 (0.253)	1.000 (0.032)	
n_horas	0.358 (0.053) 6.807	-14.312 (2.669) -5.362	5.977 (0.270) 22.153	110.070 (3.478) 31.646

PSI

Note: This matrix is diagonal.

	Mater02	Lazer03
	-----	-----
	8.284 (0.262) 31.646	10.548 (0.333) 31.646

Porque o índice de estabilidade (= 0.201) é inferior a 1 é possível calcular efeitos indirectos e totais.

Total and Indirect Effects

Total Effects of X on Y

	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
Mater02	2.575 (0.291) 8.835	0.023 (0.006)	-0.192 (0.065) -2.977	- -
Lazer03	1.154 (0.145) 7.946	0.010 (0.003) 3.897	0.199 (0.093) 2.142	-0.027 (0.008) -3.267

Indirect Effects of X on Y

	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
Mater02	- -	- -	- -	- -
Lazer03	1.154 (0.145) 7.946	0.010 (0.003) 3.897	-0.086 (0.029) -2.938	- -

168 Comandos e Outputs Anotados - Modelos Estruturais

Total Effects of Y on Y

	Mater02	Lazer03
Mater02	- -	- -
Lazer03	0.448 (0.025) 18.175	- -

Largest Eigenvalue of B*B' (Stability Index) is 0.201

Standardized Total and Indirect Effects

Standardized Total Effects of X on Y

	logRend	idade	sexo	n_horas
Mater02	0.194	0.087	-0.065	- -
Lazer03	0.073	0.033	0.057	-0.082

Standardized Indirect Effects of X on Y

	logRend	idade	sexo	n_horas
Mater02	- -	- -	- -	- -
Lazer03	0.073	0.033	-0.025	- -

Standardized Total Effects of Y on Y

	Mater02	Lazer03
Mater02	- -	- -
Lazer03	0.376	- -

B.3 Modelo Equações Estruturais Global

O ficheiro de sintaxe com os comandos LISREL que permite estimar o modelo da Figura 3.5 é o seguinte:

```
TI Modelo Global
DA NI=30 NO=2008 MA=KM
LA
pid idade sexo crianças n_horas logRend
Sat2_02 Sat3_02 Sat5_02 Sat6_02 Sat7_02 Sat8_02
```

```

Sat2_03 Sat3_03 Sat5_03 Sat6_03 Sat7_03 Sat8_03
Sat2_04 Sat3_04 Sat5_04 Sat6_04 Sat7_04 Sat8_04
Sat2_05 Sat3_05 Sat5_05 Sat6_05 Sat7_05 Sat8_05
KM FI='C:\LISREL_SPE2012\global.cor' SY
AC FI='C:\LISREL_SPE2012\global.acm
SE
7 8 9 16 17 18 6 2 3 5/
MO NX=4 NY=6 NE=2 LY=FU,FI PH=SY,FR PS=DI,FR TE=DI,FR BE=SD GA=FU,FI
LE
MATERO2 LAZER03
VA 1 LY(1,1) LY(4,2)
FR LY(2,1) LY(3,1) LY(5,2) LY(6,2)
FR GA(1,1) GA(1,2) GA(1,3) GA(2,3) GA(2,4)
PD
OU ME=WL ND=3 SC MI

```

Apresentam-se excertos de alguns dos outputs obtidos. A matriz de correlações observada entre variáveis manifestas:

Correlation Matrix

	Sat2_02	Sat3_02	Sat5_02	Sat6_03	Sat7_03	Sat8_03
Sat2_02	1.000					
Sat3_02	0.414	1.000				
Sat5_02	0.437	0.312	1.000			
Sat6_03	0.284	0.263	0.290	1.000		
Sat7_03	0.297	0.232	0.285	0.687	1.000	
Sat8_03	0.284	0.280	0.238	0.754	0.711	1.000
logRend	0.319	0.077	0.031	0.034	-0.023	0.032
idade	0.026	0.101	0.068	-0.015	0.022	0.023
sexo	-0.030	-0.006	-0.107	-0.025	-0.007	0.060
n_horas	0.014	-0.025	-0.054	-0.007	-0.093	-0.026

	logRend	idade	sexo	n_horas
logRend	1.000			
idade	-0.038	1.000		
sexo	0.091	-0.048	1.000	
n_horas	0.154	-0.121	0.570	1.000

