UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA

Modelagem de Equações Estruturais no Software R

Autora: Suziane dos Santos Pereira
Orientadora: Prof[a]. Dr[a]. Lisiane Priscila Roldão Selau

Porto Alegre, dezembro de 2013.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Modelagem de Equações Estruturais no Software R

Autora: Suziane dos Santos Pereira

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Profª. Drª. Lisiane Priscila Roldão Selau
Prof. Dr. João Riboldi

Porto Alegre, dezembro de 2013.

Agradecimentos

Agradeço a todas as pessoas que de alguma forma colaboraram com este trabalho ou estiveram comigo durante todo o processo de desenvolvimento. Em especial aos meus pais Telmo e Eliete e ao meu irmão Diego: muito obrigada por toda a paciência, apoio e carinho. Aos professores que participaram e transmitiram seus conhecimentos, principalmente à professora Lisiane por ser tão amável e prestativa na orientação deste trabalho. Ao amigo Cristiano por todo incentivo, suporte e dedicação para a concretização desta etapa.

Resumo

Modelagem de Equações Estruturais ou SEM (*Structural Equation Modeling*) é um conjunto de técnicas e procedimentos que abordam uma extensão de outras técnicas multivariadas, avaliando relações simultâneas, ou seja, relações de dependência e independência entre uma ou mais variáveis. Além disso, consegue representar variáveis, que não podem ser medidas diretamente, através de grupos de outras variáveis, denominados construtos latentes. Nesse sentido, o objetivo geral deste trabalho é apresentar os variados usos da Modelagem de Equações Estruturais a partir de um *software* livre que tenha suporte para a técnica, seja de fácil acesso e com uma linguagem já conhecida dos Estatísticos. Para isso será utilizado o *Software* R versão 3.0.1 e os pacotes *Lavaan* 0.5 e *semPlot* 0.3.2. As técnicas relacionadas à Modelagem de Equações Estruturais abordadas nesse trabalho são: Análise de Regressão Multivariada, Análise Fatorial Confirmatória e Análise de Múltiplos Grupos. É apresentada a aplicação em dados publicados na literatura e na análise de um instrumento de medida realizado para entender o relacionamento de uma empresa com seus clientes.

Palavras-chave: Modelagem de Equações Estruturais, Regressão Multivariada, Análise Fatorial Confirmatória, Análise de Múltiplos Grupos, *Software* R.

Sumário

1. INTRODUÇÃO

Modelagem de Equações Estruturais ou SEM (*Structural Equation Modeling*) é um conjunto de técnicas e procedimentos que abordam uma extensão de outras técnicas multivariadas, avaliando relações simultâneas, ou seja, relações de dependência e independência entre uma ou mais variáveis. São equações de regressão multivariada analisadas simultaneamente, em que a variável resposta em uma equação pode aparecer como preditora em outra, sendo possível as variáveis influenciar uma na outra reciprocamente, diretamente ou através de outras variáveis (HAIR *et al*., 2005).

As equações estruturais se destinam a representar relações causais entre as variáveis no modelo. Por exemplo, usando um dos bancos de dados que será analisado no Capítulo 5, onde o objetivo do estudo era medir a opinião dos clientes em relação a um provedor de serviços, algumas questões que a SEM pode ajudar a responder são:

Quais variáveis estão relacionadas com a satisfação do cliente?

Como a satisfação do cliente pode impactar no valor pago e na reputação do provedor de serviços?

Como o valor e a reputação do provedor influenciam na confiança?

A Modelagem de Equações Estruturais permite tratar de todas essas questões juntas, o que nenhuma outra técnica multivariada permite. Além disso, consegue representar variáveis, que não podem ser medidas diretamente, através de grupos de outras variáveis, denominados construtos latentes ou variáveis latentes. SEM estima separadamente uma série de equações de regressão múltipla, mas interdependentes, simultaneamente, dependendo da especificação do modelo estrutural usado pelo programa estatístico (HAIR *et al*., 2005). Segundo os autores, outra vantagem é que suas suposições são mais fáceis de serem atendidas do que a maioria das outras técnicas multivariadas. Como casos particulares da análise de equações estruturais tem-se: regressão múltipla, análise de caminhos (*path analysis*), análise fatorial (confirmatória e exploratória), análise de múltiplos grupos e modelos de crescimento latente (*latent growth modeling).* Comumente é mais utilizada quando o objetivo é testar um modelo através do conhecimento a priori do pesquisador (análise fatorial confirmatória), mas também pode ser utilizada com caráter exploratório a fim de obter o modelo que melhor se adequa aos dados.

1.1 Motivação e Objetivos

A Modelagem de Equações Estruturais fornece um método para lidar diretamente com múltiplas relações simultâneas de dependência e independência entre as variáveis, fornecendo uma transição da análise exploratória para a confirmatória. Permite também a utilização de variáveis que não podem ser medidas diretamente, os chamados construtos latentes.

É uma técnica relativamente nova, comparada às outras técnicas multivariadas, e está expandindo cada vez mais seus campos de atuação nas mais diversas áreas, como: marketing, psicologia, medicina, ciências sociais, engenharias e biologia. Devido a esse amplo espaço de utilização, é muito importante que os Estatísticos tenham conhecimento sobre essa útil ferramenta na análise de dados.

Atualmente existem alguns programas específicos para o uso de SEM, como o LISREL, EQS, AMOS, MPLUS. Todos esses são *softwares* comerciais, o que dificulta o acesso aos estudantes que querem aprender e aplicar a técnica.

Nesse sentido, o objetivo geral deste trabalho é apresentar o uso da Modelagem de Equações Estruturais a partir de um *software* livre que tenha suporte para a técnica, seja de fácil acesso e com uma linguagem já conhecida dos Estatísticos. Para isso será utilizado o *Software* R versão 3.0.1 e dois pacotes: *Lavaan* 0.5, para as análises de equações estruturais e *semPlot* 0.3.2, para a construção do diagrama de caminhos.

O objetivo principal pode ser dividido em alguns objetivos específicos:

- Explorar de forma prática as várias utilizações do uso de equações estruturais mostrando a fundamentação estatística que permite seu uso;
- Explicitar as etapas de utilização para cada uma das técnicas abordadas;
- Descrever o funcionamento do pacote *Lavaan* do *software* R;

- Apresentar a aplicação computacional de SEM em dados publicados na literatura e a análise de um instrumento de medida realizado para entender o relacionamento de uma empresa com seus clientes, ressaltando os aspectos computacionais e a interpretação dos resultados.

1.2 Estrutura do Trabalho

Este trabalho está estruturado em seis capítulos:

- No Capítulo 1 é apresentada uma breve introdução ao assunto de SEM, expondo a motivação e os objetivos do estudo.
- No Capítulo 2 serão apresentados alguns conceitos básicos e as técnicas utilizadas; variáveis latentes e observadas, variáveis endógenas e exógenas, escalas de medida e associação entre as variáveis, além dos pressupostos do modelo.
- O Capítulo 3 apresenta as etapas necessárias para a Modelagem de Equações Estruturais, bem como, a representação gráfica do modelo, a identificação do modelo, o modelo de medida e o modelo estrutural.
- O Capítulo 4 descreve os aspectos computacionais do *Software* R e do pacote *Lavaan*.
- No Capítulo 5 alguns exemplos de dados publicados na literatura serão apresentados e também uma análise do instrumento de medida utilizado na tese de doutorado “A Prática do Marketing de Relacionamento e a Retenção de Clientes: um estudo aplicado em um ambiente de serviços” de Gabriel Sperandio Milan.
- O Capítulo 6 apresenta as conclusões e considerações finais do estudo.

2. MODELAGEM DE EQUAÇÕES ESTRUTURAIS

2.1 Tipos de Variáveis

Como SEM tem uma linguagem própria, sua nomenclatura se diferencia muito da forma normalmente utilizada em outras técnicas estatísticas, tanto nas variáveis, quanto nos gráficos, e também em outros aspectos que envolvem as análises. Desta forma, neste tópico serão abordados os tipos de variáveis utilizados nesta técnica.

As variáveis observadas, ou manifestas, são variáveis que são medidas diretamente. São obtidas através de questionários, pesquisas, escores de entrevistas. Por sua vez, variáveis latentes, fatores ou construtos, são variáveis que não podem ser medidas ou observadas diretamente (como satisfação, confiança, inteligência), mas que podem ser representadas ou medidas através de uma ou mais variáveis observadas.

Por exemplo, as atitudes e impressões das pessoas em relação a algum produto não são possíveis de se medir de forma precisa e direta. Entretanto, fazendo-se várias perguntas é possível avaliar os aspectos da atitude da pessoa. Essas respostas conjuntamente formam uma medida razoavelmente precisa do construto latente (atitude) para um indivíduo (HAIR *et al*., 2005).

As variáveis observadas e latentes podem ser classificadas como exógenas (independentes) e endógenas (dependentes), respectivamente. As variáveis exógenas causam flutuações em outras variáveis no modelo. Já as flutuações nesse tipo de variável não podem ser explicadas pelo modelo, podendo, apenas, ser influenciadas por fatores externos ao modelo. Endógenas são as variáveis influenciadas pelas exógenas, e a variação destas variáveis é explicada por variáveis presentes no modelo.

2.2 Escala de Medida

De uma forma geral as variáveis em SEM sãs quantitativas, classificadas como escala intervalar ou de razão. Em aplicações de Análise de Equações Estruturais, é irrelevante se as variáveis são do tipo intervalar ou de razão, precisando apenas serem agrupadas em variáveis quantitativas (quer contínuas, quer discretas) (MARÔCO, 2010). É comum na Análise de Equações Estruturais termos variáveis ordinais, incluindo-se nesse tipo a escala de Likert. Escalas ordinais violam alguns pressupostos da SEM, como continuidade e normalidade. Porém na prática essas variáveis são usadas frequentemente, havendo simulações que indicam que os resultados obtidos são de confiança desde que se obtenha pelo menos 5 classes (ATKINSON, 1988 *apud* LEMKE, 2005[1]) e que a distribuição de frequências se aproxime da curva normal (BOLLEN, 1989 *apud* MARÔCO,2010[2]).

2.3 Associação entre as Variáveis

Um dos aspectos fundamentais da Modelagem de Equações Estruturais é a associação entre as variáveis, pois a construção do modelo depende totalmente de como as variáveis latentes e observadas estão relacionadas umas com as outras. As associações das variáveis podem ser avaliadas através da covariância. Entretanto comparações entre variáveis de diferentes magnitudes não são possíveis, pois não são diretamente comparáveis. Para acabar com essa limitação, a utilização do Coeficiente de Correlação de Pearson é uma opção (MARÔCO,2010).

O Coeficiente de Correlação de Pearson é utilizado quando as variáveis são medidas numa escala pelo menos intervalar. Se as variáveis observadas são ordinais, mas têm subjacentes variáveis continuas com distribuição normal bivariada, deve-se utilizar, preferencialmente, o coeficiente de correlação policórica. Já o coeficiente de correlação poliserial é indicado se uma das variáveis for pelo menos intervalar e a outra ordinal. Se as variáveis forem contínuas, mas dicotomizadas ao serem medidas, utiliza-se o coeficiente de correlação tetracórico (JÖRESKOG E SÖRBOM, 1996 *apud* MARÔCO, 2010[3]).

A utilização de covariância ou correlação na análise de equações estruturais ajudam a avaliar a qualidade do modelo, devendo ser adaptada a cada caso. Muitas vezes é interessante poder comparar variáveis com magnitudes diferentes e outras vezes isso pode levar a resultados errados (LEMKE,2005).

---

[1] ATKINSON, L. *The measurement –statistics controversy: Factor analysis and subinterval data.* Bulletin of the Psychometric Society, 26, p. 361-364, 1988.

[2] BOLLEN, K. A. *Structural Equations with Latent Variables*. John Wiley & Sons, New York, 1989.

[3] JORESKOG, K. G., & Sorbom, D. LISREL 8: *users's reference guide*. Licolnwood: Scientific Software International, 1996

2.4 Análise Fatorial

A Análise Fatorial foi uma das primeiras técnicas multivariadas que teve por objetivo descobrir e analisar a estrutura de um conjunto de variáveis inter-relacionadas de modo a construir uma escala de medida para os fatores (MARÔCO,2010). O principal objetivo desta análise é explicar a relação entre muitas variáveis observadas através de um número reduzido de variáveis latentes. Desta forma, a análise fatorial pode ser considerada como uma técnica de redução. Existem dois tipos de análise fatorial: Análise Fatorial Exploratória e Análise Fatorial Confirmatória.

Análise Fatorial Exploratória (EFA) é utilizada quando as variáveis observadas e latentes não possuem especificação predefinida de um modelo que as relacione. Através da EFA, o pesquisador identifica fatores que explicam correlações dentro de um conjunto de variáveis. É um procedimento de redução da dimensão dos dados originais que visa identificar um pequeno número de fatores que explique a maior parte da variação observada em um número muito maior de variáveis (LEMKE,2005). Portanto é utilizada pelos pesquisadores como uma técnica exploratória para determinar o número de fatores comuns e descobrir quais variáveis mensuradas são indicadores razoáveis das dimensões latentes.

Análise Fatorial Confirmatória (CFA) é recomendada quando o pesquisador possui uma ideia a priori sobre a estrutura relacional entre as variáveis observadas e latentes em estudo (fatores), ou seja, a intenção é testar um modelo definido previamente. Essa ideia prévia pode ser baseada na teoria ou em resultados de estudos anteriores. Devido a esta característica de permitir o teste de um modelo teórico que estruture o relacionamento entre as variáveis observadas e as variáveis latentes, a análise fatorial confirmatória é amplamente utilizada em SEM, seja simplesmente para testar um instrumento de medida ou compor um modelo completo de Equações Estruturais, em que o objetivo é verificar hipóteses de causas entre as variáveis latentes. Ambos serão detalhados nas próximas seções.

2.5 Modelo de Equações Estruturais

O modelo linear que estabelece relações entre as variáveis observadas e latentes sob estudo pode ser separado em dois modelos de acordo com a estrutura relacional entre as variáveis. O submodelo de medida e o submodelo estrutural.

2.5.1 Submodelos

O submodelo de medida relaciona as variáveis observadas e as variáveis não observadas, sendo, portanto, uma ligação entre o instrumento de medida (variáveis observadas) e o construto teórico em estudo (variáveis latentes). O submodelo estrutural define as relações causais ou de associação entre as variáveis e os construtos não observados (variáveis latentes), especificando se uma variável latente causa mudanças em outras variáveis latentes no modelo, direta ou indiretamente. Através da simbologia própria de SEM, é possível ver na Figura 1 um exemplo de um modelo estrutural com três variáveis latentes, sendo uma variável latente exógena (P2) operacionalizada por três variáveis manifestas independentes (p7, p8, p9) e duas variáveis latentes endógenas (P1, P3) operacionalizadas por (p1, p2, p3, p13, p14, p15).

