

Modelagem com Equações Estruturais: Princípios Básicos e Aplicações

Leila Denise Alves Ferreira Amorim¹, Rosemeire Leovigildo Fiaccone¹,
Carlos Antônio de Souza Teles Santos², Lia Terezinha Lana Pimenta de Moraes¹,
Nelson Fernandes de Oliveira², Silvano Barbosa Oliveira¹,
Tereza Nadya Lima dos Santos¹.

¹ Departamento de Estatística, Universidade Federal da Bahia,

² Universidade Estadual de Feira de Santana, Feira de Santana, Bahia.

Salvador, 2012

Agradecimentos

Esse projeto foi financiado pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), Termo de Outorga n.0082/2006. Agradecemos ainda a equipe coordenada pelo Prof. Dr.Maurício L. Barreto, do Instituto de Saúde Coletiva da UFBA, pela disponibilidade dos dados utilizados neste relatório. Parte desses dados é proveniente do estudo “Avaliação do Impacto Epidemiológico do Programa de Saneamento Ambiental da Baía de Todos os Santos - Bahia Azul”, que teve suporte do Programa de Núcleo de Excelência (PRONEX-CNPq/MCT, Brazil), Contrato num. 66.1086/1998-4 e da Secretaria de Desenvolvimento Urbano, Bahia.

LISTA DE ILUSTRAÇÕES

Figura 1: Filosofia da Modelagem com Equações Estruturais	5
Figura 2: Elementos básicos utilizados na construção de um diagrama de caminhos	12
Figura 3: Relações téóricas em SEM	13
Figura 4: Relações causais representadas em diagramas de caminhos	14
Figura 5: Diagrama de caminhos para um modelo convencional em SEM	15
Figura 6: Exemplo de um modelo recursivo	18
Figura 7: Exemplo de um modelo não recursivo	19
Figura 8: Etapas para a construção de um modelo de equações estruturais	32
Figura 9: Diagrama de caminhos referente ao consumo alimentar em crianças menores de 12 meses	37
Figura 10: Diagrama de caminhos referente ao consumo alimentar em crianças de 12 a 24 meses	39
Figura 11: Análise fatorial confirmatória para desenvolvimento cognitivo em menores de 48 meses	41
Figure 12: Modelo de equações estruturais para desenvolvimento cognitivo em menores de 48 meses	42
Figura 13: Diagrama de caminhos para SEM mais complexo para descrever o desenvolvimento cognitivo infantil	43
Figura 14: Estimativas para SEM mais complexo para descrever o desenvolvimento cognitivo infantil	44

SUMÁRIO

1 Introdução	4
2 Modelagem com Equações Estruturais	5
2.1 Visão histórica da Modelagem com Equações Estruturais	7
2.2 Tipos de variáveis nos modelos de equações estruturais	10
2.3 Diagramas de caminhos	12
2.4 Submodelos da SEM	14
2.5 Modelos recursivos e não recursivos	15
2.6 Especificação matemática do modelo de mensuração e do modelo estrutural	16
2.7 Estimação do modelo	19
2.8 Identificabilidade do modelo estrutural	22
2.9 Avaliação dos critérios de qualidade do ajuste	24
2.10 Efeitos direto, indireto e total	26
2.11 Estimação padronizada	27
2.12 Especificação do Modelo de Equações Estruturais Generalizado	29
2.13 Etapas da implementação da SEM	32
2.14 <i>Softwares</i> estatísticos	33
3 Aplicações da SEM	36
3.1 Padrões de consumo alimentar no interior da Bahia	36
3.2 Desenvolvimento cognitivo em menores de 48 meses	40
4. Considerações finais	45
Referências bibliográficas	46

1 Introdução

A modelagem com equações estruturais (SEM, em inglês) considera vários tipos de procedimentos estatísticos para avaliar relações entre variáveis observadas, com o objetivo de permitir a realização de análises quantitativas sobre modelos teóricos hipotetizados pelo pesquisador. A SEM tem-se mostrado um método flexível e poderoso para estimação de parâmetros em uma extensa família de modelos lineares incluindo o teste *t* de Student, ANOVA, MANOVA e modelos de regressão múltipla. O aspecto mais importante da SEM, no entanto, é sua extensão para permitir a estimação de erros de medidas através do uso de fatores ou variáveis latentes múltiplas. Nesses modelos podem-se incluir variáveis que não são medidas diretamente, mas através de seus efeitos, denominados indicadores, ou de suas causas observáveis. Essas variáveis não mensuráveis são conhecidas por variáveis latentes, construtos ou fatores. Mais especificamente, vários modelos teóricos podem ser testados sob a abordagem da SEM para avaliar como conjuntos de variáveis observadas definem construtos e como esses construtos relacionam-se entre si (Schumacker e Lomax, 2004). Além disto, este método permite a avaliação de mecanismos mediadores complexos através da decomposição dos efeitos (Bollen, 1987).

Nesta abordagem uma série de equações é definida para descrever as estruturas hipotetizadas nas relações entre as diversas variáveis incluídas na modelagem. Este método consiste basicamente na definição de modelos para a variável latente e para as medidas observadas. Uma característica interessante dessa metodologia é a possibilidade de que uma mesma variável seja resposta em uma equação e apareça como variável explanatória em outra equação. É ainda possível a especificação de um efeito recíproco, ou seja, aquele no qual duas variáveis afetam uma à outra através de um *feedback loop*. A aplicação desta metodologia é baseada na teoria utilizada pelo pesquisador para explicação das interrelações entre um conjunto de variáveis, que podem ser classificadas como endógenas (dependentes) ou exógenas (independentes). Este modelo teórico pode ser ainda expresso por meio de diagramas, que resumem um conjunto de hipóteses.

Os métodos modernos referentes à SEM representam a confluência de trabalhos que foram realizados em várias disciplinas, incluindo bioestatística, econometria, psicometria e estatística social. Neste trabalho os conceitos fundamentais relacionados aos modelos de equações estruturais são apresentados, considerando-se os métodos mais comumente descritos na literatura (Bollen, 1989; Kaplan, 2000; Farias, 2000). Na seção 2 deste trabalho

são apresentados um panorama histórico de seu desenvolvimento, definição da terminologia, especificação matemática do modelo, discussão do procedimento de estimação, bem como a definição de critérios para avaliação da bondade do ajuste. A seção 3, por sua vez, apresenta exemplos de aplicação da SEM em pesquisas relacionadas à saúde da criança, com discussão e interpretação dos resultados correspondentes. Na seção 4 são apresentadas as considerações finais.

2 Modelagem com Equações Estruturais

A modelagem com equações estruturais, denominada SEM - *Structural Equation Modeling*, em inglês, abrange técnicas multivariadas de análise de dados que combinam aspectos de regressão múltipla e de análise fatorial para estimar simultaneamente uma série de relações de dependência. Para construir um modelo de equações estruturais parte-se de um modelo teórico previamente definido que permitirá determinar as múltiplas relações de dependência (ou relações causais) entre as variáveis do modelo. Um modelo teórico consiste em um conjunto sistemático de relações que fornecem explicações consistentes e abrangentes dos fenômenos. O modelo teórico que serve de apoio à construção de um modelo de equações estruturais não é restrito a uma teoria definida no âmbito acadêmico, mas pode ser alicerçado na experiência e na prática obtidas a partir da observação do comportamento real, no sentido estrito do termo. A idéia geral da SEM pode ser representada pelo esquema apresentado na Figura 1.

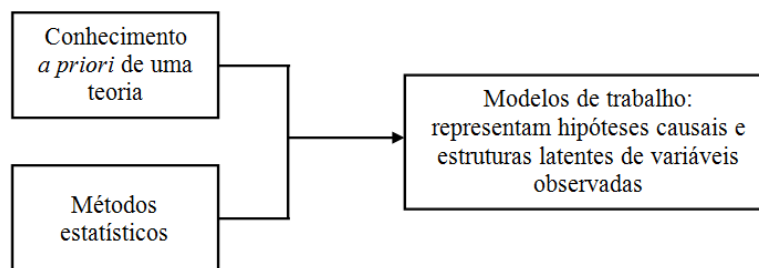


Figura 1: Filosofia da Modelagem com Equações Estruturais.

Quando existem múltiplas relações de dependência entre as variáveis, a combinação das técnicas de regressão múltipla e de análise fatorial permite definir procedimentos que visam à incorporação dos erros de medida diretamente no modelo. As variáveis utilizadas em SEM

podem ser variáveis observadas ou variáveis construídas (chamadas de construtos ou variáveis latentes) a partir das variáveis observadas. Segundo Codes (2005), essa é uma das diferenças mais importantes entre SEM e as demais técnicas de modelagem, pois os procedimentos clássicos de análise de dados modelam apenas as mensurações observáveis.

A SEM diz respeito não apenas a um modelo, mas a uma família de modelos. Além da terminologia modelos de equações estruturais, essa família é também conhecida por: análise de caminhos (*path analysis*, em inglês), análise de estrutura de covariância, análise de variáveis latentes, análise fatorial confirmatória, análise LISREL e modelo LISREL. Os nomes “análise LISREL” e “modelo LISREL” decorrem do nome de um dos primeiros *softwares* utilizados para ajustar SEM chamado LISREL (*Linear Structural Relationships*), desenvolvido por Jöreskog e Sorbon em 1993. Outros *softwares* úteis para implementação dos SEM são EQS (Bentler, 1989), AMOS (Arbuckle, 1997; SPSS, 1998), MPLUS, LVPLS, MX GRAPH, CoSan, RAMONA e SEPATH. Existem ainda pacotes no SAS, R e STATA para implementar esses modelos. As técnicas estatísticas multivariadas mais conhecidas - dentre as quais regressão múltipla, análise fatorial e análise multivariada da variância - permitem o exame de apenas uma única relação de dependência de cada vez entre as variáveis independentes e as dependentes do modelo. Nestas técnicas, um modelo é fixado e ajustado com base em dados observados e são realizados testes de significância ou a estimação dos parâmetros de interesse.

A SEM, enquanto técnica multivariada, tem como base um conjunto de relações, sendo cada uma com variáveis dependentes e independentes, apresentando algumas vantagens em relação às demais técnicas, a citar:

- i) Permite a incorporação dos erros de medição no processo de estimação do modelo;
- ii) Consiste na estimação simultânea de diversas relações de dependência interrelacionadas;
- iii) Permite que uma variável dependente em uma etapa do modelo se torne uma variável independente nas subseqüentes relações de dependência;
- iv) A capacidade de definir suposições elaboradas com base no suporte teórico e incluí-las no modelo dá à SEM flexibilidade no exame de questões analíticas dos dados.

Mais recentemente, os avanços na SEM permitiram a inclusão de novos métodos de estimação para lidar com distribuições não-normais. Devido ao trabalho de Browne (1984), Muthén (1978, 1984) e outros, agora é possível estimar os parâmetros complexos da SEM quando os dados são não-normais, incluindo a mistura de variáveis dicotômicas, categóricas,

ordinais e contínuas. Nesse trabalho, no entanto, discutiremos a SEM apenas para avaliação de relações envolvendo variáveis contínuas.

2.1 Visão Histórica da Modelagem com Equações Estruturais

Segundo Kaplan (2000), o modelo de equações estruturais originou-se da aproximação de duas áreas distintas de aplicações da Estatística: a análise fatorial, utilizada pela psicologia e pela psicomетria; e o modelo de equações simultâneas, utilizado inicialmente na econometria e posteriormente aplicado à genética.

A análise fatorial tem sua origem no trabalho de Galton (1869 apud Kaplan, 2000) e no de Pearson (Pearson e Lee, 1904 apud Kaplan, 2000), porém é no trabalho de Spearman (1904 apud Kaplan, 2000) que foi proposto o modelo de fatores comuns. Spearman propôs o uso de um fator de capacidade geral para representar as habilidades mentais dos indivíduos devido à existência de interrelações entre os testes de aptidão e capacidade específicas. De forma consistente com a definição geral da SEM dada por Kaplan (2000), a proposta de Spearman (1904, apud Kaplan, 2000) considera a correlação entre os termos de um conjunto de parâmetros estruturais. Spearman e outros pesquisadores compõem a chamada escola britânica de análise fatorial. A idéia básica da análise fatorial é que se existe um conjunto de itens correlacionados, as respostas individuais a cada um desses itens podem ser somadas para definir um escore que, por sua vez, estabeleceria ou implicaria em um construto (Schumacker e Lomax, 2004).

Na década de 1930, na Universidade de Chicago, Thurstone e colaboradores questionaram as definições de Spearman, argumentando a existência de um número maior de fatores referidos como “habilidade mental primária” (Thurstone, 1935 apud Kaplan, 2000). Muito debate surgiu em torno dessa questão, motivado pelo princípio da parcimônia (Mulaik, 1972 apud Kaplan, 2000). Pesquisadores da escola britânica, como Mulaik (1972 apud Kaplan, 2000), utilizaram a correlação existente entre os fatores para validar suas pretensões sobre um fator de habilidade geral unitário. Durante as décadas de 1950 e 1960, o acesso mais amplo à computação fez com que a análise fatorial ganhasse enorme popularidade, concomitantemente com seu desenvolvimento teórico. Os trabalhos de Jöreskog (1967 apud Kaplan, 2000), Jöreskog e Lawley (1968 apud Kaplan, 2000), Lawley (1958, 1967 apud Kaplan, 2000) e Lawley e Maxwell (1971 apud Kaplan, 2000) levaram ao uso da estimação de máxima verossimilhança na análise fatorial, o que permitiu aos pesquisadores testar a hipótese da existência de um determinado número de fatores e descrever suas

intercorrelações. A minimização da função ajustada de máxima verossimilhança levou ao teste de hipótese qui-quadrado para verificar se o modelo proposto se adequa aos dados. Uma aproximação por mínimos quadrados generalizados foi posteriormente desenvolvida por Jöreskog e Goldberger (1972 apud Kaplan, 2000). Outras pesquisas realizadas por Anderson e Rubin (1956 apud Kaplan, 2000) e mais tarde por Jöreskog (1969 apud Kaplan, 2000) levaram à construção da metodologia denominada “análise fatorial confirmatória”, que permite testar as hipóteses sobre o número de fatores e o padrão de carga. De uma perspectiva histórica, essas metodologias permitiram uma aproximação estatística rigorosa às idéias de estrutura simples de Thurstone. Foi possível, a partir desses desenvolvimentos, especificar um modelo em que certos fatores representavam as correlações de apenas um subconjunto específico de variáveis observadas. As análises fatoriais exploratória e confirmatória permanecem até hoje como metodologias muito populares em pesquisas da área de ciências sociais quantitativas.

