

MODELO DE EQUAÇÕES ESTRUTURAIS: UMA INTRODUÇÃO APLICADA

Jorge Alexandre
Barbosa Neves

COLEÇÃO

Metodologias
de Pesquisa

Modelo de equações estruturais: uma introdução aplicada

Enap Escola Nacional de Administração Pública

Presidente

Francisco Gaetani

Diretora de Formação Profissional e Especialização

Iara Cristina da Silva Alves

Diretor de Educação Continuada

Paulo Marques

Diretor de Inovação e Gestão do Conhecimento

Guilherme Alberto Almeida de Almeida

Diretor de Pesquisa e Pós-Graduação Stricto Senso

Fernando de Barros Filgueiras

Diretora de Gestão Interna

Camile Sahb Mesquita

Editor: Flávio Cireno Fernandes (Enap). *Revisão:* Luiz Augusto Barros de Matos e Renata Fernandes Mourão. *Projeto gráfico e editoração eletrônica:* Ana Carla Gualberto Cardoso.

Modelo de equações estruturais: uma introdução aplicada

Jorge Alexandre Barbosa Neves

Brasília
Enap
2018

© 2018 Enap

Ficha Catalográfica por: Daiane da Silva Yung Valadares – CRB1/2802

N518m Neves, Jorge Alexandre Barbosa.
Modelo de equações estruturais: uma introdução aplicada./
Jorge Alexandre Barbosa Neves. – Brasília: Enap, 2018.
81 p. : il. –

Inclui bibliografia.
ISBN: 978-85-256-0089-9

1. Método de Pesquisa - Estatística. 2. Pesquisa Social.
I. Título.

CDU 311.2

Catalogado na fonte pela Biblioteca Graciliano Ramos da Enap

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista da Escola Nacional de Administração Pública (Enap). É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

Enap Fundação Escola Nacional de Administração Pública
SAIS – Área 2-A
70610-900 – Brasília, DF
Telefones: (61) 2020 3096 / 2020 3102 – Fax: (61) 2020 3178
Sítio: www.enap.gov.br
Tiragem: 500 exemplares

SUMÁRIO

1 Introdução	7
1.1 Os modelos de equações estruturais	7
1.2 Conceitos fundamentais.....	14
2 Fundamentos estatísticos e substantivos	17
3 Modelos contendo apenas variáveis observadas	23
3.1 Primeiro exemplo: o modelo de realização de <i>status</i> socioeconômico	23
3.1.2 Testes de hipóteses	32
3.1.3 Análise da qualidade do ajuste	35
3.2 Segundo exemplo: a estimação do IGD-M	40
3.2.1 Testes de hipóteses	45
3.2.3 Análise da qualidade do ajuste	46
3.3 Comentários finais do capítulo.....	48
4 Modelos contendo apenas a análise de mensuração	49
4.1 Primeiro exemplo: a mensuração da origem socioeconômica	50
4.2 Segundo exemplo: a mensuração do destino socioeconômico.....	54
4.3 Terceiro exemplo: mensurando a dependência do município em relação ao Programa Bolsa Família.....	57
4.4 Comentários finais do capítulo.....	61

5 Modelos contendo variáveis observadas e construtos latentes	63
5.1 Primeiro exemplo: o modelo de realização de <i>status</i> socioeconômico contendo a mensuração de um construto	63
5.2 Segundo exemplo: modelo de determinação do IGD-M	71
5.3 Comentários finais do capítulo.....	78
6 Conclusão	79
Referências bibliográficas	80

1 INTRODUÇÃO

1.1 Os modelos de equações estruturais

A modelagem de equações estruturais, ou MEE, é uma técnica de modelagem estatística multivariada de caráter geral, que é amplamente utilizada nas Ciências Humanas e Sociais. Pode ser vista como uma combinação de análise fatorial¹ e regressão (ou a ampliação dessas para a análise de trajetórias ou caminhos²). O interesse de muitos pesquisadores e outros profissionais em MEE deriva, muitas vezes, das construções teóricas que podem ser desenvolvidas a partir dos construtos latentes. As relações entre as construções teóricas são representadas por coeficientes de regressão ou coeficientes de trajetória entre variáveis observadas e/ou latentes. O modelo de equações estruturais implica uma estrutura para as covariâncias entre as variáveis observadas.

A modelagem de equações estruturais fornece uma estrutura muito geral e conveniente para análises estatísticas que incluem vários procedimentos multivariados tradicionais, em particular, análise fatorial, análise de regressão, análise discriminante e correlação canônica, como casos especiais. Os modelos de equações estruturais são, na maioria das vezes, visualizados por um diagrama de trajetórias. O modelo estatístico geralmente pode ser representado em um conjunto de equações matriciais. No início da introdução dessa técnica nas pesquisas sociais e comportamentais, os *softwares* normalmente demandavam configurações que especificassem o modelo em termos dessas matrizes. Assim, os

¹ Análise fatorial é uma técnica estatística multivariada (não determinística) que permite a mensuração de variáveis latentes (construtos não observados de forma direta) a partir de um conjunto de variáveis manifestas (observadas diretamente).

² A análise de trajetórias ou de caminhos é uma extensão da análise de regressão linear de mínimos quadrados. Essa extensão permite a decomposição de efeitos estatísticos entre: efeito direto e efeitos indiretos.

pesquisadores tinham que destilar a representação da matriz a partir do diagrama de trajetórias e fornecer aos *softwares* uma série de matrizes para os diferentes conjuntos de parâmetros, como cargas fatoriais e coeficientes de regressão. Isso não se faz mais necessário hoje, pois os modelos podem ser definidos graficamente, a partir de um conjunto extremamente simples de comandos. Atualmente, há alguns *softwares* que permitem a construção de modelos de equações estruturais a partir do desenho de diagramas. Entre esses *softwares* está o STATA³, que foi utilizado para o desenvolvimento dos modelos que serão apresentados neste livro.

Modelos de equações estruturais são, portanto, particularmente relevantes pelas seguintes vantagens: a) permitem que se trabalhe simultaneamente com estimação e mensuração; b) permitem que sejam estimados efeitos diretos e indiretos de variáveis explicativas sobre variáveis respostas; c) são bastante robustos, em função do relaxamento de pressupostos, quando comparados, por exemplo, com o modelo de regressão de mínimos quadrados e; d) apresentam facilidade interpretativa advinda de suas interfaces gráficas. Em função dessas vantagens, os MEE conquistaram bastante espaço entre pesquisadores e profissionais das áreas de ciências humanas e sociais, em particular nas análises psicométricas, mas não apenas.

A modelagem de equações estruturais é, fundamentalmente, uma técnica de análise estatística confirmatória. Ou seja, ela não se presta à exploração de dados. Todavia, ela pode ser utilizada, em particular, de três diferentes formas:

- I. Abordagem estritamente confirmatória (AEC), na qual se testa um modelo teórico previamente especificado, concluindo-se por sua aceitação ou refutação.
- II. Abordagem de modelos alternativos (AMA), na qual se faz uma análise comparativa da qualidade de ajuste de dois ou mais modelos teóricos previamente especificados.

³ Mais especificamente, foi utilizada a versão 12.1 do STATA.

III. Abordagem de desenvolvimento de modelos (ADM), na qual há um primeiro passo semelhante à AEC, porém, no caso da refutação do modelo especificado, se parte para a busca de um modelo com melhor qualidade de ajuste, em geral mais parcimonioso do que o modelo original.

Como dito acima, os MEE combinam análise de regressão com a análise fatorial. A análise de regressão é a técnica de análise multivariada mais utilizada. Ela é uma técnica de análise determinística, na qual se busca observar as covariâncias (ou “efeitos”) de uma ou mais variáveis independentes sobre uma variável dependente.

A análise de regressão é uma metodologia estatística multivariada que mede a mudança média que ocorre em uma variável dependente que está associada às mudanças ocorridas em uma ou mais variáveis independentes. Na regressão simples, estima-se a relação entre a variável dependente e uma única variável independente, ao passo que na análise de regressão múltipla, a variável dependente associa-se com mais de uma variável independente.

O modelo de análise de regressão pode ser expresso a partir de estruturas matriciais. Assim, na forma de matrizes, um sistema de equações de regressão múltipla pode ser assim apresentado:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \cdot \\ \cdot \\ \beta_{n1} \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \end{bmatrix}$$

Isto quer dizer que as dimensões das matrizes e dos vetores envolvidos são as seguintes:

Y → (**n x 1**), vetor coluna da observação da variável dependente;

X → (**n x k**), matriz das observações dadas “n” e de “k-1” variáveis X2 até Xk, também conhecida como matriz dos dados;

$\beta \rightarrow (k \times 1)$, vetor coluna dos parâmetros desconhecidos $\beta_1, \beta_2, \beta_3, \dots, \beta_k$;

$e \rightarrow (n \times 1)$, vetor coluna de n observações.

De maneira que a representação matricial geral de k variável é dada por

$$Y = X\beta + e_i$$

As equações normais de mínimos quadrados podem ser apresentadas em notação matricial como está representado no Quadro 1.1 a seguir:

Quadro 1.1 – Modelo matricial de regressão múltipla, estatísticas e parâmetros

Equações / Fórmulas	Significado
$Y = X\beta^* + e$	Apresentação da forma matricial da regressão geral
$\beta^* = (X'X)^{-1} (X'Y)$	Estimativa dos parâmetros sob a forma matricial
$\text{Var-Cov} (\beta^*) = s^2 (X'X)^{-1}$	Matriz de variâncias e covariâncias
$e'e = (Y'Y) - b^*(XY)$	Erro da regressão
$R^2 = \frac{\beta^*(X'Y) - NY2}{Y'Y - NY2}$	Coefficiente de determinação múltipla sob a forma matricial
$F = \frac{(\beta^*(X'Y) - NY2) / (k - 1)}{(Y'Y - \beta^*(X'Y)) / (N - k)}$	Forma matricial de teste da existência da regressão linear geral, F com k-1 e n-k graus de liberdade na hipótese de $(\beta_2 = \beta_3 = \beta_4 = \dots = \beta_k = 0)$

Fonte: Gujarati (1995 p. 303).

Esse método matemático de definição do modelo de regressão é conhecido como método dos mínimos quadrados ordinários (MQO). Ele é assim denominado pois minimiza a soma dos quadrados dos erros, tornando o estimador eficiente. Ou seja, respeitados os pressupostos da

análise de regressão, uma estimação com base em MQO será eficiente, indicando que terá um nível mínimo de erro.

O modelo de regressão de k-variáveis nas unidades de origem pode ser resumido como se segue: $E(Y_i) = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{4i} + \dots + b_k X_{ki}$. Onde:

$E(Y_i)$ = valor esperado (ou esperança matemática) da variável dependente;

b_0 = intercepto (ou constante) independente dos fatores;

b_1 = coeficiente de regressão referente à primeira variável independente X_{1i} ;

b_2 = coeficiente de regressão referente à segunda variável independente X_{2i} ;

b_3 = coeficiente de regressão referente à terceira variável independente X_{3i} ;

b_4 = coeficiente de regressão referente à quarta variável independente X_{4i} ;

b_k = coeficiente de regressão referente à última variável independente X_{ki} ;

X_{1i} = primeira variável independente;

X_{2i} = segunda variável independente;

X_{3i} = terceira variável independente;

X_{4i} = quarta variável independente;

X_{ki} = última variável independente;

Note-se que na equação acima os coeficientes da equação de regressão aparecem representados por uma letra latina ("b") e não por uma letra grega ("β"). Isso indica que, nesse caso, tem-se uma equação amostral de regressão. Ou seja, quando os dados são amostrais, os coeficientes da equação de regressão são estimadores (representados por letra latina) adequados dos parâmetros populacionais (representados por letra grega). Para que os estimadores sejam adequados, eles precisam atender aos seguintes pressupostos fundamentais:

- a) Aleatoriedade, implicando que as variáveis incluídas na análise são aleatórias, ou seja, não sofrem qualquer tipo de restrição ou censura, em particular derivada de viés de seletividade amostral.
- b) Linearidade, implicando que o valor esperado da variável dependente é uma função linear das variáveis independentes.
- c) Não-tendenciosidade, implicando que o valor esperado do estimador é igual ao parâmetro populacional, ou seja, $E(b) = \beta$.
- d) Independência das observações, implicando que as observações são independentes entre si.
- e) Eficiência, implicando que a variância do estimador é mínima, ou seja, que nenhum outro método de estimação poderia prover estimações com menor variância.
- f) Homoscedasticidade, implicando que a variância dos erros é constante ao longo do contínuo da variável dependente.
- g) Independência dos erros, indicando que a covariância entre o erro e qualquer das variáveis independentes é nula, ou seja, $\sigma^2(X_i, e_i) = 0$.

Se uma estimação por um modelo de regressão de MQO satisfaz as propriedades acima, essa estimação pode ser chamada de **estimação de máxima verossimilhança**. A função de verossimilhança indica quão provável a amostra observada é como uma função de possíveis valores de parâmetro. Portanto, maximizar a função de verossimilhança determina os parâmetros que têm maior probabilidade de produzir os dados observados.

Os MEE, por sua vez, produzem estimações de máxima verossimilhança, mesmo quando alguns dos pressupostos dos modelos de MQO não são satisfeitos. Em particular, os MEE não precisam satisfazer

o pressuposto da independência dos erros. Estimacões com base em MEE que não satisfazem o pressuposto da independência dos erros são chamadas de modelos não-recursivos. Nesses casos, a existência de endogeneidade não é um problema. Quando os MEE satisfazem o pressuposto da independência dos erros, eles são chamados de modelos recursivos. Ao longo deste livro introdutório, os exemplos apresentados se basearão em casos de modelos recursivos.

Como dito acima, os MEE combinam a análise de regressão com a análise fatorial para, assim, buscar conjugar modelos determinísticos com modelos de mensuração. A análise fatorial confirmatória (AFC) é aplicada nos MEE e tem elementos em comum com a análise de componentes principais. Na AFC, a variação de cada variável é decomposta em duas partes, sendo uma comum e uma parte única. A primeira diz respeito à variação que é compartilhada com outras variáveis, ao passo que a segunda é específica de uma única variável. Portanto, uma diferença entre a ACP e a AFC diz respeito ao montante de variância analisada. Ao passo que a ACP leva em consideração a variação total presente no conjunto das variáveis utilizadas para a mensuração, a AFC só faz uso da variação comum que é partilhada por todas as variáveis (REIS, 1997). O elemento comum fundamental entre a AFC e a ACP é que as variáveis manifestas incluídas na análise podem ser transformadas em combinações lineares de um conjunto de fatores – ou construtos – hipotéticos ou latentes. As cargas fatoriais são responsáveis por relacionar a associação específica entre os fatores e as variáveis observadas diretamente (manifestas). Portanto, tanto no caso da AFC quanto da ACP, o primeiro passo é encontrar as cargas e a solução matemática para os fatores, que indicarão a associação entre as variáveis observadas e os fatores ou construtos mensurados, sendo as cargas fatoriais derivadas dos autovalores, que, por sua vez, se relacionam às variáveis observadas. Assim, uma carga fatorial é um coeficiente que indica o peso de cada variável observada para a mensuração do construto.

Os MEE fazem uso da AFC a partir de procedimentos de máxima verossimilhança. Assim, como tanto a estimação do modelo de regressão quanto a mensuração de construtos obtidos a partir da análise

fatorial confirmatória são feitas a partir de procedimentos de máxima verossimilhança, a integração dos dois métodos se faz possível. Ao contrário da ACP, a AFC não se presta a abordagens exploratórias. Isso é mais um elemento em comum desta com à análise de regressão, ou seja, ambas são estratégias confirmatórias de análise estatística. Assim, para a realização de mensurações a partir da AFC se faz necessário que se pense antecipadamente quais variáveis se espera que venham a convergir para formar um mesmo construto. Da mesma forma, antes de se realizar uma análise de regressão, se faz necessário ter, de antemão, hipóteses consistentes sobre a relação entre as variáveis independentes e a variável dependente que irão compor a equação.

1.2 Conceitos fundamentais

Finalmente, um elemento inicial importante é a sumarização de termos centrais às MEE. Segue, abaixo, uma lista de termos com uma curta descrição de cada um:

- **Análise confirmatória:** uso de técnica estatística multivariada para testar (ou confirmar ou refutar) um conjunto pré-estabelecido de relações. No caso dos MEE, a análise confirmatória é aplicada tanto na estimação (análise de regressão) quanto na mensuração (análise fatorial).
- **Análise de trajetórias:** conjunto de equações de regressão que permite estimar efeitos diretos e indiretos de variáveis independentes sobre variáveis dependentes.
- **Causalidade:** relação de causa e efeito entre variáveis, que pode ser concluída a partir da satisfação de pressupostos somados à consistência teórica da análise confirmatória proposta.
- **Coefficiente de determinação:** semelhante ao coeficiente de determinação dos modelos de regressão de MQO (R^2), indica a proporção da variância total (de todas as variáveis incluídas) explicada pelo modelo.