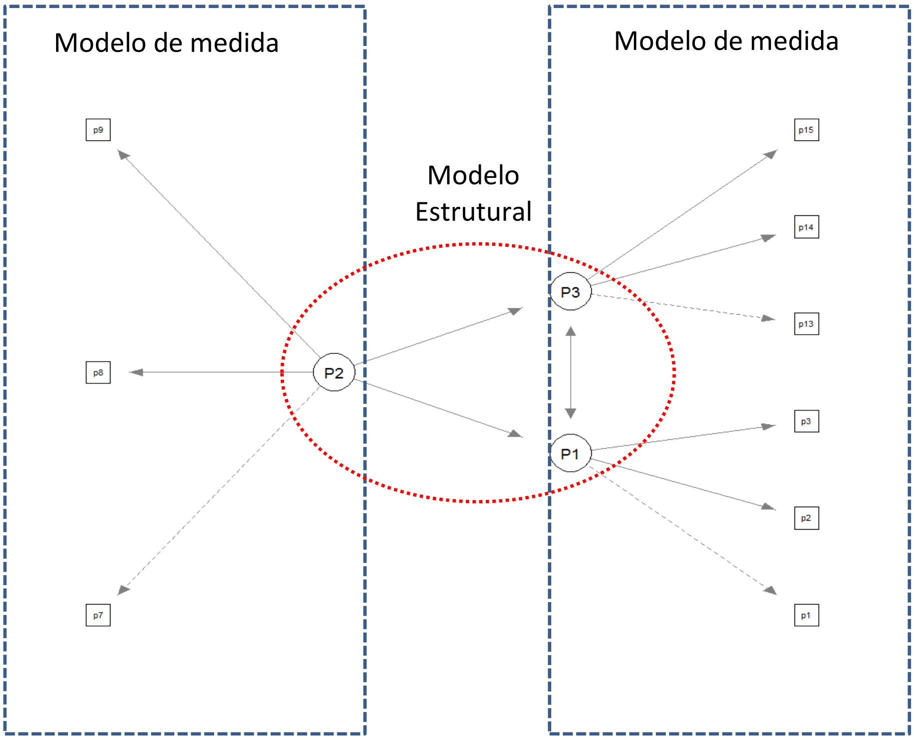


Figura 1. Submodelos
[adaptado de Byrne, 2001]

2.5.2 Representação Gráfica do Modelo

As relações entre as variáveis observadas e latentes são representadas através do diagrama de caminho (*path diagram*). SEM utiliza uma simbologia particular, por convenção, que possibilita ver de forma mais clara o modelo. As variáveis latentes são representadas dentro de círculos ou elipses; as variáveis observadas são representadas por retângulos ou quadrados; as relações de causa entre duas variáveis são representadas por setas unidirecionais e a associações correlacionais (sem um sentido causal explicitamente definido) são representadas por setas bidirecionais, conforme apresentado na Figura 2.

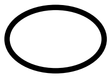
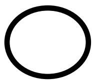
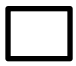
Descrição	Símbolo
Variáveis latentes (não observadas)	 ou 
Variáveis observadas	 ou 
Relação de causa	$\longrightarrow$
Associação correlacional	$\longleftrightarrow$

Figura 2. Símbolos para Representação Gráfica

[adaptado de LEMKE(2005) e SILVA(2006)]

É possível demonstrar as equações estruturais da Figura 1 separando por modelo de medida e modelo estrutural, como pode-se ver abaixo.

Modelo Estrutural:

$$P1 = P2 + r1$$

$$P3 = P2 + r2$$

onde r1 e r2 são os resíduos que representam os erros de predição dos fatores endógenos P3 e P1 a partir de fator exógeno P2.

Modelo de Medida:

$$p1 = P1 + e1$$

$$p2 = P1 + e2$$

$$p3 = P1 + e3$$

$$p7 = P2 + e4$$
$$p8 = P2 + e5$$
$$p9 = P2 + e6$$
$$p13 = P3 + e7$$
$$p14 = P3 + e8$$
$$p15 = P3 + e9$$

onde os e1 a e9 são os erros de medida das variáveis observadas. Eles refletem a adequação destas em medir os fatores em estudo.

2.6 Pressupostos para a Modelagem de Equações Estruturais

A utilização da técnica de Modelagem de Equações Estruturais exige a validação de um conjunto de suposições sem as quais os resultados do modelo testado podem ficar comprometidos. A violação dos pressupostos pode provocar resultados viesados no que diz respeito às estimativas dos parâmetros do modelo, bem como dos seus índices de ajuste associados. Desta forma, a decisão quanto à validação de um modelo teórico pode ser equivocada.

Nesta seção serão apresentados os principais pressupostos de SEM: independência das observações, normalidade multivariada, covariâncias amostrais não nulas, ausência de multicolinearidade e inexistência de *outliers* (MARÔCO, 2010). Esses pressupostos são importantes na Modelagem de Equações Estruturais, mas nem sempre é possível observar que todos estejam comtemplados.

2.6.1 Independência das Observações

As observações de sujeitos diferentes devem ser independentes entre si. Esse pressuposto pode ser assegurado com a prática de amostragem aleatória. A violação desta suposição pode aumentar a chance de o pesquisador cometer o erro tipo II (concluir pela não significância de um parâmetro que é, na população, significativo). Isso ocorre porque observações dependentes geralmente geram aumento dos erros-padrão dos parâmetros estimados.

2.6.2 Normalidade Multivariada

As variáveis observadas devem apresentar normalidade multivariada quando os métodos de estimação do modelo são o de Máxima Verossimilhança (ML) ou o de Mínimos Quadrados Generalizados (GLS). Com o pressuposto de normalidade, estes dois métodos apresentam propriedades de consistência (a estimativa do parâmetro tende ao verdadeiro valor quando a dimensão da amostra aumenta); de eficiência assintótica (a variância dos estimadores é mínima) e viés nulo assintótico (os estimadores dos parâmetros são não viesados para amostras grandes, ou seja, não subestimam nem superestimam os parâmetros populacionais) (MARÔCO, 2010).

Os testes usuais de normalidade univariada, como *Shapiro-Wilk* e *Kolmogorov-Smirnov*, podem ser utilizados e é possível assumir, com algumas exceções, que se um conjunto de variáveis apresentarem distribuição normal univariada, então a distribuição condicionada das variáveis é normal multivariada (MARÔCO, 2010). Estes testes não estão na maioria dos *softwares* de SEM em razão da sensibilidade a grandes amostras, que são corriqueiramente utilizadas na técnica. Os testes de ajustamento ficam extremamente sensíveis a pequenos desvios de normalidade, levando ao erro tipo I (concluir que a variável não tem distribuição normal quando de fato tem).

Como alternativa pode-se utilizar as medidas que avaliam a forma da distribuição, como assimetria (sk) e curtose (ku) univariados e a curtose multivariada (ku_M) para verificar a suposição de normalidade das variáveis. Valores absolutos de sk superiores a 3 e ku univariado e multivariado superiores a 10 indicam violação séria do pressuposto da normalidade e a inequação dos métodos ML e GLS para a estimação do modelo (MARÔCO, 2010).

2.6.3 Covariâncias Amostrais Não Nulas

Como em modelos de equações estruturais existem fatores latentes (construtos) que são operacionalizados por um conjunto de variáveis observadas, há a exigência de que estas apresentem associação, isto é, que sua covariância seja não nula. Este pressuposto é importante somente para a parte do modelo de medida do SEM.

2.6.4 Ausência de Multicolinearidade

Na parte estrutural do modelo é importante que as variáveis independentes (exógenas) não se encontrem fortemente correlacionadas (multicolinearidade). Na presença de multicolinearidade, pelo menos uma das variáveis independentes é redundante, inflacionando a estimativa das variâncias dos parâmetros, produzindo coeficientes padronizados superiores a 1 ou inferiores a -1 e podendo, ainda, causar estimativas de variâncias negativas (MARÔCO, 2010).

A avaliação da multicolinearidade pode ser feita através da estatística *VIF (variance inflation fator)* fornece um índice que mede o quanto a variância de um coeficiente de regressão estimado é aumentada por causa de colinearidade. Valores superiores a 5 indicam possíveis problemas de multicolinearidade (MARÔCO, 2010).

$$\boldsymbol{VIF_i = \frac{1}{1-R_i^2};}$$

onde R^2 é o coeficiente de determinação do modelo de regressão entre X_i e as variáveis independentes restantes.

2.6.5 Inexistência de *Outliers*

A presença de *outliers* pode inflacionar ou reduzir as covariâncias entre as variáveis e isto pode repercutir nas estimativas das médias, desvios-padrão e covariâncias, comprometendo a qualidade do ajustamento do modelo (MARÔCO, 2010). O diagnóstico de *outliers* pode ser feito através de dois tipos de medidas:

i. Medidas univariadas e diagnóstico visual com diagramas de extremos e quartis (*boxplot)*: uma observação é considerada *outlier* superior se o seu valor for superior a P_{75}+1,5(P_{75}-P_{25}) e é considerada *outlier* inferior se o seu valor for menor que P_{25}-1,5(P_{75}-P_{25}) onde P_{25} e P_{75} são, respectivamente, os percentis 25 e 75 da distribuição. Porém uma observação pode não ser um *outlier* univariado, e ainda assim ser um *outlier* multivariado.
ii. Medidas multivariadas para a detecção de outliers multivariados geralmente utiliza-se a Distância de Mahalanobis. Esta estatística mede a distância de uma observação x_i à média de todas as observações de todas as variáveis ($\bar{x}$). (MARÔCO, 2010).

3. ETAPAS DA MODELAGEM DE EQUAÇÕES ESTRUTURAIS

É possível dividir a Modelagem de Equações Estruturais em algumas etapas que vão ser apresentadas ao longo deste Capítulo, conforme é apresentado na Figura 3.

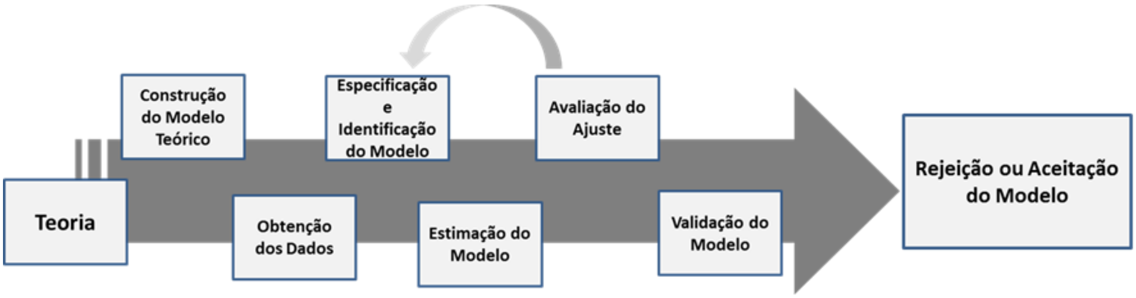


Figura 3. Etapas da Modelagem de Equações Estruturais
[MARÔCO, 2010]

Como se pode observar, o Modelo de Equações Estruturais necessita de vários passos para chegar ao resultado final, e, além disso, necessita de cálculos relativamente complexos para sua estimação. Por conta dessa complexidade, para que as análises sejam consistentes, os dados recolhidos devem obedecer a uma estratégia de análise bem definida e estabelecida a priori.

3.1 Construção do Modelo Teórico

Nesta etapa do processo é necessário o levantamento de informações que constituirão o embasamento teórico da pesquisa. Baseando-se nessas informações, o pesquisador construirá o modelo teórico que será testado e confirmará ou não a relação entre as variáveis do modelo. Essa é uma fase muito sensível, pois a intensidade e convicção com que o pesquisador pode assumir causalidade não dependem dos métodos de análise escolhidos, mas do referencial teórico utilizado.

O que acontece na construção de alguns modelos teóricos é a omissão de variáveis preditoras fundamentais, esse erro é conhecido como erro de especificação. Quando ocorre um erro de especificação, a avaliação da importância de outras variáveis fica prejudicada. Porém o desejo de incluir todas as variáveis deve ser equilibrado com as limitações práticas de SEM. Mesmo não existindo um limite teórico para o número de variáveis no modelo, deve-se preservar a parcimônia e reconhecer seus benefícios e os de modelos teoricamente concisos (HAIR *et al*., 2005).

3.2 Obtenção dos Dados

Depois de estabelecido o modelo teórico que vai ser testado, o próximo passo é obter os dados para a análise. Nessa etapa o instrumento de medida deve ser definido, bem como as variáveis que vão ser mensuradas e também o tamanho da amostra.

O tamanho de amostra ideal não é consenso na literatura, mas existem várias suposições para o tamanho de amostra que deve ser utilizada em SEM. Autores indicam que deve haver no mínimo cinco respondentes para cada parâmetro estimado, sendo considerada mais adequada uma proporção de dez respondentes para cada parâmetro estimado (HAIR *et al*., 2005). De uma forma geral, é necessária uma amostra maior para a Modelagem de Equações Estruturais do que para outras técnicas para assegurar que exista variabilidade suficiente para estimar os parâmetros do modelo.

3.3 Especificação e Identificação do Modelo

A especificação do modelo é uma das partes mais complexas em SEM. Nesta fase são decididas quais variáveis observadas constituem as variáveis latentes, as relações causais e não causais entre as variáveis que devem ser incluídas ou excluídas do modelo e a correlação dos erros ou resíduos. Tudo isso definido de forma que o modelo teórico constituído anteriormente abranja as hipóteses a priori do pesquisador e do referencial teórico (MARÔCO, 2010).

As regras de especificação que devem ser observadas variam de modelo para modelo. O modelo de medida apresenta que as variáveis latentes, ou construtos teóricos, causam as variáveis observadas. A variância das variáveis observadas que não é explicada pelas variáveis latentes é explicada pelos erros de medida ou pelos resíduos. Esses erros podem ser correlacionados, o que indica que é uma fonte de variação comum dos itens não explicada pelos construtos latentes (LEMKE, 2005).

O modelo estrutural considera que as relações são desenhadas sempre de causa para efeito e que a variância das variáveis exógenas, que não é explicada pelas variáveis endógenas, é explicada por erros associados às variáveis latentes (MARÔCO, 2010). Cada variável latente que possui uma ou mais retas que vão até ela é considerada variável dependente em uma equação separada. Isso significa que as variáveis preditoras estão todas nos terminais extremos das setas que conduzem até a variável endógena (HAIR *et al*., 2005).

Como foi comentado na seção 3.1, a inclusão ou omissão de variáveis relevantes para explicar as correlações do modelo conduz a erros de especificação. Estes erros podem ser classificados em dois tipos. O primeiro tipo é quando existem mais variáveis no modelo ou relações entre as variáveis do que aquelas que são possíveis de estimar pelos dados, e o segundo tipo quando o modelo tem menos variáveis do que as necessárias para explicar as relações entre as variáveis. Esses erros podem levar a estimativas viesadas dos parâmetros do modelo ou até mesmo impedir sua estimação.

O modelo é não identificado quando o número de parâmetros a estimar é superior ao número de dados, ou seja, quando o número de graus de liberdade do modelo é negativo. Os graus de liberdade do modelo são calculados através da fórmula:

$$\mathrm{g.l.} = \frac{(p+q)(p+q+1)}{2} - t$$

onde:

p é o número de variáveis dependentes observadas.

q é o número de variáveis independentes observadas.

t é o número de parâmetros a estimar.

Quando o modelo é não identificado, ou seja, quando têm graus de liberdade negativos, ele é não estimável. Sendo o modelo não identificado, deve-se então assumir alguma hipótese sobre as variáveis latentes ou sobre os parâmetros a estimar. Para resolver esse problema é possível alterar a definição de quais são os parâmetros livres e os fixos.

Parâmetros livres são aqueles que vão ser estimados a partir das covariâncias e variâncias das variáveis observadas. Os parâmetros fixos são definidos pelo pesquisador a partir de seu conhecimento a priori ou considerações teóricas sobre o valor do parâmetro.

Há também o tipo em que os parâmetros são estimáveis, porém definidos iguais entre si. Imaginando um modelo em que se fixem os pesos fatoriais dos erros em um (1), isso indicará que as variáveis latentes tem a mesma métrica das correspondentes variáveis observadas. Quando essas mudanças são definidas pelo pesquisador as trajetórias são consideradas significativas, ou seja, não há teste para significância. Quando o investigador não conhece a priori qual o item que deve fixar como significativo, essa prática é desaconselhada. Para explicações mais detalhadas recomenda-se a leitura de Bollen (1989).

A identificação do modelo pode ser definida em *underidentified, just-identifiede overidentified* (LEMKE, 2005). Modelos *underidentified* são modelos onde os graus de liberdade são negativos, indicando que o número de parâmetros que devem ser estimados é superior às informações presentes nas variáveis observadas (variâncias e covariâncias). Os modelos indeterminados possuem infinitas soluções possíveis e por isso não é viável.

Nos modelos *just-identified* os graus de liberdade são iguais a zero. Como o cálculo das estimativas dos parâmetros no modelo saturado utiliza todas as informações disponíveis não é possível testar hipóteses sobre o modelo, tendo ele só uma solução. Dessa forma seu ajustamento é perfeito.