As estimativas dos parâmetros da componente estrutural do modelo, na solução não estandardizada, e correspondentes valores t , indiciam que as relações estruturais postuladas são todas verificadas:

170 Comandos e Outputs Anotados - Modelos Estruturais

BETA

	MATERO2	LAZER03
	-----	-----
MATERO2	- -	- -
LAZER03	0.594	- -
	(0.036)	
	16.365	

GAMMA

	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
MATERO2	0.076	0.044	-0.067	- -
	(0.022)	(0.016)	(0.023)	
	3.505	2.778	-2.910	
LAZER03	- -	- -	0.099	-0.067
			(0.030)	(0.022)
			3.296	-2.982

As estimativas dos parâmetros do modelo de medida, na solução não estandardizada:

LAMBDA-Y

	MATERO2	LAZER03
	-----	-----
Sat2_02	1.000	- -
Sat3_02	0.784	- -
	(0.042)	
	18.485	
Sat5_02	0.839	- -
	(0.043)	
	19.404	
Sat6_03	- -	1.000
Sat7_03	- -	0.948
		(0.022)
		43.000
Sat8_03	- -	1.018
		(0.020)
		50.485

Por último, as estimativas dos ϕ , dos ψ e dos θ_ϵ :

PHI

	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
logRend	1.000 (0.022) 44.795			
idade	0.060 (0.056) 1.087	1.000 (0.022) 44.828		
sexo	0.059 (0.017) 3.551	-0.058 (0.016) -3.559	1.000 (0.022) 44.851	
n_horas	-0.023 (0.054) -0.429	0.520 (0.267) 1.948	0.557 (0.016) 34.536	1.000 (0.022) 44.812

PSI

Note: This matrix is diagonal.

MATER02	LAZER03
-----	-----
0.535 (0.032) 16.683	0.563 (0.025) 22.526

THETA-EPS

Sat2_02	Sat3_02	Sat5_02	Sat6_03	Sat7_03	Sat8_03
-----	-----	-----	-----	-----	-----
0.453 (0.040) 11.313	0.664 (0.036) 18.540	0.614 (0.037) 16.818	0.244 (0.032) 7.637	0.320 (0.031) 10.314	0.216 (0.032) 6.842

B.4 Modelo Global com Efeito Longitudinal

A seqüência de comandos de sintaxe LISREL que permite estimar o modelo da Figura 1.6 é a seguinte:

```

TI Exemplo de Modelo Global cross-lagged
DA NI=30 NO=2008 MA=KM
LA
pid idade sexo crianas n_horas logRend
Sat2_02 Sat3_02 Sat5_02 Sat6_02 Sat7_02 Sat8_02
Sat2_03 Sat3_03 Sat5_03 Sat6_03 Sat7_03 Sat8_03
Sat2_04 Sat3_04 Sat5_04 Sat6_04 Sat7_04 Sat8_04
Sat2_05 Sat3_05 Sat5_05 Sat6_05 Sat7_05 Sat8_05
KM FI='C:\LISREL_SPE2012\global.cor' SY
AC FI=C:\LISREL_SPE2012\global.acm
SE
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 6 2 3 5 /
MO NX=4 NY=24 NE=8 LY=FU,FI PH=SY,FR PS=SY,FI TE=SY,FI BE=FU,FI GA=FU,FI
LE
MATERO2 LAZER02 MATERO3 LAZER03 MATERO4 LAZER04 MATERO5 LAZER05
FR LY(1,1) LY(2,1) LY(3,1) LY(4,2) LY(5,2) LY(6,2)
FR LY(7,3) LY(8,3) LY(9,3) LY(10,4) LY(11,4) LY(12,4)
FR LY(13,5) LY(14,5) LY(15,5) LY(16,6) LY(17,6) LY(18,6)
FR LY(19,7) LY(20,7) LY(21,7) LY(22,8) LY(23,8) LY(24,8)
FR GA(1,1) GA(1,2) GA(1,3) GA(2,3) GA(2,4)
FR BE(3,1) BE(5,3) BE(7,5) BE(4,2) BE(6,4) BE(8,6)
FR BE(3,2) BE(5,4) BE(7,6) BE(4,1) BE(6,3) BE(8,5)
FR PS(1,1) PS(2,2) PS(3,3) PS(4,4)
FR PS(5,5) PS(6,6) PS(7,7) PS(8,8)
FR PS(2,1) PS(4,3) PS(6,5) PS(8,7)
FR TE(1,1) TE(2,2) TE(3,3) TE(4,4) TE(5,5) TE(6,6)
FR TE(7,7) TE(8,8) TE(9,9) TE(10,10) TE(11,11) TE(12,12)
FR TE(13,13) TE(14,14) TE(15,15) TE(16,16) TE(17,17) TE(18,18)
FR TE(19,19) TE(20,20) TE(21,21) TE(22,22) TE(23,23) TE(24,24)
FR TE(7,1) TE(13,1) TE(19,1) TE(13,7) TE(19,7) TE(19,13)
FR TE(8,2) TE(14,2) TE(20,2) TE(14,8) TE(20,8) TE(20,14)
FR TE(9,3) TE(15,3) TE(21,3) TE(15,9) TE(21,9) TE(21,15)
FR TE(10,4) TE(16,4) TE(22,4) TE(16,10) TE(22,10) TE(22,16)
FR TE(11,5) TE(17,5) TE(23,5) TE(17,11) TE(23,11) TE(23,17)
FR TE(12,6) TE(18,6) TE(24,6) TE(18,12) TE(24,12) TE(24,18)
PD
OU ME=ML ND=3 SC