- Efeito direto: coeficiente de regressão padronizado ou não-padronizado.
- Efeito indireto: produto dos coeficientes de regressão (padronizados ou não-padronizados) de uma estrutura complexa de causalidade.
- Modelo padronizado: baseado na matriz de correlação.
- Modelo não-padronizado: baseado na matriz de covariância.
- Comunalidade: quantidade de variância que uma variável observada tem em comum com um construto.
- Confiabilidade: nível de consistência interna do conjunto de indicadores (variáveis observadas) na mensuração de um construto, podendo ser entendida, ainda, como o inverso do erro de mensuração (ou seja, confiabilidade = $1 - \text{erro de mensuração}$).
- Construto: conceito latente que não pode ser observado de forma direta ou medido sem erro, dependendo, para sua mensuração, da comunalidade entre duas ou mais variáveis observadas.
- Diagrama de trajetórias: representação gráfica da relação complexa (que inclui efeitos diretos e indiretos) entre um conjunto de variáveis observadas ou mensuradas.
- Erro de estimação: diferença entre os valores estimados de uma variável dependente (a partir de uma equação de regressão) e os valores observados.
- Erro de mensuração: diferença entre a descrição real e a descrição perfeita de um construto latente a partir das variáveis observadas, podendo ser entendido, ainda, como o inverso da confiabilidade (ou seja, erro de mensuração = $1 - \text{confiabilidade}$).
- Estatística da diferença entre coeficientes de qui-quadrado: diferença entre os qui-quadrados de dois modelos alternativos, sendo o grau de liberdade a diferença entre os graus de liberdade de cada um dos modelos alternativos ($X^2\Delta = X^2_1 - X^2_2$; $GL_\Delta = GL_1 - GL_2$), representa uma medida de qualidade do ajuste.

- Estimação de máxima verossimilhança: método de estimação utilizado nos MEE.
- Matriz de covariância: matriz contendo a variância e a covariância de todas as variáveis observadas do MEE.
- Modelo causal: conjunto de equações de regressão (equações estruturais) que formam as relações de determinação a partir de efeitos diretos e indiretos de variáveis independentes sobre variáveis dependentes.
- Modelo de mensuração: análise fatorial confirmatória da mensuração de cada construto do MEE.
- Modelo nulo: modelo hipotético no qual a relação entre as variáveis é nula.
- Qualidade do ajuste: medida que indica o quão bem um modelo especificado replica a matriz de covariância entre as variáveis observadas.
- Qui-quadrado: medida estatística da diferença entre modelos.
- Relação espúria: relação falsa ou enganosa entre duas variáveis que têm uma mesma causa.
- Resíduo (ou erro): diferença entre um valor real e um valor estimado.
- Variável endógena: variável observada ou latente que é, em algum momento, dependente de outras no MEE.
- Variável exógena: variável observada ou latente que nunca é dependente de outras no MEE.
- Variável latente: variável mensurada (construto) por análise fatorial confirmatória a partir de duas ou mais variáveis observadas.
- Variável observada: variável que pode ser mensurada sem erro (observada de forma direta).

2 FUNDAMENTOS ESTATÍSTICOS E SUBSTANTIVOS

O modelo de equações estruturais (MEE) tem se tornado cada vez mais importante entre os métodos estatísticos focados na análise de relações entre variáveis. O método tem sido aplicado nas relações que são observadas e estimadas, cujos modelos podem incluir dados observados ou latentes. Os dados primários são usados para estimar a variância e a covariância que explica a direção do modelo de equações estruturais. Os métodos estatísticos tradicionais – tais como análise de variância, análise de regressão múltipla e análise fatorial confirmatória – são os métodos básicos na integração e estimação na modelagem teórica das equações estruturais.

O desenvolvimento dos computadores e dos *softwares* proporcionaram a generalização, integração e extensão dos modelos do tipo análise de variância (Anova), análise de regressão múltipla e análise fatorial confirmatória, processando, simultaneamente, a modelagem de equações estruturais, que permite estimar, simultaneamente, as variáveis dependentes e independentes e as correlações entre elas em um único sistema de equações denominado estrutural. O MEE se consolidou, assim, como uma contribuição pela busca do aperfeiçoamento de análises causais, em particular nas ciências sociais e na Psicologia.

Se um pesquisador ou analista de dados busca um maior nível de segurança em suas análises, ele desenvolverá indicadores válidos e confiáveis dos seus conceitos teóricos substantivos. Para tanto, o pesquisador ou analista deve se basear em boa teoria substantiva e rigorosas definições operacionais de suas categorias conceituais. Isso é dito pois é importante ressaltar que análises estatísticas corretas não têm a capacidade de corrigir erros teóricos ou conceituais, como, por exemplo, a omissão de variáveis centrais ou um conceito incorreto orientando a mensuração de construtos. Se, do ponto de vista conceitual ou teórico, em termos substantivos, a análise proposta é correta, a aplicação de

modelagens estatísticas multivariadas, como o MEE, é de grande valia para a pesquisa científica ou a análise técnica de dados.

Portanto, quando alguém precisa desenvolver um MEE, o melhor a fazer é começar com um modelo no qual se tenha forte confiança teórica. Essa confiança deriva de fundamentos teóricos ou empíricos sobre as ligações entre as variáveis. Por seu caráter confirmatório, o MEE requer que se inicie a análise com um nível de confiança elevado no modelo a ser estimado. Portanto, o MEE não é adequado quando se tem pouco conhecimento ou confiança no modelo inicial, independentemente de se o resultado final é teórica ou empiricamente plausível.

Um excelente exemplo de modelo teórico inicial é o de realização de *status*, que será apresentado a partir do próximo capítulo. Esse modelo tem uma longa tradição nas ciências sociais em nível internacional. Nele, as condições da chamada **teoria estatística clássica da causalidade** são atendidas, quais sejam:

- a) Se uma variável X deve ser considerada uma causa de uma variável Y, deve haver associação estatística entre elas (condição da covariância).
- b) Se uma variável X deve ser considerada uma causa de uma variável Y, X deve preceder Y no tempo (condição da precedência temporal).
- c) Se uma variável X deve ser considerada uma causa de uma variável Y, outras possíveis causas de Y devem ser mensuradas e incluídas na análise (condição da eliminação de causas concorrentes).

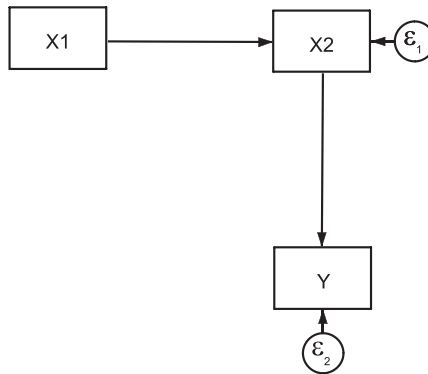
Em particular por causa da dificuldade em se obter certeza absoluta de que a terceira condição foi atendida, faz-se necessário que o modelo inicial a ser estimado seja teoricamente confiável⁴. Todavia, muitas vezes há limitações objetivas à mensuração de variáveis, em particular se o

⁴ Um nível realmente elevado de confiança no atendimento da terceira condição só é alcançado em pesquisas com dados experimentais. A utilização de variáveis instrumentais é uma segunda opção, porém com um poder inferior à análise experimental.

pesquisador ou analista faz uso de dados secundários. Portanto, com dados observacionais haverá sempre um nível relativamente elevado de incerteza no atendimento da terceira condição.

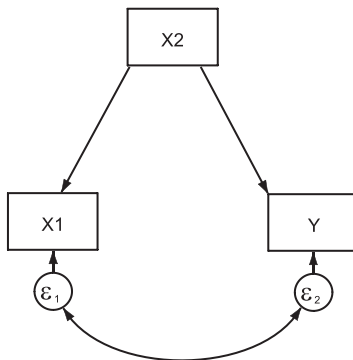
Neste livro, estudaremos apenas modelos recursivos. Ou seja, a relação entre variáveis se dará sempre em uma mesma direção. Todavia, as relações entre três ou mais variáveis poderão ser de três tipos. Imaginemos que se está estimando um modelo com apenas três variáveis (X_1 , X_2 e Y), todas observadas. As relações entre elas poderão assumir os seguintes formatos:

Figura 2.1 – Formato de efeito apenas indireto de X_1 sobre Y



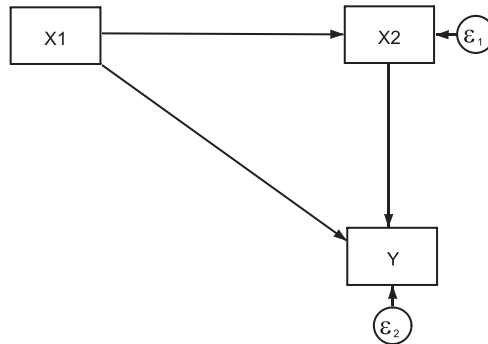
Fonte: Elaboração própria.

Figura 2.2 – Formato de relação espúria entre X_1 e Y



Fonte: Elaboração própria.

Figura 2.3 – Formato de efeitos direto e indireto de X1 sobre Y



Fonte: Elaboração própria.

Na Figura 2.1, observa-se uma situação na qual a relação entre a variável X1 e a variável Y é totalmente mediada pela variável X2. Nesse caso, X1 precede temporalmente X2, que, por sua vez, precede temporalmente Y. Portanto, todo possível efeito causal de X1 sobre Y se dá através de X2.

Na Figura 2.2, observa-se uma situação na qual a relação entre X1 e Y é espúria. Ou seja, toda a correlação entre X1 e Y se deve ao fato de que ambas as variáveis têm X2 como causa.

Na Figura 2.3, observa-se o caso mais interessante e comum de relação causal complexa entre três ou mais variáveis. Nesse caso, X1 precede temporalmente X2 que precede temporalmente Y, mas, diferentemente do que se observa na Figura 2.1, aqui o potencial efeito causal de X1 sobre Y é decomposto em um efeito direto e um efeito indireto.

Os MEE recursivos atendem a maior parte dos pressupostos dos modelos de regressão de mínimos quadrados⁵, quais sejam:

- a) Independência dos erros.
- b) Independência das observações.
- c) Linearidade.
- d) Normalidade dos erros.

⁵ Sobre os pressupostos da análise de regressão de mínimos quadrados, ver, entre outros, Gujarati (2006), além da discussão apresentada no capítulo introdutório deste livro.

- e) Homoscedasticidade.
- f) Aleatoriedade.

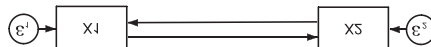
Todavia, de modo geral, os MEE permitem a estimação de modelos com elevados níveis de colineariedade entre variáveis explicativas. No caso de modelos não recursivos⁶, o pressuposto da independência dos erros pode ser relaxado.

Os MEE são uma integração da análise de regressão com a análise fatorial confirmatória. A análise fatorial usa a matriz de covariância para estimar o fator estrutural decorrente da análise do fator desenvolvida inicialmente para explicar a correlação entre as variáveis. O objetivo do uso da matriz de covariância (ou de correlação, no caso de modelos com coeficientes padronizados) é reduzir os tipos de variáveis padronizadas no modelo.

Os modelos de fatores são conhecidos como subidentificados, identificados ou sobreidentificados. Se o modelo é subidentificado, não se pode estimar por não ter uma solução única. Quando o modelo for identificado, existe uma solução exata para a matriz de covariância dos dados, implicando em um ajuste perfeito. Se o modelo for sobreidentificado, não ocorre o ajuste perfeito do modelo (SHARMA, 1996).

A análise fatorial pressupõe que as covariâncias entre um conjunto de variáveis observadas podem ser explicadas por um menor número de fatores latentes subjacentes. No modelo de fator exploratório, procede-se como se não houvesse hipóteses sobre o número de fatores latentes e as relações entre os fatores latentes e as variáveis observadas. Os procedimentos estatísticos são utilizados para estimar o número de fatores subjacentes e para estimar o fator de carga. Na análise fatorial exploratória, o modelo é arbitrário: todas as variáveis são carregadas em todos os fatores. Os MEE, por sua vez, fazem uso de modelos de análise fatorial confirmatória. Ou seja, na modelagem de equações estruturais,

⁶ Modelos não recursivos são aqueles nos quais a direção da causalidade não é única. Por exemplo, neste caso a relação entre duas variáveis pode ser recíproca, como na figura abaixo:



o modelo de fator confirmatório é imposto aos dados. Nesse caso, o propósito da modelagem de equação estrutural é duplo. Primeiro, pretende obter estimativas dos parâmetros do modelo, ou seja, as cargas fatoriais, as variâncias e covariâncias do fator e as variâncias de erro residual das variáveis observadas. O segundo objetivo é avaliar o ajuste do modelo, ou seja, avaliar se o próprio modelo fornece um ajuste adequado aos dados (HOX; BECHGER, 1998).

Os MEE completos, portanto, são compostos da integração de modelos de regressão com análises fatoriais confirmatórias. Os modelos de regressão fornecem a parte determinística dos MEE (o modelo causal), ao passo que as análises fatoriais confirmatórias fornecem a parte referente às mensurações. No próximo capítulo, será demonstrada e discutida a estimação de MEE compostos apenas pela parte determinística. No quarto capítulo, serão demonstrados e discutidos os MEE voltados apenas à mensuração de variáveis latentes. Finalmente, no quinto capítulo, serão demonstrados e discutidos MEE completos, ou seja, contendo a parte determinística e a parte referente às mensurações. Ao final, há uma breve conclusão do livro.

3 MODELOS CONTENDO APENAS VARIÁVEIS OBSERVADAS

Os modelos de equações estruturais que incluem apenas variáveis observadas são expansões das análises de regressão. Mais especificamente, eles são um conjunto de equações de regressão, às vezes chamado de sistema de equações estruturais.

3.1 Primeiro exemplo: o modelo de realização de *status* socioeconômico

Vamos, inicialmente, desenvolver um modelo com base na teoria de estratificação social, conhecido como modelo de realização de *status* socioeconômico. Esse modelo se estabeleceu a partir do trabalho seminal de Blau e Duncan (1967). No presente capítulo, será estimado um modelo semelhante àquele utilizado por Blau e Duncan (1967), a partir da análise de uma subamostra de homens de 30 a 50 anos da Pesquisa Nacional por Amostra de Domicílios de 2014 (PNAD-2014)⁷. Depois de 18 anos, o IBGE realizou uma nova PNAD que possibilita a estimação dos modelos de realização de *status* socioeconômico e das análises de mobilidade social⁸.

As variáveis incluídas no modelo foram:

- *escmãe*: anos de escolaridade completados com sucesso pela mãe do indivíduo incluído na amostra (escala discreta entre 0 e 16 pontos);
- *escpai*: anos de escolaridade completados com sucesso pelo pai do indivíduo incluído na amostra (escala discreta entre 0 e 16 pontos);

⁷ Esses dados foram preparados por Neves e Lima (2017).

⁸ As PNADs de 1973, 1976, 1982, 1988, 1996 e 2014 contêm dados que permitem tais tipos de análise.

- iseopai: índice socioeconômico da ocupação⁹ do pai do indivíduo incluído na amostra (escala contínua entre 0 e 100 pontos);
- esco: anos de escolaridade completados com sucesso pelo próprio indivíduo incluído na amostra (escala discreta entre 0 e 16 pontos);
- iseoi: índice socioeconômico da ocupação do próprio indivíduo incluído na amostra (escala contínua entre 0 e 100 pontos).