Overidentified são modelos cujo número de graus de liberdade é maior que zero e por isso a qualidade do ajustamento pode ser avaliada. Porém os modelos devem passar por alguma modificação ou restrição teórica para viabilizar essa avaliação. Em SEM o objetivo é especificar esse tipo de modelo.

O número de parâmetros do modelo é uma informação de extrema importância na Modelagem de Equações Estruturais. É importante ter essas informações bem definidas antes de proceder a análise, pois através dessa informação é possível saber se o modelo testado é ou não estatisticamente identificável.

3.4 Estimação do Modelo

Nessa fase os parâmetros do modelo são estimados de forma que representem da melhor maneira possível os dados observados. Tem-se uma matriz de covariâncias que é constituída pelas variáveis observadas e através dessa matriz são feitas as estimativas. A estimação na Modelagem de Equações Estruturais tem o objetivo de encontrar estimativas para os pesos fatoriais, covariâncias, médias e parâmetros do modelo, de uma forma que maximizem a probabilidade de observar a estrutura correlacional da amostra. (MARÔCO, 2010).

Segundo o referencial teórico, o modelo de medida e o modelo estrutural são definidos. Se esses modelos estiverem corretos, os dados gerados serão próximos dos dados observados e o erro será mínimo, indicando que a diferença entre a matriz de covariâncias encontrada na amostra e a matriz de covariâncias predita pelo modelo será mínima. O objetivo é, portanto, encontrar o vetor θ dos parâmetros do modelo que melhor reproduza a matriz de covariâncias observadas. O vetor de parâmetros θ pode ser estimado por métodos exatos ou quase exatos, porém essa estimação é um tanto complexa. Alguns *softwares* utilizam algoritmos iterativos como Newton-Raphson, mínimos quadrados ponderados, Fletcher & Powell entre outros (BOLLEN, 1989 *apud* MARÔCO, 2010).

3.4.1 Método de Máxima Verossimilhança (ML)

É o método mais utilizado, estimando os parâmetros que maximizam a verossimilhança de observar a matriz de covariâncias amostrais S. Ele produz estimativas dos parâmetros centradas e consistentes, à medida que a dimensão da amostra aumenta, aproximando-se do verdadeiro valor do parâmetro, com distribuição normal. O método ML é robusto quanto à suposição de normalidade se a assimetria e curtose das distribuições das variáveis observadas não forem grandes (MARÔCO,2010).

3.4.2 Mínimos Quadrados Generalizados (GLS)

É um método iterativo que estima os parâmetros, ponderando os erros da matriz residual E (que é a diferença entre a matriz de covariâncias amostral (S) e a matriz estimada de covariâncias ($\Sigma(\theta)$)). Os maiores valores da matriz E, ou seja, os que têm maior variância amostral têm menor peso no modelo. Assim, as estimativas são mais eficientes do que as que seriam obtidas sem a ponderação.

O método dos mínimos quadrados generalizados possui as mesmas propriedades assintóticas do método ML, porém pode ser utilizado com suposições menos restritivas quanto à normalidade. Entretanto não é mais utilizado que o método ML, pois tem maiores probabilidades do erro tipo I no teste χ^2 e produz estimativas incorretas dos parâmetros com maior frequência do que o método ML (OLSSON, TROYE, & HOWELL, 1999).

3.4.3 Distribuição Assintótica Livre ou Mínimos Quadrados Ponderados

É um método que não exige que as variáveis observadas apresentem normalidade multivariada. Em função disto, é um tanto atrativo justamente por não precisar cumprir essa suposição, principalmente em aplicações nas ciências sociais e humanas, onde muitas vezes é difícil verificar a normalidade, em função do tipo de dados coletados.

Fornece estatísticas de teste à qualidade global do modelo com distribuição assintótica χ^2 e erros padrão dos parâmetros não viesados. Porém, apesar de ser atrativo, o cálculo é relativamente complexo, pois sua dimensão aumenta exponencialmente em relação ao número de variáveis observadas no modelo.

3.5 Avaliação do Ajuste do Modelo e Validação do Modelo

Nesta etapa é analisado o quão bem o modelo teórico consegue reproduzir a estrutura correlacional das variáveis observadas na amostra. Existindo inúmeros índices de avaliação do modelo e estatísticas que podem ser utilizadas.

A avaliação do modelo é realizada através do teste χ^2 de ajustamento e também através de alguns índices que em sua maioria são utilizados em todos os *softwares* de Modelagem de Equações Estruturais.

3.5.1 Teste χ^2de Ajustamento

É um teste de ajustamento da função de discrepância minimizada durante o ajustamento do modelo. As hipóteses estatísticas do teste são:

H_0: $\Sigma = \Sigma\left(\hat{\theta}\right)$ (a matriz de covariância populacional é igual à matriz de covariância estimada pelo modelo)

H_1: $\Sigma \neq \Sigma\left(\hat{\theta}\right)$ (a matriz de covariância populacional é diferente da matriz de covariância estimada pelo modelo)

A utilização do teste χ^2 requer alguns cuidados, pois é um teste altamente sensível ao tamanho da amostra. Para amostras pequenas, o teste tem probabilidades maiores de não rejeitar a hipótese de que o modelo se ajusta bem aos dados quando o ajustamento é ruim (erro tipo II). Para amostras grandes tem probabilidades maiores de rejeitar a hipótese de que o modelo se ajusta bem aos dados, quando o ajustamento é bom (erro tipo I).

O teste do χ^2 pode ser corrigido de forma a considerar a distribuição amostral não central da estatística do teste quando a normalidade multivariada não é valida. Essa correção é conhecida como Correção de Satorra-Bentler, porém ainda não se encontra disponível em todos os *softwares* de SEM. Para maiores informações sobre a Correção Satorra-Bentler recomenda-se a leitura de Marôco (2010) ou Bentler (1995a e 1995b).

3.5.2 Índices de Avaliação do Ajuste do Modelo

Por conta das limitações do teste χ^2, os pesquisadores desenvolveram alguns índices de avaliação do ajuste do modelo (*goodness-of-fit*). Os índices aqui apresentados são baseados em Byrne (2001), Lemke (2005) e Silva (2006).

CMIN (χ^2) é o valor mínimo de discrepância.

CMIN/DF é a razão χ^2 com os graus de liberdade, que deve ser < 5.

Root Mean Square Residual (RMR) é a raiz quadrada da matriz dos erros dividida pelos graus de liberdade, assumindo que o modelo ajustado é o correto (Joreskog & Sorbom,1996 *apud* MARÔCO,2010). Quanto menor o RMR melhor será o ajustamento, com RMR = 0 tem-se um ajustamento perfeito. Para um modelo bem ajustado, o valor deve ser pequeno, 0,05 ou menos.

Goodness-of- Fit Index (GFI) explica a proporção da covariância observada entre as variáveis observadas, explicada pelo modelo ajustado (MARÔCO, 2010). Portanto é a medida da quantidade relativa de variância e covariância em S (matriz amostral de covariância dos escores das variáveis observadas) que é conjuntamente explicada por Σ (matriz populacional de covariância). O índice tem amplitude de 0 a 1, sendo que valores perto de 1 são indicativos de bom ajuste. Pode ser escrito como indicado abaixo.

$$GFI = 1 - \frac{\hat{F}}{\hat{F}_b}$$

onde $\hat{F}$ é o mínimo da função de discrepância generalizada depois do modelo ter sido ajustado e $\hat{F}_b$é a função de discrepância antes de qualquer modelo ser ajustado.

*Adjusted Goodness-of-fit (*AGFI) é um índice parecido com o GFI, porém esse índice é ajustado pelo número de graus de liberdade do modelo que foi especificado, ou seja, ela acaba punindo parâmetros adicionais no modelo. A equação do AGFI é:

$$AGFI = 1 - (1 - GFI)\frac{d_b}{d}$$

onde:

$d_b = \Sigma_{(g=1)}^{G} p^{*(g)}$

$p^{*(g)}$ é o número de momentos amostrais no grupo g

$d = p - q$, sendo o número de graus de liberdade do modelo

Parsimony Goodness-of-fit Index (PGFI) é um índice que leva em conta o número de parâmetros estimados do modelo teórico na avaliação geral do modelo. Geralmente seus valores são baixos. Pode-se obter o PGFI através da equação:

$$PGFI = GFI\,\frac{d}{d_b}$$

onde:

d é o número de graus de liberdade do modelo avaliado

$d_b = \Sigma_{(g=1)}^{G} p^{*(g)}$ são os graus de liberdade do modelo de referência

Normed Fit Index (NFI) avalia a percentagem de incremento na qualidade do ajustamento do modelo ajustado relativamente ao modelo totalmente independente ou ao pior modelo possível (MARÔCO, 2010). É um índice que varia de 0 a 1, e é decorrente da comparação entre o modelo teórico e o modelo de independência (no modelo de independência todas as correlações entre as variáveis são zero). Valor acima de 0,90 indica bom ajuste do modelo. Pode-se obter o NFI através da equação:

$$NFI = 1 - \frac{\hat{C}}{\hat{C}_b} = 1 - \frac{\hat{F}}{\hat{F}_b}$$

onde:

$\hat{C} = n\hat{F}$ é a mínima discrepância do modelo avaliado

$\hat{C}_b = n\hat{F}_b$ é a mínima discrepância do modelo de referência

Comparative Fit Index (CFI) foi criado para corrigir o erro do NFI que apresentava certa tendência de subestimar o ajuste em amostras pequenas. O CFI leva em conta o tamanho da amostra. É a razão entre o ajustamento do modelo em estudo e o ajustamento do pior modelo possível. Valor acima de 0,90 indica bom ajuste do modelo. Obtém-se o CFI através da equação:

$$CFI = 1 - \frac{\max\left(\hat{C} - d, 0\right)}{\max\left(\hat{C}_b - d, 0\right)} = 1 - \frac{NCP}{NCP_b}$$

Relative Fit Index (RFI) também varia de 0 a 1, valores preferencialmente acima de 0,95 indicam um bom ajuste.

Tucker-Lewis Index (TLI) também conhecido com NNFI, assume valores entre 0 e 1, com valores próximos a 0,95 (para amostras grandes) indicando bom ajuste.

Já os índices de parcimônia levam em consideração a complexidade do modelo. Abaixo são apresentados esses índices.

Parsimony ratio (PRATIO) é a razão de parcimônia inicial. Não é um teste, mas é utilizado na construção de outras medidas, como PNFI e PCFI. É obtido através da equação:

$$PRATIO = \frac{d}{di}$$

onde:

$di =$ número de graus de liberdade do modelo de independência

PNFI é um índice contado de forma relativa ao índice NFI. Penaliza o NFI pela razão de parcimônia.

$$PNFI = NFI\ x\ {}^{d}/{}_{d_b}$$

PCFI é um índice contado de forma relativa ao índice CFI. Penaliza o CFI pela razão de parcimônia.

$$PCFI = CFI\ x\ {}^{d}/{}_{d_b}$$

Outros índices comparam o ajustamento do modelo obtido com os momentos amostrais relativamente ao ajustamento do modelo que se obteria com os momentos populacionais. A função de discrepância para o modelo ajustado é estimada assumindo que os momentos populacionais são conhecidos através dos momentos amostrais. Quando o ajustamento do modelo a partir dos momentos amostrais não é perfeito, obtém-se uma função χ^2 não central. Esses índices avaliam se o modelo ajustado é aproximadamente correto comparando o ajustamento obtido na amostra com o ajustamento que se obteria se o mínimo da função de discrepância fosse obtido a partir dos momentos populacionais (MARÔCO, 2010)

Noncentrality Parameter (NCP) é valor do parâmetro de não centralidade que representa o χ^2 menos seus graus de liberdade. Quando o modelo teórico está incorreto o χ^2 tem uma distribuição não central, com um parâmetro de não centralidade λ. O intervalo de confiança indica que com 90% de confiança o intervalo contém verdadeiro valor (populacional) do parâmetro de não centralidade (λ).

FMIN é a função de discrepância mínima.

FO é a discrepância populacional, fornecendo um intervalo de 90% de confiança em torno de FO.

RMSEA é um dos critérios reconhecidos como mais explicativos na modelagem em estruturas de covariâncias, levando em conta o erro de aproximação na população. Essa discrepância medida pelo RMSEA é expressa por graus de liberdade, sendo sensível ao número de parâmetros estimados no modelo, dependendo da sua complexidade. Valores que indicam bom ajuste ficam abaixo de 0,05. Valores que representam erros razoáveis na aproximação com a população são maiores que 0,08. Já entre 0,08 e 0,10, os valores indicam um ajuste medíocre, e maiores que 0,10, um ajuste pobre. O RMSEA tende a rejeitar modelos verdadeiros em função de a amostra ser pequena, por isso esses critérios devem ser utilizados com cuidado e não podem ser considerados infalíveis. RMSEA pode ser obtido através das equações:

$$RMSEA = \sqrt{\frac{F_0}{d}} \text{ populacional}$$

$$RMSEA = \sqrt{\frac{\hat{F}_0}{d}} \text{ estimado}$$

onde:

PCLOSE é um índice que testa a proximidade do ajuste do RMSEA, ou seja, testa a H_0: RMSEA $\leq 0{,}05$. Preferencialmente o índice deve ser maior que 0,50.

Akaike's Information Criteria (AIC) é um critério que utiliza a parcimônia na avaliação do modelo, levando em conta o número de parâmetros estimados. É usado quando são comparando dois ou mais modelos. O modelo que apresenta melhor ajuste do modelo é o que possuir menor AIC.

$$AIC = \hat{C} + 2q$$

onde q é o número de parâmetros do modelo.

CAIC é um critério que leva em conta o tamanho da amostra, operando da mesma forma que o AIC.

Browne-Cudeck Criterion (BCC) é um critério que funciona da mesma maneira que o AIC e CAIC com diferença que impõem grandes penalidades para a complexidade do modelo.

Expected Cross-Validation Index (ECVI) é um índice que fornece uma medida da discrepância entre a matriz de covariância ajustada na amostra e a matriz de covariância esperada que poderia ser obtida em outra amostra de mesmo tamanho.

Critical N (CN) é também chamado índice de Hoelter. Neste critério, o foco é a adequação do tamanho da amostra, indicando se o tamanho da amostra é suficientemente adequado ao ajuste o modelo para o teste χ^2. Um valor que exceda 200 é indicativo que o modelo representa adequadamente os dados amostrais.

Nesta seção forma abordados vários índices que indicam o ajuste do modelo, porém não sendo necessária a utilização de todos na análise de um modelo. Eles servem para dar uma ideia de quão bem o modelo teórico se ajusta com os dados amostrais. Todos os índices operam diferentemente de acordo com o tamanho da amostra, parcimônia do modelo, método de estimação e violações dos pressupostos, tornando difícil a tarefa de escolha dos índices que se deve utilizar (BYRNE, 2001).

Mesmo com a natureza problemática da estatística χ^2de ser sensível a amostras grandes, não se deve utilizar exclusivamente os índices de ajuste. O modelo pode apresentar um bom ajuste segundo os índices, porém, mesmo assim, pode ter sido especificado de maneira incorreta, podendo não refletir a extensão da aceitabilidade do modelo, sendo de inteira responsabilidade do pesquisador julgar essa aceitabilidade. Assim, pode-se concluir que a avaliação da adequação do modelo deve ser baseada em múltiplos critérios, que envolvem considerações teóricas, estatísticas e práticas (BYRNE, 2001).

Existem algumas formas de verificar a existência da falta na especificação do modelo. Como exemplo de possíveis formas de verificar, pode-se utilizar os Resíduos Padronizados e os Índices de Modificação. Os resíduos padronizados indicam as diferenças entre a matriz de covariância implícita do modelo teórico e a matriz de covariância amostral, capturadas pela matriz residual de covariância. Os resíduos devem ser utilizados sempre na forma padronizada para não tornar sua interpretação complexa por dependerem da unidade de medida das variáveis observadas. Os resíduos padronizados são os resíduos divididos pelos seus erros padrões assintóticos. Eles representam uma estimativa do número de desvios padrões que os resíduos observados estão do resíduo zero, que existiria caso o modelo fosse perfeitamente ajustado.