Em relação à SEM, a análise fatorial constitui parte de uma metodologia mais ampla. A SEM, em sua forma mais simples, reúne a análise fatorial e a análise de caminhos em um modelo onde estão contidas relações de dependência.

A análise de caminhos teve origem histórica econométrica e biométrica, sendo apresentada pelo biólogo Sewell Wright (1918, 1921, 1934, 1960 apud Kaplan, 2000). Wright mostrou como as correlações entre variáveis constituem os parâmetros do modelo representado pelo diagrama de caminhos, que consiste em um dispositivo pictórico que foi creditado a Wright. Este pesquisador também mostrou como o modelo de equações estruturais poderia estimar os efeitos diretos, os indiretos e os totais. A análise de caminhos usa coeficientes de correlação e análise de regressão para modelar relações mais complexas entre variáveis observadas (Schumacker e Lomax, 2004). Inicialmente Wright (1918 apud Kaplan, 2000) aplicou a análise de caminhos ao problema da estimação do número de componentes das medições dos ossos. Essa primeira aplicação da análise de caminhos foi estatisticamente equivalente à análise fatorial e existem indícios de que tenha sido desenvolvida sem conhecimento do trabalho de Spearman (Bollen, 1989). A análise de caminhos também foi utilizada por Wright na estimação das equações de oferta e demanda, que ainda abordou o problema de identificabilidade do modelo. Estas questões constituíram o núcleo inicial da contribuição econométrica na SEM (Goldberger, 1972 apud Kaplan, 2000). As equações simultâneas foram também utilizadas em 1918 para estudar as influências genéticas através de gerações. No campo da econometria, o uso de modelos matemáticos para descrição de fenômenos

econômicos começou com Petty, em 1676 (Spanos, 1986 apud Kaplan, 2000). Porém, com respeito à SEM, a contribuição inicial encontra-se no trabalho de Haavelmo (1943 apud Kaplan, 2000). Haavelmo utilizou o sistema de equações simultâneas para modelar a interdependência entre variáveis econômicas utilizando a expressão $y = \mathbf{B}y + \mathbf{\Gamma}x + \zeta$, onde y é o vetor de variáveis endógenas que o modelo pretende explicar; x o vetor de variáveis exógenas, que são propostas para explicar y mas cujo comportamento não é explicado; ζ o vetor que representa a perturbação; e \mathbf{B} e $\mathbf{\Gamma}$ representam as matrizes dos coeficientes do modelo. Este modelo foi a maior inovação na modelagem econométrica. Seu desenvolvimento e refinamento foi objeto de pesquisa de estatísticos e econometricistas na Universidade de Chicago, em 1945, e depois em Yale (Berndt, 1991 apud Kaplan, 2000), com ênfase no estudo do recém-desenvolvido modelo de equações simultâneas com o método de estimação de máxima verossimilhança, associado às metodologias de testes de hipóteses (Hood e Koopmans, 1953; Koopmans, 1950 apud Kaplan, 2000). Nos 25 anos seguintes, as pesquisas econométricas caminharam no sentido do refinamento dos procedimentos de uso das equações simultâneas (Fisher, 1966 apud Kaplan, 2000). O modelo de equações simultâneas apresentado, embora utilizado por um tempo longo, foi objeto de diversas críticas. Seus críticos afirmavam que um grave problema com grandes modelos macroeconômicos de equações simultâneas era que não poderiam competir com métodos de distribuições livres nos modelos de séries temporais de Box-Jenkins para previsões precisas (Cooper, 1972 apud Kaplan, 2000).

SEM é definida pela combinação das duas técnicas brevemente historiadas. A combinação dessas metodologias na forma corrente teve início no trabalho de Jöreskog (1973 apud Kaplan, 2000), Keesling (1972 apud Kaplan, 2000) e Wiley (1973 apud Kaplan, 2000). O modelo geral de equações estruturais, delineado por Jöreskog (1973, apud Kaplan, 2000), consiste em dois submodelos: (a) o modelo de medição, onde as variáveis observadas são relacionadas às variáveis latentes através de análise fatorial confirmatória; e (b) o modelo estrutural, que se refere à parte do modelo que relaciona as variáveis latentes umas as outras por meio de sistemas de equações simultâneas. Esses modelos são discutidos mais detalhadamente nas seções 2.4 e 2.6.

O avanço nos modelos de equações estruturais levou à ampliação de sua aplicação para diferentes áreas do conhecimento. Atualmente tem-se observado o uso da SEM em:

- Genética comportamental - é bastante utilizada na observação de variáveis fenotípicas com o objetivo de obter estimativas seguras das contribuições genéticas não-

observadas e de fatores ambientais para as variações nas variáveis resposta observadas.

- Ciências sociais - a justificativa de seu uso baseia-se no argumento de que a SEM permite a formulação e teste de modelos de estruturas causais correlacionadas. Entretanto, existem diversas críticas a esta justificativa argumentando que a existência de correlação não implica em relação causal entre variáveis e, portanto, não seria possível realizar uma inferência causal. Vale salientar que correlação é uma condição necessária mas não suficiente para a existência de uma relação causal. Argumenta-se ainda que as inferências causais só poderiam ser obtidas a partir de experimentos controlados. Apesar das posições críticas, a SEM tem sido utilizada em ciências sociais em problemas relacionados à teoria do comportamento, em análises de padrões de estabilidade e em mudanças em dados longitudinais.
- Na área educacional - diversos pesquisadores têm utilizado SEM para a análise de questões importantes no campo da educação, como a reforma escolar, por exemplo (Kaplan, 2000).

A crescente popularidade da SEM está relacionada a diversos fatores, incluindo o fato dos pesquisadores estarem tornando-se mais conscientes da necessidade de uso de múltiplas variáveis observadas para entender melhor sua área de investigação científica. Os métodos estatísticos tradicionais podem ser utilizados apenas para um número limitado de variáveis, que podem não ser capazes de dar conta das sofisticadas teorias que vêm sendo desenvolvidas. Outra razão refere-se ao crescente reconhecimento da validade e confiabilidade de escores observados a partir de instrumentos de mensuração, sendo que o SEM permite a inclusão explícita de erros de mensuração. Além disso, a disponibilidade de vários *softwares* de fácil uso pelos pesquisadores, permitindo a implementação desde os modelos mais simples até SEM mais sofisticada, tem sido um motivo adicional para o amplo uso dessas metodologias nos dias atuais.

2.2 Tipos de variáveis nos modelos de equações estruturais

As variáveis na SEM podem ser classificadas em relação a diversos aspectos do modelo. Quanto ao aspecto de serem mensuráveis ou não, elas podem ser classificadas como variáveis latentes, variáveis de medição e variáveis indicadoras. Variáveis que não são mensuráveis diretamente são chamadas de variáveis latentes ou construtos e dizem respeito a conceitos teóricos que não podem ser observados diretamente. Se construtos forem observados

diretamente, não será possível medi-los sem a ocorrência de erros. Na abordagem com SEM os construtos permitem a formação das relações causais a serem estimadas pelos modelos e são medidos, aproximadamente, por um conjunto de variáveis observadas. Segundo Hair e colaboradores (2005), construtos ou variáveis latentes podem também estar relacionados com variáveis de medição (ou variáveis observadas) em uma relação de dependência. Em SEM, convencionalmente, assume-se que as variáveis de medição são dependentes dos construtos. Assim, em um diagrama de caminhos a seta apresenta o seguinte sentido: parte do construto em direção às variáveis de medição que foram utilizadas para a construção da variável latente. Uma variável latente ou construto é resultado da combinação de diversas variáveis de medição. Assim, aquelas variáveis observadas que são utilizadas para a construção de uma variável latente são chamadas de variáveis indicadoras. O pesquisador deve justificar a base teórica das variáveis indicadoras porque a SEM examina apenas as características empíricas das variáveis. Para cada construto que aparece no modelo é necessário determinar quais são as variáveis indicadoras que estão relacionadas com ele.

O uso de variáveis latentes em SEM permite a melhoria da estimação estatística por dois motivos: representa os conceitos teóricos de forma mais adequada; e incorpora o erro de mensuração. Os erros de mensuração resultam de respostas imprecisas e do uso de conceitos teóricos - construtos. Entretanto, se a magnitude dos problemas que geram os erros é conhecida, é possível incorporar a confiabilidade na estimação do modelo. Na estimação das relações entre variáveis dependentes e independentes, o modelo de medição permite avaliar a contribuição de cada item da escala de mensuração, bem como incorporar ao modelo a escala que melhor mede o conceito.

Quanto à influência que uma variável exerce sobre outras, as variáveis são classificadas como exógenas e endógenas. São chamadas de variáveis exógenas aquelas variáveis que não são influenciadas ou não sofrem efeito de outras variáveis do modelo, sendo também chamadas de independentes ou preditoras. Como no modelo de regressão tradicional, assume-se que essas variáveis são mensuradas sem erro. Essas variáveis podem ser quantitativas ou qualitativas. As variáveis endógenas ou variáveis dependentes são aquelas que recebem influência de outras variáveis presentes no modelo. Os erros estruturais (ou *disturbances*) representam as causas omitidas agregadas das variáveis endógenas, juntamente com o erro de mensuração. Assim, haverá um erro associado a cada variável endógena do modelo. Para os modelos discutidos neste trabalho estaremos considerando que essas variáveis são contínuas.

2.3 Diagramas de caminhos

Como em SEM, em geral, os modelos são bastante complexos, muitos pesquisadores acham mais conveniente retratá-los na forma de um diagrama. O diagrama permite a rápida visualização das relações, consistindo em um gráfico pictórico que representa as relações de interdependência consideradas no modelo. Esta apresentação visual é chamada de diagrama de caminhos (*path diagram*). O modelo teórico, que é a base da SEM, deve primeiramente ser explicitado através de um diagrama de caminhos.

O diagrama de caminhos é representado por um conjunto de figuras geométricas e setas que servem para evidenciar o tipo de variável (observada ou latente) e o tipo de relação entre elas. A Figura 2 ilustra as convenções usadas para representação das relações entre um construto e uma ou mais variáveis de medição e a relação entre construtos.






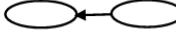
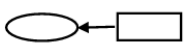


Descrição	Elemento básico
Variável latente ou construto	
Variável observada ou indicadora	
Relação causal direta ou direcional entre duas variáveis	
Relação não recursiva ou não direcional entre duas variáveis	
Correlação entre duas variáveis	
Relação entre duas variáveis latentes	
Relação entre uma variável observada e uma variável latente	
Erro de mensuração na variável observada	
Erro na predição da variável latente	

Figura 2: Elementos básicos utilizados na construção de um diagrama de caminhos.

Os princípios básicos na construção de um diagrama de caminhos no modelo de medição, são os seguintes:

- Representa-se por X as variáveis indicadoras dos construtos exógenos e por Y as variáveis indicadoras de construtos endógenos;
- Construtos são representados normalmente por círculos ou figuras ovais;
- As variáveis de medição são representadas por retângulos ou quadrados;
- As variáveis indicadoras X e/ou Y são associadas com seus respectivos construtos por uma seta que parte do construto para a variável de medição (ou variáveis indicadoras).

Quando duas variáveis não estão ligadas através de uma seta não implica necessariamente que uma não afete a outra. Essa relação pode ocorrer indiretamente, podendo ser identificada através de caminhos mais complexos.

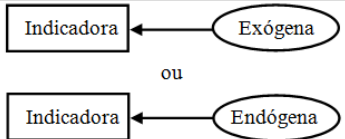
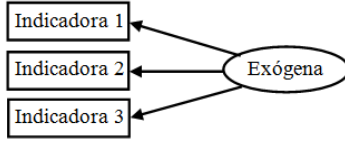

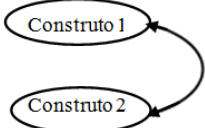
Tipo de relação	Representação no diagrama de caminhos
Entre um construto (variável exógena ou endógena) e uma variável indicadora	
Entre um construto (variável exógena ou endógena) e diversas variáveis indicadoras	
Estrutural: dependência entre dois construtos	
Correlacional: construtos correlacionados	

Figura 3: Relações teóricas em SEM.

A Figura 3 ilustra diversas formas de representação das relações entre um construto e uma ou mais variáveis de medição e a relação entre construtos. Na Figura 4 são ilustrados três tipos de relação que podem ser descritos através de um diagrama de caminhos e suas correspondentes equações:

- Figura (a): É um modelo simples que mostra a relação entre duas variáveis exógenas (independentes) X_1 e X_2 e a variável endógena ou dependente Y_1 . A curva entre X_1 e X_2 identifica a correlação existente entre essas variáveis. Esta relação pode ser mostrada com uma única equação, de forma similar a uma equação de regressão.
- Figura (b): Uma segunda variável dependente Y_2 foi adicionada ao modelo descrito em (a). Além da equação definida para (a), será necessária uma segunda equação para representar a relação entre X_2 e Y_1 (variáveis exógenas) com a variável endógena Y_2 . Observa-se que a variável X_2 é compartilhada pelas duas equações e que Y_1 é variável endógena na primeira equação e variável exógena na segunda.
- Figura (c): Nesta figura são apresentadas três variáveis dependentes ou endógenas, Y_1 ,

Y_2 e Y_3 , que se relacionam entre si e com as variáveis independentes X_1 , X_2 e X_3 . Observa-se ainda uma relação de reciprocidade entre Y_2 e Y_3 , isto é, em uma equação Y_2 aparece como variável preditora (exógena) de Y_3 e em outra equação como variável dependente de Y_3 . Para representar todas as relações apresentadas nesta figura são necessárias diversas equações, uma para cada variável dependente.