```

Seguem-se as estimativas obtidas, respectivos erros padrão e valores-*t* para os efeitos cruzados e desfasados (coeficientes β) e para os efeitos

das variáveis explicativas da satisfação (coeficientes γ):

BETA

	MATER02	LAZER02	MATER03	LAZER03	MATER04	LAZER04
	-----	-----	-----	-----	-----	-----
MATER02	--	--	--	--	--	--
LAZER02	--	--	--	--	--	--
MATER03	0.726 (0.048) 15.188	0.047 (0.042) 1.121	--	--	--	--
LAZER03	0.010 (0.037) 0.259	0.704 (0.036) 19.485	--	--	--	--
MATER04	--	--	0.805 (0.043) 18.535	-0.002 (0.038) -0.054	--	--
LAZER04	--	--	0.105 (0.036) 2.946	0.661 (0.033) 20.307	--	--
MATER05	--	--	--	--	0.798 (0.046) 17.229	0.002 (0.041) 0.047
LAZER05	--	--	--	--	0.027 (0.041) 0.655	0.688 (0.039) 17.589

GAMMA

	logRend	idade	sexo	n_horas
	-----	-----	-----	-----
MATER02	0.248 (0.041) 5.991	0.102 (0.163) 0.626	-0.098 (0.034) -2.853	--
LAZER02	--	--	0.058 (0.125) 0.460	-0.089 (0.215) -0.413

Apêndice C

Comandos e Outputs

Anotados - Outros Tópicos

C.1 Análise Multi-Grupos

C.1.1 Estrutura de Comandos de Sintaxe LISREL em Análise Multi-Grupos

Há que escrever, separadamente para cada grupo, uma estrutura de comandos. Há que declarar, para cada grupo, a dimensão da amostra e a matriz de variância-covariância amostral. O número de grupos (NG) tem que ser definido na linha DA do primeiro grupo. Importa especificar todos os comandos para o primeiro grupo. Para os restantes grupos só é necessário repetir os comandos que forem diferentes dos do primeiro grupo. Contudo, comandos posteriores não podem contrariar comandos previamente definidos. Por exemplo, se uma matriz foi especificada como *DIagonal* para o grupo 1, não pode ser especificada como *FULL* ou como *SYmmetric* nos grupos seguintes. Assim, a forma de cada matriz de parâmetros deve ser especificada na linha MO do primeiro grupo e não pode depois ser contrariada nas linhas MO dos restantes grupos, mas o modo (*FIXed* ou *FRee*) pode ser alterado.

A designação $BE(4,3)$, por exemplo, refere-se ao parâmetro β_{43} da matriz \mathbf{B} do grupo em causa. Para fazer referência a um parâmetro de um outro grupo há que usar um terceiro índice. Assim, $BE(2,4,3)$ refere-se ao parâmetro β_{43} da matriz $\mathbf{B}^{(2)}$, a matriz \mathbf{B} do grupo 2.