A maioria dos *softwares* apresenta um módulo chamado *Modification Indices* ou Índices de Modificação, que auxiliam na reespecificação do modelo, indicando quanto o modelo hipotético está adequadamente descrito. Esses índices podem ser vistos como a estatística χ^2 com um grau de liberdade. Para cada parâmetro especificado, fixo, um valor é dado, e representa a queda esperada no valor do χ^2 geral se o parâmetro não fosse mais fixo, e sim livremente estimável, numa próxima vez em que o modelo fosse executado. Todos os parâmetros livremente estimados automaticamente têm valores iguais a zero. Na Figura 4 é apresentado um resumo dos principais índices descritos e seus valores de referência.

Estatística	**Valores de Referência**
Qui-quadrado e p-valor	Quanto menor, melhor $p > 0{,}05$
Qui-quadrado/ graus de liberdade	>5 - Ajustamento mau 2 a 5 -Ajustamento razoável 1 a 2 - Ajustamento bom ~ 1 - Ajustamento muito bom
CFI GFI TLI	< 0,8 - Ajustamento mau 0,8 a 0,9 - Ajustamento razoável 0,9 a 0,95 - Ajustamento bom ≥ 0,95 - Ajustamento muito bom
PGFI PCFI	<0,6 - Ajustamento mau 0,6 a 0,8 - Ajustamento bom ≥0,8 - Ajustamento muito bom
RMSEA	> 0,10 - Ajustamento inaceitável 0,05 a 0,10 - Ajustamento bom ≤ 0,05 - Ajustamento muito bom
AIC BCC ECVI	Comparação de modelos, quanto menor o valor, melhor o modelo.

Figura 4 – Estatísticas e índices de qualidade do ajustamento
[adaptado de Byrne (2001) e Mâroco (2010)]

3.5.3 Validação do Modelo

Os modelos ajustados com a utilização do recurso dos índices de modificação necessitam validação numa amostra independente daquela utilizada para o ajuste do modelo. Mas muitas vezes esse teste não é viável pela dimensão dos dados. Uma outra alternativa é a utilização do ECVI que foi proposta por Browne e Cudeck (1989). O intervalo de confiança do ECVI é utilizado como indicador da validade dos modelos alternativos na população a partir de uma amostra. Esses autores indicam que os modelos com menor valor de ECVI são os mais estáveis na população. Outra forma de realizar a validação do modelo é através da simulação por Bootstrap.

4. O SOFTWARE R E O PACOTE LAVAAN

Existem alguns *softwares* construídos especialmente para realizar a Modelagem de Equações Estruturais, sendo, em sua maioria, comerciais. Porém, também existem *softwares* livres. O R é um *software* livre que abrange muitas técnicas estatísticas e que está sempre em crescimento por contar com a contribuição contínua dos seus usuários.

Lavaan é um pacote construído para ser utilizado através do *software* R. Foi desenvolvido para fornecer aos pesquisadores aplicados, professores e estatísticos, uma ferramenta totalmente livre com tanta qualidade quanto os *softwares* comerciais, para modelagem de variáveis latentes. O pacote está disponível no *Comprehensive R Archive Network* (CRAN) em http://CRAN.R-project. org / package = *lavaan* e também pelo site http://*lavaan*.org/.

4.1 O Pacote *Lavaan*

Lavaan é a sigla para *latent variable analysis* (análise de variável latente). O pacote tem o objetivo de fornecer um conjunto de ferramentas que podem ser usadas para explorar, avaliar e compreender modelos que envolvem variáveis latentes, incluindo análise fatorial, de equações estruturais, longitudinal, entre outros (ROSSEEL, 2012).

O pacote foi construído visando alguns aspectos que muitas vezes são essenciais tanto na Modelagem de Equações Estruturais, quanto na escolha do *software*. Um dos aspectos que procura suprir é que geralmente os programas livres para a modelagem não são intuitivos ou não possuem muitos recursos de análise. O pacote *Lavaan* tenta preencher esses objetivos. Como é de fácil obtenção e possui fácil instalação, pode ser usado por qualquer estudante ou professor em suas aulas. Outro fator atraente é que por ser um programa livre é interessante para estatísticos que trabalham na área de SEM e gostariam de programar novas metodologias, pois permite acesso direto ao código do pacote.

Segundo Rosseel (2012), a necessidade dos pesquisadores de obter um *software* comercial para realizar as análises acabou levando muitos a desistir ou mudar completamente seu objetivo de pesquisa. É lamentável que novos desenvolvimentos na área possam ter se prejudicado pela falta de um *software* livre ou que pudesse ser adaptado. É essa a justificativa que o autor utiliza para manter o pacote *Lavaan* totalmente livre.

4.2 Especificando Modelos no *Lavaan*

Os modelos são especificados através de uma sintaxe simples e fácil de entender. Considere um modelo de regressão simples com uma variável dependente y e quatro variáveis independentes x_1, x_2, x_3, x_4. O modelo de regressão pode ser escrito como:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \epsilon_i$$

onde:

β_0 é o intercepto, os betas de 1 a 4 são os coeficientes de regressão para cada uma das quatro variáveis, e

ϵ_i é o erro residual para a observação i.

Ao construir a mesma equação de regressão acima no ambiente R, obtém-se então:

```
y ~ x1+x2+x3+x4
```

O sinal de til (~) é o operador de regressão. Do lado esquerdo do operador está a variável dependente (y) e no lado direito as variáveis independentes. O intercepto e o termo de erro residual não estão explicitamente inclusos na equação. Porém quando uma função é utilizada ambos são estimados. A lógica é que o termo de erro e o intercepto são quase sempre parte de um modelo de regressão e não há necessidade de mencioná-los na fórmula de regressão (ROSSEEL, 2012).

Como na Modelagem de Equações Estruturais tem-se um conjunto de equações de regressão em que as variáveis independentes (exógenas) em uma equação podem se tornar dependentes (endógenas) em outra equação, pode-se utilizar o mesmo método para especificar as equações do modelo de SEM, com uma diferença, as variáveis latentes devem ser construídas com o operador =~. Na equação abaixo f1, seria uma variável latente.

```
y1 ~ x1+x2+x3+x4

f1 =~ item1+item2+item3
```

Lavaan fornece o operador ~~ para indicar variâncias e covariâncias na sintaxe do modelo. Se o nome ao lado esquerdo do operador é igual ao nome ao lado direto é uma variância, se os nomes diferem tem-se uma covariância.

```
Item1 ~~ Item1 # variância

Item1 ~~ Item2 # covariância
```

A Figura 5 apresenta os operadores utilizados na definição do modelo no pacote *Lavaan*.

Tipo	Operador
Variável Latente	=~
Regressão	~
(Co) Variância Residual	~~
Intercepto	~1
Parâmetro Definido	:=
Restrição de Igualdade	==
Restrição de Desigualdade	<
Restrição de Desigualdade	>

Figura 5 – Operadores utilizados na definição do modelo

[adaptado de Rosseel (2012)]

A função *cfa()* é uma função específica para a análise de modelos fatoriais confirmatórios. O primeiro objeto da função *cfa()* é o objeto que contém a sintaxe do modelo; o segundo é o conjunto de dados que contém as variáveis observadas. A função *cfa()* define automaticamente as variâncias e covariâncias das variáveis observadas e latentes. No exemplo a seguir, a variável latente ‘Visual’ é explicada pelas variáveis observadas x1, x2 e x3 enquanto a variável latente ‘ Textual’ é explicada pelas variáveis observadas x4, x5 e x6.

```
HS.model = ~ ‘Visual=~ x1+x2+x3

Textual =~x4+x5+x6’

fit = cfa(HS.model, data=dados)
```

A função *sem()* é muito semelhante a função *cfa()*. Na verdade, as duas são quase idênticas; mas, segundo Rosseel (2012), isto pode mudar no futuro. Já a função *lavaan()* não acrescenta parâmetros adicionais para o modelo, tornando-se responsabilidade do usuário especificar o modelo corretamente, podendo levar a extensas sintaxes.

4.3 Resultados dos Modelos

Tanto para a função *cfa()*, *Lavaan()* ou *sem()* existem diversas maneiras para examinar as estatísticas e parâmetros estimados do modelo ajustado. Talvez a função *sumary()* seja a mais útil. Se for chamada sem argumentos extras apresentará um breve resumo do modelo ajustado juntamente com estimativas dos parâmetros (Rosseel, 2012). Abaixo são abordadas brevemente as funções que apresentam os resultados do modelo:

summary(): apresenta um longo resumo dos resultados do modelo.

show(): apresenta um breve resumo dos resultados do modelo.

coef(): apresenta as estimativas dos parâmetros livres no modelo.

fitted(): apresenta os momentos implícitos (matriz de covariância e vetor médio) do modelo.

resid():apresenta os resíduos brutos, normalizados ou padronizados (diferencia entre momentos implícitos e observados).

vcov(): apresenta a matriz de covariância dos parâmetros estimados.

AIC() e BIC(): apresenta os critérios de informação (se a estimativa de máxima verossimilhança foi utilizada).

inspect(): mostra a representação interna do modelo, retornando, por padrão, uma lista de matrizes contendo os parâmetros livres do modelo, mas pode também ser usada para extrair os valores iniciais, os valores de gradiente, e outros.

As funções de ajuste podem ser enriquecidas com mais informações. Os argumentos extras que podem ser utilizados são: *fit.measures*, *standardized* e *rsquare*. Se definidos como *TRUE* a saída será enriquecida com medidas adicionais, estimativas padronizadas e os valores de R^2 para as variáveis dependentes, respectivamente. Na Figura 6 e 7 pode-se ver um exemplo onde apenas *fit.measures* foi definido como *TRUE*.

```
R> HS.model <- 'visual  =~ x1 + x2 + x3
+               textual =~ x4 + x5 + x6
+               speed   =~ x7 + x8 + x9'
R> fit <- cfa(HS.model, data = HolzingerSwineford1939)
R> summary(fit, fit.measures = TRUE)

lavaan (0.4-14) converged normally after 41 iterations

  Number of observations                           301

  Estimator                                         ML
  Minimum Function Chi-square                   85.306
  Degrees of freedom                                24
  P-value                                        0.000

Chi-square test baseline model:

  Minimum Function Chi-square                  918.852
  Degrees of freedom                                36
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    0.931
  Tucker-Lewis Index (TLI)                       0.896

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -3737.745
  Loglikelihood unrestricted model (H1)      -3695.092

  Number of free parameters                         21
  Akaike (AIC)                                7517.490
  Bayesian (BIC)                              7595.339
  Sample-size adjusted Bayesian (BIC)         7528.739

Root Mean Square Error of Approximation:

  RMSEA                                          0.092
  90 Percent Confidence Interval          0.071  0.114
  P-value RMSEA <= 0.05                          0.001

Standardized Root Mean Square Residual:
```

Figura 6. Saída da função *summary()*

[ROSSEEL,2012].

A saída é constituída por três seções. A primeira seção (as primeiras seis linhas) contém o número da versão do pacote, uma indicação sobre a convergência do modelo (e em quantas iterações) e o número de observações utilizadas na análise. Em seguida, a estatística de teste χ^2, graus de liberdade e o p-valor são impressos. Se *fit.measures* = *TRUE*, a segunda seção é impressa contendo a estatística de teste do modelo de base (onde todas as variáveis observadas são assumidas como não correlacionadas) e vários índices de ajustamento populares. Se a estimativa de máxima verossimilhança é usada, também haverá informações sobre o *Loglikelihood*, o *AIC* e o *BIC*.

```
  SRMR                                           0.065

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                1.000
    x2                0.553    0.100    5.554    0.000
    x3                0.729    0.109    6.685    0.000
  textual =~
    x4                1.000
    x5                1.113    0.065   17.014    0.000
    x6                0.926    0.055   16.703    0.000
  speed =~
    x7                1.000
    x8                1.180    0.165    7.152    0.000
    x9                1.082    0.151    7.155    0.000

Covariances:
  visual ~~
    textual           0.408    0.074    5.552    0.000
    speed             0.262    0.056    4.660    0.000
  textual ~~
    speed             0.173    0.049    3.518    0.000

Variances:
    x1                0.549    0.114
    x2                1.134    0.102
    x3                0.844    0.091
    x4                0.371    0.048
    x5                0.446    0.058
    x6                0.356    0.043
    x7                0.799    0.081
    x8                0.488    0.074
    x9                0.566    0.071
    visual            0.809    0.145
    textual           0.979    0.112
    speed             0.384    0.086
```

Figura 7. Continuação da saída da função *summary()*

[ROSSEEL,2012].

A terceira seção apresenta uma visão geral das estimativas de parâmetros, incluindo o tipo de erros padrão utilizado e se a matriz de informação observada ou esperada foi utilizada para calcular os erros padrão. Então, para cada um dos parâmetros do modelo, a estimativa do erro padrão é exibida, os valores de z baseados no teste de Wald e p-valor.

Embora a função *summary()* forneça um bom resumo dos resultados do modelo, é útil apenas para exibição. Uma alternativa é a função *parameterEstimates()*, que retorna as estimativas de parâmetros como um *data frame*, tornando a informação facilmente acessível para posterior processamento (ROSSEEL,2012). Por padrão, a função *parameterEstimates()* inclui as estimativas, erros padrão, valor z, p-valor e os intervalos de confiança de 95% para todos os parâmetros do modelo.

4.4 Medidas de Ajuste

A função *summary()* quando utilizada com o argumento *fit.measures* = *TRUE* irá imprimir uma série de medidas de ajuste. Se são necessárias estatísticas de ajuste para posterior processamento, utilizar diretamente a função *fitMeasures()* é a melhor alternativa. O primeiro argumento para *fitMeasures()* é o objeto contendo o modelo ajustado e o segundo argumento é um vetor de caracteres contendo os nomes das medidas de ajuste que se deseja extrair. Por exemplo, se somente é necessário o CFI e os valores de RMSEA, pode-se utilizar:

```
R> fitMeasures(fit, c("cfi", "rmsea"))

  cfi rmsea
0.995 0.035
```

Figura 8. Saída da função *fit.measures()* para os índices de ajuste específicos (CFI, RMSEA)
[ROSSEEL,2012].

4.5 Análise de Múltiplos Grupos

O pacote *Lavaan* tem suporte completo para vários grupos SEM. Para solicitar uma análise em vários grupos, a variável que define os grupos nos dados pode ser chamada nas funções *cfa()*, *sem()* ou *Lavaan()*. Por padrão, o mesmo modelo é montado em todos os grupos, sem quaisquer restrições de igualdade sobre os parâmetros do modelo (ROSSEEL,2012).

A função *summary()* mostra, nesses casos, um conjunto de estimativas dos parâmetros para os grupos. Se for preciso impor restrições de igualdade em parâmetros do modelo através de grupos, pode-se utilizar o argumento *group.equal*. Por exemplo, *group.equal* = *c ("loadings", "intercepts")* restringe tanto as cargas fatoriais, quanto os interceptos das variáveis observadas para serem iguais em todos os grupos. Outra alternativa é utilizar o pacote *semTools* 0.4 como auxílio, pois facilita a restrição dos parâmetros.