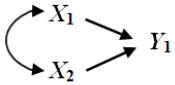
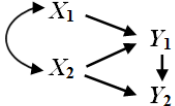
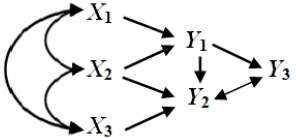
Tipo	Diagrama de caminhos	Variáveis independentes	Variáveis dependentes
(a)		X_1 X_2	Y_1
(b)		i) X_1 e X_2 ii) X_2 e Y_1	i) Y_1 ii) Y_2
(c)		i) X_1 e X_2 ii) X_2 , X_3 , Y_2 e Y_3 iii) Y_1 e Y_2	i) Y_1 ii) Y_2 iii) Y_3

Figura 4: Relações causais representadas em diagramas de caminhos.

As relações de dependência e as relações correlacionais que foram explicitadas na Figura 4 ilustram a importância dos modelos de equações estruturais como método para estimar simultaneamente um conjunto grande de equações com as características citadas (Hair et al, 2005).

Na construção de um diagrama de caminhos duas suposições são assumidas. A primeira é que todas as relações causais são mostradas no diagrama, e a escolha dessas relações está de acordo com a teoria que dá suporte à construção do modelo a ser ajustado. O objetivo da SEM é modelar as relações com o menor número de caminhos causais ou correlações entre as variáveis que possam ser justificadas teoricamente. A segunda suposição diz respeito à natureza das relações entre as variáveis, latentes ou observadas, que são assumidas como lineares ou podem ser linearizadas por transformação.

2.4 Submodelos da SEM

Um modelo completo na terminologia SEM é composto por dois submodelos: o modelo de medição e o modelo estrutural. O modelo de medição mostra como as variáveis de medição

representam construtos ou variáveis latentes. Esta parte do modelo está relacionada ao uso de análise fatorial confirmatória, que determina a forma como as variáveis latentes são construídas a partir das variáveis observadas. O modelo de medição oferece ainda uma descrição das propriedades de mensuração (validade e confiabilidade) dessas variáveis.

O modelo estrutural, por sua vez, mostra como os construtos estão associados uns com os outros. Seu desenvolvimento fundamenta-se no cálculo de sistemas de equações simultâneas. Nessa etapa da SEM estão os procedimentos de especificação e estimação das associações das variáveis latentes entre si ou com outras variáveis observáveis, descrevendo seus efeitos e suas magnitudes. Incluem também as informações sobre a variância explicada e a não explicada de cada termo endógeno presente no modelo.

Ambos os submodelos podem ser bastante complexos. O esquema de um modelo convencional de SEM na forma de diagrama de caminhos é apresentado na Figura 5. O modelo de medição pode ser especificado em termos das variáveis exógenas ou em termos das variáveis endógenas. No diagrama de caminhos existem uma ou mais setas que conduzem até uma variável endógena.

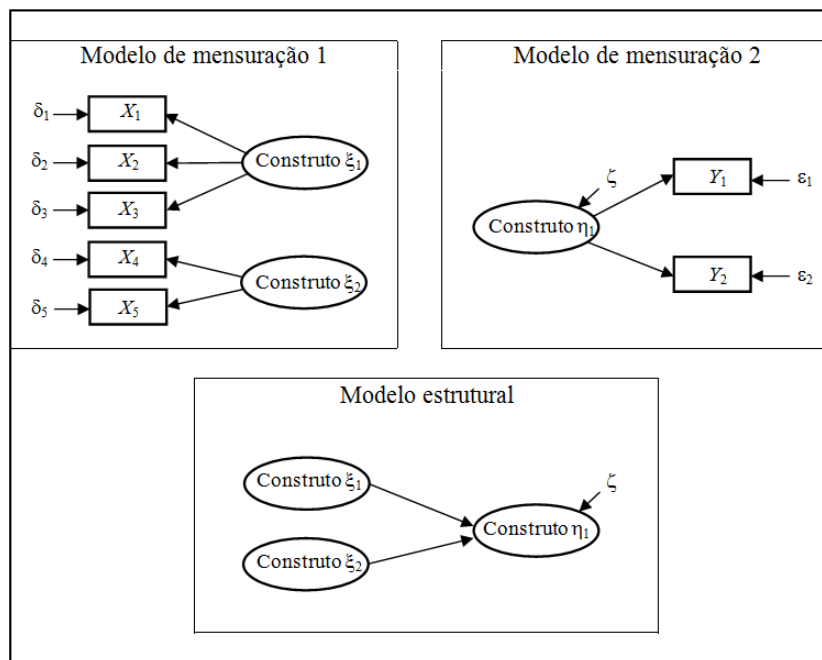


Figura 5: Diagrama de caminhos para um modelo convencional em SEM.

2.5 Modelos recursivos e não recursivos

Os modelos teóricos especificados na SEM e apresentados em diagramas de caminhos podem ser classificados em dois tipos a depender do sentido das relações causais representadas por

setas no diagrama de caminhos: modelo recursivo e modelo não recursivo.

Nos modelos recursivos as covariâncias são nulas entre os termos de perturbação do modelo e as relações causais entre as variáveis endógenas são propostas em uma única direção, isto é, cada variável tem efeito direto nas outras (Figura 6). Algumas vezes os requerimentos para a causalidade unidirecional e erros independentes são apropriados para subconjuntos ou blocos de variáveis endógenas e seus erros associados, mas não entre todas as variáveis incluídas no modelo. Nesse caso, o modelo é dito bloco recursivo. Quando a relação causal apresenta mais de uma direção o modelo é denominado não recursivo. Na área da Economia os modelos não recursivos são denominados modelos de equações simultâneas. Neste modelo existe um movimento de ida e volta (*feedback loop*) entre duas variáveis endógenas especificadas e as covariâncias são não nulas entre os termos de perturbação das variáveis endógenas do modelo, representadas pela seta curva entre ε_1 e ε_2 na Figura 7. Os modelos não recursivos são úteis na análise da maioria dos sistemas dinâmicos presentes nas ciências sociais. Esses modelos são ilustrados e discutidos em termos do processo de estimação na seção 2.6.

2.6 Especificação matemática do modelo de mensuração e do modelo estrutural

Na modelagem por equações estruturais, deseja-se descrever médias, variâncias e covariâncias através de um conjunto de variáveis em termos de um reduzido número de parâmetros estruturais. Os conceitos sobre esses parâmetros estruturais podem ser introduzidos pelo seu significado a respeito das interrelações entre um conjunto de variáveis observadas, que usualmente denomina-se análise de caminhos (*path analysis*).

O sistema de equações estruturais para a *path analysis* pode ser representado compactamente por:

$$y = \alpha + \mathbf{B}y + \mathbf{\Gamma}x + \zeta \quad (1)$$

onde y é um vetor $p \times 1$ de variáveis endógenas observadas, x é um vetor $q \times 1$ de variáveis exógenas observadas; α é um vetor $p \times 1$ de interceptos estruturais; \mathbf{B} é uma matriz $p \times p$ que relaciona as variáveis endógenas entre si, $\mathbf{\Gamma}$ é uma matriz $p \times q$ de coeficientes que relaciona variáveis endógenas a variáveis exógenas; e ζ é um vetor $p \times 1$ de termos de ruídos (*disturbances*), sendo $cov(\zeta) = \Psi$ a matriz de covariância entre os termos de ruído. Assume-se que os erros têm esperança zero e são independentes em relação (ou pelo menos não correlacionados com) às variáveis exógenas. Os erros para diferentes observações são considerados independentes entre si, com variância constante entre as observações. Considere

$cov(x) = \Phi$ a matriz $q \times q$ de covariância para as variáveis exógenas (Kaplan, 2000).

Pode-se distinguir três tipos de parâmetros a serem estimados para representar as interrelações entre as variáveis. O primeiro tipo de parâmetro é denominado “parâmetro livre” e refere-se a aqueles parâmetros que serão estimados pelo modelo. O segundo conjunto de parâmetros refere-se a aqueles que serão considerados constantes (ou com valores fixados *a priori*) durante o procedimento de estimação. Geralmente os parâmetros fixos são aqueles que representam ausência de relação, logo são considerados fixos e iguais a zero. No entanto, outro valor pode ser considerado como fixo se houver teoria para suportar tal escolha. O último grupo de parâmetros é denominado parâmetros restritos (*constraint parameters*), referindo-se a aqueles que são considerados iguais a outros parâmetros do modelo. Por exemplo, pode-se requerer que os efeitos de duas relações representadas no diagrama de caminhos sejam iguais.

A equação (1) pode ser reescrita isolando-se as variáveis exógenas e endógenas em lados diferentes da equação, de modo que:

$$y - \mathbf{B}y = \alpha + \Gamma x + \zeta \quad (2)$$

$$y(\mathbf{I} - \mathbf{B}) = \alpha + \Gamma x + \zeta \quad (3)$$

Assumindo-se que $(\mathbf{I} - \mathbf{B})$ não é singular, então sua inversa pode ser definida e a equação (3) pode ser reescrita como:

$$y = (\mathbf{I} - \mathbf{B})^{-1}\alpha + (\mathbf{I} - \mathbf{B})^{-1}\Gamma x + (\mathbf{I} - \mathbf{B})^{-1}\zeta \quad (4)$$

$$= \Pi_0 + \Pi_1 x + \zeta^*, \quad (5)$$

sendo $\Pi_0 = (\mathbf{I} - \mathbf{B})^{-1}\alpha$ o vetor de interceptos na forma reduzida; $\Pi_1 = (\mathbf{I} - \mathbf{B})^{-1}\Gamma$ o vetor de coeficientes angulares (*slopes*) na forma reduzida e $\zeta^* = (\mathbf{I} - \mathbf{B})^{-1}\zeta$ o vetor de ruídos na forma reduzida, tal que $cov(\zeta^*) = \Psi^*$. Note que a expressão anterior representa diretamente um modelo de regressão multivariado de y em x .

O sistema de equações em (5) pode ser representado em termos de modelagem de médias, variâncias e covariâncias. Os parâmetros estruturais a serem modelados podem ser representados pelo vetor $\Omega = (\mathbf{B}, \Gamma, \Psi, \Phi)$, onde Φ contém as variâncias e covariâncias das variáveis exógenas; Ψ as variâncias e covariâncias dos termos de ruído; \mathbf{B} e Γ os coeficientes de regressão. Além disso, considera-se $E(x) = \mu_x$ o vetor de médias para x ; $cov(x) = E(x'x) = \Phi$ e $E(\zeta) = 0$. Assim,

$$\begin{aligned}
E(y) &= E[(\mathbf{I} - \mathbf{B})^{-1}\alpha + (\mathbf{I} - \mathbf{B})^{-1}\Gamma x + (\mathbf{I} - \mathbf{B})^{-1}\zeta] \\
&= (\mathbf{I} - \mathbf{B})^{-1}\alpha + (\mathbf{I} - \mathbf{B})^{-1}\Gamma E(x) + (\mathbf{I} - \mathbf{B})^{-1}E(\zeta) \\
&= (\mathbf{I} - \mathbf{B})^{-1}\alpha + (\mathbf{I} - \mathbf{B})^{-1}\Gamma\mu_x
\end{aligned} \tag{6}$$

e

$$\begin{aligned}
E(y, x) &= \Sigma_{yx} \\
&= \begin{pmatrix} E(y'y) & E(yx') \\ E(x'y) & E(x'x) \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\Gamma\Phi\Gamma' + \Psi)(\mathbf{I} - \mathbf{B})^{-1}' & (\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi \\ \Phi\Gamma'[(\mathbf{I} - \mathbf{B})^{-1}]' & \Phi \end{pmatrix}
\end{aligned} \tag{7}$$

onde (6) e (7) mostram que a modelagem por equações estruturais representa o vetor de média e a matriz de covariância. Essa estrutura é definida em termos dos parâmetros do modelo.

Usualmente para simplificar as equações estruturais, considera-se o uso de variáveis centradas em sua média (Fox, 2008; Mueller, 1999). Com isso, os termos contendo os interceptos nas equações anteriores tornam-se zero, reduzindo-se a forma matricial desse modelo para:

$$y = \mathbf{B}y + \Gamma x + \zeta \tag{8}$$

Para facilitar a notação, considere de agora em diante o uso de variáveis centradas.

A especificação dos elementos da matriz \mathbf{B} permitirá a distinção entre duas classificações analíticas dos modelos de caminhos (*path analysis*): (a) recursivos e (b) não recursivos ou simultâneos, conforme discutido na seção 2.5. Para ilustrar através de diagramas de caminhos a diferença entre esses dois tipos de modelos, considere as Figuras 6 e 7.

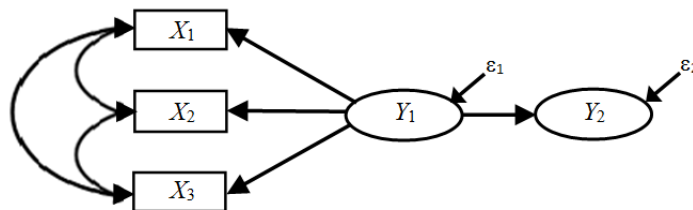


Figura 6: Exemplo de um modelo recursivo.