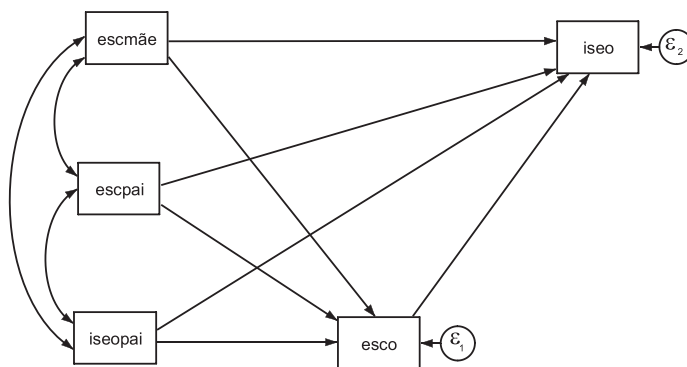
O sistema de equações para a realização da análise é:

$$\text{esco}_i = \beta_0 + \beta_1(\text{escmãe})_i + \beta_2(\text{escpai})_i + \beta_3(\text{iseopai})_i + \varepsilon_{1i} \quad (3.1)$$

$$\text{iseoi}_i = \beta_0 + \beta_4(\text{escmãe})_i + \beta_5(\text{escpai})_i + \beta_6(\text{iseopai})_i + \beta_7(\text{iseoi})_i + \varepsilon_{2i} \quad (3.2)$$

Esse sistema de equações gera o seguinte diagrama de equações estruturais, exposto na Figura 3.1:

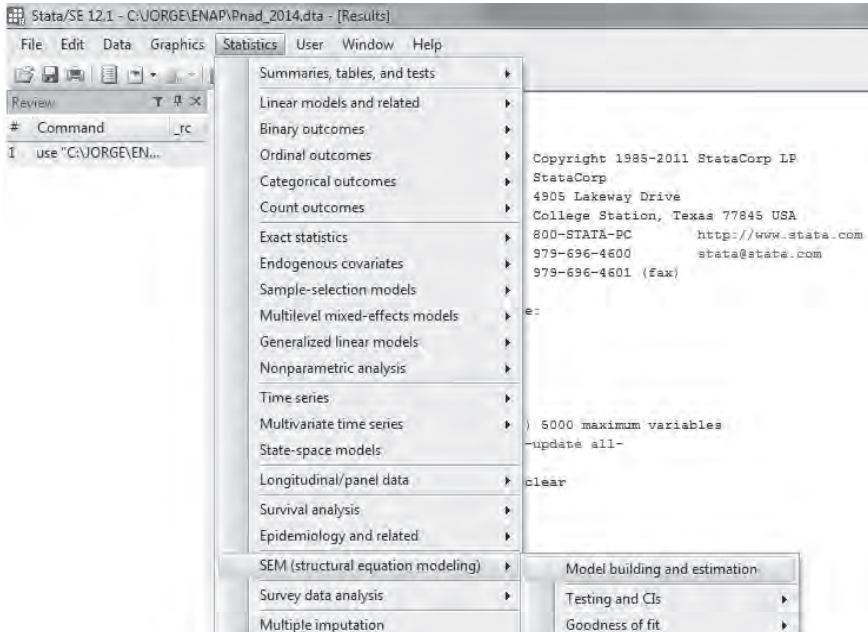
Figura 3.1 – Diagrama de equações estruturais do modelo de realização de *status* socioeconômico com dados da PNAD-2014



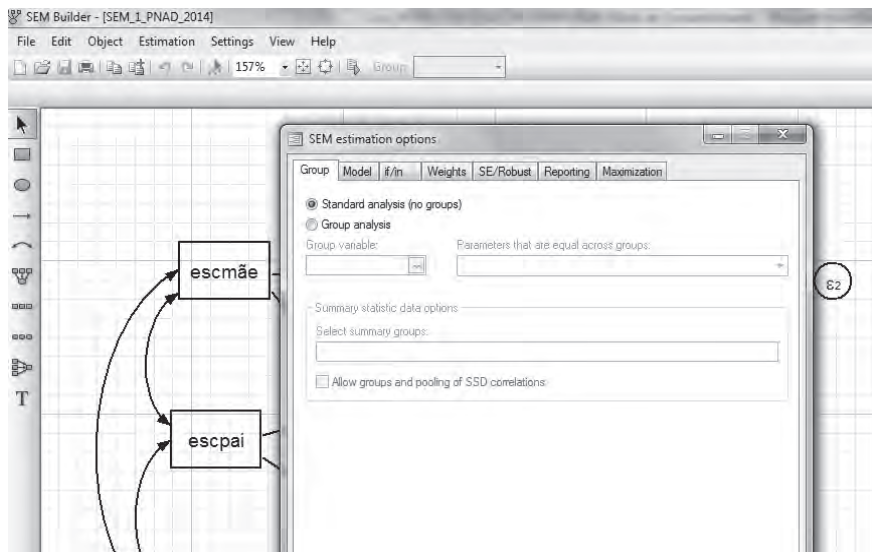
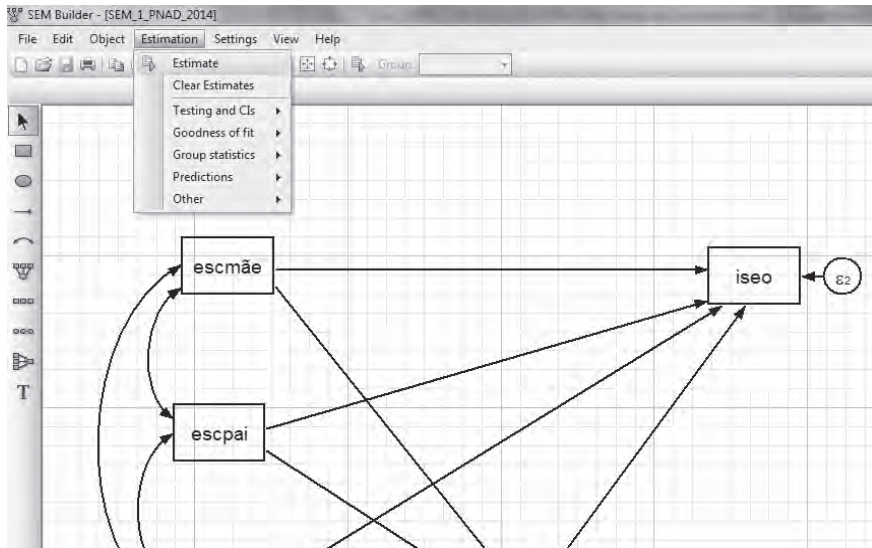
Fonte: Elaboração própria.

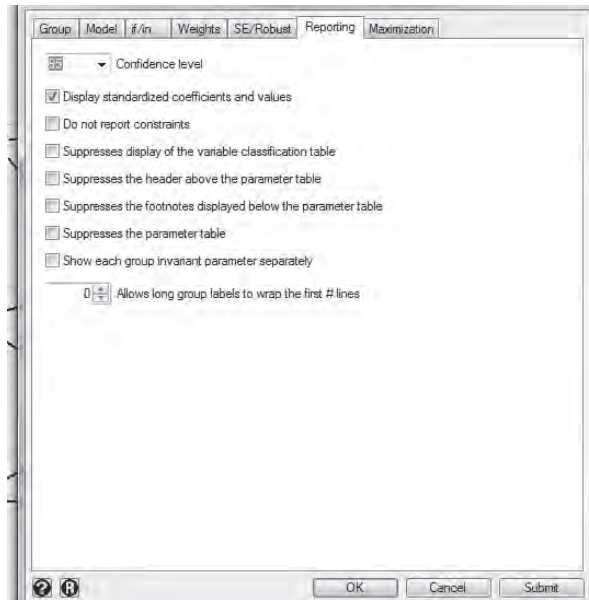
⁹ O índice socioeconômico da ocupação foi também criado por Blau e Duncan (1967). Atualmente, há um índice internacional (*International Socioeconomic Index of Occupational Status – ISEI*) desenvolvido por Ganzeboom e Treiman (1996), que foi utilizado nas análises deste livro. Nesse tipo de índice, o *status* socioeconômico das ocupações varia numa escala entre 0 e 100 pontos.

Para desenhar o diagrama, utiliza-se a seguinte sequência de comandos do STATA: *Statistics -> SEM (structural equation modeling) -> Model building and estimation*.



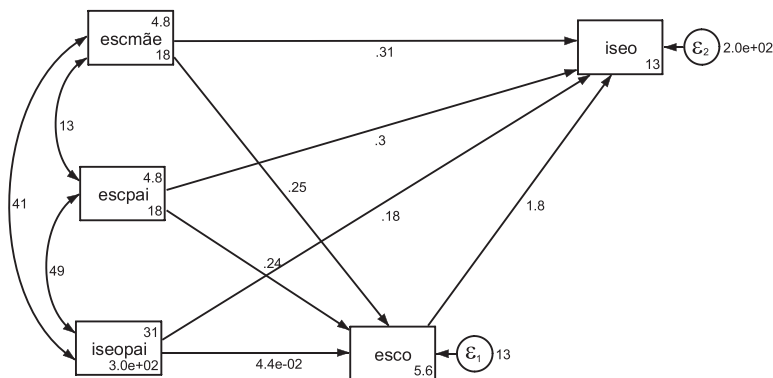
Dentro da janela para desenvolvimento do MEE, desenha-se o diagrama com o uso dos ícones do lado esquerdo da janela e, depois, parte-se para a estimação dos coeficientes com o seguinte conjunto de comandos: *Estimation -> Estimate -> Reporting -> Display standardized coefficients and values -> Ok*. A sintaxe gerada pelo STATA é: *sem (escmãe -> esco) (escmãe -> iseo) (escpai -> esco) (escpai -> iseo) (iseopai -> esco) (iseopai -> iseo) (esco -> iseo), cov(escpai*esc) > mãe iseopai*escmãe iseopai*escpai nocapslatent*.





As estimações aparecerão sobre o próprio diagrama, na forma da Figura 3.2:

Figura 3.2 – Diagrama de equações estruturais do modelo de realização de *status* socioeconômico com dados da PNAD-2014, com a estimação dos coeficientes não padronizados

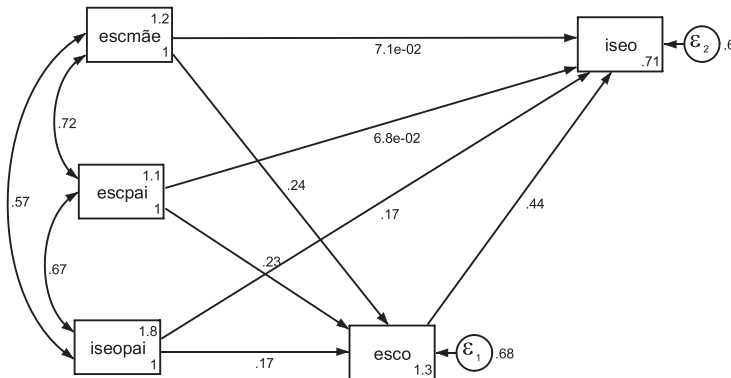


Fonte: Elaboração própria.

Em geral, é de interesse observar também os coeficientes padronizados. Para tanto, basta utilizar a seguinte sequência de comandos

na janela de MEE do STATA: View -> Standardized Estimates. O resultado será este, conforme a Figura 3.3:

Figura 3.3 – Diagrama de equações estruturais do modelo de realização de *status* socioeconômico com dados da PNAD-2014, com a estimação dos coeficientes padronizados



Fonte: Elaboração própria.

O modelo de realização de *status* levou à consolidação de toda uma área de pesquisa nas ciências sociais sobre o processo de estratificação social. Ele permitiu que se pudessem testar empiricamente as hipóteses que fundaram a análise de estratificação e mobilidade social nos tempos atuais.

A partir da década de 1930, houve um importante debate teórico entre dois importantes nomes da sociologia americana da Universidade Harvard: Talcott Parsons e Pitirim Sorokin. O primeiro, entre outras importantes contribuições, é um dos pais da chamada Teoria da Modernização nas Ciências Sociais, ao passo que o segundo é um dos pais da moderna Teoria da Reprodução Social.

Para Parsons, as sociedades modernas tenderiam a se tornar mais fluidas com o processo de modernização, indicando que o efeito de variáveis referentes à origem socioeconômica sobre a alocação ocupacional dos indivíduos tenderia a diminuir, ao passo que o efeito da educação tenderia a aumentar. Isso decorreria de uma “revolução educacional”, ou seja, de um processo de universalização do acesso à educação pública

que derivaria da modernização das sociedades. O fenômeno descrito por Parsons transformaria as sociedades no que depois passou a ser denominado de sistema meritocrático.

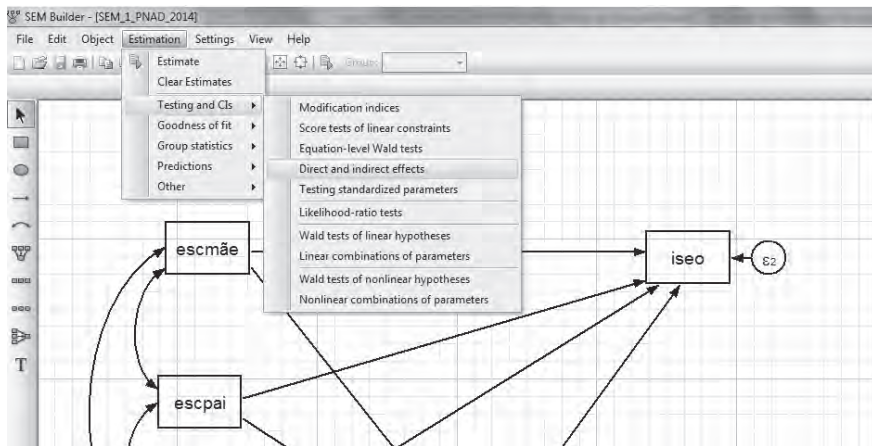
Sorokin, por sua vez, argumentou que sociedades tendem a substituir mecanismos de reprodução social antigos por mecanismos novos. Assim, para ele, a expansão dos sistemas educacionais públicos que decorre da modernização não levaria a uma sociedade mais fluida, mas apenas criaria um mecanismo diferente de reprodução social intergeracional.

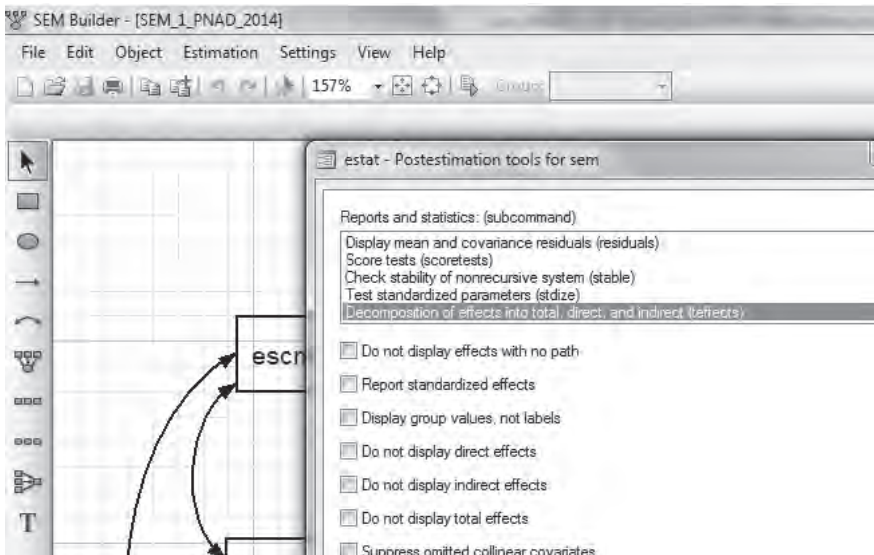
As pesquisas sociológicas desenvolvidas a partir do modelo alocação de *status* ocupacional referido acima têm trazido fortes evidências a favor da hipótese de Sorokin. O Modelo de alocação de *status* é o mais adequado para confrontar as hipóteses de Parsons e Sorokin, pois permite observar os efeitos diretos e indiretos de variáveis de origem socioeconômica (características socioeconômicas dos pais e das mães dos indivíduos) sobre o destino socioeconômico medido pelo índice socioeconômico da ocupação. O que se tem observado em nível mundial é que o processo de modernização leva a uma transformação do processo de transmissão intergeracional da desigualdade, fazendo com que o efeito de variáveis de origem socioeconômica sobre a alocação de *status* ocupacional se dê cada vez menos a partir de efeitos diretos e mais a partir de efeitos indiretos.

O modelo estimado acima permite a observação dos efeitos diretos e indiretos. Os coeficientes expostos no diagrama de trajetórias representam os efeitos diretos. Para se estimar os efeitos indiretos, é necessário multiplicar os efeitos diretos. A análise comparativa entre efeitos diretos e indiretos é muito melhor de ser entendida quando se estão utilizando coeficientes padronizados. Assim, por exemplo, o efeito direto padronizado da escolaridade da mãe (*escmãe*) sobre o índice socioeconômico da ocupação do entrevistado (*iseo*) é 0,071 ($7,1 * 10^{-2}$). Por sua vez, o efeito direto padronizado da escolaridade da mãe (*escmãe*) sobre a escolaridade do entrevistado é 0,24. Finalmente, o efeito padronizado direto da escolaridade do entrevistado (*esco*) sobre o índice socioeconômico da ocupação do entrevistado (*iseo*) é 0,44. Assim, o efeito

indireto padronizado da mãe (escmãe) sobre o índice socioeconômico da ocupação do entrevistado (iseo) é o produto dos dois efeitos diretos, ou seja, $0,24 * 0,44 = 0,106$. Por sua vez, o efeito total padronizado é a soma do efeito direto com o efeito indireto, ou seja, $0,071 + 0,106 = 0,177$. Observa-se, assim, que a maior parte do efeito da escolaridade da mãe sobre o nível socioeconômico da ocupação do entrevistado se dá de forma indireta (aproximadamente 60% do efeito total padronizado é indireto, ao passo que 40% é direto). O efeito total pode ser interpretado da seguinte forma: a elevação de um desvio padrão na escolaridade da mãe é acompanhada pela elevação de 0,177 desvio padrão no *status* socioeconômico da ocupação do entrevistado.

A decomposição em efeitos diretos, indiretos e totais pode ser obtida a partir da seguinte sequência de comandos: Estimation -> Testing and CIs -> Direct and indirect effects -> Decomposition of effects into total, direct, and indirect (teffects) -> Report standardized effects -> Ok.