176 Comandos e Outputs Anotados - Outros Tópicos

Para definir restrições de igualdade entre grupos há que especificar os parâmetros como livres (sujeitos a estimação) no 1º grupo e depois escrever as restrições de igualdade nos comandos dos outros grupos. Por exemplo, para especificar β_{43} como invariante nos grupos 2 e 3 há que definir:

- no grupo 1: *FR BE*(4,3)
- no grupo 2: *EQ BE*(1,4,3) *BE*(4,3)
- no grupo 3: *EQ BE*(1,4,3) *BE*(4,3)

Na linha *MO*, de definição do modelo, dos grupos 2, 3, . . . , *G* é possível usar as seguintes opções:

- *SP* (*same pattern*) significa que determinada matriz tem a mesma estrutura de elementos *fixed* e *free* que a correspondente matriz no grupo anterior;
- *IN* (*invariant*) significa que determinada matriz é invariante para todos os grupos, i.e., tem o mesmo *pattern* de elementos *fixed* e *free* em todos os grupos e todos os elementos declarados como *free* no grupo 1 (e sujeitos por isso a estimação) assumem iguais valores em todos os grupos.

O ficheiro de output fornece soluções para cada grupo, bem como um conjunto de medidas de qualidade do ajustamento global do modelo. O valor global da estatística de χ^2 deve ser entendido como uma medida da qualidade do ajustamento de todos os modelos em todos os grupos, incluindo restrições. Os valores individuais das estatísticas χ^2 deverão ser entendidos como o contributo (em valor absoluto e em valor relativo ou percentual) de cada grupo para o valor global da estatística de χ^2 , não devendo ser interpretados isoladamente para um determinado grupo. Deve pois olhar-se ao valor das medidas de qualidade do ajustamento global do modelo multi-grupos em análise.

C.1.2 Exemplo de AFC multi-grupos

```

GRUPO: SEM CRIANCAS
DA NI=29 NO=1233 MA=KM NG=2
LA
pid idade sexo n_horas logRend
Sat2_02 Sat3_02 Sat5_02 Sat6_02 Sat7_02 Sat8_02
Sat2_03 Sat3_03 Sat5_03 Sat6_03 Sat7_03 Sat8_03
Sat2_04 Sat3_04 Sat5_04 Sat6_04 Sat7_04 Sat8_04
Sat2_05 Sat3_05 Sat5_05 Sat6_05 Sat7_05 Sat8_05
KM FI='C:\LISREL_SPE2012\global_SEMcriancas.cor' SY
AC FI=C:\LISREL_SPE2012\global_SEMcriancas.acm
SE
    Sat2_02 Sat3_02 Sat5_02 Sat6_03 Sat7_03 Sat8_03/
MO NX=6 NK=2 LX=FU,FI PH=SY,FR TD=DI,FR
LK
MATERO2 LAZER03
VA 1 LX(1,1) LX(4,2)
FR LX(2,1) LX(3,1) LX(5,2) LX(6,2)
PD
OU ME=ML ND=3 SC
GRUPO: COM CRIANCAS
DA NI=29 NO=775 MA=KM
KM FI='C:\LISREL_SPE2012\global_COMcriancas.cor' SY
AC FI=C:\LISREL_SPE2012\global_COMcriancas.acm
SE
    Sat2_02 Sat3_02 Sat5_02 Sat6_03 Sat7_03 Sat8_03/
MO NX=6 NK=2 LX=FU,FI PH=SY,FR TD=DI,FR
LK
MATERO2 LAZER03
VA 1 LX(1,1) LX(4,2)
FR LX(2,1) LX(3,1) LX(5,2) LX(6,2)
PD
OU ME=ML ND=3 SC

```

C.1.3 Exemplo de modelo global multi-grupos

Comandos que permitem estimar o modelo global da Figura 1.5 sem restrições de igualdade entre parâmetros dos modelos dos dois grupos:

```

GRUPO: SEM CRIANCAS
DA NI=29 NO=1233 MA=KM NG=2