4.6 Modificação do Modelo

Se o modelo de ajuste não é excelente, pode ser informativo olhar os índices de modificação e suas correspondentes alterações de parâmetros esperados (EPCs). Em essência, os índices de modificação fornecem uma estimativa aproximada de como o teste χ^2 iria melhorar no modelo se um determinado parâmetro fosse alterado. O método *modificationIndices()* (ou *modindices()*) irá imprimir uma longa lista de parâmetros como *data frame* (ROSSEEL,2012). Na saída apresentada na Figura 9 são mostrados apenas os parâmetros para os quais a modificação índice é de 10 ou superior.

```
R> MI <- modificationIndices(fit)
R> subset(MI, mi > 10)

     lhs op rhs     mi    epc sepc.lv sepc.all sepc.nox
1 visual =~  x7 18.631 -0.422  -0.380   -0.349   -0.349
2 visual =~  x9 36.411  0.577   0.519    0.515    0.515
3     x7 ~~  x8 34.145  0.536   0.536    0.488    0.488
4     x8 ~~  x9 14.946 -0.423  -0.423   -0.415   -0.415
```

Figura 9. Saída da função *modificationIndices()* [ROSSEEL,2012].

4.7 Outros Recursos

O pacote Lavaan 0.5 possui suporte para variáveis categóricas, dados não normais (Augmented Dickey-Fuller, Satorra-Bentler, bootstrapping), ANOVA, dados faltantes (missings), igualdade linear e não linear e restrições de desigualdade, efeitos indiretos e análise de mediação, entre outros. Recomenda-se a leitura de Rosseel (2012) para obter mais informações de todos os pontos abordados neste Capítulo 4.

5. APLICAÇÕES DA MODELAGEM DE EQUAÇÕES ESTRUTURAIS

Nesta seção serão apresentados alguns exemplos de aplicações de Equações estruturais utilizando o *Software* R. São exemplificados os usos das técnicas de Regressão Linear Multivariada, Análise Fatorial Confirmatória, Análise de Múltiplos Grupos e, por fim, a Análise de um Instrumento de Medida.

5.1 Regressão Linear Multivariada

Em SEM, Modelos de Regressão Linear tratam-se de modelos hipoteticamente causais e que relacionam as variáveis observadas do modelo. Os erros de medida, nesse caso, são considerados desprezíveis, ou seja, as variáveis preditoras são medidas sem erro. Já as variáveis preditas podem ser medidas com erro, sendo os erros independentes e identicamente distribuídos. Na análise dos modelos de regressão, assume-se que as variáveis tem distribuição normal multivariada. Nestas condições, e com uma amostra de tamanho grande, as estimativas dos parâmetros são eficientes e apresentam distribuição normal. Nos modelos de Regressão Linear Multivariada assumem-se relações lineares entre duas ou mais variáveis dependentes (preditas) e duas ou mais variáveis independentes (preditoras).

Para demonstrar como funciona a Regressão Linear Múltivariada, será utilizado o banco de dados Amitriptyline adaptado de Marôco (2010) que encontra-se no anexo A. A amitriptyline é um antidepressivo que está associado a alguns efeitos colaterais, como arritmia cardíaca, pressão sanguínea anormal e curvas irregulares no eletrocardiograma. Foram coletadas informações de 17 pacientes que apresentaram overdose de amitriptyline. Entre as informações coletadas estavam o nível total de antidepressivos (TTC) e concentração de amitriptyline no plasma (AMI) como variáveis dependentes. As variáveis independentes são gênero (feminino, masculino), quantidade de antidepressivos ingeridos (AntiDepr), pressão diastólica (PrDiast), curva PR e curva QRS (obtidas no exame de eletrocardiograma). Deseja-se avaliar a significância das variáveis independentes e a capacidade explicativa do modelo.

Com o Pacote *Lavaan* instalado e o banco de dados ativos no *software* R, se define o modelo com a seguinte linha de comando:

```
modelo=' TotTCAD ~ Genero+AntiDepr+PR+PrDiast+QRS

AMI~ Genero+AntiDepr+PR+PrDiast+QRS'
```

O modelo de regressão deve ficar semelhante ao apresentado na Figura 10. Para a construção dos gráficos foi utilizado o Pacote *semPlot* 0.3.3. Como setas unilaterais indicam relação de causa, pode-se ver na Figura 10 que o modelo definido indica que cada uma das variáveis dependentes é explicada por todas as variáveis independentes, ou seja, tanto AMI (concentração de amitriptyline no plasma) quanto TTC (nível total de antidepressivos) são explicadas por QRS, PrD (pressão diastólica), PR, AnD (quantidade de antidepressivos ingeridos) e Gnr (gênero).

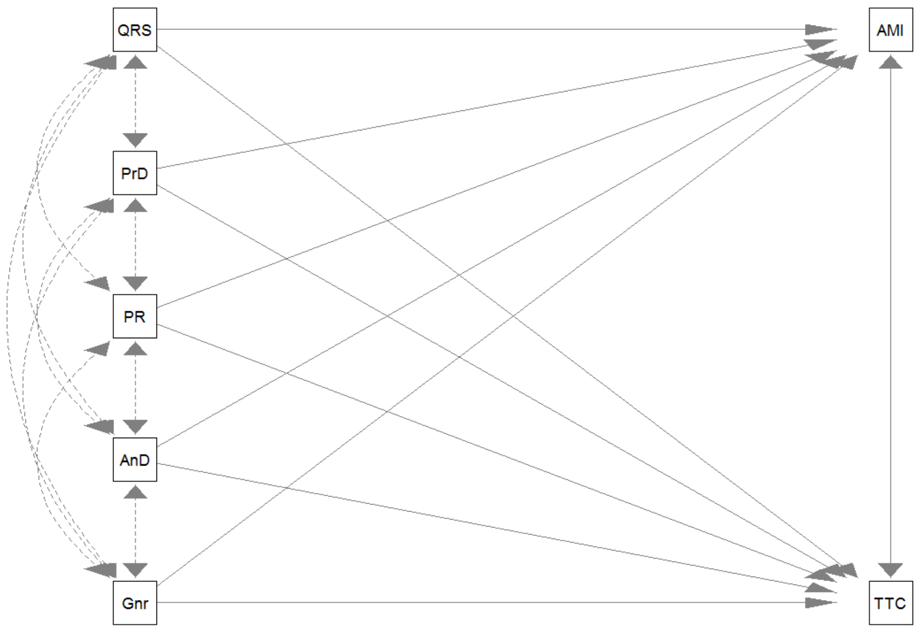


Figura 10. Modelo de regressão multivariada entre as variáveis dependentes e independentes.

Para ajustar o modelo utiliza-se a função *sem()* como pode-se ver abaixo.

```
fit=sem(modelo, data = dados)
```

Com o modelo ajustado, é possível obter as estimativas dos coeficientes de regressão, os valores de R^2 para as variáveis dependentes e a correlação entre as variáveis. Para obter essas estatísticas necessita-se de um resumo do modelo que pode ser obtido através da definição apresentada a seguir.

```
summary(fit, standardized=TRUE, fit.measures=TRUE, rsquare=TRUE)
```

Como *standardized*, *fit.measures* e *rsquare* foram definidos com *TRUE*, medidas adicionais, estimativas padronizadas e os R^2 são apresentados. Na Figura 11 e Figura 12, pode-se ver a extensa saída que o pacote apresenta e, no caso de Regressão Linear Multivariada, a parte mais importante é apresentada a partir de *Regressions*.

```
lavaan (0.5-14) converged normally after 148 iterations

  Number of observations                            17

  Estimator                                         ML
  Minimum Function Test Statistic                0.000
  Degrees of freedom                                 0
  P-value (Chi-square)                           1.000

Model test baseline model:

  Minimum Function Test Statistic               94.041
  Degrees of freedom                                11
  P-value                                        0.000

Full model versus baseline model:

  Comparative Fit Index (CFI)                    1.000
  Tucker-Lewis Index (TLI)                       1.000

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)               -605.479
  Loglikelihood unrestricted model (H1)       -605.479

  Number of free parameters                         13
  Akaike (AIC)                                1236.958
  Bayesian (BIC)                              1247.789
  Sample-size adjusted Bayesian (BIC)         1207.921

Root Mean Square Error of Approximation:

  RMSEA                                          0.000
  90 Percent Confidence Interval          0.000  0.000
  P-value RMSEA <= 0.05                          1.000

Standardized Root Mean Square Residual:

  SRMR                                           0.000

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard
```

Figura 11. Saída da função *summary()* para a Regressão

```
                   Estimate   Std.err  Z-value  P(>|z|)   Std.lv  Std.all
Regressions:
  TotTCAD ~
    Genero          675.651   130.357    5.183    0.000  675.651    0.457
    AntiDepr          0.285     0.049    5.814    0.000    0.285    0.751
    PR               10.272     3.423    3.001    0.003   10.272    0.344
    PrDiast           7.251     2.594    2.795    0.005    7.251    0.328
    QRS               7.598     3.096    2.454    0.014    7.598    0.237
  AMI ~
    Genero          763.030   135.551    5.629    0.000  763.030    0.520
    AntiDepr          0.306     0.051    6.013    0.000    0.306    0.812
    PR                8.896     3.559    2.500    0.012    8.896    0.299
    PrDiast           7.206     2.698    2.671    0.008    7.206    0.328
    QRS               4.987     3.219    1.549    0.121    4.987    0.157

Covariances:
  TotTCAD ~~
    AMI           45039.792 16908.929    2.664    0.008 45039.792    0.846

Variances:
    TotTCAD       51176.959 17553.552                   51176.959    0.113
    AMI           55335.817 18980.029                   55335.817    0.124

R-Square:

    TotTCAD           0.887
    AMI               0.876
```

Figura 12. Continuação da saída da função *summary()* para a Regressão

Na Figura 12, em R-Square pode-se ver o quanto o modelo ajustado explica das variáveis dependentes, nesse caso, o modelo explica 88,7% da variabilidade de TotCAD e 87,6% da variabilidade de AMI. Na última coluna observa-se as estimativas dos coeficientes de regressão padronizados. Nenhum valor é elevado, o que sugeriria a existência de multicolinearidade. Deve-se observar que os testes exigidos para que os resultados da regressão sejam válidos precisam ser realizados, como o cálculo do VIF, a presença de outliers, normalidade, entre outros.

Na Figura 13 são apresentadas as estimativas dos coeficientes de regressão e as correlações nos seus devidos caminhos.

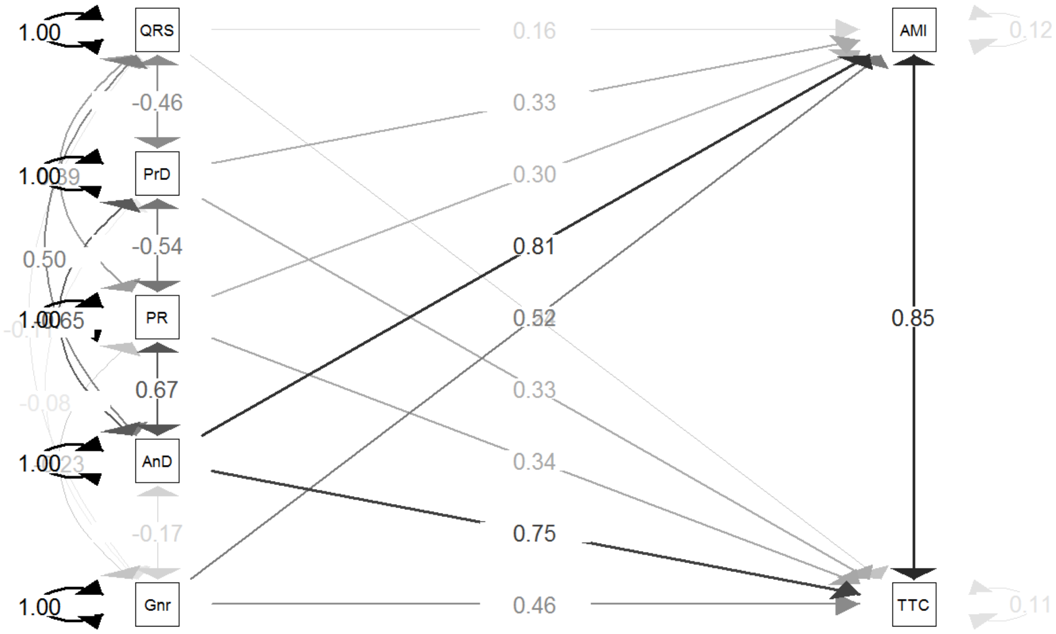


Figura 13. Estimativas dos coeficientes de regressão padronizados e correlações no modelo de regressão linear múltivariada da TotCad e AMI.

Para verificar a significância das trajetórias no modelo utiliza-se a função *parameterEstimates()*, cujo resultado é apresentado na Figura 14. Utilizando um α = 0,05, apenas QRS com AMI não teve um coeficiente significativo (p-valor = 0,121). Por uma questão de parcimônia os caminhos não significativos podem ser removidos do modelo (MARÔCO, 2010). Define-se então:

```
parameterEstimates(fit).
```

	lhs	op	rhs	est	se	z	pvalue	ci.lower	ci.upper
1	TotTCAD	~	Genero	675.651	130.357	5.183	0.000	420.155	931.147
2	TotTCAD	~	AntiDepr	0.285	0.049	5.814	0.000	0.189	0.381
3	TotTCAD	~	PR	10.272	3.423	3.001	0.003	3.564	16.980
4	TotTCAD	~	PrDiast	7.251	2.594	2.795	0.005	2.166	12.336
5	TotTCAD	~	QRS	7.598	3.096	2.454	0.014	1.530	13.666
6	AMI	~	Genero	763.030	135.551	5.629	0.000	497.355	1028.704
7	AMI	~	AntiDepr	0.306	0.051	6.013	0.000	0.207	0.406
8	AMI	~	PR	8.896	3.559	2.500	0.012	1.921	15.872
9	AMI	~	PrDiast	7.206	2.698	2.671	0.008	1.918	12.493
10	AMI	~	QRS	4.987	3.219	1.549	0.121	-1.323	11.297
11	TotTCAD	~~	TotTCAD	51176.958	17553.552	2.915	0.004	16772.629	85581.287
12	AMI	~~	AMI	55335.816	18980.028	2.915	0.004	18135.644	92535.988
13	TotTCAD	~~	AMI	45039.791	16908.928	2.664	0.008	11898.901	78180.681
14	Genero	~~	Genero	0.208	0.000	NA	NA	0.208	0.208
15	Genero	~~	AntiDepr	-139.533	0.000	NA	NA	-139.533	-139.533
16	Genero	~~	PR	-2.388	0.000	NA	NA	-2.388	-2.388
17	Genero	~~	PrDiast	-1.176	0.000	NA	NA	-1.176	-1.176
18	Genero	~~	QRS	-1.066	0.000	NA	NA	-1.066	-1.066
19	AntiDepr	~~	AntiDepr	3147002.595	0.000	NA	NA	3147002.595	3147002.595
20	AntiDepr	~~	PR	26752.422	0.000	NA	NA	26752.422	26752.422
21	AntiDepr	~~	PrDiast	-35170.588	0.000	NA	NA	-35170.588	-35170.588
22	AntiDepr	~~	QRS	18762.543	0.000	NA	NA	18762.543	18762.543
23	PR	~~	PR	507.280	0.000	NA	NA	507.280	507.280
24	PR	~~	PrDiast	-370.353	0.000	NA	NA	-370.353	-370.353
25	PR	~~	QRS	183.785	0.000	NA	NA	183.785	183.785
26	PrDiast	~~	PrDiast	925.176	0.000	NA	NA	925.176	925.176
27	PrDiast	~~	QRS	-293.765	0.000	NA	NA	-293.765	-293.765
28	QRS	~~	QRS	441.792	0.000	NA	NA	441.792	441.792

Figura 14. Coeficientes não padronizados (est), erros padrão (se), valores z (z) e p-valor (*pvalue*) dos testes de significância dos coeficientes do modelo de regressão e covariâncias.

A significância da covariância entre as variáveis também é merecedora de atenção, principalmente as covariâncias entre os erros das variáveis dependentes. Na Figura 14, a covariância entre os erros das variáveis dependentes é mostrado em TotCad~~AMI, indicando que está é estatisticamente significativa (z = 2,644, p-valor = 0,008) e a correlação, apresentada na Figura 12, também é significativa.

É possível, então, concluir que o modelo ajustado explica aproximadamente 89% da variabilidade da variável TotCad e 88% da variável AMI. Apenas o coeficiente de regressão entre a variável explicativa QRS com a variável dependente AMI não apresentou coeficiente significativo (p-valor = 0,121).