Uma característica dos sistemas recursivos é que os elementos de \mathbf{B} que representam as relações entre as variáveis endógenas do modelo encontram-se na porção triangular inferior

de \mathbf{B} . Além disso, esse modelo não contém covariâncias entre os termos de ruído. Assim, Ψ é uma matriz diagonal cujos elementos representam as variâncias dos ruídos. Tipicamente os coeficientes estruturais das matrizes \mathbf{B} e Γ contém alguns zeros, e os elementos da diagonal de \mathbf{B} são iguais a 1. Pode-se escrever esse modelo para a Figura 6 como:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \beta_{21} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \quad (9)$$

Nos modelos não recursivos, por sua vez, existe um *loop* entre variáveis endógenas do modelo, o que implica que \mathbf{B} não é triangular inferior. Além disso, muitas vezes considera-se a existência de um termo de covariância entre os ruídos das variáveis endógenas envolvidas no *loop*. Nesse caso, Ψ é uma matriz simétrica, com elementos não nulos fora da diagonal principal. Este tipo de modelo implica em uma especificação dinâmica do modelo estrutural, que pode causar problemas de instabilidade do procedimento de estimação. O modelo para a Figura 7 pode ser definido como:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 \\ 0 & 0 & \gamma_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \quad (10)$$

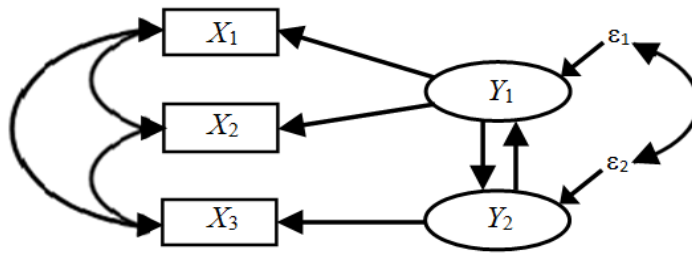


Figura 7: Exemplo de um modelo não recursivo.

2.7 Estimação do modelo

O procedimento de estimação em SEM tem como objetivo obter estimativas para o vetor de parâmetros Ω , denominado $\hat{\Omega}$, que minimizem a função de discrepância $F(S, \hat{\Sigma})$. Os parâmetros a serem estimados pelo modelo são:

- as variâncias e covariâncias das variáveis exógenas em Φ ;
- as variâncias e covariâncias dos termos de ruído em Ψ ;
- os coeficientes de regressão em \mathbf{B} e Γ .

Assim, deseja-se estimar os parâmetros contidos no vetor $\boldsymbol{\Omega} = (\mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$ através do estimador $\hat{\boldsymbol{\Omega}}$ que minimize a função de discrepância $F(S, \hat{\boldsymbol{\Sigma}})$, onde $\hat{\boldsymbol{\Sigma}} = \Sigma(\hat{\boldsymbol{\Omega}})$ é a matriz de covariância baseada nas estimativas obtidas pelo modelo (matriz de covariância ajustada) (Kaplan, 2000).

A função $F(S, \hat{\boldsymbol{\Sigma}})$ é um escalar que mensura a distância entre a matriz de covariância amostral (S) e a matriz de covariância ajustada ($\Sigma(\hat{\boldsymbol{\Omega}})$). Essa função pode ser caracterizada pelas seguintes propriedades:

- $F(S, \hat{\boldsymbol{\Sigma}}) \geq 0$,
- $F(S, \hat{\boldsymbol{\Sigma}}) = 0$, se e apenas se $\hat{\boldsymbol{\Sigma}} = S$,
- $F(S, \hat{\boldsymbol{\Sigma}})$ é uma função contínua em $\hat{\boldsymbol{\Sigma}}$ e S .

Os dois métodos de estimação mais comumente utilizados são o método de máxima verossimilhança (ML, em inglês) e os mínimos quadrados generalizados (GLS, em inglês).

Para discussão do método de máxima verossimilhança no contexto da modelagem com equações estruturais considere um vetor de respostas z (que contém y e x), baseado em uma amostra com $n = N - 1$ observações, com matriz de covariância amostral não-viciada S , onde N representa o número total de observações. Para os desenvolvimentos desse método, assume-se que as observações provenientes da população seguem uma distribuição normal multivariada, cuja função de densidade é dada por:

$$f(z) = (2\pi)^{-(p+q)/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2} z' \boldsymbol{\Sigma}^{-1} z\right]$$

Assumindo-se que N observações são independentes entre si, a função de densidade conjunta pode ser escrita como o produto de suas densidades individuais:

$$f(z_1, z_2, \dots, z_N) = f(z_1) f(z_2) \dots f(z_N)$$

A verossimilhança da amostra pode ser então definida por:

$$L(\boldsymbol{\Omega}) = (2\pi)^{-N(p+q)/2} |\boldsymbol{\Sigma}(\boldsymbol{\Omega})|^{-N/2} \exp\left[-\frac{1}{2} \sum_{i=1}^N z_i' \boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1} z_i\right]$$

Para facilitar o processo de derivação, considere o logaritmo da verossimilhança dado por:

$$\begin{aligned} \log L(\boldsymbol{\Omega}) &= \frac{-N(p+q)}{2} \log(2\pi) - \frac{N}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^N \text{tr}[z_i' \boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1} z_i] \\ &= \frac{-N(p+q)}{2} \log(2\pi) - \frac{N}{2} \log |\boldsymbol{\Sigma}| - \frac{N}{2} \sum_{i=1}^N \text{tr}[N^{-1} z_i z_i' \boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}] \end{aligned}$$

$$= -\frac{N(p+q)}{2} \log(2\pi) - \frac{N}{2} \log |\boldsymbol{\Sigma}| - \frac{N}{2} \sum_{i=1}^N \text{tr}[\mathbf{T}\boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}]$$

onde \mathbf{T} é a matriz de covariância amostral baseada em N observações no lugar de $n = N - 1$. No processo da maximização de $\log L(\boldsymbol{\Omega})$ alguns termos que não possuem os parâmetros de interesse podem ser ignorados. Além disso, salienta-se que a diferença entre \mathbf{T} (baseada em N) e \mathbf{S} (baseada em $n = N - 1$) é negligenciável em amostras grandes. Assim, a expressão anterior pode ser reescrita como:

$$\log L(\boldsymbol{\Omega}) = -\frac{N}{2} \left\{ \log |\boldsymbol{\Sigma}| + \text{tr}[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}] \right\}$$

Um problema com esta expressão é que não possui as propriedades referentes à função de discrepância conforme descrito anteriormente. Para que a expressão anterior possa ser considerada uma função de discrepância apropriada, será necessário adicionar termos que não dependam dos parâmetros do modelo. Além disso, o termo $-\frac{N}{2}$ será removido, o que significa que a função será minimizada invés de maximizada. Assim, tem-se:

$$F_{ML} = \log |\boldsymbol{\Sigma}| + \text{tr}[\mathbf{S}\boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}] - \log |S| - s$$

onde s é o número total de variáveis em y e x , tal que $(p+q) = s$. Nesse caso, se o modelo ajustar perfeitamente, o primeiro e terceiro termos somam zero, ao mesmo tempo em que o segundo e quarto termos somam zero, o que resulta em uma função de ajuste apropriada.

Além das estimativas dos parâmetros do modelo, pode-se ainda estimar a matriz de covariância dos estimadores dos parâmetros. Assim, a matriz de covariância assintótica para $\boldsymbol{\Omega}_{r \times 1}$ pode ser escrita por:

$$\text{cov}(\hat{\boldsymbol{\Omega}}) = \left\{ -E \left[\frac{\partial^2 \log L(\boldsymbol{\Omega})}{\partial \boldsymbol{\Omega} \partial \boldsymbol{\Omega}'} \right] \right\}$$

onde $I(\hat{\boldsymbol{\Omega}}) = \frac{\partial^2 \log L(\boldsymbol{\Omega})}{\partial \boldsymbol{\Omega} \partial \boldsymbol{\Omega}'}$ é a matriz de informação de Fisher. Os erros padrão podem ser

obtidos pela raiz quadrada dos elementos da diagonal principal da matriz de covariância assintótica das estimativas.

O estimador de mínimos quadrados generalizados (GLS) também assume que os dados são provenientes de uma distribuição normal multivariada. A forma geral da função de discrepância ajustada pelo GLS é dada por:

$$F_{GLS} = [S - \Sigma(\Omega)]' W^{-1} [S - \Sigma(\Omega)]$$

onde W^{-1} uma matriz de pesos que pondera os desvios $S - \Sigma(\Omega)$ em termos de suas variâncias e covariâncias com outros elementos.

O estimador GLS é parte do grupo de estimadores de mínimos quadrados ponderados (WLS, em inglês). As duas escolhas mais comuns para W são a matriz identidade, $W^{-1} = \mathbf{I}$, e a matriz de covariância amostral, $W^{-1} = S^{-1}$. Quando $W^{-1} = S^{-1}$, então a função de discrepância ajustada pelo GLS pode ser definida por:

$$\begin{aligned} F_{GLS} &= \frac{1}{2} tr[S^{-1}(S - \Sigma)]^2 \\ &= \frac{1}{2} tr(\mathbf{I} - S^{-1}\Sigma)^2 \end{aligned}$$

Sob a suposição de normalidade multivariada, o estimador GLS é assintoticamente normal e eficiente.

Um número mínimo de observações deve ser requerido para que SEM possa ser utilizado. De acordo com Mueller (1996), recomenda-se que a razão entre o tamanho da amostra e o número de parâmetros a ser estimado pelo modelo seja de 10:1 ou mesmo de 20:1, se testes de significância estatística são de interesse.

2.8 Identificabilidade do modelo estrutural

O problema de identificabilidade do modelo refere-se à existência de solução única na estimação dos parâmetros. Logo, se os parâmetros do modelo não são identificáveis, sua estimação não pode ser realizada. Embora o problema de identificabilidade do modelo exista em quase todos os modelos estatísticos, o seu papel é talvez mais claro nos modelos de equações estruturais (Kaplan, 2000).

Um modelo é dito identificado quando teoricamente é possível obter uma única estimativa para cada parâmetro. Dentro da abordagem com equações estruturais, a definição de identificabilidade está relacionada à modelagem da estrutura de covariância. Assim, a identificabilidade do modelo geralmente ocorre quando o número de elementos da matriz de covariância entre as variáveis observadas for maior ou igual ao número de parâmetros a serem estimados (Kline, 2005). Os modelos identificados que possuem o mesmo número de parâmetros e observações são chamados de exatamente identificados, enquanto que os modelos que possuem menos parâmetros que observações são chamados de super-

identificados.

De modo geral, algumas restrições precisam ser impostas aos modelos para que não haja problemas de identificabilidade. Essas restrições incluem:

- Normalização: requer que os elementos da diagonal de \mathbf{B} sejam zero (ou seja, uma variável endógena não pode ter um efeito direto sobre ela mesma);
- Definição de uma métrica para os termos de ruído ζ . A forma mais comum é feita fixando-se os coeficientes que relacionam as variáveis endógenas aos termos de ruído em 1 (um).

Vale lembrar que em SEM os parâmetros do modelo contidos no vetor $\mathbf{\Omega} = (\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Psi}, \mathbf{\Phi})$ são estimados com base na matriz de variância-covariância $\mathbf{\Sigma}$.

Existem na literatura algumas regras para verificação da identificabilidade do modelo, que incluem a regra da contagem (*counting rule*) e a regra recursiva (*recursive rule*). A primeira regra é geralmente usada em aplicações da área de ciências sociais, enquanto que a outra regra é mais comum em aplicações econométricas com uso de modelos não recursivos.

A regra da contagem é a mais simples para avaliação da identificabilidade do modelo. Seja $s = p + q$ o número total de variáveis, onde p e q representam o número de variáveis endógenas e exógenas, respectivamente. Assim, o número de elementos não redundantes em $\mathbf{\Sigma}$ é igual a $\frac{1}{2}s(s+1)$. Considere ainda que seja t o número de parâmetros a serem estimados pelo modelo. Dessa forma, uma condição necessária para a identificabilidade do modelo é que $t \leq \frac{1}{2}s(s+1)$.

Pela regra da contagem, tem-se que:

- Se $t = \frac{1}{2}s(s+1)$ então o modelo é dito *just identified* [exatamente identificável];
- Se $t < \frac{1}{2}s(s+1)$ então o modelo é dito *overidentified* [super-identificável];
- Se $t > \frac{1}{2}s(s+1)$ então o modelo é dito *not identified* [não identificável].

Para ilustração do uso da regra da contagem, suponha, por exemplo, que os parâmetros a serem estimados pelo modelo são 10 variâncias e covariâncias de variáveis exógenas e 8 coeficientes de regressão ($t = 18$). Nesse exemplo, considere que $s = 7$ e o número de elementos não redundantes em $\mathbf{\Sigma}$ é igual a 28. Nesse caso, como $t < 28$, o modelo é super-

identificável.

A maior vantagem da regra da contagem é sua simplicidade. No entanto, essa regra é necessária, mas não suficiente. Nesse caso, pode-se considerar a regra recursiva para a qual uma condição suficiente é que \mathbf{B} seja triangular e que Ψ seja uma matriz diagonal. Isso implica que todo modelo recursivo é identificável, pois Ψ não contém covariância entre os termos de ruído. Verifique que estas condições são satisfeitas no exemplo apresentado em (9).

Outra condição que é necessária e suficiente para identificabilidade do modelo é denominada condição de posto da matriz (*rank condition*). De acordo com esse procedimento, para que o modelo de equação estrutural seja identificável é necessário ter pelo menos o mesmo número de variáveis exógenas e parâmetros a serem estimados na equação.

Suponha que existam q variáveis exógenas no modelo. Pela condição anterior tem-se que:

- Se um modelo tem exatamente q parâmetros estruturais então o modelo é dito *just identified* [exatamente identificável];
- Se um modelo tem menos que q parâmetros estruturais então o modelo é dito *overidentified* [super-identificável];
- Se um modelo tem mais que q parâmetros estruturais então o modelo é dito *not identified* [não identificável] e não pode ser estimado.

Assim, os modelos recursivos podem ser apenas de dois tipos: exatamente identificáveis ou super-identificáveis. Outros tipos de restrições devem ser definidos para os modelos não-recursivos, mas não são descritos neste trabalho.