Os resultados calculados pelo STATA estão reportados a seguir:

```
: estat teffects, standardized
```

Direct effects

	OIM					Std. Coef.
	Coef.	Std. Err.	z	P> z		
Structural						
escb <-						
escmae	.2522337	.0186668	13.51	0.000	.2425067	
escpai	.2358551	.0205711	11.47	0.000	.229248	
iseopai	.0439897	.0042686	10.31	0.000	.1738263	
isec <-						
escb	1.842101	.059412	31.01	0.000	.437889	
escmae	.3108105	.0756305	4.11	0.000	.0710341	
escpai	.2961333	.0828775	3.57	0.000	.0684224	
iseopai	.178752	.0171502	10.42	0.000	.1679057	

Indirect effects

	OIM					Std. Coef.
	Coef.	Std. Err.	z	P> z		
Structural						
escb <-						
escmae	0	(no path)			0	
escpai	0	(no path)			0	
iseopai	0	(no path)			0	

Total effects						
	OIM					
	Coef.	Std. Err.	z	P> z	Std. Coef.	
Structural						
esco <-						
escmãe	.2522337	.0186688	13.51	0.000	.2425067	
escpai	.2358551	.0205711	11.47	0.000	.229248	
iseopai	.0439897	.0042686	10.31	0.000	.1738263	
iseo <-						
esco	1.842101	.059412	31.01	0.000	.437889	
escmãe	.7754506	.0817194	9.43	0.000	.1772251	
escpai	.7306024	.090046	8.11	0.000	.1688076	
iseopai	.2597855	.018685	13.90	0.000	.2440224	

A última coluna à direita apresenta os coeficientes padronizados. Pode-se observar que os efeitos direto, indireto e total padronizados da relação entre escmãe e iseo são idênticos aos reportados previamente.

3.1.2 Testes de hipóteses

O modelo proposto gera sete estimadores. Assim, há a necessidade de realizar o mesmo número de testes de hipóteses, nos quais as hipóteses nulas serão sempre de que os parâmetros populacionais são iguais a zero ($\beta = 0$), assumindo-se um teste bilateral. Os resultados (gerados a partir da sequência de comandos reportada no início deste capítulo) calculados pelo STATA estão reportados a seguir:

```
. sem (escmãe -> esco) (escmãe -> iseo) (escpai -> esco) (escpai -> iseo) (iseopai -> esco) (iseopai -> iseo) (esco -> iseo), cov(escpai*esc  
> mãe iseopai*escmãe iseopai*escpai) nocaplatent
```

Endogenous variables

Observed: esco iseo

Exogenous variables

Observed: escmãe escpai iseopai

Fitting target model:

```
Iteration 0: log likelihood = -71837.512  
Iteration 1: log likelihood = -71837.512
```

```
Structural equation model          Number of obs   =    4467  
Estimation method = ml  
Log likelihood      = -71837.512
```

	OIM				[95% Conf. Interval]	
	Coef.	Std. Err.	z	P> z		
Structural						
esco <-						
escmãe	.2522337	.0186688	13.51	0.000	.2156435	.288824
escpai	.2358551	.0205711	11.47	0.000	.1955366	.2761737
iseopai	.0439897	.0042686	10.31	0.000	.0356234	.052356
_cons	5.563614	.1109121	50.16	0.000	5.34623	5.780398
iseo <-						
esco	1.842101	.059412	31.01	0.000	1.725656	1.958547
escmãe	.3108105	.0756305	4.11	0.000	.1625775	.4590435
escpai	.2961333	.0828775	3.57	0.000	.1336964	.4585702
iseopai	.178752	.0171502	10.42	0.000	.1451381	.2123658
_cons	13.04828	.5506584	23.70	0.000	11.96901	14.12755
Mean						
escmãe	4.830535	.0628219	76.89	0.000	4.707406	4.953664
escpai	4.787105	.0635113	75.37	0.000	4.662626	4.911585
iseopai	30.74323	.2581993	119.07	0.000	30.23717	31.24929

Variance						
e.esco	12.91255	.2732241			12.38799	13.45932
e.iseo	203.5993	4.308076			195.3283	212.2205
escmãe	17.62944	.3730316			16.91327	18.37594
escpai	18.01848	.3812634			17.2865	18.78145
iseopai	297.8009	6.301341			285.7031	310.4109
Covariance						
escmãe						
escpai	12.81505	.3284445	39.02	0.000	12.17131	13.45879
iseopai	41.07021	1.246156	32.96	0.000	38.62779	43.51263
escpai						
iseopai	49.30665	1.321167	37.32	0.000	46.71721	51.89609

LR test of model vs. saturated: $\chi^2(0) = 0.00$, Prob > $\chi^2 = .$

```

sem (escmãe -> esco) (escmãe -> isec) (escpai -> esco) (escpai -> isec) (isecpai -> esco) (isecpai -> isec) (esco -> isec), standardized co
> v( escpai*escmãe isecpai*escpai) mocallatent

Endogenous variables:
Observed: esco isec

Exogenous variables:
Observed: escmãe escpai isecpai

Fitting target model:
Iteration 0: Log likelihood = -71837.512
Iteration 1: Log likelihood = -71837.512

Structural equation model. Number of obs = 4467
Estimation method = ml
Log likelihood = -71837.512
    
```

Standardized	OIM				[95% Conf. Interval]	
	Coeff.	Std. Err.	z	P> z		
Structural						
esco <-						
escmãe	.2425067	.0177238	13.68	0.000	.2077687	.2772447
escpai	.229248	.01981	11.57	0.000	.1904211	.268075
isecpai	.1738263	.0167526	10.38	0.000	.1409918	.2066608
_cons	1.273967	.0339682	37.50	0.000	1.207391	1.340544
isec <-						
esco	.437889	.0132579	33.03	0.000	.4113039	.4638741
escmãe	-.0710341	.017272	4.11	0.000	-.0371817	-.1048865
escpai	.0684224	.0191376	3.58	0.000	.0309133	.1059314
isecpai	.1679057	.0160278	10.48	0.000	.1364918	.1993197
_cons	-.7102396	.0349412	-20.33	0.000	-.6417561	-.7787231
Mean						
escmãe	1.150472	.0192877	59.65	0.000	1.112669	1.186275
escpai	1.127753	.0191369	58.93	0.000	1.090245	1.165261
isecpai	1.781503	.0240647	74.03	0.000	1.734337	1.828669

Variance						
e.esco	.6770403	.0115136			.654846	.6999868
e.isec	.6032256	.0113703			.5813468	.6259279
escmãe	1	.			.	.
escpai	1	.			.	.
isecpai	1	.			.	.

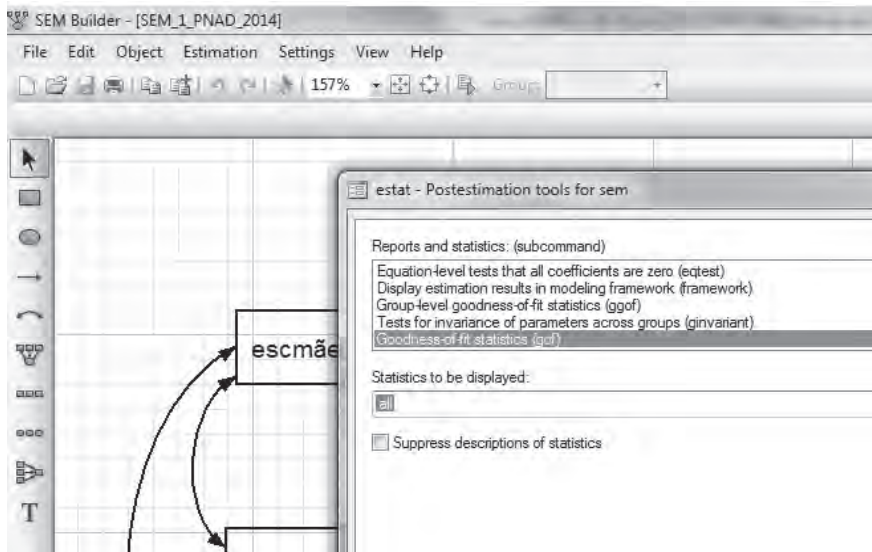
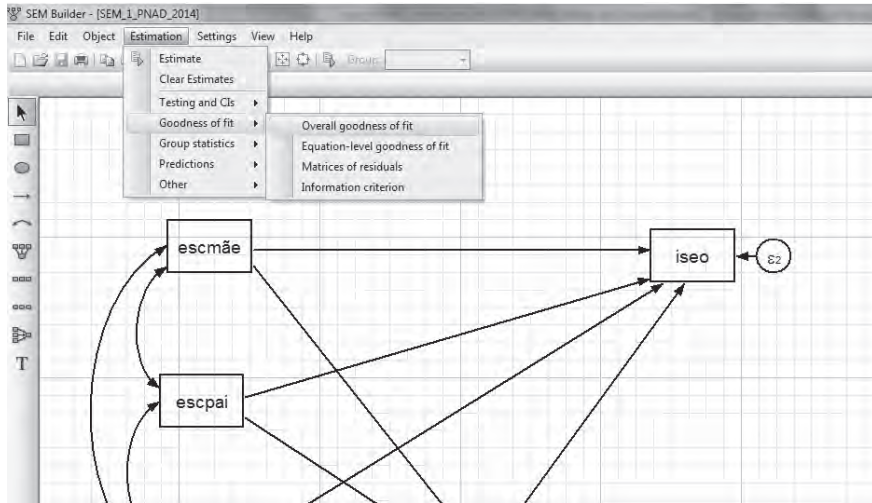
Covariance						
escmãe						
escpai	.7190218	.0072268	99.49	0.000	.7048575	.7331861
isecpai	.5668193	.010155	55.82	0.000	.5469159	.5867228
escpai						
isecpai	.673106	.0081832	82.25	0.000	.6570673	.6891448

LR test of model vs. saturated: chi2(0) = 0.00, Prob > chi2 =

Os resultados do STATA reportados mostram que, para todos os testes de hipóteses, é possível a rejeição da hipótese nula, visto que todos apresentam p-valores $< 0,001$. Obviamente, como se está trabalhando com uma amostra bastante grande ($n = 4467$), a rejeição da hipótese nula tende a ser bastante facilitada. Assim, é cada vez mais consensual que se deve aplicar uma lógica típica da estatística bayesiana e se buscar analisar se os coeficientes são não apenas significantes, mas também substantivamente relevantes. Essa análise exige, do pesquisador ou de qualquer outro profissional que esteja desenvolvendo os procedimentos analíticos, um conhecimento aprofundado do que está sendo analisado para poder ter uma boa ideia da relevância dos resultados encontrados. Para tanto, é recomendável que se observem os resultados com base nos coeficientes padronizados. Esses resultados indicam que dois trajetos têm coeficientes padronizados bastante baixos e, talvez, possam ser omitidos do modelo. Esses trajetos são os efeitos diretos da escolaridade da mãe e da escolaridade do pai sobre o *status* socioeconômico da ocupação do entrevistado. Para se decidir se os referidos trajetos devem ser mantidos no modelo, faz-se necessária uma análise comparativa da qualidade de ajuste dos modelos com e sem o referido trajeto.

3.1.3 Análise da qualidade do ajuste

O modelo estimado acima é comumente denominado de modelo saturado, pois ele tem todos os trajetos possíveis. Para se obter as principais estatísticas referentes à qualidade de ajuste de um modelo, utiliza-se a seguinte sequência de comandos do STATA: Estimation -> Goodness of fit -> Overall goodness of fit -> Goodness of fit statistics (gof) -> All of above.



Os resultados encontrados estão reportados a seguir:

```

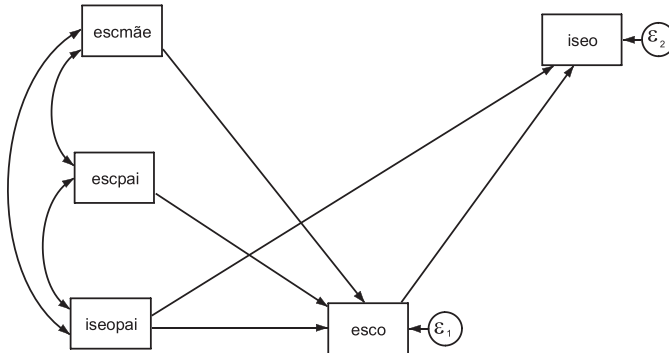
: estat gof, stats(all)

```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(0)	0.000	model vs. saturated
p > chi2	.	
chi2_bs(7)	4000.147	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.000	Root mean squared error of approximation
90% CI, lower bound	0.000	
upper bound	0.000	
pclose	1.000	Probability RMSEA <= 0.05
Information criteria		
AIC	143715.024	Akaike's information criterion
BIC	143843.114	Bayesian information criterion
Baseline comparison		
CFI	1.000	Comparative fit index
TLI	1.000	Tucker-Lewis index
Size of residuals		
SRMR	0.000	Standardized root mean squared residual
CD	0.376	Coefficient of determination

Agora, deve-se estimar um modelo sem os dois trajetos que têm coeficientes com valores baixos e que talvez sejam irrelevantes. Esse modelo tem o seguinte diagrama:

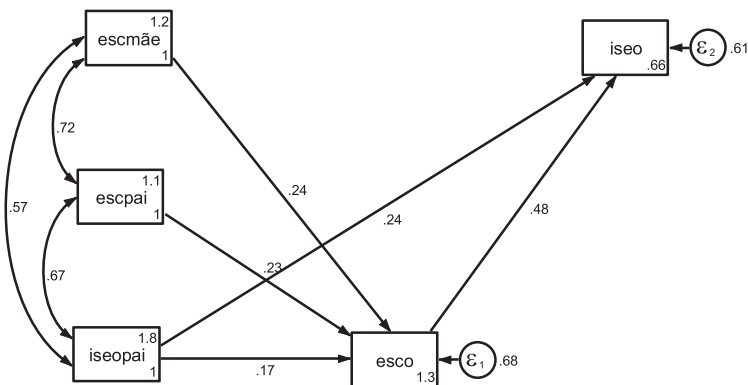
Figura 3.4 – Variação mais parcimoniosa do diagrama de equações estruturais do modelo de realização de *status* socioeconômico



Fonte: Elaboração própria.

Os resultados com coeficientes não-padronizados e padronizados desse modelo mais parcimonioso (ou seja, com um menor número de parâmetros para ser estimado) encontram-se a seguir:

Figura 3.5 – Resultados da variação mais parcimoniosa do diagrama de equações estruturais do modelo de realização de *status* socioeconômico com dados da PNAD-2014, com a estimação dos coeficientes padronizados



Fonte: Elaboração própria.