5.2 Análise Fatorial Confirmatória

Como já foi comentado anteriormente a Análise Fatorial Confirmatória na Modelagem de Equações Estruturais é usada, geralmente, para avaliar a qualidade de um modelo de medida teórico e a estrutura correlacional observada entre as variáveis manifestas. De um ponto de vista formal, o modelo geral da Análise Fatorial Confirmatória é simplesmente o modelo de medida do modelo de equações estruturais.

O banco de dados que será utilizado está no anexo B e foi adaptado de Marôco (2010) e apresenta uma proposta da medida das preocupações parentais em crianças. Por se tratar de um banco de dados didático a construção do modelo teórico e obtenção dos dados ficaram por conta dos autores. Segundo Algarvio, Leal e Marôco (2008), os principais fatores que dividem essas preocupações são cinco: Problemas familiares e escolares (PrFE); Problemas para comer, dormir e físicos (PrCD); Preparação para mudanças (Prpr); Medo; e Pensamentos negativos (PnsN). O objetivo é testar se o modelo proposto para medir as preocupações parentais se ajusta bem aos dados. A Figura 15 apresenta a especificação do modelo que se deseja testar.

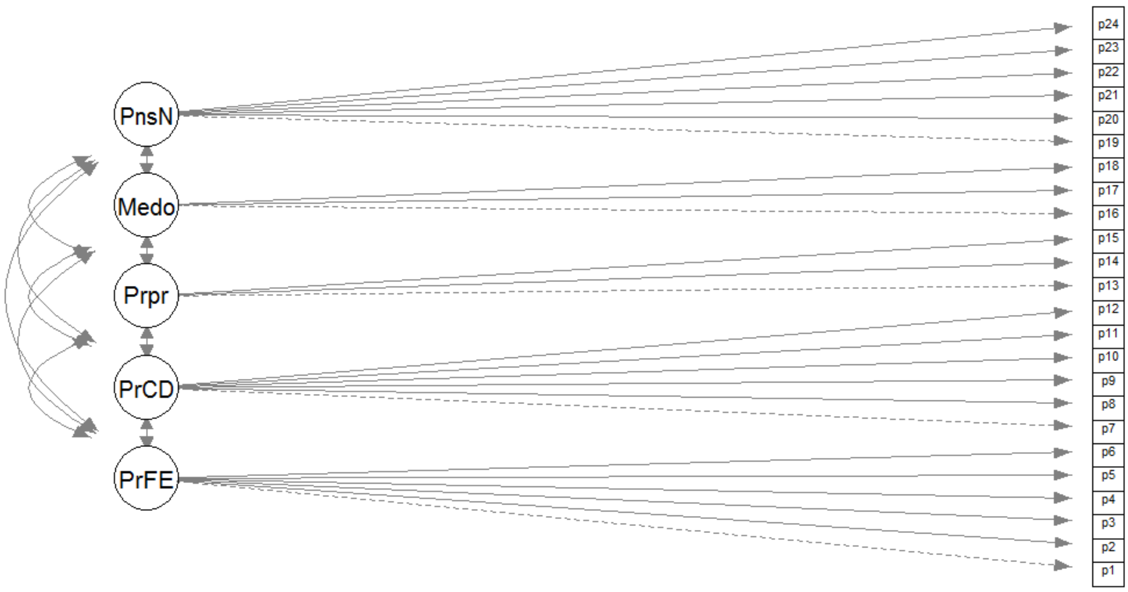


Figura 15. Modelo Fatorial de Preocupações Parentais em Crianças.

Examinando a Figura 15 podem-se listar algumas características:

1- Existem 5 fatores latentes, indicados pelos círculos: PnsN, Medo, Prpr, PrCD e PrFE;

2- Os cinco fatores são intercorrelacionados, indicados pelas setas bidirecionais;

3- Existem 24 variáveis observadas, indicados pelos 24 retângulos (p1-p24);

4- Cada variável observada carrega um e apenas um fator.

As setas unilaterais representam coeficientes de regressão estrutural, que indicam o impacto de uma variável na outra. Na Figura 15, essas setas apontando dos fatores para as suas variáveis observadas sugerem que seus valores são influenciados por seus respectivos fatores. Já as setas bidirecionais representam as correlações ou covariâncias entre os pares de variáveis.

Com o propósito da identificação do modelo e escala da variável latente, a primeira de cada variável observada em cada jogo congenérico (um jogo de medidas é dito congenérico se cada medida acessa o mesmo construto) de medidas associadas a cada fator de AC tem como carga fatorial definida o valor 1,0.

Depois de ativar o banco de dados no *software* R deve-se prosseguir com a definição do modelo, com os seguintes comandos:

```
modelo=' PrFE =~ p1+p2+p3+p4+p5+p6

PrCD=~p7+p8+p9+p10+p11+p12

Prpr=~p13+p14+p15

Medo=~p16+p17+p18

PnsN=~p19+p20+p21+p22+p23+p24
```

Pensamentos negativos (PnsN) é explicado pelas variáveis p24, p23, p22, p21, p20 e p19. Medo é medido pelas variáveis p18, p17 e p16. Preparação para mudanças (Prpr) é medida pelas variáveis p15, p14 e p13. Problemas para comer, dormir e físicos (PrCD) e Problemas familiares e escolares (PrFE) são explicados por p12, p11, p10, p9, p8, p7 e p6, p5, p4, p3, p2, p1 respectivamente. Assim, o modelo especificado a priori indica que as preocupações parentais podem ser explicadas por cinco fatores, que cada item de medida tem carga não zero no fator que foi designado para medir (carga alvo) e carga zero nos demais fatores. Também indica que os cinco fatores de preocupações parentais são correlacionados.

Para ajustar o modelo utiliza-se a função *cfa()* por se tratar de uma análise fatorial confirmatória.

```
fit = cfa(modelo,data = dados)
```

Após o ajuste dos dados, obtêm-se, então, as estimativas. Para obter essas estatísticas necessita-se de um resumo do modelo que pode ser obtido através da função a seguir. A saída dos dados também é apresentada na Figura 16.

```
summary(fit, standardized = TRUE)
```

Quanto à identificação do modelo, obteve-se 48 pesos de regressão, 29 que foram fixados e 19 livres para serem estimados. Os 29 fixados compreendem o peso de regressão de cada fator (linha pontilhada de cada fator) mais os 24 termos de erro. Obteve-se, também, 10 covariâncias (entre os fatores) e 29 variâncias (5 dos fatores e 24 dos erros). No total existem 87 parâmetros dos quais 58 foram estimados. A Figura 16 mostra um resumo geral do modelo que foi testado. Em SEM, os únicos dados com os quais trabalha-se são as variáveis observadas, e neste caso há 24 variáveis. Baseado na fórmula [p(p+1)/2], chega-se ao número 300, referente aos momentos amostrais, obtido pelo cálculo [24(25)/2]. Tem 300 momentos distintos da amostra com os quais são computadas as estimativas do modelo e 58 parâmetros para serem estimados, deixando o modelo com 242 graus de liberdade. A partir desta informação conclui-se que o modelo é *overidentified* com 242 graus de liberdade.

Ainda na Figura 16, existe uma coluna com as estimativas, erros padrão, valor z, p-valor, estimativas padronizadas das variáveis latentes e estimativas padronizadas de todas as variáveis. O método usado na estimação é o Método da Máxima Verossimilhança. Outra maneira de ver as estimativas de todas as variáveis padronizadas é olhando o diagrama de caminhos que é apresentado na Figura 17.

```
lavaan (0.5-14) converged normally after  63 iterations

  Number of observations                           302

  Estimator                                         ML
  Minimum Function Test Statistic              394.437
  Degrees of freedom                               242
  P-value (Chi-square)                           0.000

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)   Std.lv  Std.all
Latent variables:
  ProblemasFamiliaresEscolares =~
    p1                1.000                               0.828    0.796
    p2                0.857    0.068   12.564    0.000    0.709    0.678
    p3                1.073    0.073   14.764    0.000    0.888    0.771
    p4                1.108    0.068   16.179    0.000    0.917    0.827
    p5                0.854    0.051   16.603    0.000    0.707    0.843
    p6                1.029    0.066   15.698    0.000    0.852    0.808
  ProblemasComerDormir =~
    p7                1.000                               0.696    0.694
    p8                0.946    0.090   10.494    0.000    0.658    0.638
    p9                1.399    0.101   13.901    0.000    0.973    0.860
    p10               1.158    0.087   13.246    0.000    0.805    0.816
    p11               1.245    0.090   13.881    0.000    0.866    0.858
    p12               1.024    0.090   11.361    0.000    0.712    0.693
  Preparacao =~
    p13               1.000                               0.713    0.714
    p14               1.166    0.100   11.617    0.000    0.831    0.785
    p15               1.234    0.106   11.654    0.000    0.880    0.790
  Medo =~
    p16               1.000                               0.810    0.782
    p17               1.096    0.066   16.592    0.000    0.887    0.878
    p18               1.126    0.068   16.576    0.000    0.911    0.877
  PensamentoNegativo =~
    p19               1.000                               0.756    0.706
    p20               1.021    0.080   12.831    0.000    0.772    0.787
    p21               0.791    0.080    9.902    0.000    0.598    0.602
    p22               1.041    0.079   13.104    0.000    0.787    0.804
    p23               0.962    0.078   12.322    0.000    0.728    0.754
    p24               1.064    0.081   13.062    0.000    0.805    0.801

Covariances:
  ProblemasFamiliaresEscolares ~~
    PrblmsCmrDrmr     0.499    0.059    8.478    0.000    0.866    0.866
    Preparacao        0.375    0.052    7.167    0.000    0.636    0.636
    Medo              0.477    0.059    8.144    0.000    0.712    0.712
    PensamentNgtv     0.462    0.058    7.944    0.000    0.738    0.738
  ProblemasComerDormir ~~
    Preparacao        0.291    0.044    6.560    0.000    0.586    0.586
    Medo              0.451    0.056    8.125    0.000    0.802    0.802
    PensamentNgtv     0.411    0.053    7.708    0.000    0.781    0.781
  Preparacao ~~
    Medo              0.329    0.050    6.635    0.000    0.570    0.570
    PensamentNgtv     0.323    0.049    6.635    0.000    0.599    0.599
  Medo ~~
    PensamentNgtv     0.489    0.060    8.134    0.000    0.798    0.798

Variances:
    p1                0.397    0.037                      0.397    0.367
    p2                0.590    0.051                      0.590    0.540
    p3                0.537    0.049                      0.537    0.405
    p4                0.390    0.038                      0.390    0.317
    p5                0.204    0.020                      0.204    0.290
    p6                0.386    0.037                      0.386    0.347
    p7                0.522    0.046                      0.522    0.519
    p8                0.631    0.054                      0.631    0.593
    p9                0.334    0.034                      0.334    0.261
    p10               0.326    0.031                      0.326    0.335
    p11               0.268    0.027                      0.268    0.263
    p12               0.549    0.048                      0.549    0.520
    p13               0.490    0.051                      0.490    0.491
    p14               0.430    0.053                      0.430    0.384
    p15               0.468    0.059                      0.468    0.376
    p16               0.417    0.040                      0.417    0.389
    p17               0.235    0.029                      0.235    0.230
    p18               0.250    0.031                      0.250    0.231
    p19               0.577    0.052                      0.577    0.502
    p20               0.367    0.036                      0.367    0.381
    p21               0.628    0.054                      0.628    0.637
    p22               0.338    0.034                      0.338    0.353
    p23               0.402    0.038                      0.402    0.432
    p24               0.361    0.036                      0.361    0.358
    PrblmsFmlrsEs     0.686    0.084                      1.000    1.000
    PrblmsCmrDrmr     0.484    0.072                      1.000    1.000
    Preparacao        0.508    0.077                      1.000    1.000
    Medo              0.655    0.083                      1.000    1.000
    PensamentNgtv     0.572    0.084                      1.000    1.000
```

Figura 16. Saída da função *summary()* para o modelo de preocupações parentais

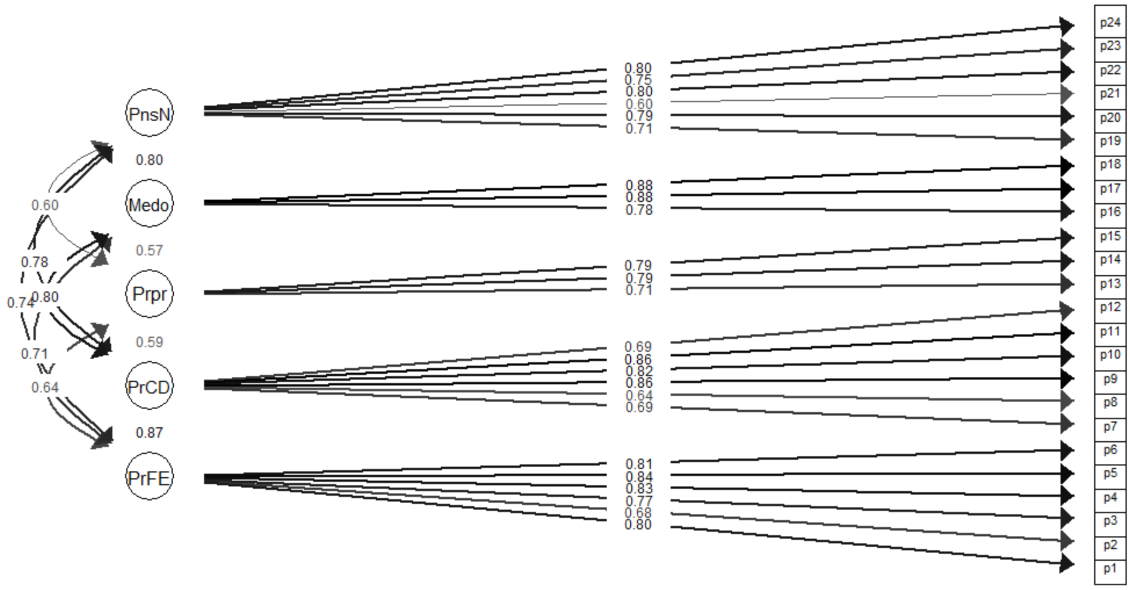


Figura 17. Modelo Fatorial de Preocupações Parentais com as estimativas padronizadas e as correlações entre os fatores

Na avaliação do modelo é analisado o quão bem o modelo teórico consegue reproduzir a estrutura correlacional das variáveis observadas na amostra. Existem inúmeros índices de avaliação do modelo e estatísticas que podem ser utilizadas e podem vir de diversas perspectivas. Primeiramente analisa-se a adequação dos parâmetros estimados.

A plausibilidade é um tópico relevante dentro da análise da adequação dos parâmetros estimados. Estimativas que saem da amplitude admissível (conforme os dados) são um claro indicador de que o modelo está errado ou que os dados não trazem informações suficientes. Conforme a Figura 16, a plausibilidade do modelo está garantida, pois não existem dados discrepantes (correlações >1 e variâncias negativas). Quanto à significância das estimativas, a estatística apresentada é a estatística z, juntamente com o p-valor. Baseado no nível de significância 0,05, o teste estatístico precisa ser $|z| > 1.96$ para que a hipótese (de que o parâmetro seja igual à zero) possa ser rejeitada. Parâmetros que não foram significantes, com exceção das variâncias, podem ser considerados não importantes ao modelo. Nesse momento é importante ter cuidado, pois parâmetros não significativos pode ser um indício de que o tamanho da amostra é muito pequeno (BYRNE, 2001). A Figura 16 indica que as estimativas são todas razoáveis e estatisticamente significativas.

Ainda na Figura 16 é apresentado um valor do χ^2, os graus de liberdade e o valor de probabilidade com a intenção de fornecer uma ideia geral sobre o ajuste do modelo. Existem ainda muitas outras informações que devem ser analisadas. A hipótese nula (H_0) do processo de ajuste é que a matriz populacional de covariância (Σ) e a matriz de covariância restringida pelo modelo ($\Sigma(\theta)$) são iguais ($\Sigma=\Sigma(\theta)$), ou seja, se o modelo testado condiz com a população.