2.9 Avaliação dos critérios de qualidade do ajuste

Vários critérios podem ser usados para avaliação da bondade do ajuste do modelo. A determinação do ajuste do modelo é complicada nesse contexto porque vários critérios de bondade do ajuste foram desenvolvidos para avaliação dos modelos de equações estruturais sob diferentes suposições. Segundo Strumacker e Lomax (2004), a determinação da bondade do ajuste na SEM não é tão direta quanto em outros procedimentos multivariados porque nesses métodos assume-se que as variáveis observadas são mensuradas sem erro e existem testes estatísticos com distribuições conhecidas. Os índices de ajuste da SEM, por sua vez, não têm um teste de significância estatística que identifique o modelo correto considerando-se os dados amostrais.

O qui-quadrado é o único teste estatístico usado para avaliação do modelo teórico. Neste

caso, o valor do qui-quadrado varia entre zero, quando se considera o modelo saturado, e um valor máximo, quando não se considera nenhuma relação entre as variáveis (modelo independente). Um resultado estatisticamente não significativo para o qui-quadrado indica que a matriz de covariância amostral e a matriz de covariância estimada pelo modelo são similares. Muitos dos critérios de ajuste são calculados com base no conhecimento do modelo saturado, no modelo independente, no tamanho amostral, nos graus de liberdade e nos valores do qui-quadrado para formular um índice que varia entre 0 (ajuste inadequado) e 1 (ajuste perfeito).

Tabela 1. Principais critérios de ajuste para o SEM.

Medida de ajuste	Critério para avaliação	Comentários
Qui-quadrado	$p > 0,05$	Com amostras grandes e menor desvios, o ajuste pode ser estatisticamente significativo.
Qui-quadrado/gl	< 2: excelente ajuste 3 a 5: ajuste bom > 5: ajuste pobre	Compensa para amostras grandes
Jöreskog Sorbom Índice de Bondade (GFI)	1 refere-se ao ajuste perfeito Objetivo é > 0,9	Refere-se à proporção da covariância observada que é explicada pela covariância do modelo
Jöreskog Sorbom Índice de Bondade Ajustado (AGFI)	1 refere-se ao ajuste perfeito Objetivo é > 0,9	Inclui ajuste para modelos complexos
Índice de Comparação do Ajuste de Bentler (CFI)	1 refere-se ao ajuste perfeito Objetivo é > 0,9	Corresponde melhor ajuste do modelo aos dados quando as variáveis são independentes
Índice de Tucker-Lewis	1 refere-se ao ajuste perfeito Objetivo é > 0,9	Inclui ajuste para modelos complexos
Raiz do Erro Quadrático Médio de Aproximação (RMSEA)	< 0,05	Reflete a diferença média entre a covariância observada e a do modelo
Hoelter's n	> 200	Estima o tamanho da amostra suficiente para ajuste adequado do modelo

Recomenda-se que vários critérios de bondade do ajuste sejam usados em combinação como medida de ajuste global (Hair et al, 2005). Dentre os critérios de bondade de ajuste, ressaltam-se:

- o índice de qualidade do ajuste (GFI, em inglês): valor 1 indica ajuste perfeito,
- a raiz do erro quadrático médio (RMSEA, em inglês), em que valores inferiores a 0,05 indicam um bom ajuste do modelo,

- o índice de ajuste comparativo (CFI, em inglês), para o qual se esperam valores superiores a 0,9.

Matematicamente, esses critérios podem ser definidos como:

$$GFI = 1 - \frac{\chi_{\text{modelo}}^2}{\chi_{\text{nulo}}^2}, \quad RMSEA = \sqrt{\frac{\chi_{\text{modelo}}^2 - gl_{\text{modelo}}}{(N-1)gl_{\text{modelo}}}} \quad \text{e} \quad CFI = 1 - \frac{\chi_{\text{modelo}}^2 - gl_{\text{modelo}}}{\chi_{\text{nulo}}^2 - gl_{\text{nulo}}}$$

onde gl denota os graus de liberdade do modelo.

A Tabela 1 sintetiza alguns dos principais critérios para avaliação da qualidade do ajuste segundo Jackson e colaboradores (2005).

2.10 Efeitos direto, indireto e total

As suposições estatísticas, as equações e as interpretações associadas aos coeficientes em uma análise de caminhos são similares às dos modelos de regressão tradicionais, mas as relações estruturais são especificadas *a priori*.

Na terminologia associada à *path analysis*, os elementos em \mathbf{B} e em $\mathbf{\Gamma}$ representam os efeitos diretos do modelo. Nesse caso, a interpretação é feita similarmente a qualquer tipo de coeficiente de regressão. Além dos efeitos diretos, a *path analysis* permite a decomposição dos efeitos em totais e indiretos.

O efeito indireto é definido quando uma variável exógena influencia uma variável endógena através da mediação de pelo menos uma outra variável. Na existência de efeito indireto, o efeito total é definido pela soma do efeito direto e de todos os efeitos indiretos.

Considere para a *path analysis* o modelo definido por:

$$y = \mathbf{B}y + \mathbf{\Gamma}x + \zeta$$

Já vimos que assumindo-se que $(\mathbf{I} - \mathbf{B})$ não é singular e com alguma manipulação algébrica, esse modelo pode ser reescrito como:

$$y = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}x + (\mathbf{I} - \mathbf{B})^{-1}\zeta$$

Nesse caso, tem-se:

- O efeito total é definido por $(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}$;
- $\mathbf{\Gamma}$ é a matrix de efeitos diretos das variáveis exógenas sobre as endógenas;
- $(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma} - \mathbf{\Gamma}$ contém os efeitos indiretos.

Geralmente os *softwares* estatísticos provêm os efeitos totais. No entanto, a modelagem de equações estruturais permite a estimação automática de todos os efeitos que estejam

contemplados no diagrama de caminhos correspondente.

2.11 Estimação padronizada

A interpretação dos coeficientes estimados pelos modelos de regressão, bem como em modelos mais complexos, como os modelos de equações estruturais, é fundamental para o entendimento do pesquisador sobre as relações existentes no fenômeno em estudo. A complexidade dos fatores que influenciam esses coeficientes geralmente dificulta a interpretação dos mesmos. Uma questão interessante diz respeito a diferentes tipos de coeficientes que podem ser obtidos durante a modelagem e suas correspondentes interpretações (Grace, Bollen, 2005).

Quando as variáveis observadas provêm de diferentes escalas arbitrárias, é geralmente necessário o uso de coeficientes padronizados para ajudar na interpretação dos resultados. Uma forma de resolver o problema de métricas distintas e/ou arbitrárias das variáveis é padronizar os parâmetros estruturais do modelo.

No caso da regressão linear simples, sabe-se que os coeficientes não padronizados representam a inclinação da reta, enquanto que o coeficiente padronizado representa a raiz quadrada da variância explicada da variável resposta. No contexto da modelagem com equações estruturais, os coeficientes padronizados são amplamente utilizados, sendo disponibilizados por todos os *softwares* estatísticos.

Na literatura encontram-se descritos diversos métodos para obtenção de coeficientes padronizados para modelos de regressão lineares, incluindo métodos usando medidas de variabilidade, usando correlações e usando variáveis padronizadas. Considerando-se o modelo de regressão linear múltiplo:

$$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i; \quad \varepsilon_i \sim N(0, \sigma^2),$$

onde $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$ representa o vetor de parâmetros ou coeficientes não padronizados do modelo. O coeficiente padronizado para a k -ésima variável independente, usando o primeiro método, é definido multiplicando-se a estimativa não padronizada pela razão entre o desvio padrão da variável independente e da variável resposta (Kaplan, 2000):

$$\beta'_k = \beta_k \times \frac{S_{xk}}{S_y}$$

onde β'_k representa o k -ésimo coeficiente padronizado no modelo de regressão linear múltiplo. Neste caso, o erro padrão do coeficiente padronizado é dado por:

$$S_{bk} = S_{bk} \times \frac{S_{xk}}{S_y} \sqrt{\frac{1 - R_{YH}^2}{(1 - R_{X_k G_k}^2) \times (N - K - 1)}},$$

onde H representa todas as variáveis independentes, G_k representa o grupo de todas as variáveis X exceto a variável X_k e R^2 o coeficiente de determinação do modelo.

Outro método utiliza as correlações entre as variáveis resposta e independentes e as correlações entre as variáveis independentes (Grace e Bollen, 2005). No modelo de regressão linear múltiplo com apenas duas variáveis independentes, os coeficientes padronizados podem ser definidos por:

$$\beta'_1 = \frac{r_{y1} - (r_{12} \times r_{y2})}{1 - r_{12}^2},$$

$$\beta'_2 = \frac{r_{y2} - (r_{12} \times r_{y1})}{1 - r_{12}^2}$$

onde r_{y1} e r_{y2} representam os coeficientes de correlação de Pearson entre as variáveis independentes e dependentes e r_{12} representa o coeficiente de correlação de Pearson entre as variáveis independentes. No caso bivariado (modelo linear simples), ou seja, quando só existe uma variável independente, o coeficiente padronizado pode ser definido por:

$$\beta' = r_{xy},$$

onde r_{xy} é o coeficiente de correlação de Pearson entre a variável independente e a variável dependente.

Um terceiro método consiste em padronizar todas as variáveis incluídas na análise através da estatística x_{z_i} dada por:

$$x_{z_i} = \frac{x_i - \bar{X}}{S}.$$

O ajuste do modelo é feito então utilizando-se as variáveis padronizadas. O modelo de regressão linear múltiplo com variáveis padronizadas pode ser definido por:

$$y_{z_i} = \beta'_1 x_{z_{1i}} + \beta'_2 x_{z_{2i}} + \dots + \beta'_p x_{z_{pi}} + \epsilon'_i; \quad \epsilon'_i \sim N(0,1),$$

onde β'_k representa o k -ésimo coeficiente padronizado.

Vale lembrar que quando é feita a padronização o parâmetro β'_0 torna-se igual a zero, sendo por este motivo retirado da equação do modelo. A equação anterior pode ser escrita da seguinte forma:

$$\frac{y_i - \bar{Y}}{S_y} = \beta'_1 \frac{x_{1i} - \bar{X}_1}{S_{x_1}} + \beta'_2 \frac{x_{2i} - \bar{X}_2}{S_{x_2}} + \dots + \beta'_p \frac{x_{pi} - \bar{X}_p}{S_{x_p}} + \epsilon'_i; \quad \epsilon'_i \sim N(0,1)$$

No contexto de *path analysis*, considere que γ_{pq} representa um dos coeficientes não padronizados para um elemento de $\mathbf{\Gamma}$. O coeficiente padronizado é obtido, então, por:

$$\hat{\gamma}_{pq}^* = \frac{\hat{\sigma}_{x_q}}{\hat{\sigma}_{y_p}} \hat{\gamma}_{pq}$$

onde $\hat{\sigma}_{y_p}$ e $\hat{\sigma}_{x_q}$ são as variâncias de y_p e x_q , respectivamente. Note que $\hat{\sigma}_{y_p}$ é obtido através do elemento estimado na diagonal superior à esquerda da matriz $\mathbf{\Sigma}_{yx}$, dado por $(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}' + \mathbf{\Psi})(\mathbf{I} - \mathbf{B})^{-1}$; enquanto $\hat{\sigma}_{x_q}$ é obtido pelo elemento estimado na diagonal inferior à direita da matriz $\mathbf{\Sigma}_{yx}$. Analogamente, para cada elemento de \mathbf{B} , o coeficiente padronizado é dado por:

$$\hat{\beta}_{q'}^* = \frac{\hat{\sigma}_{x_{q'}}}{\hat{\sigma}_{y_p}} \hat{\beta}_{q'}$$

Coefficientes padronizados também podem ser obtidos para efeitos diretos e indiretos do modelo.

2.12 Especificação do Modelo de Equações Estruturais Generalizado

Nesta seção o modelo de equações estruturais generalizado, que consiste dos submodelos de mensuração e estrutural, é definido. Esse modelo foi introduzido por Karl Jöreskog e colaboradores, sendo uma generalização do modelo discutido previamente. O modelo estrutural é similar a aquele definido para variáveis observadas, sendo denotado por:

$$\eta = \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta \quad (11)$$

onde η representa um vetor $m \times 1$ de variáveis latentes endógenas, ξ representa um vetor $k \times 1$ de variáveis latentes exógenas, \mathbf{B} é uma matriz $m \times m$ de coeficientes relacionando as variáveis latentes endógenas entre si; $\mathbf{\Gamma}$ é uma matriz $m \times k$ de coeficientes relacionando variáveis endógenas a variáveis exógenas; e ζ é um vetor $m \times 1$ de ruídos (*disturbances*) estruturais. Usando-se essa notação \mathbf{B} apresenta zeros em sua diagonal principal (Kaplan, 2000).

As variáveis latentes são relacionadas com as variáveis observadas através do modelo de mensuração, que é definido separadamente para variáveis endógenas e exógenas por:

$$y = \Lambda_y \eta + \varepsilon \quad \text{e} \quad x = \Lambda_x \xi + \delta$$

onde Λ_y e Λ_x são matrizes $p \times m$ e $q \times k$, respectivamente, de cargas fatoriais, e ε e δ são vetores $p \times 1$ e $q \times 1$, respectivamente, de erros de mensuração em y e em x . Cada coluna das matrizes Λ geralmente contém um valor que é fixado em 1 para estabelecer a escala da variável latente correspondente. Alternativamente, isso pode ser feito fixando-se em 1 as variâncias das variáveis latentes exógenas em Φ .

Nesse modelo assume-se que os erros de mensuração ε e δ têm esperança zero, cada um com distribuição normal multivariada, independentes entre si e independentes das variáveis exógenas latentes (ξ), das variáveis endógenas latentes (η) e dos ruídos (ζ). Além disso, assume-se que as observações são amostradas independentemente e que as variáveis exógenas latentes (ξ) têm distribuição normal multivariada. Essa última suposição é desnecessária para variáveis exógenas que são medidas sem erro.