Para decidir sobre qual modelo devemos utilizar, vamos pedir as estatísticas de qualidade do ajustamento do segundo modelo. Essas estatísticas encontram-se a seguir:

```
. estat gof, stata(all)
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(2)	58.923	model vs. saturated
p > chi2	0.000	
chi2_bs(7)	4000.147	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.080	Root mean squared error of approximation
90% CI, lower bound	0.063	
upper bound	0.098	
pclose	0.002	Probability RMSEA <= 0.05
Information criteria		
AIC	143769.947	Akaike's information criterion
BIC	143885.228	Bayesian information criterion
Baseline comparison		
CFI	0.986	Comparative fit index
TLI	0.950	Tucker-Lewis index
Size of residuals		
SRMR	0.018	Standardized root mean squared residual
CD	0.368	Coefficient of determination

Os resultados das estatísticas de ajustamento dos dois modelos são bastante bons. Todavia, a parcimônia é um princípio importante da análise estatística. Assim, deve-se buscar identificar se a queda no poder explicativo ocorrida com o segundo modelo – que é mais parcimonioso – é pequena o suficiente para que se faça a opção por ele. Os resultados de ajuste dos dois modelos são bastante bons, indicando que ambos são adequados, pois têm estatísticas CFI, TLI e CD bastante próximas. Todavia, um teste mais objetivo (embora afetado pelo tamanho da amostra) pode ser realizado pela comparação das estatísticas de qui-quadrado das razões de verossimilhança dos dois modelos. O que se observa é que o segundo modelo, quando comparado ao modelo saturado, tem uma estatística qui-

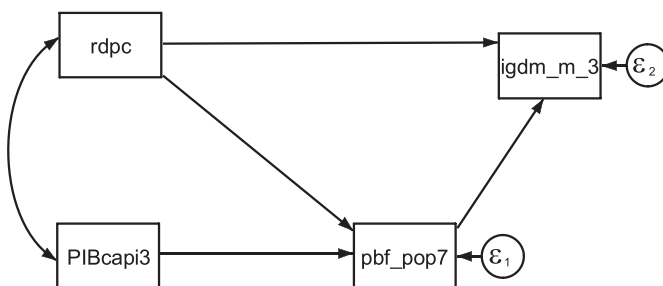
quadrado de 58,923, com 2 graus de liberdade. Visto que o valor crítico da estatística qui-quadrado com 2 graus de liberdade e nível de significância de 0,01 é 9,21, observa-se, assim, que, quando comparado ao modelo saturado (ou seja, o primeiro modelo), o segundo modelo não tem uma boa qualidade de ajuste (o ideal seria um qui-quadrado com um p-valor > 0,05, quando se tem um p-valor < 0,001). Conclui-se, portanto, que o modelo saturado (o primeiro modelo) deve ser o escolhido.

3.2 Segundo exemplo: a estimação do IGD-M

Em 2006, o então Ministério do Desenvolvimento Social e Combate à Fome (MDS) criou o índice de gestão descentralizada (IGD), para medir e acompanhar a gestão do Programa Bolsa Família e do Cadastro Único por parte dos entes federados. No caso dos municípios, o índice se chama IGD-M. Nesta parte do livro, vamos desenvolver um primeiro MEE para explicar a determinação do referido índice. Para tanto, são utilizados os dados referentes ao ano de 2014.

O modelo proposto é o seguinte:

Figura 3.6 – Diagrama de equações estruturais do modelo para explicação do IDH



Fonte: Elaboração própria

Onde:

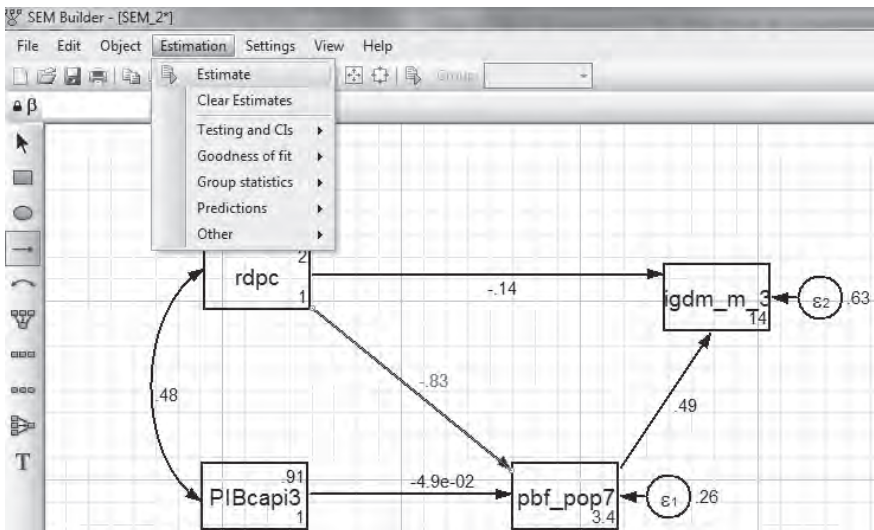
rdpc: diz respeito à renda *per capita* do município.

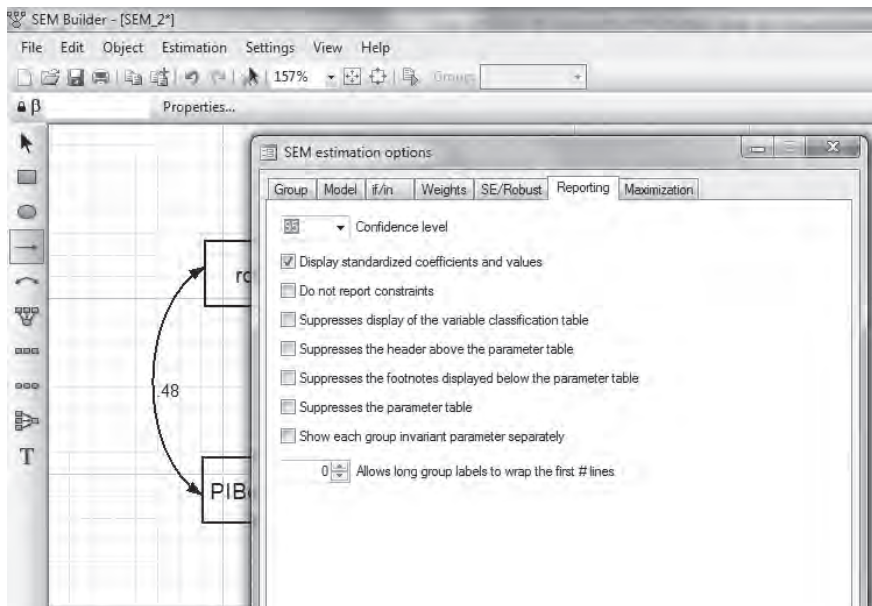
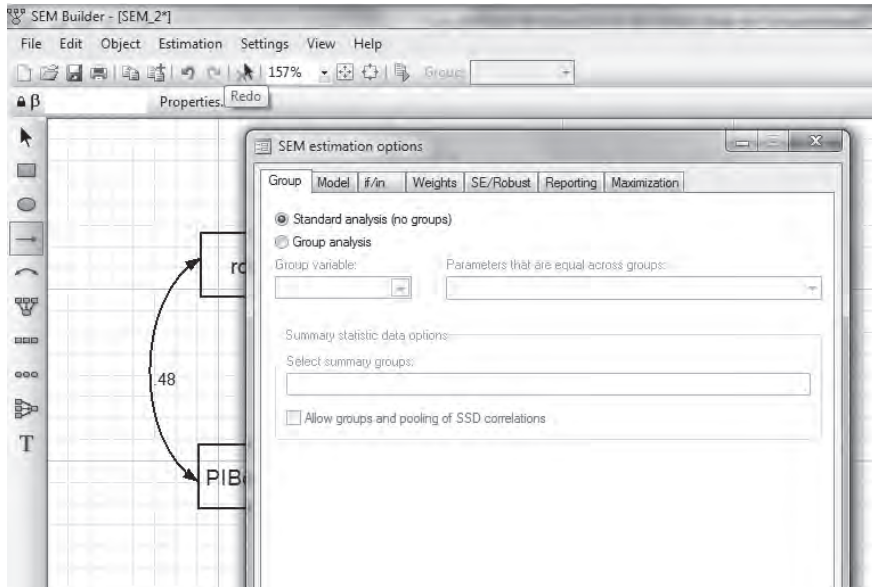
PIBcapi3: diz respeito ao PIB *per capita* do município.

pbf_pop7: é a proporção de beneficiários do Programa Bolsa Família em relação ao tamanho da população.

lgdm_m_3: diz respeito ao índice de gestão descentralizada municipal.

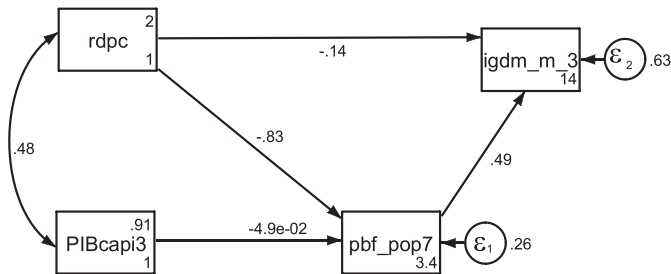
A estimação do modelo se dá a partir da seguinte sequência de comandos no STATA:





Após sua estimação, o modelo apresenta os seguintes resultados com coeficientes padronizados:

Figura 3.7 – Diagrama de equações estruturais com os resultados do modelo para explicação do IDH



Fonte: Elaboração própria.

Structural equation model		Number of obs		=		5565	
Estimation method		= ml					
Log likelihood		= -98514.333					
Standardized	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
Structural							
pbf_pop7 <-							
rdpc	-.8334903	.0053654	-155.34	0.000	-.8440064	-.8229743	
PIBcapi3	-.0494409	.0078453	-6.30	0.000	-.0648173	-.0340644	
_cons	3.40322	.0269021	126.50	0.000	3.350493	3.455947	
igdm_m_3 <-							
pbf_pop7	.4854643	.0199965	24.28	0.000	.4462718	.5246568	
rdpc	-.1390791	.0205991	-6.75	0.000	-.1794525	-.0987056	
_cons	13.50853	.1533944	88.06	0.000	13.20788	13.80917	
Mean							
rdpc	2.029232	.023446	86.55	0.000	1.983281	2.075184	
PIBcapi3	.9111131	.0169461	57.14	0.000	.8798593	.942367	
Variance							
e_pbf_pop7	.2633566	.00606			.2517432	.2755059	
e_igdm_m_3	.6292312	.0102721			.6094169	.6496897	
rdpc	1	-			-	-	
PIBcapi3	1	-			-	-	
Covariance							
rdpc							
PIBcapi3	.4791835	.010327	46.40	0.000	.4589429	.499424	
LR test of model vs. saturated:		chi2(1)		=		0.53, Prob > chi2 = 0.4680	

Indirect effects

	OIM				Std. Coef.
	Coef.	Std. Err.	z	P> z	
Structural					
pbf_pop7 <-					
rdpe	0	(no path)			0
PIBcapi3	0	(no path)			0
igdm_m_3 <-					
pbf_pop7	0	(no path)			0
rdpe	-.0100186	.0004364	-22.96	0.000	-.4046298
PIBcapi3	-.0093419	.0015333	-6.09	0.000	-.0240018

Total effects

	OIM				Std. Coef.
	Coef.	Std. Err.	z	P> z	
Structural					
pbf_pop7 <-					
rdpe	-.070715	.000665	-106.34	0.000	-.8334903
PIBcapi3	-.0659384	.0104529	-6.31	0.000	-.0494409
igdm_m_3 <-					
pbf_pop7	.1416756	.0060255	23.51	0.000	.4854643
rdpe	-.0134622	.0002799	-48.10	0.000	-.5437089
PIBcapi3	-.0093419	.0015333	-6.09	0.000	-.0240018

Os resultados reportados para o modelo da Figura 3.7, quando a variável resposta é a proporção de beneficiários do Programa Bolsa Família em relação à população total do município, são compatíveis com o que se poderia esperar. Mais especificamente, se observa que a renda *per capita* e o PIB dos municípios têm um efeito negativo sobre a proporção de beneficiários do Programa Bolsa Família em relação à população total do município. Obviamente, se espera que, em municípios mais ricos, ou seja, com maior renda e maior produção, se observe uma menor proporção de beneficiários do Programa Bolsa Família.

Por outro lado, quando a variável resposta é o IGD-M, os resultados talvez sejam surpreendentes para alguns. Observa-se que a renda *per capita* e o PIB dos municípios têm um efeito negativo sobre IGD-M, indicando que, em municípios mais ricos, a qualidade da política de assistência social tende a ser pior. Por sua vez, a variável referente à

proporção de beneficiários do Programa Bolsa Família em relação à população total do município apresenta um efeito positivo sobre o IGD-M. A estimação do modelo indica, portanto, que o IGD-M é bastante dependente da proporção de beneficiários do Programa Bolsa Família em relação à população total do município. Ou seja, o que impulsiona a elevação da qualidade da política de assistência social dos municípios é sua dependência dos recursos oriundos do Programa Bolsa Família.

Outro ponto que chama atenção nos resultados acima é o efeito indireto da variável referente à renda *per capita* sobre o IGD-M. Mais especificamente, observa-se que a maior parte do efeito da renda *per capita* sobre o IGD-M se dá através da proporção de beneficiários do Programa Bolsa Família em relação à população total do município. Ou seja, uma maior renda *per capita* tem um efeito negativo forte sobre o IGD-M, ao reduzir a dependência do município dos recursos do Programa Bolsa Família.

3.2.1 Testes de hipóteses

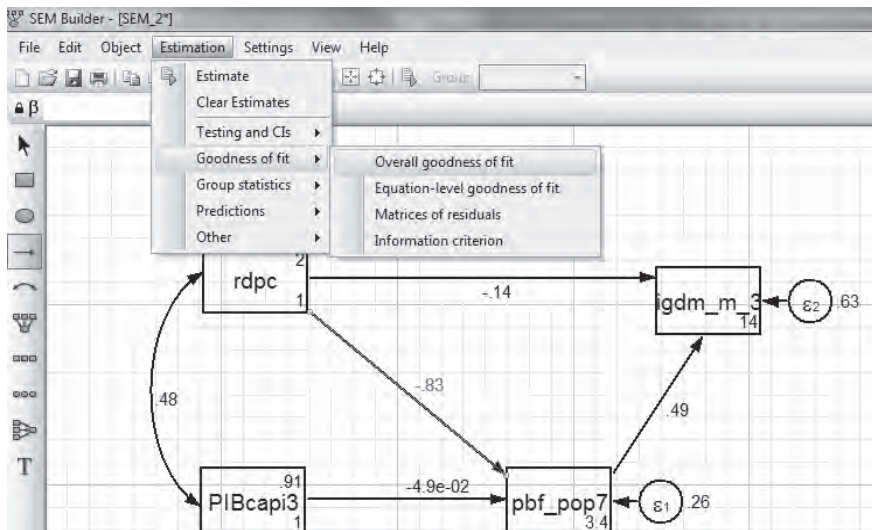
O modelo proposto gera quatro estimadores. Assim, há a necessidade de realizar o mesmo número de testes de hipóteses, nos quais as hipóteses nulas serão sempre de que os parâmetros populacionais são iguais a zero ($\beta = 0$), assumindo-se um teste bilateral. Os resultados (gerados a partir da sequência de comandos reportada no início deste capítulo) calculados pelo STATA estão reportados nas tabelas anteriormente apresentadas.

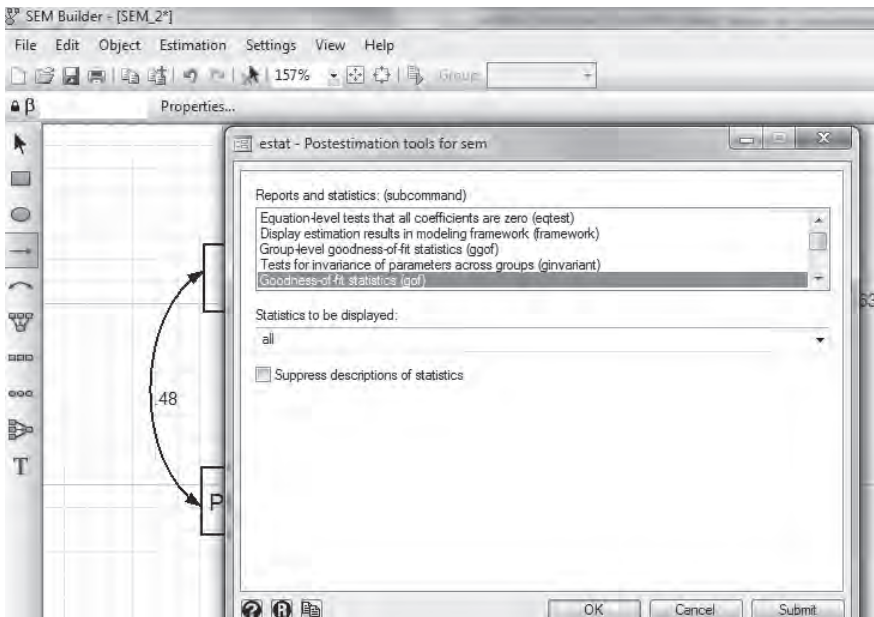
Os resultados do STATA mostram que, para todos os testes de hipóteses, é possível a rejeição da hipótese nula, visto que todos apresentam p-valores $< 0,001$. Obviamente, como se está trabalhando com uma amostra bastante grande ($n = 5565$), a rejeição da hipótese nula tende a ser bastante facilitada. Assim, é cada vez mais consensual que se deve aplicar uma lógica típica da estatística bayesiana e se buscar analisar se os coeficientes são não apenas significantes, mas também substantivamente relevantes. Essa análise exige, do pesquisador ou de qualquer outro profissional que esteja desenvolvendo os procedimentos

analíticos, um conhecimento aprofundado do que está sendo analisado para poder ter uma boa ideia da relevância dos resultados encontrados. Para tanto, é recomendável que se observem os resultados com base nos coeficientes padronizados. Os resultados parecem indicar que todos os coeficientes são relevantes. Todavia, para se concluir de forma mais segura que o modelo proposto é adequado, faz-se necessária a realização de uma análise sobre a qualidade de ajuste do modelo.

3.2.3 Análise da qualidade do ajuste

Para se realizar a análise da qualidade do ajuste do modelo, deve-se seguir os mesmos procedimentos desenvolvidos para o exemplo anterior, quais sejam:





Os resultados obtidos são os seguintes:

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(1)	0.527	model vs. saturated
p > chi2	0.468	
chi2_bs(5)	10003.629	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.000	Root mean squared error of approximation
90% CI, lower bound	0.000	
upper bound	0.032	
pclose	0.999	Probability RMSEA <= 0.05
Information criteria		
AIC	197054.666	Akaike's information criterion
BIC	197140.782	Bayesian information criterion
Baseline comparison		
CFI	1.000	Comparative fit index
TLI	1.000	Tucker-Lewis index
Size of residuals		
SPMR	0.002	Standardized root mean squared residual
CD	0.739	Coefficient of determination

Os resultados de ajuste do modelo são bastante bons, indicando sua adequabilidade, pois tem estatísticas CFI, TLI e CD com valores bastante elevados. Todavia, um teste mais objetivo (embora afetado pelo tamanho da amostra) pode ser dado pela comparação das estatísticas de qui-quadrado das razões de verossimilhança do modelo. O que se observa é que o modelo, quando comparado ao modelo saturado, tem uma estatística qui-quadrado de 0,527, com 1 grau de liberdade. Visto que o valor crítico da estatística qui-quadrado com 1 grau de liberdade e nível de significância de 0,05 é 3,85, observa-se, assim, que, quando comparado ao modelo saturado (ou seja, o primeiro modelo), o segundo modelo tem uma boa qualidade de ajuste (o ideal é justamente um qui-quadrado com um p-valor > 0,05). Conclui-se, portanto, que o modelo estimado tem uma boa qualidade de ajuste.

3.3 Comentários finais do capítulo

Neste capítulo, foram estimados MEE contendo apenas variáveis observadas. Primeiramente, foi utilizado um modelo clássico da análise de estratificação social, o chamado modelo de alocação de *status* socioeconômico. Os resultados indicaram que o modelo saturado foi o que apresentou o melhor ajuste e que os efeitos indiretos de duas das variáveis exógenas foram mais relevantes do que os efeitos diretos, o que indica a adequabilidade da estimação de um MEE. Posteriormente, foi estimado um modelo para a explicação do índice de gestão descentralizada municipal (IGD-M). O modelo proposto apresentou um elevado nível de qualidade do ajuste e uma estrutura complexa e reveladora de relação entre as variáveis.

A leitura deste capítulo, portanto, permite que se aprenda:

- a) como realizar a estimação de MEE contendo apenas variáveis observadas (ou seja, apenas o modelo causal);
- b) como realizar os testes de hipóteses;
- c) como realizar a análise sobre a qualidade de ajuste dos modelos.

4 MODELOS CONTENDO APENAS A ANÁLISE DE MENSURAÇÃO

Os MEE que incluem apenas análise de mensuração representam um subtipo dos modelos completos, que serão estudados no próximo capítulo. No caso das análises de mensuração, o objetivo é apenas o de mensurar construtos latentes a partir de variáveis observadas. Ou seja, ao passo que, no capítulo anterior, o objetivo da análise era propor e testar sistemas complexos de causalidade ou de determinação entre variáveis observadas, neste capítulo utilizam-se variáveis observadas para mensurar construtos (ou variáveis) latentes que devem ser fortemente associadas, mas não contam com relações de determinação entre elas.

As análises de mensuração dos MEE requerem a aplicação de análises fatoriais confirmatórias. Neste caso, como foi visto anteriormente, a análise fatorial é utilizada para confirmar uma expectativa de que duas ou mais variáveis observadas irão convergir para formar um mesmo construto latente e que tal construto é consistente e confiável. Ao passo que, na análise fatorial exploratória, os construtos (ou fatores) são formados a partir apenas de procedimentos estatísticos e suas constituições são desconhecidas pelo pesquisador ou analista em momentos prévios à análise, a análise fatorial confirmatória se presta à confirmação ou não de uma prévia teoria de mensuração. Ou seja, é necessário que o pesquisador ou analista tenha uma expectativa prévia de quais variáveis devem formar um construto.

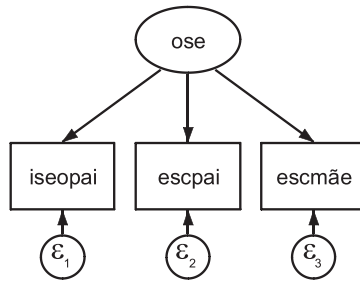
A definição sobre a adequabilidade do modelo de mensuração proposto é dada por análises de confiabilidade e/ou de ajuste. É comum o uso do coeficiente Alfa de Cronbach para análises de confiabilidade de escalas ou construtos latentes. Todavia, esses coeficientes são muito afetados pelo número de itens (variáveis observadas) da escala (construto). Mais especificamente, mantido tudo o mais constante, quanto maior o número de itens da escala mensurada, maior tenderá a ser o valor do coeficiente Alfa de Cronbach. Quando o modelo de mensuração conta

com um número pequeno de itens, o valor do Alfa de Cronbach tende a ser muito baixo. No caso das análises que serão utilizadas como exemplo, neste capítulo, ter-se-á justamente esse tipo de situação, pois todos os modelos de mensuração contarão com apenas duas ou três variáveis observadas (itens). Assim, fica ainda mais evidente a utilidade de se realizar mensurações a partir de MEE, pois é possível obter-se uma análise de ajuste do modelo mais adequada e ampla do que com o coeficiente Alfa de Cronbach, em particular quando se tem um número pequeno de variáveis observadas incluídas no modelo.

4.1 Primeiro exemplo: a mensuração da origem socioeconômica

Para que se possa propor um modelo de mensuração com base em uma análise fatorial confirmatória, deve-se partir de uma teoria da mensuração. Como visto no capítulo anterior, a teoria da estratificação social propõe que variáveis tais como a escolaridade dos pais e das mães e o *status* socioeconômico da ocupação dos pais são componentes da origem socioeconômica dos indivíduos. No capítulo anterior, essas variáveis foram incluídas de forma independente em um modelo de caminhos. Todavia, é possível que um analista ou pesquisador queira analisar se essas variáveis poderiam ser operacionalizadas a partir de um único construto. Assim, com base na teoria da estratificação social, propõe-se uma teoria da mensuração na qual se espera que as variáveis referentes à educação do pai, à educação da mãe e ao *status* socioeconômico do pai converjam para formar um único construto. Para tanto, mais uma vez, será utilizada a subamostra da PNAD-2014, a mesma usada no capítulo anterior. O diagrama de equações estruturais proposto é, portanto, o seguinte:

Figura 4.1: Diagrama do modelo de equações estruturais para a mensuração da variável latente referente à origem socioeconômica



Fonte: Elaboração própria

Onde:

ose = variável latente referente à origem socioeconômica dos indivíduos da subamostra;

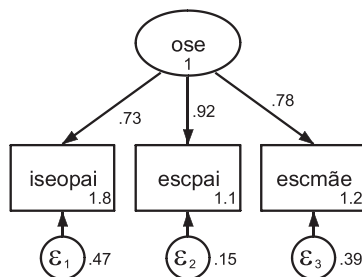
iseopai = índice socioeconômico da ocupação do pai do indivíduo da subamostra;

escpai = anos de escolaridade do pai do indivíduo da subamostra;

escmãe = anos de escolaridade da mãe do indivíduo da subamostra.

Os resultados encontrados (coeficientes padronizados) foram os seguintes:

Figura 4.2: Diagrama do modelo de equações estruturais para a mensuração da variável latente referente à origem socioeconômica com os resultados padronizados a partir dos dados da PNAD-2014



Fonte: Elaboração própria