```

178 Comandos e Outputs Anotados - Outros Tópicos

```
LA
pid idade sexo n_horas logRend
Sat2_02 Sat3_02 Sat5_02 Sat6_02 Sat7_02 Sat8_02
Sat2_03 Sat3_03 Sat5_03 Sat6_03 Sat7_03 Sat8_03
Sat2_04 Sat3_04 Sat5_04 Sat6_04 Sat7_04 Sat8_04
Sat2_05 Sat3_05 Sat5_05 Sat6_05 Sat7_05 Sat8_05
KM FI='C:\LISREL_SPE2012\global_SEMcriancas.cor' SY
AC FI=C:\LISREL_SPE2012\global_SEMcriancas.acm
SE
  6 7 8 15 16 17 5 2 3 4/
MO NX=4 NY=6 NE=2 LY=FU,FI PH=SY,FR PS=DI,FR TE=DI,FR BE=SD GA=FU,FI
LE
MATERO2 LAZER03
VA 1 LY(1,1) LY(4,2)
FR LY(2,1) LY(3,1) LY(5,2) LY(6,2)
FR GA(1,1) GA(1,2) GA(1,3) GA(2,3) GA(2,4)
PD
OU ME=ML ND=3
GRUPO: COM CRIANCAS
DA NI=29 NO=775 MA=KM
KM FI='C:\LISREL_SPE2012\global_COMcriancas.cor' SY
AC FI=C:\LISREL_SPE2012\global_COMcriancas.acm
SE
  6 7 8 15 16 17 5 2 3 4/
MO NX=4 NY=6 NE=2 LY=FU,FI PH=SY,FR PS=DI,FR TE=DI,FR BE=SD GA=FU,FI
LE
MATERO2 LAZER03
VA 1 LY(1,1) LY(4,2)
FR LY(2,1) LY(3,1) LY(5,2) LY(6,2)
FR GA(1,1) GA(1,2) GA(1,3) GA(2,3) GA(2,4)
PD
OU ME=ML ND=3
```

Para impôr restrição de igualdade das relações estruturais nos dois grupos há que declarar as matrizes B e Γ como invariantes nos dois grupos, bastando para tal substituir a linha de definição do modelo do 2º grupo pelos comandos:

```
MO NX=4 NY=6 NE=2 LY=FU,FI PH=SY,FR PS=DI,FR TE=DI,FR BE=IN GA=IN
```

e remover a linha

```
FR GA(1,1) GA(1,2) GA(1,3) GA(2,3) GA(2,4)
```

C.2 Modelo de Análise Factorial de Segunda Ordem

A estrutura de comandos LISREL que permite estimar o modelo de análise factorial de segunda ordem da Figura 5.1 é a seguinte:

```

TI Modelo AF 2 ORDEM - Psychol Morbidity
DA NI=24 NO=4562 MA=KM
LA
GHQ_B_06 GHQ_E_06 GHQ_F_06 GHQ_I_06 GHQ_A_06 GHQ_C_06 GHQ_D_06 GHQ_G_06 GHQ_H_06
GHQ_L_06 GHQ_J_06 GHQ_K_06 GHQ_B_09 GHQ_E_09 GHQ_F_09 GHQ_I_09 GHQ_A_09 GHQ_C_09
GHQ_D_09 GHQ_G_09 GHQ_H_09 GHQ_L_09 GHQ_J_09 GHQ_K_09
KM FI='C:\AF2ordem\AF2_Ordem.cor' SY
AC FI=C:\AF2ordem\AF2Ordem.acm
SE
13 14 15 16 17 18 19 20 21 22 23 24 /
MO NY=12 NX=0 NK=1 NE=3 BE=FU,FI GA=FU,FR PS=DI,FR TE=DI,FR PH=SY,FI LY=FU,FI
LE
AD SD LC
LK
PM
VA 1 PH(1,1) LY(1,1) LY(5,2) LY(11,3)
FR LY(2,1) LY(3,1) LY(4,1)
FR LY(6,2) LY(7,2) LY(8,2) LY(9,2) LY(10,2)
FR LY(12,3)
PD
OU ME=ML SC ND=3
    
```



XX Congresso da Sociedade Portuguesa de Estatística
Católica Porto | Campus Foz
26 a 29 de Setembro de 2012

telefone | 93 203 09 00

emàll | spe2012@porto.ucp.pt

website | www.porto.ucp.pt/spe2012



FUNDAÇÃO
CALOUSTE
GULBENKIAN