Os ruídos estruturais (ζ) têm esperança zero, têm distribuição normal multivariada e são independentes das variáveis exógenas latentes (ξ). Sob essas suposições, os indicadores observados, x e y , têm uma distribuição normal multivariada

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N_{p+q}(0, \Sigma)$$

onde Σ representa a matriz de covariância populacional dos indicadores. Essa matriz é função dos parâmetros do modelo $\Omega = (\mathbf{B}, \Gamma, \Lambda_x, \Lambda_y, \Psi, \Theta_\delta, \Theta_\varepsilon \text{ e } \Phi)$, podendo ser expressa por:

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} = \begin{pmatrix} \Lambda_y (\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) [(\mathbf{I} - \mathbf{B})^{-1}]' \Lambda_y' + \Theta_\varepsilon & \Lambda_y (\mathbf{I} - \mathbf{B})^{-1} \Gamma \Phi \Lambda_x' \\ \Lambda_y (\mathbf{I} - \mathbf{B})^{-1} \Gamma \Phi \Lambda_x' & \Lambda_x \Phi \Lambda_x' + \Theta_\delta \end{pmatrix}$$

onde Φ é a matriz de covariância $k \times k$ das variáveis exógenas latentes, Ψ é a matriz de covariância $m \times m$ de termos de ruído, e Θ_ε e Θ_δ são as matrizes de covariância dos erros de mensuração ε e δ , respectivamente.

A análise de caminhos clássica, conforme discutida na seção 2.6, pode ser vista como um submodelo nessa estrutura mais geral, considerando-se $\Lambda_x = \mathbf{I}$, $\Theta_\delta = 0$, $\Lambda_y = \mathbf{I}$, $\Theta_\varepsilon = 0$.

Substituindo esses resultados em Σ obtém-se a mesma estrutura definida em (7). Similarmente, para obtenção da matriz Σ associada à análise fatorial confirmatória, considere $\mathbf{B} = 0$, $\Gamma = 0$, $\Psi = 0$, $\Lambda_y = 0$ e $\Theta_\varepsilon = 0$.

Em qualquer modelo em particular serão necessárias algumas restrições em alguns elementos

dessa matriz. Mais comumente, algumas restrições incluem fixar alguns desses parâmetros como zeros. Além disso, algumas das matrizes Λ contêm valores fixos conforme discutido anteriormente. Se as restrições desse modelo são suficientes, então estimativas de máxima verossimilhança podem ser obtidas para esses parâmetros.

A log-verossimilhança associada a esse modelo é dada por:

$$\log_e L(\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Lambda}_x, \mathbf{\Lambda}_y, \mathbf{\Psi}, \mathbf{\Theta}_\delta, \mathbf{\Theta}_\epsilon; \mathbf{\Phi}) = -\frac{N(p+q)}{2} \log_e 2\pi - \frac{N}{2} [\log_e \det \mathbf{\Sigma} + \text{trace}(S\mathbf{\Sigma}^{-1})] \quad (12)$$

onde S é a matriz de covariância amostral entre as variáveis observadas. Como o logaritmo da verossimilhança pode ser pensado como uma medida de proximidade entre $\mathbf{\Sigma}$ e S , então as estimativas de máxima verossimilhança dos parâmetros são definidas de modo que essas duas matrizes sejam o mais próximo possível. Novamente os erros padrão assintóticos para as estimativas dos parâmetros podem ser obtidos pela raiz quadrada das diagonais da matriz de informação. Assume-se neste contexto que as relações estruturais entre as variáveis latentes exógenas e endógenas são lineares, bem como as relações entre as variáveis indicadoras (observadas) e os construtos associados às mesmas.

A verificação da identificabilidade desses modelos com variáveis latentes é um problema complexo que não tem solução simples. Uma condição geral necessária para identificabilidade é o número de parâmetros livres no modelo, que não pode ser superior ao número de variâncias e covariâncias entre as variáveis observáveis, que é dado por:

$$\frac{(p+q) \times (p+q+1)}{2}$$

Esta condição, no entanto, não é suficiente. É fácil atender a esta condição e ainda ter um modelo não identificável. Uma regra que às vezes pode ajudar na verificação da identificabilidade do modelo é a seguinte:

- todos os erros de mensuração são não correlacionados com os outros,
- existem pelo menos dois indicadores exclusivos para cada variável latente, ou se há apenas um único indicador para uma variável latente, essa é mensurada sem erro,
- o modelo estrutural contém apenas variáveis observadas.

Os *softwares* estatísticos geralmente detectam as tentativas de ajuste de um modelo subidentificado. Neste caso, a matriz de informação será singular. Outra forma de verificar problemas de identificabilidade do modelo é observar se as estimativas das variâncias são muito grandes.

2.13 Etapas da implementação da SEM

A Figura 8 ilustra as etapas para a construção de um modelo de equações estruturais. De acordo com esse esquema, primeiramente identifica-se uma teoria que será utilizada como base para a especificação do modelo. A especificação do modelo consiste na definição das relações causais. De posse do modelo e a partir de dados de uma amostra, obtêm-se as medidas referentes às variáveis observáveis do modelo, que permitirão a construção das variáveis latentes especificadas no modelo. Com essas medidas (das variáveis observadas e dos construtos) são estimados os parâmetros do modelo de medição e do estrutural. Geralmente os modelos de mensuração e estrutural são ajustados simultaneamente, podendo, no entanto, ser o modelo de mensuração estimado inicialmente. A avaliação da bondade do ajuste pode ser feita após a estimação dos parâmetros da SEM. Índices de modificação do modelo são produzidos pela maioria dos *softwares*, podendo servir de suporte para potenciais alterações do modelo teórico. Com o modelo definitivo, após as modificações necessárias e a estimação final dos parâmetros, segue uma discussão sobre o modelo encontrado.

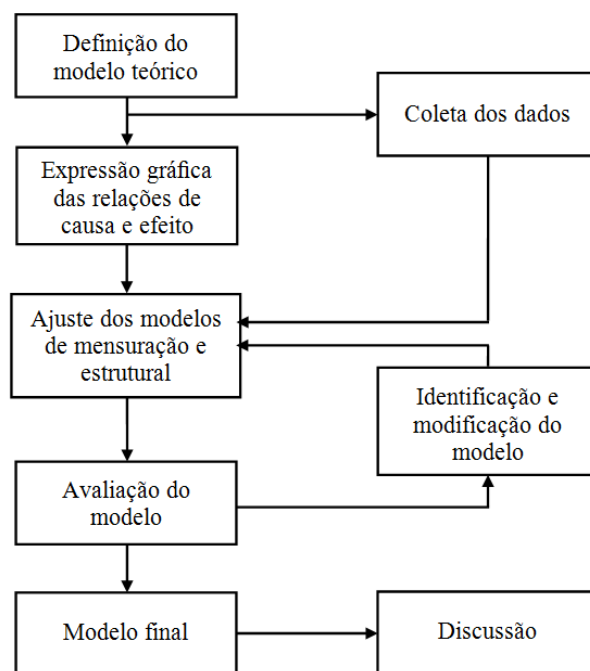


Figura 8: Etapas para a construção de um modelo de equações estruturais.

Essas etapas foram sumarizadas por Hair e colaboradores (2005) em estágios sequenciados, que podem ser descritos como:

- Estágio 1. Desenvolvimento do modelo teórico - consiste no pesquisador expressar suas hipóteses baseando-se nas relações causais entre as variáveis selecionadas para a pesquisa com fundamentos teóricos. O investigador precisa basear-se em pesquisas e teorias prévias para escolher e especificar um modelo plausível. A especificação do modelo envolve determinar cada relação e parâmetro do modelo de interesse. Alguns autores afirmam que essa é a etapa mais árdua da SEM (Strumacker e Lomax, 2004).
- Estágio 2. Construção do diagrama de caminhos das relações definidas - expressão gráfica de causa e efeito desenvolvida no modelo teórico.
- Estágio 3. Conversão do diagrama de caminhos em modelo de mensuração e modelo estrutural - consiste na formação do modelo teórico através das equações propostas. Esta etapa envolve também a validação dos construtos por intermédio da análise fatorial confirmatória. A partir do modelo teórico são criadas as hipóteses a serem testadas na aplicação dos sistemas de equações estruturais.
- Estágio 4. Verificação da identificabilidade do modelo estrutural - consiste em verificar a existência de limitações explanatórias do modelo para gerar soluções únicas.
- Estágio 5. Avaliação dos critérios de ajuste do modelo - etapa responsável pela avaliação da adequação e ajuste geral do modelo através do uso das medidas de adequação absolutas, medidas de ajuste incremental e medidas de ajustes de parcimônia.
- Estágio 6. Interpretação e modificação do modelo - consiste na comparação dos resultados obtidos no desenvolvimento do modelo com os objetos e hipóteses estabelecidas no estudo a partir da teoria, comparando estatisticamente essas afirmações. Podem também ser estabelecidos parâmetros para a reespecificação do modelo, com a inclusão ou não dos parâmetros estimados no modelo teórico original.

2.14 Softwares estatísticos

Existem diversos programas estatísticos que implementam os modelos de equações estruturais. Esses programas podem ser: programáveis - o pesquisador necessita conhecer a linguagem de programação para efetuar suas análises; ou não programáveis - que têm uma apresentação interativa com o usuário. A seguir alguns desses programas são descritos, apresentando-se suas características principais e um pouco de sua história.

– AMOS (*Analysis of Moment Structures*)

O Amos é um módulo de extensão do programa estatístico SPSS, desenvolvido para aplicação de SEM. Ele é subdividido em dois módulos: *Amos Graphics* e *Amos Basic*. No *Amos Graphics* o usuário pode especificar o modelo através de diagramas de caminhos, não precisando fazer nenhum tipo de programação devido à existência de um grupo de ferramentas disponíveis para analisar, modificar ou alinhar elementos do modelo. O *Amos Basic* utiliza a linguagem de programação, ou seja, o modelo e os dados são especificados por sintaxe em todas as análises. A vantagem deste modo é a possibilidade de programação adicional não disponível no *Amos Graphic* (Kline, 2005).

– LISREL (*Linear Structural Relationships*)

O LISREL é considerado sinônimo de SEM por ter sido desenvolvido especificamente para o emprego desta técnica e por ser o pioneiro em sua aplicação. Atualmente o LISREL inclui outras aplicações como manipulação de dados, estatísticas básicas, modelos lineares generalizados e ainda modelos multiníveis. Foi desenvolvido há cerca de 30 anos para permitir a avaliação empírica das teorias de pesquisadores da área de ciências sociais através do uso da SEM. O LISREL é subdividido em três módulos: PRELIS, SIMPLIS e *Path Diagram*. O PRELIS é responsável pela manipulação dos dados e cálculos de medidas estatísticas. O SIMPLIS é o modo de programação, onde o usuário necessita especificar suas análises através de sintaxe. O *Path Diagram* é o modo que o usuário utiliza para realizar a análise a partir do diagrama de caminhos, consistindo em desenhar as relações entre as variáveis e a partir daí criar a sintaxe para posteriormente ser executada.

– MPLUS

O MPlus é também um programa desenvolvido para a implementação da Modelagem de Equações Estruturais, que foi desenvolvido há cerca de 10 anos pelos pesquisadores L. Muthén e B. Muthén na versão do Windows. Atualmente encontra-se na versão 6.1. É um pacote de análise de dados poderoso no sentido que foi construído sobre a estrutura de modelagem geral, mas com flexibilidade para usar combinações de variáveis latentes contínuas e categóricas. Tal estrutura integrada considera como casos especiais a SEM, os modelos de curva de crescimento misto e latente e os modelos multiníveis. Recentemente incorporou a questão de lidar com

amostras complexas nesses modelos. O MPlus possui a desvantagem de não construir o diagrama de caminhos, dificultando a visualização das relações entre as variáveis, necessitando dessa forma que o diagrama seja construído separadamente para servir como base (Kline, 2004). O MPlus foi desenvolvido na forma de linguagem de programação, mas caso o usuário desconheça o uso de sintaxes poderá, através do uso das ferramentas do programa, obter as sintaxes apenas direcionando ao tipo de análise escolhida.

– MX GRAPH

Até o momento foram apresentados programas que são pagos (AMOS, LISREL, MPlus), mas o Mx Graph é um programa completamente livre, podendo ser obtido pela internet sem a necessidade de licença. O Mx Graph é considerado um processador de matriz e um otimizador numérico que pode analisar modelos de equações estruturais e outros tipos de modelos estatísticos multivariados. Existem dois tipos de módulos de utilização do Mx Graph (Kline, 2004): o primeiro refere-se ao uso de linguagem de programação para descrever os dados e definir o modelo a ser utilizado na análise. Sua sintaxe representa o modelo de equação estrutural em três tipos de matrizes: a simétrica (para análise de associações), a assimétrica (para efeitos diretos) e a filtrada (para identificar as variáveis observadas). Para o uso do segundo módulo não é necessário o conhecimento de linguagem de programação, pois as relações são estabelecidas através do diagrama de caminhos utilizando as ferramentas do Mx Graph. Uma característica especial do Mx Graph para SEM é sua habilidade em calcular intervalos de confiança e poder estatístico para estimar parâmetros individuais, e em analisar modelos com variáveis contínuas ou categóricas.

– R

O R é um programa estatístico livre desenvolvido em ambiente de linguagem de programação R. Foi criado por Ross Ihaka e Robert Gentleman com a colaboração de vários pesquisadores ao redor do mundo. É muito utilizado por pesquisadores, especialmente pelos Estatísticos. Atualmente encontra-se na versão 2.15, estando disponível tanto para Windows quanto para Mac e Unix. O R é também altamente expansível com o uso dos seus pacotes, que são bibliotecas com funções específicas ou áreas de estudo específicas, como o pacote SEM. O pacote do R com aplicação em SEM possibilita ajuste de modelos com variáveis observadas e latentes através do uso da técnica da máxima verossimilhança, assumindo multinormalidade (Fox, 2006).

3 Aplicações da SEM

Para ilustração da aplicabilidade dos métodos discutidos neste trabalho são apresentadas duas análises que foram conduzidas usando esta metodologia. Essas aplicações incluem desde o uso do modelo mais simplificado na abordagem SEM, contendo apenas análise fatorial confirmatória, até situações mais complexas envolvendo modelos não recursivos. A interpretação dos resultados em cada uma das aplicações é enfatizada.