```
Iteration 0: log likelihood = -41573.346
Iteration 1: log likelihood = -41573.346

Structural equation model          Number of obs   =    4467
Estimation method = ml
Log likelihood   = -41573.346
```

(1) [iseopai]_cse = 1

Standardized	DIM		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
Measurement						
iseopai <-						
_cse	.7284387	.0086166	84.54	0.000	.7115506	.7453269
_cons	1.781503	.0240647	74.03	0.000	1.734337	1.828669
escpai <-						
_cse	.9240393	.0066999	137.92	0.000	.9109077	.9371709
_cons	1.127753	.0191369	58.93	0.000	1.090245	1.165261
escmãe <-						
_cse	.778129	.0079551	97.82	0.000	.7625373	.7937208
_cons	1.150472	.0192877	59.65	0.000	1.112669	1.188275
Variance						
e.iseopai	.469377	.0125533			.4454068	.4946372
e.escpai	.1461514	.012382			.1237909	.1725508
e.escmãe	.3945152	.0123802			.3709616	.4195417
_cse	1					

Os resultados do modelo de mensuração da variável referente à origem socioeconômica mostram que as cargas fatoriais são todas bastante elevadas (superiores a 0,70; a menor é igual a 0,73 e a maior é igual a 0,92) e estatisticamente significantes.

A seguir, têm-se os resultados do ajuste do modelo. Pode-se observar que o *software* não gerou um coeficiente de qui-quadrado referente à comparação com o modelo saturado, pois o modelo analisado é exatamente igual ao modelo saturado (ou seja, ele contém todas as relações possíveis dadas as variáveis observadas utilizadas). Portanto, o coeficiente de qui-quadrado que deve ser observado é aquele referente à comparação entre o modelo nulo (*baseline*) e o modelo saturado. Mais uma vez, um coeficiente que tenha um p-valor < 0,05 indica que o primeiro modelo da comparação (no caso, o modelo nulo) tem um ajuste inferior ao segundo modelo (no caso, o modelo saturado). Portanto, visto que o

modelo utilizado na análise é equivalente ao modelo saturado, conclui-se que ele tem um ajuste adequado, quando se toma como referência a análise do qui-quadrado. Outras medidas de ajuste do modelo utilizadas com frequência (CFI, TLI e CD) – todas com valor igual ou próximo de 1,00 – também apresentam resultados que corroboram a qualidade de ajuste do modelo de mensuração proposto.

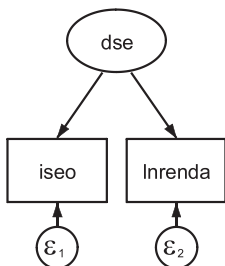
Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(0)	0.000	model vs. saturated
g > chi2	.	
chi2_bs(3)	5063.873	baseline vs. saturated
g > chi2	0.000	
Population error		
RMSEA	0.000	Root mean squared error of approximation
90% CI, lower bound	0.000	
upper bound	0.000	
pclose	1.000	Probability RMSEA <= 0.05
Information criteria		
AIC	83164.693	Akaike's information criterion
BIC	83222.333	Bayesian information criterion
Baseline comparison		
CFI	1.000	Comparative fit index
TLI	1.000	Tucker-Lewis index
Size of residuals		
SRMR	0.000	Standardized root mean squared residual
CD	0.895	Coefficient of determination

Assim, com base nos valores elevados das cargas fatoriais, nos testes de significância dessas cargas e nos indicadores de ajuste, pode-se concluir que o modelo apresenta um ajuste adequado. Assim, do ponto de vista substantivo, a conclusão deve ser a de que a teoria de mensuração proposta está correta, ou seja, as variáveis referentes à origem socioeconômica dos indivíduos de fato convergem adequadamente para formar um único construto latente.

4.2 Segundo exemplo: a mensuração do destino socioeconômico

No capítulo anterior, a variável resposta final do modelo foi o *status* socioeconômico da ocupação dos indivíduos da subamostra. Todavia, pode ser desejável a estimação de um modelo de mensuração no qual a variável latente é um construto que componha o índice de *status* socioeconômico da ocupação e o rendimento do trabalho. A composição dessas duas variáveis permite a mensuração do destino socioeconômico considerando tanto o elemento referente ao *status* ocupacional – que é uma variável de caráter mais estrutural – como também o rendimento do trabalho, que tem um caráter mais individual. Utilizando-se a mesma subamostra do exemplo anterior – e com base na teoria de mensuração de que o destino socioeconômico pode ser mensurado a partir da composição das variáveis observadas referentes ao *status* socioeconômico da ocupação e ao rendimento do trabalho –, propõe-se um modelo de mensuração com base no seguinte diagrama:

Figura 4.3: Diagrama do modelo de equações estruturais para a mensuração da variável latente referente ao destino socioeconômica



Fonte: Elaboração própria

Onde:

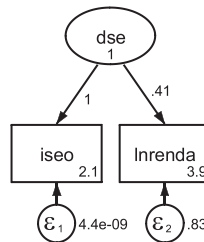
dse = construto latente referente ao destino socioeconômico dos indivíduos da subamostra;

iseo = índice socioeconômico da ocupação dos indivíduos da subamostra;

Inrenda = logaritmo natural (ou neperiano) do rendimento do trabalho (de todas as ocupações) dos indivíduos da subamostra.

Os resultados encontrados (coeficientes padronizados) foram os seguintes:

Figura 4.4: Diagrama do modelo de equações estruturais para a mensuração da variável latente referente ao destino socioeconômico com resultados padronizados a partir dos dados da PNAD-2014



Fonte: Elaboração própria

Fitting target model:

```
Iteration 0: log likelihood = -28044.383 (not concave)
Iteration 1: log likelihood = -28044.383 (not concave)
Iteration 2: log likelihood = -28044.383 (not concave)
Iteration 3: log likelihood = -28044.383 (not concave)
Iteration 4: log likelihood = -28044.383 (backed up)
```

```
Structural equation model          Number of obs   =   4467
Estimation method = ml
Log likelihood      = -28044.383
```

| 1) [iseo]dae = 1

Standardized	OIM					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Measurement						
iseo <-						
dse	1	161.3873	0.01	0.995	-315.3134	317.3134
_cons	2.097088	.0267603	78.37	0.000	2.044639	2.149537
Inrenda <-						
dse	.4129058	66.63776	0.01	0.995	-130.1947	131.0205
_cons	3.89199	.0438105	88.84	0.000	3.806123	3.977857
Variance						
e_iseo	4.37e-09	322.7747				
e_inrenda	.8295088	55.03023			2.81e-57	2.44e+56
dse	1					

Os resultados da mensuração do destino socioeconômico mostram que as cargas fatoriais são todas adequadas (superiores a 0,40) e estatisticamente significantes. Abaixo, têm-se os resultados do ajuste do modelo. Pode-se observar que, mais uma vez, o *software* não gerou um coeficiente de qui-quadrado referente à comparação com o modelo saturado, pois o modelo analisado é exatamente igual ao modelo saturado (ou seja, ele contém todas as relações possíveis dadas as variáveis observadas utilizadas). Portanto, o coeficiente de qui-quadrado que deve ser observado é aquele referente à comparação entre o modelo nulo (*baseline*) e o modelo saturado. Mais uma vez, um coeficiente que tenha um p-valor < 0,05 indica que o primeiro modelo da comparação (no caso, o modelo nulo) tem um ajuste inferior ao segundo modelo (no caso, o modelo saturado). Portanto, visto que o modelo utilizado na análise é equivalente ao modelo saturado, conclui-se que ele tem um ajuste adequado, quando se toma como referência a análise do qui-quadrado. Outras medidas de ajuste do modelo utilizadas com frequência (CFI, TLI e CD) – todas com valor igual a 1,00 – também apresentam resultados que corroboram a qualidade de ajuste do modelo de mensuração proposto.

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(0)	0.000	model vs. saturated
p > chi2	.	
chi2_bs(1)	834.978	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.000	Root mean squared error of approximation
90% CI, lower bound	0.000	
upper bound	0.000	
pclose	1.000	Probability RMSEA <= 0.05
Information criteria		
AIC	56098.766	Akaike's information criterion
BIC	56130.788	Bayesian information criterion
Baseline comparison		
CFI	1.000	Comparative fit index
TLI	1.000	Tucker-Lewis index
Size of residuals		
SPMR	0.000	Standardized root mean squared residual
CD	1.000	Coefficient of determination

Assim, com base nos valores adequados das cargas fatoriais, nos testes de significância dessas cargas e nos indicadores de ajuste, pode-se concluir que o modelo apresenta um ajuste adequado. Assim, do ponto de vista substantivo, a conclusão deve ser a de que a teoria de mensuração proposta está correta, ou seja, as variáveis referentes ao destino socioeconômico dos indivíduos de fato convergem adequadamente para formar um único construto latente.

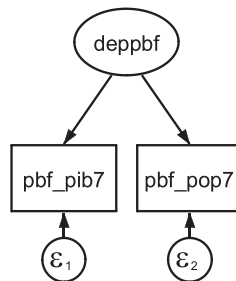
4.3 Terceiro exemplo: mensurando a dependência do município em relação ao Programa Bolsa Família

No capítulo anterior, foi estimado um modelo de determinação do IGD-M. Observou-se que ele era influenciado pela dependência que o município tem dos recursos financeiros do Programa Bolsa Família. Todavia, naquele momento, essa dependência financeira foi mensurada por apenas uma variável, a saber, a proporção da população beneficiária

em relação à população total do município. Ocorre que, um pesquisador ou analista pode querer testar um construto para a dependência do município em relação ao Programa Bolsa Família. Para tanto, propõe-se uma teoria de mensuração de que a dependência se refere não apenas à proporção da população beneficiária do programa, mas também à proporção da soma dos recursos totais repassados pelo Programa Bolsa Família com relação ao PIB do município.

O modelo de mensuração proposto foi:

Figura 4.5: Diagrama do modelo de equações estruturais para a mensuração da variável latente referente à dependência do programa bolsa família



Fonte: Elaboração própria

Onde:

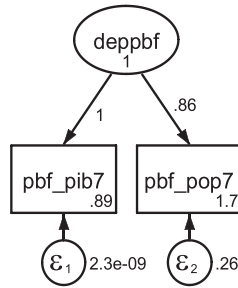
pbf_pib7: é o valor da soma de benefícios do Programa Bolsa Família recebido pelos moradores do município em relação ao PIB municipal.

pbf_pop7: é a proporção de beneficiários do Programa Bolsa Família em relação ao tamanho da população.

deppbf: é o construto latente que representa a dependência que o município tem do Programa Bolsa Família e é derivado das duas variáveis acima.

Os resultados encontrados (coeficientes padronizados) foram os seguintes:

Figura 4.6: Diagrama do modelo de equações estruturais para a mensuração da variável latente referente à dependência do programa bolsa família com resultados padronizados



Fonte: Elaboração própria

```
Structural equation model          Number of obs   =   5565
Estimation method   = ml
Log likelihood      = -35849.491