Nesta etapa, devem-se verificar as estatísticas de ajuste do modelo (*Goodness-of-Fit)* que são apresentadas pelo pacote *Lavaan.* A linha de comando utilizada e a saída são apresentadas abaixo.

```
fitMeasures (fit)
```

```
           fmin           chisq              df          pvalue  baseline.chisq
          0.653         394.437         242.000           0.000        4995.685
    baseline.df baseline.pvalue             cfi             tli            nnfi
        276.000           0.000           0.968           0.963           0.963
            rfi             nfi            pnfi             ifi             rni
          0.910           0.921           0.808           0.968           0.968
           logl unrestricted.logl         npar             aic             bic
      -8173.528       -7976.310          58.000       16463.057       16678.262
         ntotal            bic2           rmsea  rmsea.ci.lower  rmsea.ci.upper
        302.000       16494.317           0.046           0.037           0.054
   rmsea.pvalue             rmr      rmr_nomean            srmr     srmr_nomean
          0.806           0.046           0.046           0.044           0.044
          cn_05           cn_01             gfi            agfi            pgfi
        214.836         227.709           0.900           0.876           0.726
            mfi            ecvi
          0.777           1.690
```

Figura 18. Índices de Ajuste do Modelo Fatorial de Preocupações Parentais.

Observa-se na Figura 18 que a informação do χ^2 (chisq=394,437), graus de liberdade (df=242) e probabilidade (pvalue<0,001) são apresentados novamente, com isso rejeita-se a hipótese nula de que a matriz de covariância populacional é igual à matriz de covariância estimada pelo modelo. Entretanto, como comentado na Seção 3.5.1, o teste χ^2 é altamente sensível ao tamanho da amostra. Assim, conseguir modelos hipotéticos bem ajustados onde os valores do χ^2 são próximos aos graus de liberdade é muito raro em SEM. Uma alternativa, segundo Arbuckle (2008), é, ao invés de olhar somente os graus de liberdade, observar a relação χ^2/gl, onde se H_0 do teste χ^2 de ajustamento for verdadeira, o valor esperado dos graus de liberdade é igual ao valor esperado da estatística de teste. Assim, na situação de ajustamento perfeito, $\chi^2/gl = 1$.

Nesse caso, $\chi^2/gl = 1{,}63$, o que, segundo os valores de referência da Figura 4, é considerado um bom ajustamento.

Outra maneira de contornar essas limitações do teste χ^2 é verificar os índices de ajustamento do modelo. Aborda-se os principais índices apresentados na Seção 3.5.2.

GFI (*Goodness-of- Fit Index*) que explica a proporção da covariância observada entre as variáveis observadas, explicada pelo modelo ajustado. No caso, gfi = 0,900, que conforme os valores de referência (0,9 a 0,95) é classificado com um bom ajustamento.

AGFI (*Adjusted Goodness-of-fit*) difere do GFI apenas pelo fato de ser ajustado pelo número de graus de liberdade do modelo especificado (agfi =0,876).

PGFI *(Parsimony Goodness-of-fit Index*) é um índice que leva em conta o número de parâmetros estimados do modelo teórico na avaliação geral do modelo. O modelo apresentou pgfi = 0,726, que conforme os valores de referência (0,6 a 0,8) são considerados um bom ajustamento.

NFI (*Normed Fit Index*) avalia a percentagem de incremento na qualidade do ajustamento do modelo relativamente ao modelo totalmente independente ou ao pior modelo possível. Valor acima de 0,90 indica bom ajuste do modelo e obteve-se nfi= 0,921.

CFI (*Comparative Fit Index*) foi criado para corrigir o erro do NFI que apresentava certa tendência de subestimar o ajuste em amostras pequenas. O modelo apresentou cfi = 0,968, que conforme os valores de referência (0,9 a 0,95) é considerado um bom ajustamento.

TLI (*Tucker-Lewis Index*) apresentou um valor igual a 0,963, que conforme os valores de referência (0,9 a 0,95) é considerado um bom ajustamento.

RMSEA é um dos critérios reconhecidos como mais explicativos na modelagem em estruturas de covariâncias, levando em conta o erro de aproximação na população. Obteve-se RMSEA= 0,046, que conforme os valores de referência (menor ou igual a 0,05) é considerado um bom ajustamento.

Lembra-se que esses índices servem para dar uma ideia de quão bem o modelo hipotético se ajusta aos dados reais. Deve-se sempre analisar quais índices são mais indicados para o tipo de amostra em analise e que eles sozinhos não garantem a plausibilidade do modelo. Esse julgamento se dá através de um conjunto de considerações, teóricas, estatísticas e práticas (LEMKE, 2005).

Para ter certeza de que os dados não estão incorretamente especificados há algumas maneiras, entre elas, verificar os resíduos padronizados e os índices de modificação. Como o objetivo é determinar o ajuste entre a matriz de covariância restrita pelo modelo teórico e a matriz de covariância amostral, pode-se visualizar a discrepância entre essas matrizes através da matriz residual de covariâncias. A matriz residual contém [24(25)/2] elementos, ou seja, 300 elementos. Geralmente os resíduos padronizados são mais utilizados, pois como o valor dos resíduos depende da unidade de medida das variáveis sua interpretação se torna mais difícil. A fórmula utilizada para o cálculo dos resíduos padronizados no pacote *Lavaan*, também utilizada no *software Mplus*, é feito segundo a proposta de Hausman's (1978).

Entretanto, segundo a documentação do *Mplus* há um problema com a abordagem de Hausman's para calcular a variância residual, pois por vezes a variância apresenta-se negativa e, portanto, o resíduo padronizado não é calculado. Tipicamente, em tal situação, o resíduo normalizado pode ser usado. Valores absolutos superiores a 1,96 indicam, com 95% de confiança, observações muito discrepantes das outras observações. As covariância das variáveis p12 e p13, p14 e p9, p13 e p7, p21 e p6, p14 e p2 apresentam discrepância estatisticamente significativa (ver Figura 19).

```
Resid(fit, type = “normalized”)
```

$cov

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14
p1	0.000													
p2	-0.101	0.000												
p3	-0.312	0.175	0.000											
p4	0.230	0.015	-0.337	0.000										
p5	0.055	0.228	0.031	0.041	0.000									
p6	-0.270	-0.290	0.042	-0.215	0.434	0.000								
p7	-0.969	-0.400	0.042	0.316	-0.913	0.258	0.000							
p8	1.288	1.905	1.296	0.991	0.639	0.157	-0.085	0.000						
p9	0.122	0.099	0.277	0.621	0.321	0.380	0.187	-0.192	0.000					
p10	-0.341	0.210	0.289	-0.109	-0.873	-0.589	0.364	-0.504	-0.306	0.000				
p11	-0.050	-0.275	-0.292	-0.329	-0.303	-0.457	-0.289	-0.596	0.436	0.141	0.000			
p12	-0.045	0.124	0.537	-0.502	-0.465	0.400	-0.243	1.216	-0.603	0.294	-0.011	0.000		
p13	0.736	1.185	1.348	0.996	-0.017	1.263	2.190	1.110	0.736	1.929	1.586	2.400	0.000	
p14	0.411	-2.607	0.144	-0.641	-1.537	-0.016	-1.875	-0.843	-2.238	-1.076	-1.165	0.038	-0.112	0.000
p15	0.345	-0.636	0.874	-1.006	-0.394	0.856	-0.002	0.511	-0.196	0.252	0.161	0.974	-0.561	0.453
p16	-0.090	0.216	-0.534	-0.148	-0.471	-0.775	-0.548	-0.448	-1.239	0.276	-0.758	0.162	1.043	-1.173
p17	0.360	0.078	0.249	0.393	0.193	0.040	0.337	-0.079	-0.479	0.976	0.404	-0.121	1.681	-0.685
p18	0.287	-0.035	-0.153	-0.079	-0.233	-0.076	0.237	-0.656	-0.246	0.572	0.189	0.122	1.386	-1.050
p19	-0.726	0.155	-0.305	0.510	-1.169	-0.579	1.594	0.167	0.440	0.132	-0.833	-0.539	0.706	-1.347
p20	0.858	-0.156	0.585	0.583	-0.314	-0.643	0.542	0.344	-0.693	-0.066	0.160	0.276	0.865	-0.433
p21	-0.529	-0.327	-0.702	-0.283	-2.067	-1.669	-0.432	-1.515	-1.458	-0.630	-1.885	0.465	-0.032	-0.080
p22	0.353	-0.621	0.047	0.410	-1.519	-0.404	0.066	-0.036	-0.515	-0.284	-0.788	0.235	1.013	0.466
p23	0.600	-0.146	-0.404	-0.257	-0.777	-0.215	0.026	0.089	-0.220	0.437	0.278	1.085	0.873	0.639
p24	1.040	1.363	0.793	1.586	0.394	0.874	0.550	0.735	0.904	0.354	0.588	0.389	0.848	-0.834

	p15	p16	p17	p18	p19	p20	p21	p22	p23	p24
p1										
p2										
p3										
p4										
p5										
p6										
p7										
p8										
p9										
p10										
p11										
p12										
p13										
p14										
p15	0.000									
p16	0.103	0.000								
p17	0.029	0.023	0.000							
p18	-0.399	0.261	-0.146	0.000						
p19	-0.220	-0.477	-0.732	-0.391	0.000					
p20	-0.844	-0.203	0.051	-0.275	0.754	0.000				
p21	-1.196	-0.530	-1.152	-0.204	0.684	0.119	0.000			
p22	-0.505	-0.433	-0.400	-0.011	0.238	-0.244	1.269	0.000		
p23	0.448	0.550	0.445	0.224	-0.699	0.065	0.297	0.010	0.000	
p24	-0.038	-0.133	1.327	0.338	-0.222	-0.233	-0.917	-0.114	-0.128	0.000

Figura 19. Matriz de covariância residual normalizada

Os Índices de Modificação estimam a redução χ^2do modelo, se um parâmetro fixo ou uma restrição de igualdade entre parâmetros for libertado, se os erros de medida forem correlacionados, se novas trajetórias estruturais forem adicionadas, depois de considerada a reestimação do modelo e a variação dos graus de liberdade. Refletem a extensão com que o modelo hipotético está adequadamente descrito. Os Índices de Modificação podem ser vistos como a estatística χ^2 com um grau de liberdade. Todos os parâmetros livremente estimados automaticamente têm valores MI iguais a zero. Existe ainda a estatística EPC (*expected parameter change*), que representa a mudança predita do parâmetro estimado, nas direções positiva ou negativa, para cada parâmetro fixo no modelo. Porém seus valores brutos são difíceis de interpretar.

Arbuckle (2009) recomenda que a significância dos índices seja avaliada para um α=0,05, sendo assim, um Índice de Modificação superior a 4 ($\chi2_{0.95;(1)}$ =3,84) indica uma alteração de um parâmetro do modelo que permite melhorar o ajustamento com a probabilidade do erro tipo I (concluir que o modelo se ajusta bem quando o ajustamento é ruim) de 0,001. Porém, por segurança, é melhor modificar os índices superiores a 11 ($\chi2_{0.999;(1)} = 10{,}82$)

O *Lavaan* permite pedir os Índices de modificação superiores a 11, conforme indicado. O resultado é apresentado na Figura 20.

```
MI=modificationIndices(fit)

subset(MI, mi > 11)
```

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
1	PrFE	=~	p8	11.120	0.495	0.410	0.397	0.397
2	PrFE	=~	p24	14.914	0.325	0.269	0.268	0.268
3	PrCD	=~	p13	19.727	0.434	0.302	0.303	0.303
4	PrCD	=~	p14	20.931	-0.488	-0.339	-0.320	-0.320
5	Medo	=~	p13	15.761	0.328	0.266	0.266	0.266
6	p2	~~	p14	14.175	-0.132	-0.132	-0.119	-0.119
7	p14	~~	p15	12.802	0.233	0.233	0.197	0.197
8	p17	~~	p24	14.771	0.085	0.085	0.083	0.083
9	p21	~~	p22	12.766	0.110	0.110	0.113	0.113

Figura 20. Índices de Modificação

Na Figura 20 pode-se ver os parâmetros com Índice de Modificação maior que 11, também é apresentado os valores de epc, epc padronizado (variáveis latentes (sepc.lv), todas variáveis (sepc.all), variáveis observadas exógenas (sepc.nox)). O índice mais alto é a carga fatorial entre Problemas para comer, dormir e físicos (PrCD) com a variável observada p14 (mi = 20,931 e epc = -0,488). Entretanto a modificação do modelo só deve ser feita se existirem fundamentos teóricos para isso (MARÔCO, 2010). Olhando apenas para o ponto de vista estatístico, o modelo pode ser modificado até apresentar um ajustamento quase perfeito, mas sua teoria perde o fundamento. Sendo assim essas modificações sem fundamento podem fazer com que haja uma redução da validade do modelo para a população. Deve-se considerar também a possibilidade de que o modelo reespecificado torne o modelo super-ajustado. O conhecimento do pesquisador deverá indicar quando parar de ajustar o modelo. A utilização dos Índices de Modificação deve ser sempre encarada com uma estratégia exploratória de modelos.

Após o processo de avaliação do modelo, pode-se concluir que o modelo hipotético com cinco fatores representa adequadamente as preocupações parentais em crianças. À plausibilidade e significância estatística de todos os parâmetros estimados foram garantidos, ao bom ajuste dos índices principalmente CFI (0,968) e RMSEA (0,046).

5.3 Análise de Múltiplos Grupos (*Multiple-Group Analyses*)

A análise de múltiplos grupos serve para testar se a estrutura do modelo de medida ou modelo estrutural é equivalente, ou seja, invariante em diferentes grupos ou populações. Os grupos devem ser mutuamente exclusivos para que não haja repetição de indivíduos. Muitos pesquisadores utilizam a técnica para testar uma escala psicométrica em uma população diferente do instrumento que foi validado. O objetivo é confirmar que o fator se mantém na nova população e que as cargas fatoriais não diferem significativamente entre as duas populações. Segundo Byrne (2010), outros objetivos são ver se instrumentos de medida são equivalentes em pessoas com sexo, idade e habilidades diferentes; identificar se os caminhos definidos em uma estrutura causal são invariantes entre populações; e verificar se as variáveis latentes são medidas da mesma forma em diferentes populações.

Segundo Marôco (2010), a análise de múltiplos grupos inicia-se com uma comparação do ajustamento do modelo aos diferentes grupos simultaneamente, impondo-se algumas restrições:

i. Pesos fatoriais e covariâncias dos fatores de um modelo de medida para avaliar se o modelo de medida é igual (invariante) entre os grupos.

ii. Coeficientes estruturais para avaliar se as pretensas relações causais são invariantes entre os grupos.

iii. Variâncias e covariâncias para avaliar se a estrutura dos resíduos do modelo se mantém invariante nos grupos (pouco frequente).

Se não existirem restrições aos parâmetros do modelo, a análise de múltiplos grupos pode ser feita individualmente, um para cada grupo, e a comparação dos parâmetros pode ser feita através de um teste t-student (diferença entre médias) ou um teste Z para a diferença dos parâmetros considerando as estimativas e erros padrão assintóticos (MARÔCO, 2010). Com maior frequência na Modelagem de Equações Estruturais existe mais de um parâmetro restrito no modelo. Deve haver uma manutenção das restrições dos parâmetros, considerando todos os dados dos grupos em simultâneo para obter estimativas eficientes dos parâmetros. A análise de variância do modelo fatorial pretende demonstrar que o modelo é invariante (pesos fatorais e covariância entre os fatores não diferem significativamente entre os grupos).