3.1 Padrões de consumo alimentar no interior da Bahia

Utilizou-se análise fatorial confirmatória (AFC) para definir grupos de alimentos (construtos) baseados no consumo de itens alimentares individualizados por crianças residentes na zona urbana de municípios do interior da Bahia em duas faixas etárias: menor de 12 meses (Figura 9) e entre 12 e 24 meses (Figura 10). Análise fatorial confirmatória é frequentemente usada no processo de avaliação de modelos de equações estruturais mais gerais. Contudo, AFC, por ela mesma, tem se tornado uma ferramenta poderosa para avaliação da validade e confiabilidade de instrumentos acadêmicos, do comportamento e da personalidade, por exemplo (Mueller, 1996). No contexto dessa aplicação estaremos usando-a para avaliação de dados nutricionais.

O objetivo do estudo, denominado VIDACRI, é definir perfis de consumo alimentar em crianças das mencionadas faixas etárias na população de estudo. Este estudo foi conduzido por pesquisadores da Escola de Nutrição da UFBA. Nas análises apresentadas neste relatório considerou-se somente informações das crianças que tinham dados sobre hemoglobina ($n = 1116$), sendo aproximadamente 26,97% menores de 12 meses. As variáveis observadas utilizadas na análise foram as calorias provenientes do consumo de: açúcares e doces, leite e derivados, leite materno, cereais, feijão, carne e arroz. Usando AFC testou-se se o modelo de consumo alimentar hipotetizado pelos pesquisadores é confirmado pelos dados do estudo VIDACRI. Essas análises foram realizadas usando-se o *software* LISREL.

Nas Figuras 9 e 10 estão representadas as relações teóricas entre os fatores e as variáveis observadas (indicadores). Essas relações são interpretadas como cargas fatoriais. Os erros de mensuração estão representados pelas setas ao lado esquerdo dos indicadores e indicam qual a porção de cada indicador que está mensurando alguma outra coisa além do fator hipotetizado. Para avaliar o erro de mensuração, a variância de cada erro de mensuração é estimada.

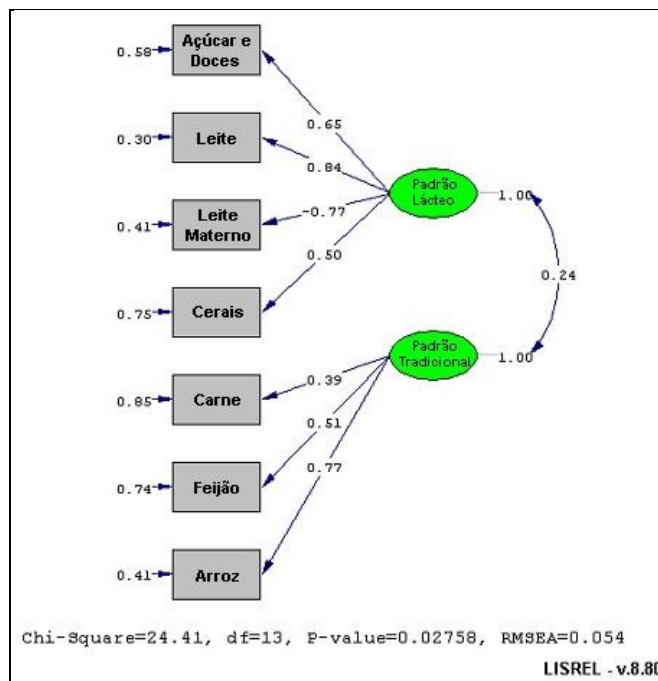


Figura 9: Diagrama de caminhos referente ao consumo alimentar em crianças menores de 12 meses.

Para verificar a identificabilidade dos modelos representados pelos diagramas de caminhos das Figuras 9 e 10, vamos utilizar a regra da contagem através da qual temos $\frac{7(7+1)}{2} = 28$ elementos distintos na matriz de covariância S . O número de parâmetros a ser estimado é dado por 7 cargas fatoriais, 7 variâncias dos erros de mensuração e uma correlação entre os construtos, resultando em um total de 15 parâmetros livres. Como o número de elementos em S é maior do que o de parâmetros livres, então tem-se $gl = 28 - 15 = 13$ graus de liberdade para o modelo especificado. De acordo com este critério o modelo é dito superidentificável. Vale lembrar que esta condição é necessária, mas não suficiente para estabelecer a identificabilidade do modelo. No entanto, considerando-se as condições listadas na seção 2.12, pode-se corroborar a identificabilidade deste modelo.

O modelo de mensuração representado da Figura 9 pode ser definido por:

$$x = \Lambda_x \xi + \delta$$

que neste caso é dado por:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{12} & 0 \\ \lambda_{13} & 0 \\ \lambda_{14} & 0 \\ 0 & \lambda_{25} \\ 0 & \lambda_{26} \\ 0 & \lambda_{27} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \end{pmatrix}$$

Analisando-se a Figura 9, pode-se notar que a variável observada *leite e derivados* (carga fatorial padronizada $\hat{\lambda}_{12} = 0,84$), referente ao consumo de calorias em crianças com menos de 12 meses, possui a maior contribuição na formação do construto *Fator 1* (ξ_1), enquanto que no construto *Fator 2* (ξ_2) a variável que exerce a maior contribuição é o *arroz* (carga fatorial padronizada $\hat{\lambda}_{27} = 0,77$). Pode-se ainda perceber que existe uma correlação positiva entre os *Fatores 1 e 2* ($r = 0,24$).

Modelo de mensuração similar ao definido anteriormente é considerado para ajuste usando dados de consumo de calorias na dieta de crianças com faixa etária de 12 a 24 meses, conforme apresentado na Figura 10. Nesse caso, verifica-se que a formação do construto *Fator 1* (ξ_1^*) deve-se principalmente ao consumo de *leite e derivados* (carga fatorial padronizada $\hat{\lambda}_{12}^* = 0,65$), enquanto que no construto *Fator 2* (ξ_2^*) novamente o consumo de *arroz* (carga fatorial padronizada = $\hat{\lambda}_{27}^* = 0,61$) é o que exerce uma maior contribuição na formação do mesmo. Nota-se, no entanto, que existe uma fraca correlação negativa entre os construtos *Fator 1 e Fator 2* ($r = -0,05$).

Tabela 2. Estimativas não padronizadas para os dois modelos de perfil de consumo alimentar segundo a faixa etária das crianças.

Variáveis observadas	< 12 meses			12 - 24 meses		
	$\hat{\lambda}$	$\hat{se}(\hat{\lambda})$	t	$\hat{\lambda}$	$\hat{se}(\hat{\lambda})$	t
Açúcar e doces	58,1	5,6	10,3	80,2	6,7	12,0
Leite e derivados	163,6	12,2	13,4	213,3	13,6	15,7
Leite materno	-56,6	4,5	12,5	-141,0	10,0	14,1
Cereais	70,8	8,6	8,3	73,8	8,8	8,4
Carne	15,8	2,6	6,2	4,4	0,8	5,5
Feijão	15,3	2,5	6,2	5,9	0,9	6,6
Arroz	32,0	4,6	7,0	22,3	3,0	7,4

Comparando as Figuras 9 e 10, pode-se notar que apesar das distintas faixas etárias das crianças, não parece haver diferença nos seus perfis alimentares, sendo que os fatores definidos indicam dieta láctea (*Fator 1*) e uma dieta tradicional (*Fator 2*).

A Tabela 2 apresenta as estimativas não padronizadas dos modelos apresentados nas Figuras 9 e 10. De acordo com esses resultados, todos os parâmetros estruturais (os coeficientes não padronizados) na matriz Λ_x são significativos, bem como as variâncias dos erros de mensuração associadas aos indicadores do modelo (dados não apresentados). Contudo, as correlações múltiplas ao quadrado (R^2) para os indicadores variaram entre 0,15 e 0,70 para os dados de crianças menores de 12 meses, e entre 0,13 e 0,42 para os dados de crianças entre 12 e 24 meses, indicando confiabilidade questionável para o padrão alimentar nessas amostras.

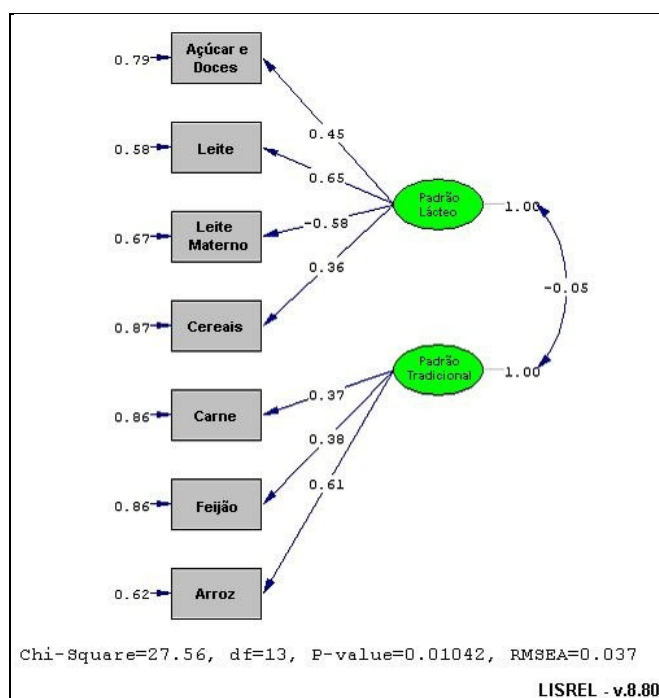


Figura 10: Diagrama de caminhos referente ao consumo alimentar em crianças de 12 a 24 meses.

CrITÉRIOS de ajuste do modelo são apresentados na Tabela 3 para os dados das crianças das duas faixas etárias. Pode-se verificar que o modelo encontra-se bem ajustado considerando-se os critérios definidos na Tabela 1.

Tabela 3. Critérios de bondade de ajuste para os dois modelos de padrão de consumo alimentar.

Medida de ajuste	Valor das estatísticas	
	<12 meses	12-24 meses
Qui-quadrado/ <i>gl</i>	1,92	2,14
GFI	0,98	0,99
AGFI	0,95	0,98
CFI	0,98	0,97
NFI	0,95	0,94
RMSEA	0,05	0,04

3.2 Desenvolvimento cognitivo em menores de 48 meses

Estudo foi realizado para descrever a relação entre o estado antropométrico, as condições socioeconômicas, a qualidade do lar e o desenvolvimento cognitivo de 320 crianças de 20 a 42 meses de idade residentes na cidade de Salvador - Bahia (Santos *et al*, 2008). Um critério de inclusão foi selecionar as crianças com menos de 42 meses nascidas no período de janeiro a julho de 1999. O desenvolvimento cognitivo (IDM) das crianças foi medido utilizando-se a escala de Bayley e as características do lar (HOME) foram mensuradas pelos objetos nele existentes. Para avaliar o estado antropométrico, utilizaram-se os escores antropométricos peso/idade e altura/idade, além do peso ao nascer. Os dados socioeconômicos foram coletados através de um questionário padrão. O SEM foi utilizado nessa aplicação para:

- i) definir construtos (variáveis latentes) que representem características do ambiente em que a criança vive, além de características relacionais e nutricionais da criança;
- ii) avaliar o impacto desses construtos no desenvolvimento cognitivo da criança.

Para a implementação do SEM foram utilizadas as seguintes variáveis: responsividade emocional e verbal (homei), ausência de punição e restrição (homeii), organização do ambiente físico e temporal (homeiii), disponibilidade de brinquedos (homeiv), envolvimento materno com a criança (homev), oportunidade de variação na estimulação (homevi), escore altura/idade (haz), peso ao nascer, em quilogramas (peso), sexo, idade em meses (idmes), e escore do desenvolvimento cognitivo infantil (IDM).

Conforme o diagrama de caminhos que representa a AFC considerada nesta aplicação, o construto *psico-social* é formado pelas variáveis observadas que se referem à responsividade emocional e verbal, à ausência de punição e restrição e ao envolvimento materno com a criança (Figura 11). O construto *físico-ambiental* é formado pelas variáveis referentes à organização do ambiente físico e temporal, à disponibilidade de brinquedos e à oportunidade de variação na estimulação, enquanto que o construto *nutricional* é formado pelas variáveis

referentes ao escore altura/idade e ao peso da criança ao nascer. A implementação dessas análises foi realizada no *software* Amos.

De acordo com resultados apresentados na Figura 11, pode-se notar que a variável referente ao envolvimento materno com a criança (carga fatorial padronizada $\hat{\lambda}_{13} = 0,74$) exerce uma maior contribuição na formação do construto *psico-social*. A variável referente à disponibilidade de brinquedos (carga fatorial padronizada $\hat{\lambda}_{22} = 0,73$) é a que exerce uma maior contribuição na formação do construto relacionado ao *ambiente* da criança, enquanto que no construto *nutricional* a variável com maior contribuição é o peso ao nascer (carga fatorial padronizada $\hat{\lambda}_{32} = 0,95$). Pode-se também notar na Figura 11 que os construtos *psico-social* e *físico-ambiental* apresentam uma correlação positiva significativa ($r = 0,67$), os construtos *físico-ambiental* e *nutricional* apresentam uma correlação positiva fraca ($r = 0,08$), enquanto que os construtos *psico-social* e *físico-ambiental* se correlacionam negativamente ($r = -0,04$).