( 1) [pbf_pib7]deppbf = 1
```

Standardized	OIM				
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Measurement					
pbf_pib7 <-					
deppbf	1	28.93244	0.03	0.972	-55.70653 57.70653
_cons#	.8857461	.0158172	56.00	0.000	.8547449 .9167472
pbf_pop7 <-					
deppbf	.8609789	24.91022	0.03	0.972	-47.96215 49.68411
_cons#	1.666828	.02072	80.45	0.000	1.626218 1.707439
Variance					
e.pbf_pib7	2.29e-09	57.86487			
e.pbf_pop7	.2587153	42.89434			1.9e-142 3.5e+140
deppbf	1				

LR test of model vs. saturated: chi2(-1) = 0.00, Prob > chi2 =

Os resultados do modelo de mensuração da dependência do programa bolsa família mostram que as cargas fatoriais são bastante elevadas (superiores a 0,80) e estatisticamente significantes. A seguir, tem-se os resultados do ajuste do modelo. Pode-se observar que, mais uma vez, o *software* não gerou um coeficiente de qui-quadrado referente à comparação com o modelo saturado, pois o modelo analisado é

exatamente igual ao modelo saturado (ou seja, ele contém todas as relações possíveis dadas as variáveis observadas utilizadas). Portanto, o coeficiente de qui-quadrado que deve ser observado é aquele referente à comparação entre o modelo nulo (*baseline*) e o modelo saturado. Mais uma vez, um coeficiente que tenha um p-valor < 0,05 indica que o primeiro modelo da comparação (no caso, o modelo nulo) tem um ajuste inferior ao segundo modelo (no caso, o modelo saturado). Portanto, visto que o modelo utilizado na análise é equivalente ao modelo saturado, conclui-se que ele tem um ajuste adequado, quando se toma como referência a análise do qui-quadrado. Outras medidas de ajuste do modelo utilizadas com frequência (CFI, TLI e CD) – todas com valor igual a 1,00 – também apresentam resultados que corroboram a qualidade de ajuste do modelo de mensuração proposto.

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(-1)	0.000	model vs. saturated
p > chi2	.	
chi2_bs(1)	7524.030	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.000	Root mean squared error of approximation
90% CI, lower bound	0.000	
upper bound	0.000	
pclose	1.000	Probability RMSEA <= 0.05
Information criteria		
AIC	71710.982	Akaike's information criterion
BIC	71750.728	Bayesian information criterion
Baseline comparison		
CFI	1.000	Comparative fit index
TLI	1.000	Tucker-Lewis index
Size of residuals		
SPMR	0.000	Standardized root mean squared residual
CD	1.000	Coefficient of determination

4.4 Comentários finais do capítulo

Neste capítulo, foram testados MEE para mensuração de variáveis latentes. Primeiramente, foram utilizadas variáveis do modelo clássico da análise de estratificação social, o chamado modelo de alocação de *status* socioeconômico. Os resultados indicaram que os modelos de mensuração apresentaram bons ajustes. Posteriormente, foi testado um modelo para a mensuração da dependência dos municípios em relação aos recursos financeiros do Programa Bolsa Família. O modelo proposto apresentou um elevado nível de qualidade do ajuste.

A leitura deste capítulo, portanto, permite que se aprenda:

- a) como realizar a estimação de MEE para mensuração de variáveis latentes;
- b) como realizar os testes de hipóteses;
- c) como realizar a análise sobre a qualidade de ajuste dos modelos.

5 MODELOS CONTENDO VARIÁVEIS OBSERVADAS E CONSTRUTOS LATENTES

Neste capítulo, serão estimados MEE com a introdução de variáveis observadas e de variáveis mensuradas (latentes). Nesses casos, os MEE são uma composição de análises de regressão e de análises fatoriais confirmatórias. Serão apresentados dois exemplos diferentes. O primeiro será uma extensão do exemplo utilizado no capítulo anterior, baseado no modelo de realização de *status* socioeconômico. O segundo será uma aplicação na área de políticas sociais.

Para se estimar modelos completos de equações estruturais, é necessário que se desenvolva uma teoria estrutural. Uma teoria estrutural é uma representação conceitual das relações entre variáveis observadas e construtos. Ela pode ser expressa em termos de um MEE que represente adequadamente a partir de um sistema de equações estruturais e pode ser representada por um diagrama gráfico. Modelos estruturais são conhecidos por diversos nomes, em particular modelo causal. Um modelo causal infere que as relações atendem às condições necessárias para causalidade, que foram discutidas anteriormente.

5.1 Primeiro exemplo: o modelo de realização de *status* socioeconômico contendo a mensuração de um construto

No terceiro capítulo, vimos a estimação de um modelo clássico de realização de *status* socioeconômico no qual todas as variáveis incluídas eram observadas. A variável resposta final do modelo diz respeito ao *status* socioeconômico da ocupação. Todavia, pode ser desejável a estimação de um modelo no qual a variável resposta final seja um construto que componha o índice de *status* socioeconômico da ocupação e o rendimento do trabalho. A composição dessas duas variáveis permite a mensuração do destino socioeconômico considerando tanto o elemento referente ao *status* ocupacional – que é uma variável de

caráter mais estrutural – como também o rendimento do trabalho, que tem um caráter mais individual.

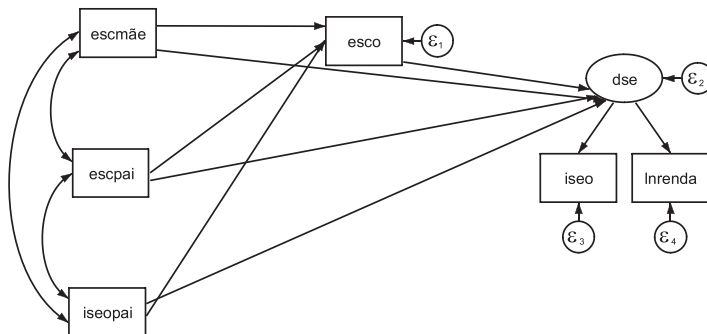
O sistema de equações para a realização da análise é:

$$\text{escoi} = \beta_0 + \beta_1(\text{escm\~{a}e})_i + \beta_2(\text{escpai})_i + \beta_3(\text{iseopai})_i + \epsilon_{1i} \quad (4.1)$$

$$\text{dsei} = \beta_0 + \beta_4(\text{escm\~{a}e})_i + \beta_5(\text{escpai})_i + \beta_6(\text{iseopai})_i + \beta_7(\text{iseopai})_i + \epsilon_{2i} \quad (4.2)$$

Esse sistema de equações gera o seguinte diagrama de equações estruturais:

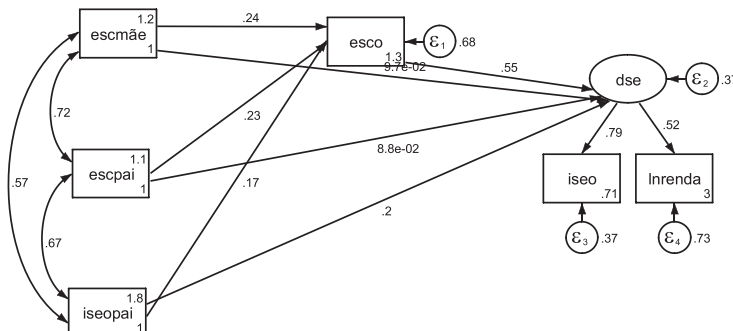
Figura 5.1: Diagrama de modelo de equações estruturais completo para a explicação do destino socioeconômico



Fonte: Elaboração própria

Os resultados da estimação do modelo foram os seguintes:

Figura 5.2: Diagrama de modelo de equações estruturais completo para a explicação do destino socioeconômico com os resultados padronizados a partir dos dados da PNAD-2014



Fonte: Elaboração própria

O modelo estimado é agora composto por duas partes: um modelo causal e um modelo de mensuração. Esse último traz uma variável latente (um construto referente ao destino socioeconômico do entrevistado, *dse*) gerada a partir de análise fatorial confirmatória com duas variáveis observadas: nível socioeconômico (*iseo*) e o logaritmo natural do rendimento do trabalho (*lnrenda*). O modelo padronizado mostra cargas fatoriais de 0,79 e 0,52, indicando que as variáveis observadas convergem de forma satisfatória para formar o construto.

Os resultados da estimação do modelo foram os seguintes:

```
. sem (escm5e -> esc0) (escm5e -> dse) (escpai -> esc0) (escpai -> dse) (iseopai -> esc0) (iseopai -> dse) (esc0 -> dse) (dse -> iseo) (dse -> lnrenda), latent(dse) cov(escpai*escm5e isepai*escm5e isepai*escpai) nocaplatent

Endogenous variables

Observed:   esc0
Measurement: isec lnrenda
Latent:     dse

Exogenous variables

Observed:   escm5e escpai isepai

Fitting target model:

Iteration 0:  log likelihood = -98975.115 (not converged)
Iteration 1:  log likelihood = -81753.898
Iteration 2:  log likelihood = -81322.503
Iteration 3:  log likelihood = -81158.676
Iteration 4:  log likelihood = -80666.76
Iteration 5:  log likelihood = -80455.208
Iteration 6:  log likelihood = -80435.158
Iteration 7:  log likelihood = -80435.036
Iteration 8:  log likelihood = -80435.036

Structural equation model          Number of obs   =   4467
Estimation method   = ml
Log likelihood      = -80435.036
```

```

1) [iseoldse = 1

```

	OIM					[95% Conf. Interval]	
	Coef.	Sdd. Brs.	z	P> z			
Structural							
esco <-							
escmãe	-.2522337	.0186688	13.51	0.000	.2156435	.288824	
escpai	-.2358551	.0205711	11.47	0.000	.1955366	.2761737	
iseopai	-.0439897	.0042686	10.31	0.000	.0356234	.052356	
_cons	5.563614	.1109121	50.16	0.000	5.34623	5.780998	
dae <-							
esco	1.838876	.0572196	32.14	0.000	1.726727	1.951024	
escmãe	-.3371437	.0713986	4.72	0.000	.1972049	.4770824	
escpai	-.3039651	.0782406	3.89	0.000	.1506164	.4573138	
iseopai	-.1716657	.0162509	10.56	0.000	.1398145	.2035163	
Measurement							
iseo <-							
dae	1 (constrained)						
_cons	13.13132	.5480972	23.96	0.000	12.05707	14.20557	
Intenda <-							
dae	-.0664392	.0022724	29.24	0.000	.0619854	.0708929	
_cons	5.568671	.0623034	89.38	0.000	5.446558	5.690783	

Mean							
escmãe	4.830535	.0628219	76.89	0.000	4.707406	4.953664	
escpai	4.787105	.0635113	75.37	0.000	4.662626	4.911585	
iseopai	30.74323	.2581993	119.07	0.000	30.23717	31.24929	

Variance							
e.esco	12.91255	.2732241			12.38799	13.45932	
e.iseo	124.6563	6.53271			112.4881	138.1409	
e.Intenda	2.536105	.059797			2.421572	2.656056	
e.dae	78.9555	6.204352			67.68535	92.10222	
escmãe	17.62944	.3730316			16.91327	18.37594	
escpai	18.01848	.3812634			17.2865	18.78145	
iseopai	297.8009	6.301341			285.7031	310.4109	

Covariance							
escmãe							
escpai	12.81605	.3284445	39.02	0.000	12.17131	13.45879	
iseopai	41.07021	1.246156	32.96	0.000	38.62779	43.51263	
escpai							
iseopai	49.30665	1.321167	37.32	0.000	46.71721	51.89609	

LR test of model vs. saturated: $\chi^2(3) = 2.51$, Prob > $\chi^2 = 0.4726$

```

LD test of model vs. saturated: chi2(3) = 2.51, Prob > chi2 = 0.4726

. sem (escmãe -> esco) (escmãe -> dse) (escpai -> esco) (escpai -> dse) (iseopai -> esco) (iseopai -> dse) (esco -> dse) (dse -> iseo) (dse -> lnrenda), standardized latent(dse) cov(escpai*escmãe iseopai*escmãe iseopai*escpai) noconstant

Endogenous variables

Observed: esco
Measurement: iseo lnrenda
Latent: dse

Exogenous variables

Observed: escmãe escpai iseopai

Fitting target model:

Iteration 0: log likelihood = -88575.115 (not concave)
Iteration 1: log likelihood = -81759.898
Iteration 2: log likelihood = -81322.808
Iteration 3: log likelihood = -81168.676
Iteration 4: log likelihood = -80866.756
Iteration 5: log likelihood = -80485.208
Iteration 6: log likelihood = -80435.158
Iteration 7: log likelihood = -80435.036
Iteration 8: log likelihood = -80435.036