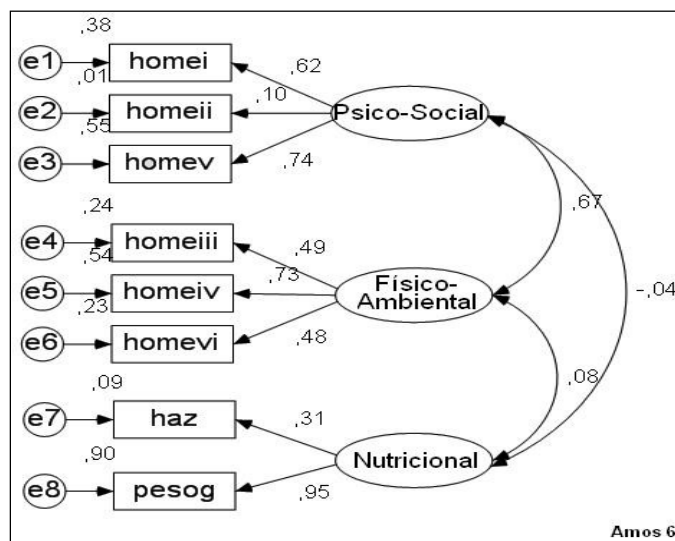


Figura 11: Análise fatorial confirmatória para desenvolvimento cognitivo em menores de 48 meses.

A análise das estimativas não padronizadas desse modelo indica que a maior parte dos parâmetros estruturais da matriz Λ_x são significativos (dados não apresentados). Ressalta-se, no entanto, o fato dos parâmetros associados ao construto *nutricional* não serem significativos, indicando potencialmente a necessidade de melhor definição deste construto. As variâncias dos erros de mensuração associadas aos indicadores do modelo foram todas

significativas (dados não apresentados). Diversos critérios indicam um adequado ajuste dos dados a este modelo (dados não apresentados).

Diferentemente do exemplo anterior, a variável endógena é observada e mensurada diretamente, referindo-se ao escore de desempenho cognitivo da criança (IDM). Aqui o modelo pode ser definido pelas seguintes equações:

$$x = \Lambda_x \xi + \delta \quad (20)$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{12} & 0 & 0 \\ \lambda_{13} & 0 & 0 \\ 0 & \lambda_{24} & 0 \\ 0 & \lambda_{25} & 0 \\ 0 & \lambda_{26} & 0 \\ 0 & 0 & \lambda_{37} \\ 0 & 0 & \lambda_{38} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \end{pmatrix}$$

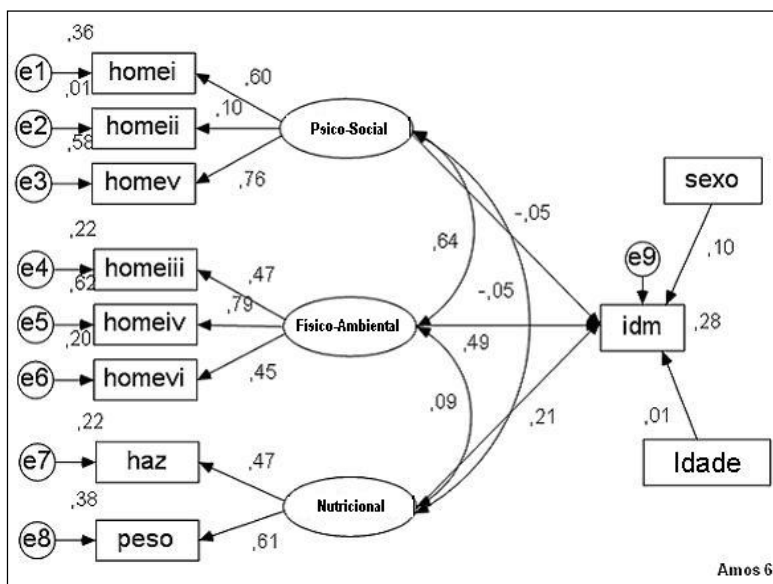


Figure 12: Modelo de equações estruturais para desenvolvimento cognitivo em menores de 48 meses.

O desempenho cognitivo é modelado pela equação

$$IDM = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \beta_1 SEXO + \beta_2 IDADE + \zeta_1$$

A Figura 12 representa o diagrama de caminhos que foi utilizado para avaliação do desempenho cognitivo infantil através da modelagem de equações estruturais utilizando o

software Amos. Pode-se identificar que a variável observada IDM está sendo explicada pelas variáveis latentes *físico-ambiental*, *psico-social* e *nutricional* e ainda pelas variáveis observadas idade e sexo, sendo que a variável latente *físico-ambiental* ($\hat{\gamma}_2$ padronizada = 0,49) é aquela que exerce a maior influência no desenvolvimento cognitivo infantil (IDM), enquanto que a variável latente *psico-social* não se encontra significativamente associada ao IDM ($p = 0,69$).

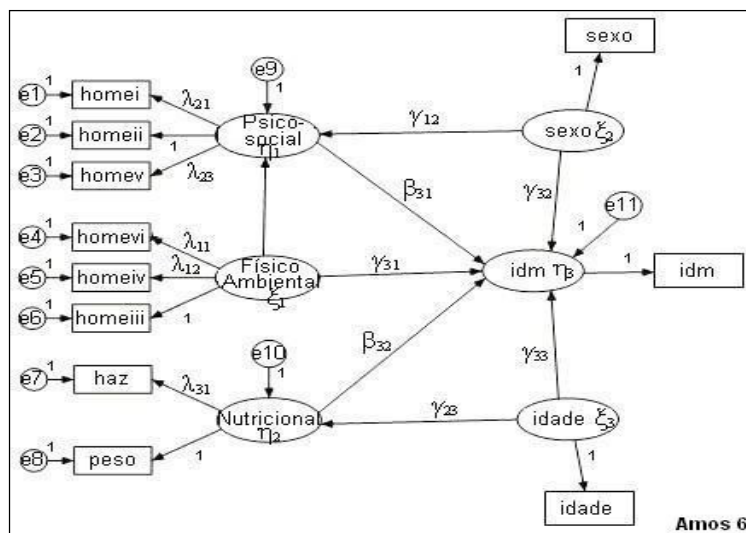


Figura 13: Diagrama de caminhos para SEM mais complexo para descrever o desenvolvimento cognitivo infantil.

Considerando-se uma estrutura um pouco mais complexa dessas interações e usando a notação da SEM, pode-se definir o diagrama de caminhos representado pela Figura 13.

Matricialmente, essas relações podem ser definidas por:

$$x = \Lambda_x \xi + \delta \quad (13)$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{12} & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{pmatrix}$$

e

$$y = \Lambda_y \eta + \varepsilon \quad (14)$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} \lambda_{21} & 0 & 0 \\ 1 & 0 & 0 \\ \lambda_{23} & 1 & 0 \\ 0 & \lambda_{31} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

As relações estruturais do SEM podem ser representadas por:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} & 0 \\ 0 & 0 & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix} \quad (15)$$

As estimativas padronizadas para o SEM descrito em (13), (14) e (15) são apresentadas na Figura 14. De acordo com esses resultados, o indicador físico ambiental é o que mais fortemente influencia o índice de desenvolvimento cognitivo (IDM) de crianças menores de 48 meses (γ_{31} padronizado = 0,51). Além do fator ambiental, o fator nutricional e sexo encontram-se significativamente associados com IDM.

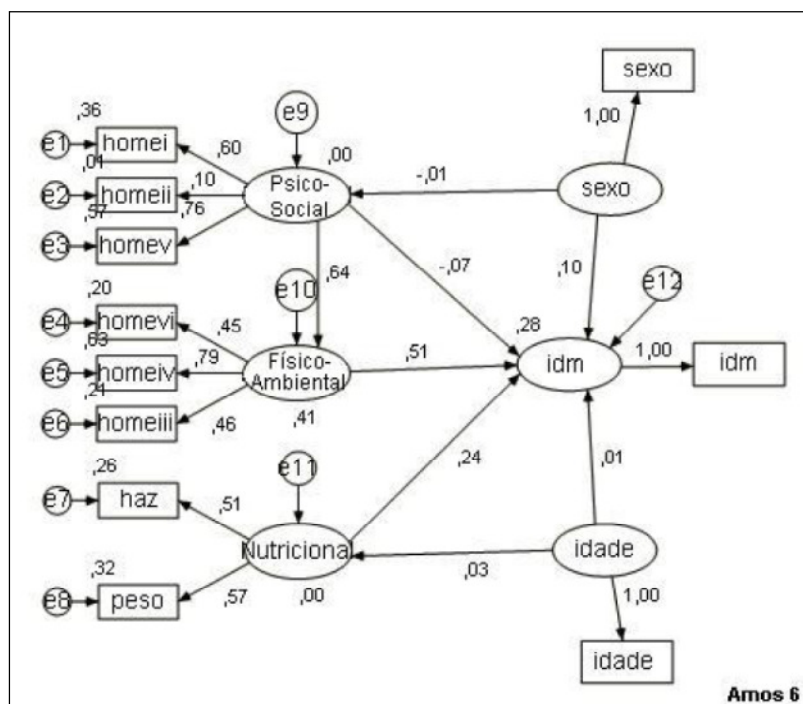


Figura 14: Estimativas para SEM mais complexo para descrever o desenvolvimento cognitivo infantil.

Analisando-se os critérios de ajuste do modelo apresentados na Tabela 4, pode se notar que

ambos os modelos encontram-se bem ajustados, não havendo diferenças importantes nos dois modelos.

Tabela 4. Critérios de bondade de ajuste para os dois SEMs sobre desenvolvimento cognitivo.

Medida de ajuste	Valor das estatísticas	
	Modelo 1 (simples)	Modelo 2 (complexo)
Qui-quadrado/ <i>gl</i>	1,576	1,667
GFI	0,967	0,967
AGFI	0,943	0,940
CFI	0,930	0,924
NFI	0,837	0,837
RMSEA	0,043	0,046

4. Considerações finais

Este trabalho abordou a modelagem com equações estruturais envolvendo variáveis endógenas contínuas, que possui teoria estatística já consolidada e tem sido amplamente utilizada em diversas áreas do conhecimento. No entanto, vale ressaltar que existe também literatura a respeito de SEM para respostas binárias ou, de maneira mais abrangente, envolvendo contextos em que o pressuposto de multinormalidade não se aplica (Muthén, 1984).

Outra área em crescente expansão é o uso de SEM para situações envolvendo estruturas complexas de dados e que muitas vezes envolve correlação entre múltiplas observações de um mesmo indivíduo ou entre indivíduos pertencentes a um mesmo grupo ou *cluster*. A aplicação do SEM em curvas de crescimento (Fergusson, 1997) ou usando a abordagem de modelagem multinível (Bauer, 2003) é objeto de investigação deste grupo de pesquisa.

Referências Bibliográficas

- AMORIM, Leila D. A. F. Structural equation modeling in epidemiology. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 26, n. 12, p. 2251-2262, 2010.
- BAUER, Daniel J.. Estimating Multilevel Linear Models as Structural Equation Models, *Journal of Educational and Behavioral Statistics*, [S.l.], n. 28, 135-167, 2003.
- BOLLEN, Kenneth. Total, Direct and Indirect Effects in Structural Equation Models, *Sociological Methodology*, [S.l.], n. 17, 37-69, 1987.
- _____. *Structural equations with latent variables*. New York: Wiley, 1989.
- CODES, Ana Luiza Machado de. Modelagem de equações estruturais: um método para a análise de fenômenos complexos. *Caderno CRH*, Salvador, v. 18, n. 45, p. 471-484, set./dez. 2005.
- FARIAS, Salomão A.; SANTOS, Rubens da C. Modelagem de Equações Estruturais e Satisfação do Consumidor: uma Investigação Teórica e Prática. *RAC*, v. 4, n. 3, 107-132, 2000.
- FERGUSON, David M. *Annotation: structural equation models in developmental research*. Great Britain: Cambridge University Press, 1997.
- FERREIRA, H. S. et al. Saúde de populações marginalizadas: desnutrição, anemia e enteroparasitoses em crianças de uma favela do "Movimento dos Sem Teto", Maceió, Alagoas. *Revista Brasileira de Saúde Materno Infantil*, Recife, n. 2, 177-185, 2002.
- FOX, John. Structural Equation Modeling with the SEM package in R. *Structural Equation Modeling: A Multidisciplinary Journal*. v.13, n. 3, 465-486, 2006.**
- HAIR Jr., JOSEPH F.; Anderson, ROLPH E.; TATHAN, Ronald L.; BLACK, William C. *Análise multivariada de dados*. Tradução Adonai Schlup Sant'Anna e Anselmo Chaves Neto. 5. ed. Porto Alegre: Bookman, 2005.
- JACKSON, J. A.; DEZEE, K.; DOUGLAS, K.; SHIMEALL, W. *Introduction to Structural Equation Modeling (Path Analysis)*. [S.l.]: SGIM Precourse PA08, 2005.
- KAPLAN, David. *Structural equation modeling. Foundations and extensions*. USA: Sage Publications, 2000. Advanced Quantitative Techniques in the Social Sciences Series, v. 10.
- KLINE, Rex B. *Principles and practice of structural equation modeling*. 2. ed. New York: The Guilford Press, 2004.
- MATOS, Sheila M. A. *Prevalência de enteroparasitoses e sua relação com o estado antropométrico na infância, Salvador-BA*. Salvador: ISC/UFBA, 2006. Dissertação de mestrado em Saúde Coletiva. Instituto de Saúde Coletiva. Universidade Federal da Bahia.
- MUELLER, R. O. *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York: Springer-Verlag, 1996.
- MUTHÉN, B. O. A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. *Psychometrika*, [S.l.], n. 49, 115-132, 1984.
- SANTOS, L. M. dos et al. Determinantes do desenvolvimento cognitivo na primeira infância: análise hierarquizada de um estudo longitudinal. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 24, n. 2, 427-437, fev. 2008.

SCHURMACKER, R. E.; LOMAX, R. G. *A Beginner's Guide to Structural Equation Modeling*. 2nd ed. New Jersey: Lawrence Erlbaum Associates, 2004.

SOARES, F. *Modelo de equações estruturais*. Rio de Janeiro: PUC, 2008. Certificação digital no 0321214/CA. Disponível em: <http://www.maxwell.lambda.ele.puc-rio.br/7630/7630_7.PDF>. Acesso em 27 de julho de 2012.