Structural equation model Number of obs = 4467
    
```

(1) [iseo]dse = 1

Standardised	OIM				
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Structural					
esco <-					
escmãe	.2425067	.0177238	13.68	0.000	.2077687 .2772447
escpai	.229248	.01981	11.57	0.000	.1904211 .268075
iseopai	.1738263	.0167526	10.38	0.000	.1409918 .2066608
_cons	1.273957	.0339682	37.50	0.000	1.207391 1.340544
dse <-					
esco	.5504308	.0164112	33.54	0.000	.5182654 .5825962
escmãe	.0970255	.0205729	4.72	0.000	.0567033 .1373476
escpai	.0884371	.0227662	3.88	0.000	.0438162 .133058
iseopai	.2030477	.0191167	10.62	0.000	.1655797 .2405157
Measurement					
iseo <-					
dse	.7941456	.012443	63.82	0.000	.7697577 .8185335
_cons	.7147597	.034844	20.51	0.000	.6464666 .7830528
lnrenda <-					
dse	.5199371	.0131326	39.59	0.000	.4941976 .5456765
_cons	2.386963	.0558567	53.48	0.000	2.877486 3.09644

Mean						
escmãe	1.150472	.0192877	59.65	0.000	1.112669	1.188275
escpai	1.127753	.0191369	58.93	0.000	1.090245	1.165261
iseopai	1.781503	.0240647	74.03	0.000	1.734337	1.828669
Variance						
e.escc	.6770403	.0115136			.654846	.6999868
e.isec	.3693327	.0197632			.3325597	.410172
e.lnrenda	.7296654	.0136563			.7033846	.7569282
e.dae	.3709245	.0197298			.3342023	.4116818
escmãe	1	.			.	.
escpai	1	.			.	.
iseopai	1	.			.	.
Covariance						
escmãe						
escpai	.7190218	.0072268	99.49	0.000	.7048575	.7331861
iseopai	.5668193	.010155	55.82	0.000	.5469159	.5867228
escpai						
iseopai	.673106	.0081832	82.25	0.000	.6570673	.6891448

LR test of model vs. saturated: chi2(3) = 2.51, Prob > chi2 = 0.4726

. estat teffects, standardized

Direct effects

	OIM				Std. Coef.
	Coef.	Std. Err.	z	P> z	
Structural					
dse <-					
escmãe	.2522337	.0713986	3.53	0.000	.2425067
escpai	.2358551	.0782406	3.01	0.003	.229248
iseopai	.0439897	.0162509	2.71	0.007	.1738263
Measurement					
isec <-					
esco	0 (no path)				0
dae	1 (constrained)				.7941456
escmãe	0 (no path)				0
escpai	0 (no path)				0
iseopai	0 (no path)				0
lnrenda <-					
esco	0 (no path)				0
dae	.0664392	.0022724	29.24	0.000	.5199371
escmãe	0 (no path)				0
escpai	0 (no path)				0
iseopai	0 (no path)				0

Structural						
esco <-						
esco	1.838876	(no path)				.5504308
escmêe	.3371437	.0186688	18.06	0.000		.0970255
escpai	.3039651	.0205711	14.78	0.000		.0884371
iseopai	.1716657	.0042686	40.22	0.000		.2030477

Indirect effects

	OIM					Std. Coef.
	Coef.	Std. Err.	z	P> z		
Structural						
dse <-						
escmêe	0	(no path)				0
escpai	0	(no path)				0
iseopai	0	(no path)				0
Measurement						
iseo <-						
esco	1.838876	.0572196	32.14	0.000		.4371222
dse	0	(no path)				0
escmêe	.8009702	.0780298	10.26	0.000		.1830574
escpai	.7376734	.0859803	8.58	0.000		.1704413
iseopai	.2525573	.0179242	14.09	0.000		.2372328

Inrenda <-						
esco	.1221734	.0038016	32.14	0.000		.1861894
dse	0	(no path)				0
escmêe	.0532158	.0054143	9.83	0.000		.11985
escpai	.0490104	.0058753	8.34	0.000		.1115901
iseopai	.0167797	.001266	13.25	0.000		.1553193

Structural						
esco <-						
esco	0	(no path)				0
escmêe	.4638265	(constrained)				.1334831
escpai	.4337083	(constrained)				.1261852
iseopai	.0808916	(constrained)				.0556794

Total effects

	OIM					Std. Coef.
	Coef.	Std. Err.	z	P> z		
Structural						
dse <-						
escmêe	.2522337	.0780298	3.23	0.001		.2425067
escpai	.2358551	.0859803	2.74	0.006		.229248
iseopai	.0439897	.0179242	2.45	0.014		.1738263

Measurement						
iseco <-						
escc	1.838876	.0572196	32.14	0.000		.4371222
dse		1 (constrained)				.7941456
escmãe	.8009702	.0780298	10.26	0.000		.1830574
escpai	.7376734	.0859803	8.58	0.000		.1704413
isecpai	.2525573	.0179242	14.09	0.000		.2372328
Inrenda <-						
escc	.1221734	.0038016	32.14	0.000		.2861894
dse	.0664392	.0022724	29.24	0.000		.5199371
escmãe	.0532158	.0054143	9.83	0.000		.11985
escpai	.0490104	.0058753	8.34	0.000		.1115901
isecpai	.0167797	.001266	13.25	0.000		.1553193
Structural						
escc <-						
escc	1.838876	(constrained)				.5504308
escmãe	.8009702	.0186688	42.90	0.000		.2305086
escpai	.7376734	.0205711	35.86	0.000		.2146222
isecpai	.2525573	.0042686	59.17	0.000		.2987271

. estat gof, atata(all)

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(3)	2.515	model vs. saturated
p > chi2	0.473	
chi2_bs(12)	5049.395	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.000	Root mean squared error of approximation
90% CI, lower bound	0.000	
upper bound	0.024	
pclose	1.000	Probability RMSEA <= 0.05
Information criteria		
AIC	160918.072	Akaike's information criterion
BIC	161071.779	Bayesian information criterion
Baseline comparison		
CFI	1.000	Comparative fit index
TLI	1.000	Tucker-Lewis index
Size of residuals		
SRMR	0.003	Standardized root mean squared residual
CD	0.445	Coefficient of determination

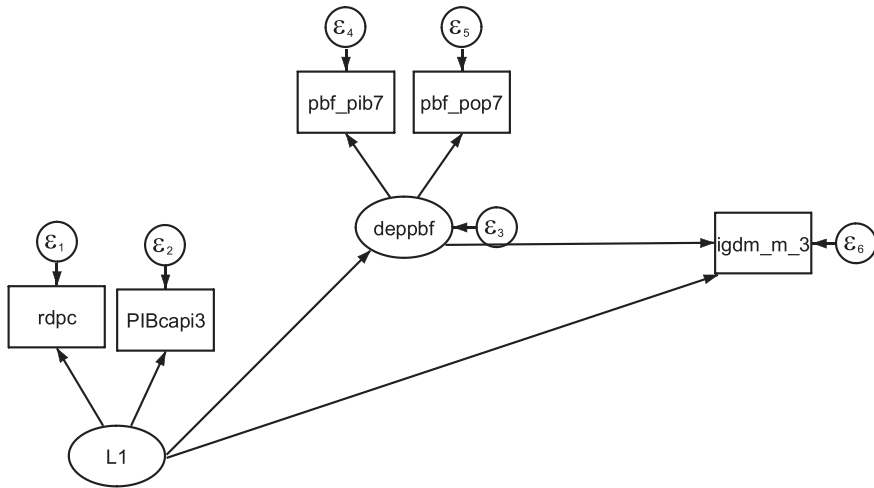
Os resultados acima são semelhantes àqueles do capítulo anterior. Mais uma vez, se observa a relevância do MEE, ao se perceber que os efeitos indiretos das variáveis referentes à origem socioeconômica – em particular da escolaridade do pai e da mãe – sobre o destino socioeconômico são tão ou mais relevantes do que os efeitos diretos. Observa-se, ainda, que o ajuste do modelo com a introdução da mensuração do construto referente ao destino socioeconômico é ainda melhor do que aquele do modelo estimado no terceiro capítulo. O coeficiente de determinação (CD) cresceu e agora é 0,445, e o qui-quadrado é 2,515, o que dá um p-valor de 0,473, bem maior do que um nível de significância de 0,05, mostrando que não se deve rejeitar a hipótese nula de que o modelo tem um bom ajustamento.

5.2 Segundo exemplo: modelo de determinação do IGD-M

Em 2006, o então Ministério do Desenvolvimento Social e Combate à Fome (MDS) criou o índice de gestão descentralizada (IGD) para medir e acompanhar a gestão do Programa Bolsa Família e do Cadastro Único por parte dos entes federados. No caso dos municípios, o índice se chama IGD-M. Nesta parte do livro, vamos desenvolver um MEE para explicar a determinação do referido índice. Para tanto, são utilizados os dados referentes ao ano de 2014.

O modelo proposto originalmente é o seguinte:

Figura 5.3: Diagrama do modelo completo para explicação do IDH-M



Fonte: Elaboração própria

Onde:

rdpc: diz respeito à renda *per capita* do município.

PIBcapi3: diz respeito ao PIB *per capita* do município.

L1: é um construto derivado das duas variáveis acima e representa o nível econômico do município.

pbf_pib: é o valor da soma de benefícios do Programa Bolsa Família recebido pelos moradores do município em relação ao PIB municipal.

pbf_pop7: é a proporção de beneficiários do Programa Bolsa Família em relação ao tamanho da população.

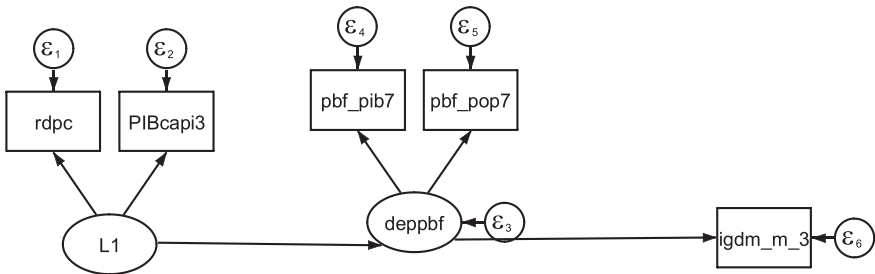
dep_pbfb: é o construto latente que representa a dependência que o município tem do Programa Bolsa Família e é derivado das duas variáveis acima.

lgdm_m_3: diz respeito ao índice de gestão descentralizada municipal.

O modelo acima pressupõe efeitos diretos e indiretos. Todavia, esse modelo teórico proposto originalmente não pode ser estimado, pois não foi possível se encontrar uma solução aceitável, mesmo após dezenas

de iterações computacionais. Então, foi necessário buscar um modelo mais parcimonioso. Assim, evoluiu-se para um modelo no qual a variável exógena só tem efeito indireto sobre a variável resposta final, a saber:

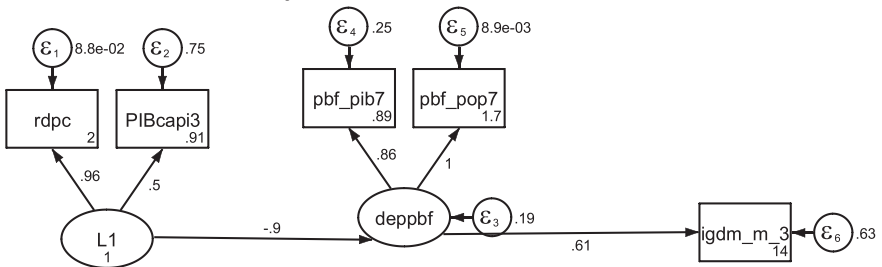
Figura 5.4: Diagrama do modelo mais parcimonioso para explicação do IDH-M



Fonte: Elaboração própria

Os resultados da estimação do modelo acima foram os seguintes:

Figura 5.5: Diagrama do modelo mais parcimonioso para explicação do IDH-M com resultados padronizados



Fonte: Elaboração própria

Structural equation model		Number of obs		=		5565	
Estimation method		= ml					
Log likelihood		= -109641.62					
(1) [pbf_pib7]deppbf = 1							
(2) [rdpc]l1 = 1							
Standardized	OIM			P> z	[95% Conf. Interval]		
	Coeff.	Std. Err.	z				
Structural							
deppbf <-							
L1	-.9015917	.0070505	-127.88	0.000	-.9154104	-.8877729	
Measurement							
rdpc <-							
L1	.9550053	.0066289	144.07	0.000	.9420129	.9679977	
_cons	2.029232	.023445	86.58	0.000	1.983291	2.075183	
PIBcap13 <-							
L1	.5017602	.0105995	47.34	0.000	.4809856	.5225349	
_cons	-.9111131	.0159461	-57.14	0.000	-.8798593	-.942367	
pbf_pib7 <-							
deppbf	.8645581	.0037573	230.10	0.000	.857194	.8719223	
_cons	-.8857461	.0158172	-56.00	0.000	-.854745	-.9167473	
pbf_pop7 <-							
deppbf	.9955595	.0019871	501.00	0.000	.9916648	.9994542	
_cons	1.666829	.02072	80.45	0.000	1.626219	1.70744	
igdm_m_3 <-							
deppbf	.6079198	.0086115	70.59	0.000	.5910415	.6247981	
_cons	14.03549	.1337129	104.97	0.000	13.77342	14.29756	

Variance				
e_rdpcc	.0879649	.0126612	.0663425	.1166346
e_PIBscapi3	.7482367	.0106368	.7276766	.7693776
e_pbf_pib7	.2525392	.0064968	.2401215	.2655991
e_pbf_pop7	.0088613	.0039566	.0036934	.0212601
e_igdm_m_3	.6304335	.0104702	.6102426	.6512924
e_deppbf	.1871325	.0127134	.1638025	.2137853
L1	1	.	.	.

LR test of model vs. saturated: $\chi^2(4) = 128.59$, Prob > $\chi^2 = 0.0000$

. estat teffects, standardized

Indirect effects

		OIM				
		Coef.	Std. Err.	z	P> z	Std. Coef.
Measurement						
rdpcc <-						
L1		0	(no path)			0
PIBscapi3 <-						
L1		0	(no path)			0
pbf_pib7 <-						
deppbf		0	(no path)			0
L1		-.0117466	.0002125	-55.28	0.000	-.7794784
pbf_pop7 <-						
deppbf		0	(no path)			0
L1		-.0797411	.0012425	-64.18	0.000	-.8975881
igdm_m_3 <-						
deppbf		0	(no path)			0
L1		-.0142102	.0003421	-41.54	0.000	-.5480954
Structural						
deppbf <-						
L1		0	(no path)			0

Total effects

	OIM				Std. Coef.
	Coef.	Std. Err.	z	P> z	
Measurement					
rdpc <-					
L1	1 (constrained)				.9550053
PIBcapi3 <-					
L1	.0334233	.0008757	38.17	0.000	.5017602
pbf_pib7 <-					
deppbf	1 (constrained)				.8645581
L1	-.0117466	.0002125	-55.28	0.000	-.7794784
pbf_pop7 <-					
deppbf	6.788453	.0594816	114.13	0.000	.9955595
L1	-.0797411	.0012425	-64.18	0.000	-.8975881
igdm_m_3 <-					
deppbf	1.209728	.0232913	51.94	0.000	.6079198
L1	-.0142102	.0003421	-41.54	0.000	-.5480954
Structural					
deppbf <-					
L1	-.0117466	.0002125	-55.28	0.000	-.9015917

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(4)	128.590	model vs. saturated
p > chi2	0.000	
chi2_bs(10)	19067.278	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.075	Root mean squared error of approximation
90% CI, lower bound	0.064	
upper bound	0.086	
pclose	0.000	Probability RMSEA <= 0.05
Information criteria		
AIC	219315.238	Akaike's information criterion
BIC	219421.226	Bayesian information criterion
Baseline comparison		
CFI	0.993	Comparative fit index
TLI	0.984	Tucker-Lewis index
Size of residuals		
SRMR	0.014	Standardized root mean squared residual
CD	0.938	Coefficient of determination

Os resultados acima mostram a adequabilidade da estimação de um MEE com a inclusão de análise de determinação e de modelos de mensuração. A maior parte dos indicadores de ajuste do modelo (CFI, TLI e CD) obtiveram coeficientes bastante elevados (próximos de 1). O único indicador de ajuste que não obteve resultado favorável à qualidade do modelo foi a comparação entre o qui-quadrado do modelo estimado em relação ao modelo saturado¹⁰. Observe-se, porém, que o modelo saturado não pode ser estimado, como se viu acima. Os coeficientes padronizados de regressão e as cargas fatoriais apresentaram todos os valores bastante elevados ou pelo menos satisfatórios (a menor carga fatorial foi maior do que 0,5). O modelo estimado, portanto, justifica a busca de uma estrutura complexa de relação entre construtos latentes e variáveis observadas considerando efeitos indiretos.

¹⁰ Dada a sensibilidade da estatística do qui-quadrado para o tamanho da amostra é que os especialistas propuseram uma variedade de índices de ajuste alternativos para avaliar o ajuste do modelo. Todas as medidas de bondade são algumas das funções do qui-quadrado e dos graus de liberdade. A maioria desses índices de ajuste não só considera o ajuste do modelo, mas também a sua simplicidade. Um modelo saturado, que especifica todos os caminhos (ou ligações) possíveis entre todas as variáveis, sempre se adapta perfeitamente aos dados, mas é tão complexo quanto os dados observados. Mas, se dois modelos tiverem o mesmo grau de ajuste, deve-se preferir o mais simples dos dois (princípio da parcimônia). Em geral, há uma compensação entre o ajuste de um modelo e sua simplicidade. Vários índices de qualidade de ajuste foram propostos para avaliar simultaneamente o ajuste e a simplicidade de um modelo. O objetivo é produzir um índice de qualidade de ajuste que não dependa do tamanho da amostra ou da distribuição dos dados. De fato, a maioria dos índices de qualidade de ajuste ainda depende do tamanho e da distribuição da amostra, mas a dependência é muito menor que a do teste de Qui-quadrado tradicional.

5.3 Comentários finais do capítulo

Neste capítulo, foram estimados MEE completos, ou seja, que incluem determinação e mensuração. Primeiramente, foram utilizadas variáveis do modelo clássico da análise de estratificação social, o chamado modelo de alocação de *status* socioeconômico. Os resultados indicaram que os modelos alcançaram bons ajustes. Posteriormente, foi estimado um modelo para a explicação do IGD-M. O modelo proposto apresentou um elevado nível de qualidade do ajuste.

Modelos estruturais completos diferem de modelos de mensuração no sentido de que a ênfase deixa de ser apenas a relação entre construtos latentes e variáveis medidas para a natureza e magnitude das relações entre construtos e variáveis observadas. Modelos de mensuração são testados usando-se apenas análise fatorial confirmatória. O resultado é a especificação de um modelo estrutural que é usado para testar o modelo teórico proposto.

Por outro lado, um MEE completo inclui não apenas a determinação entre variáveis, mas também a mensuração. Isso permite que se possam desenvolver modelos complexos de relação causal nos quais são controlados tanto erros de estimação quanto erros de mensuração.

A leitura deste capítulo, portanto, permite que se aprenda:

- a) como realizar a estimação de MEE completo, ou seja, com mensuração de variáveis latentes e sistemas complexos de relação entre variáveis;
- b) como realizar os testes de hipóteses;
- c) como realizar a análise sobre a qualidade de ajuste dos modelos.

6 CONCLUSÃO

Um MEE completo envolve tanto o teste da teoria de mensuração quanto da teoria estrutural (ou causal) que conecta construtos entre si de uma maneira teoricamente fundamentada. A modelagem de equações estruturais não é apenas mais uma técnica estatística multivariada: é uma maneira de testar teorias. Modelagens estatísticas muito mais fáceis estão disponíveis para a exploração de relações. Mas quando um pesquisador ou analista conhece suficientemente bem a questão que está analisando e que requer a observação de um conjunto de relações entre variáveis observadas e construtos latentes, além da maneira como tais construtos são medidos, o MEE passa a ser a melhor opção.

Neste livro, são apresentadas as principais variações de MEE. A aplicação dessas variações permite que o pesquisador ou analista consiga desenvolver análises multivariadas de dados tanto com perspectiva causal quanto focada na mensuração, ou ainda, o que é o potencial máximo da MEE, a combinação das duas.

A aplicação das técnicas expostas neste livro permite ao pesquisador ou ao analista de dados que desenvolva trabalhos que tanto podem contribuir para o avanço do conhecimento científico em diversas áreas das Ciências Humanas ou Sociais Aplicada, quanto realizar análises que podem subsidiar o processo de tomada de decisão de gestores públicos. As potencialidades das técnicas estudadas aqui são múltiplas e bastante significativas.

REFERÊNCIAS BIBLIOGRÁFICAS

- BLAU, Peter; DUNCAN, Otis. *The American Occupational Structure*. Nova York: The Free Press, 1967.
- FERNANDES, Flávio. *Federalismo, políticas sociais e sistemas de incentivo: o caso do Programa Bolsa Família (2006-2014)*. Belo Horizonte: PPGCP/FAFICH/UFMG, 2016. Tese (Doutorado em Ciência Política) – Universidade Federal de Minas Gerais, Belo Horizonte.
- GANZEBOOM, Harry; Treiman, Donald. Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, v. 25, p. 201-239, 1996.
- GUJARATI, Damodar. *Econometria Básica*. São Paulo: Makron Books, 1995.
- GUJARATI, Damodar. *Econometria Básica*. Rio de Janeiro: Elsevier, 2006.
- Hox, Joop; Bechger, Timo. An introduction to structural equation modeling. *Family Science Review*, v. 11, p. 354-373, 1998.
- NEVES, Jorge; Lima, Luciana. Industrialização, políticas de bem estar e fluidez social no Brasil: de 1973 a 2014. In: *Congresso Brasileiro de Sociologia, 18., 2017, Brasília. Anais...* Brasília: Sociedade Brasileira de Sociologia, 2017.
- REIS, Elizabeth. *Estatística multivariada aplicada*. Lisboa: Edições Sílabo, 1997.
- SHARMA, Subhash. *Applied multivariate techniques*. Nova York: John Wiley & Sons, 1996.

Esta obra foi impressa pela Imprensa Nacional
SIG, Quadra 6, Lote 800
70610-460 – Brasília-DF