Joreskog (1970) propôs a estratégia do teste *omnibus*[4] à invariância do modelo. Segundo ele, apenas quando rejeita-se o teste é que deve-se verificar a invariância nos grupos individuais. Se não for rejeitada, o próximo passo é testar a invariância dos pesos fatoriais do modelo de medida. Esse teste é realizado impondo-se restrições de igualdade aos pesos fatoriais de todos os grupos. Se a hipótese da invariância fatorial for rejeitada o processo de comparação dos grupos termina. Caso contrário, deve-se avaliar a invariância dos fatores específicos. Sendo a estatística de teste, nesse caso, dada pela diferença dos χ^2 do modelo com fatores específicos fixos e fatores específicos livres com graus de liberdade definidos anteriormente.

Segundo Byrne (2001), resultados recentes demonstram que mesmo quando não se rejeita H_0 do teste *omnibus* possivelmente hipóteses subsequentes referentes aos modelos de medida ou estruturais de grupos individuais devam ser rejeitadas. Porém quando se rejeita H_0 do teste *omnibus* pode não ser possível rejeitar H_0 das comparações posteriores. Assim, uma estratégia de analise adequada para confirmar ou não à invariância do modelo fatorial, segundo Marôco (2010), é:

1. Analisar individualmente os modelos fatoriais de cada grupo, sendo os parâmetros livres.
2. Analisar conjuntamente os grupos, restringindo os parâmetros do modelo selecionado no primeiro passo de forma a testar a hipótese de invariância.

[4] Testes *Omnibus* é uma espécie de teste estatístico para saber se a variância explicada em um conjunto de dados é significativamente maior do que a variação não explicada, em geral. Um exemplo é o teste F na análise de variância. Por exemplo, num modelo de duas variáveis independentes, se apenas uma variável exerce um efeito significativo sobre a variável dependente e o outro não, então o teste *omnibus* pode ser não significativo. Este fato não altera as conclusões que podem ser tiradas a partir da uma variável significante.

Essa estratégia permite a estimação eficiente dos parâmetros, impondo restrições de complexidade crescentes à estrutura fatorial.

Utiliza-se o mesmo banco do exemplo anterior que apresenta uma proposta da medida das preocupações parentais em crianças, onde os principais fatores que dividem essas preocupações são cinco: Problemas familiares e escolares (PrFE); Problemas para comer, dormir e físicos (PrCD); Preparação para mudanças (Prpr); Medo; e Pensamentos negativos (PnsN), que já foram comentados anteriormente. A intenção é testar se a estrutura fatorial proposta é invariante para meninos e meninas. Para isso avalia-se a plausibilidade do modelo de medida ajustado aos dois grupos e estuda-se a invariância em relação aos pesos fatoriais e correlações. O banco de dados possui informações para 153 meninos e 149 meninas. O modelo é o mesmo apresentado na Figura 15.

Depois de ativar o banco de dados no *software* R, deve-se prosseguir com a definição do modelo, que será dada abaixo e tem a mesma definição do exemplo da Analise Fatorial Confirmatória, feito anteriormente.

```
modelo=' ProblemasFamiliaresEscolares =~ p1+p2+p3+p4+p5+p6

ProblemasComerDormir=~p7+p8+p9+p10+p11+p12

Preparacao=~p13+p14+p15

Medo=~p16+p17+p18

PensamentoNegativo=~p19+p20+p21+p22+p23+p24'
```

Sendo assim, como anteriormente, tem-se que o modelo especificado a priori indica que as preocupações parentais podem ser explicadas por cinco fatores, que cada item de medida tem carga não nula no fator que foi designado para medir (carga alvo) e carga nula nos demais fatores. Também indica que os cinco fatores de preocupações parentais são correlacionados.

Para ajustar o modelo utiliza-se a função *cfa()* por se tratar de uma análise fatorial confirmatória entre grupos, conforme apresentado abaixo.

```
fit = cfa(modelo,data = dados, group = "Sexo")
```

Como quer se analisar a invariância entre os grupos, deve-se definir a separação dos dados (*group*), no caso, a variável utilizada para separação é a variável “Sexo”. É importante ressaltar que no banco de dados o número 1 indica sexo masculino e o número 2 sexo feminino, portanto o grupo 1 é o grupo dos meninos e o grupo 2 é o grupo das meninas. Com o modelo ajustado pode-se obter as estimativas do modelo. Para obter essas estatísticas necessita-se de um resumo do modelo que pode ser obtido através da definição a seguir. A saída é apresentada a seguir, na Figura 21.

```
summary(fit, standardized = TRUE)
```

A Figura 21 apresenta o número de iterações necessárias para o modelo convergir, o número de observações por grupo (153 meninos e 149 meninas), o método de estimação, no caso, método de máxima verossimilhança, a estatística de teste χ^2(705,573), os graus de liberdade (484) e o p-valor (<0,001). Essas informações são dadas para o modelo de meninos e meninas em simultâneo. Logo em seguida as informações das estimativas, erros padrão, valor z, p-valor, estimativas padronizadas, covariâncias e variâncias primeiro para o grupo 1 (meninos) e depois para o grupo 2 (meninas), observado na sequência dos resultados, apresentados nas Figura 22 e 23.

```
lavaan (0.5-14) converged normally after  82 iterations

  Number of observations per group
  1                                                153
  2                                                149

  Estimator                                         ML
  Minimum Function Test Statistic              705.573
  Degrees of freedom                               484
  P-value (Chi-square)                           0.000

Chi-square for each group:

  1                                            323.425
  2                                            382.148

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

Group 1 [1]:

                   Estimate  Std.err  Z-value  P(>|z|)   Std.lv  Std.all
Latent variables:
  ProblemasFamiliaresEscolares =~
    p1                1.000                               0.825    0.817
    p2                0.798    0.099    8.031    0.000    0.658    0.611
    p3                1.095    0.103   10.609    0.000    0.903    0.758
    p4                1.139    0.097   11.787    0.000    0.939    0.818
    p5                0.947    0.076   12.516    0.000    0.781    0.852
    p6                1.069    0.094   11.409    0.000    0.881    0.799
  ProblemasComerDormir =~
    p7                1.000                               0.589    0.599
    p8                1.123    0.173    6.483    0.000    0.661    0.619
    p9                1.714    0.205    8.371    0.000    1.009    0.899
    p10               1.354    0.171    7.929    0.000    0.797    0.823
    p11               1.489    0.179    8.302    0.000    0.876    0.886
    p12               1.320    0.181    7.276    0.000    0.777    0.724
  Preparacao =~
    p13               1.000                               0.661    0.650
    p14               1.314    0.174    7.541    0.000    0.869    0.796
    p15               1.324    0.175    7.547    0.000    0.875    0.797
  Medo =~
    p16               1.000                               0.786    0.754
    p17               1.201    0.105   11.491    0.000    0.944    0.892
    p18               1.144    0.099   11.573    0.000    0.899    0.899
  PensamentoNegativo =~
    p19               1.000                               0.745    0.660
    p20               1.056    0.126    8.402    0.000    0.786    0.791
    p21               0.799    0.120    6.684    0.000    0.595    0.604
    p22               1.017    0.123    8.268    0.000    0.757    0.775
    p23               0.869    0.112    7.784    0.000    0.647    0.720
    p24               1.148    0.133    8.628    0.000    0.854    0.818

Covariances:
  ProblemasFamiliaresEscolares ~~
    PrblmsCmrDrmr     0.416    0.074    5.593    0.000    0.857    0.857
    Preparacao        0.347    0.070    4.932    0.000    0.637    0.637
    Medo              0.462    0.080    5.791    0.000    0.713    0.713
    PensamentNgtv     0.443    0.081    5.448    0.000    0.721    0.721
  ProblemasComerDormir ~~
    Preparacao        0.209    0.051    4.130    0.000    0.537    0.537
    Medo              0.370    0.070    5.307    0.000    0.800    0.800
    PensamentNgtv     0.313    0.065    4.835    0.000    0.715    0.715
  Preparacao ~~
    Medo              0.299    0.066    4.555    0.000    0.575    0.575
    PensamentNgtv     0.275    0.064    4.293    0.000    0.559    0.559
  Medo ~~
    PensamentNgtv     0.448    0.082    5.427    0.000    0.765    0.765
```

Figura 21. Saída da função *summary()* para o modelo de preocupações parentais de meninos e meninas.

```
Intercepts:
    p1                2.451    0.082   30.035    0.000    2.451    2.428
    p2                2.261    0.087   25.967    0.000    2.261    2.099
    p3                2.281    0.096   23.696    0.000    2.281    1.916
    p4                2.458    0.093   26.457    0.000    2.458    2.139
    p5                1.601    0.074   21.598    0.000    1.601    1.746
    p6                2.137    0.089   23.969    0.000    2.137    1.938
    p7                3.085    0.079   38.812    0.000    3.085    3.138
    p8                2.248    0.086   26.037    0.000    2.248    2.105
    p9                2.307    0.091   25.438    0.000    2.307    2.057
    p10               2.739    0.078   34.972    0.000    2.739    2.827
    p11               2.810    0.080   35.168    0.000    2.810    2.843
    p12               3.000    0.087   34.598    0.000    3.000    2.797
    p13               3.464    0.082   42.157    0.000    3.464    3.408
    p14               3.176    0.088   36.001    0.000    3.176    2.911
    p15               2.863    0.089   32.279    0.000    2.863    2.610
    p16               3.196    0.084   37.940    0.000    3.196    3.067
    p17               2.922    0.085   34.177    0.000    2.922    2.763
    p18               3.229    0.081   39.938    0.000    3.229    3.229
    p19               2.980    0.091   32.660    0.000    2.980    2.640
    p20               2.869    0.080   35.680    0.000    2.869    2.885
    p21               3.124    0.080   39.207    0.000    3.124    3.170
    p22               3.268    0.079   41.378    0.000    3.268    3.345
    p23               3.366    0.073   46.341    0.000    3.366    3.746
    p24               3.007    0.084   35.597    0.000    3.007    2.878
    PrblmsFmlrsEs     0.000                               0.000    0.000
    PrblmsCmrDrmr     0.000                               0.000    0.000
    Preparacao        0.000                               0.000    0.000
    Medo              0.000                               0.000    0.000
    PensamentNgtv     0.000                               0.000    0.000
```

```
Variances:
    p1                0.339    0.046                      0.339    0.332
    p2                0.728    0.087                      0.728    0.627
    p3                0.602    0.077                      0.602    0.425
    p4                0.438    0.059                      0.438    0.332
    p5                0.231    0.033                      0.231    0.274
    p6                0.440    0.058                      0.440    0.361
    p7                0.620    0.074                      0.620    0.642
    p8                0.704    0.084                      0.704    0.617
    p9                0.241    0.038                      0.241    0.192
    p10               0.303    0.040                      0.303    0.323
    p11               0.210    0.032                      0.210    0.214
    p12               0.547    0.067                      0.547    0.476
    p13               0.596    0.081                      0.596    0.577
    p14               0.437    0.081                      0.437    0.367
    p15               0.438    0.082                      0.438    0.364
    p16               0.468    0.061                      0.468    0.431
    p17               0.228    0.041                      0.228    0.204
    p18               0.192    0.036                      0.192    0.192
    p19               0.720    0.090                      0.720    0.565
    p20               0.371    0.052                      0.371    0.375
    p21               0.617    0.075                      0.617    0.635
    p22               0.381    0.052                      0.381    0.400
    p23               0.389    0.050                      0.389    0.481
    p24               0.361    0.053                      0.361    0.331
    PrblmsFmlrsEs     0.680    0.113                      1.000    1.000
    PrblmsCmrDrmr     0.346    0.087                      1.000    1.000
    Preparacao        0.437    0.106                      1.000    1.000
    Medo              0.617    0.116                      1.000    1.000
    PensamentNgtv     0.554    0.125                      1.000    1.000
```

Figura 22. Continuação da saída da função *summary()* para o modelo de preocupações parentais de meninos e meninas.

Deve-se atentar para a avaliação do modelo, garantindo a plausibilidade dos parâmetros estimados, e também a plausibilidade do modelo como um todo, como visto e garantido no exemplo anterior. Na Figura 23 tem-se a continuação da saída *summary ()* para o grupo 2 (meninas).

```
Group 2 [2]:

                   Estimate  Std.err  Z-value  P(>|z|)   Std.lv  Std.all
Latent variables:
  ProblemasFamiliaresEscolares =~
    p1                1.000                              0.835    0.785
    p2                0.909    0.092    9.869    0.000   0.759    0.750
    p3                1.048    0.100   10.531    0.000   0.875    0.789
    p4                1.074    0.094   11.421    0.000   0.896    0.841
    p5                0.753    0.066   11.353    0.000   0.628    0.837
    p6                0.982    0.089   11.037    0.000   0.820    0.819
  ProblemasComerDormir =~
    p7                1.000                              0.805    0.789
    p8                0.808    0.096    8.428    0.000   0.650    0.655
    p9                1.173    0.104   11.266    0.000   0.944    0.826
    p10               0.996    0.093   10.746    0.000   0.802    0.797
    p11               1.062    0.094   11.339    0.000   0.854    0.830
    p12               0.813    0.094    8.626    0.000   0.654    0.668
  Preparacao =~
    p13               1.000                              0.763    0.778
    p14               1.045    0.116    9.033    0.000   0.797    0.782
    p15               1.165    0.129    9.062    0.000   0.889    0.785
  Medo =~
    p16               1.000                              0.829    0.807
    p17               1.000    0.084   11.968    0.000   0.829    0.864
    p18               1.121    0.094   11.973    0.000   0.930    0.864
  PensamentoNegativo =~
    p19               1.000                              0.773    0.766
    p20               0.991    0.097   10.193    0.000   0.766    0.794
    p21               0.778    0.104    7.449    0.000   0.602    0.603
    p22               1.061    0.098   10.847    0.000   0.820    0.837
    p23               1.028    0.104    9.891    0.000   0.795    0.774
    p24               0.975    0.097   10.058    0.000   0.754    0.785
```

```
Covariances:
  ProblemasFamiliaresEscolares ~~
    PrblmsCmrDrmr     0.591    0.093    6.356    0.000   0.880    0.880
    Preparacao        0.405    0.078    5.216    0.000   0.635    0.635
    Medo              0.487    0.085    5.724    0.000   0.703    0.703
    PensamentNgtv     0.487    0.083    5.842    0.000   0.755    0.755
  ProblemasComerDormir ~~
    Preparacao        0.382    0.074    5.141    0.000   0.622    0.622
    Medo              0.536    0.087    6.132    0.000   0.803    0.803
    PensamentNgtv     0.529    0.086    6.176    0.000   0.850    0.850
  Preparacao ~~
    Medo              0.361    0.074    4.865    0.000   0.571    0.571
    PensamentNgtv     0.371    0.072    5.116    0.000   0.628    0.628
  Medo ~~
    PensamentNgtv     0.534    0.087    6.146    0.000   0.833    0.833
```

```
Variances:
    p1                0.435    0.057                     0.435    0.384
    p2                0.449    0.057                     0.449    0.438
    p3                0.463    0.061                     0.463    0.377
    p4                0.333    0.047                     0.333    0.293
    p5                0.169    0.024                     0.169    0.300
    p6                0.330    0.045                     0.330    0.330
    p7                0.392    0.052                     0.392    0.377
    p8                0.562    0.069                     0.562    0.571
    p9                0.414    0.057                     0.414    0.317
    p10               0.369    0.049                     0.369    0.365
    p11               0.329    0.045                     0.329    0.311
    p12               0.531    0.065                     0.531    0.554
    p13               0.379    0.062                     0.379    0.395
    p14               0.404    0.067                     0.404    0.389
    p15               0.492    0.083                     0.492    0.383
    p16               0.367    0.053                     0.367    0.348
    p17               0.234    0.039                     0.234    0.254
    p18               0.293    0.049                     0.293    0.253
    p19               0.422    0.055                     0.422    0.414
    p20               0.344    0.046                     0.344    0.370
    p21               0.633    0.077                     0.633    0.636
    p22               0.288    0.042                     0.288    0.300
    p23               0.423    0.056                     0.423    0.401
    p24               0.354    0.047                     0.354    0.384
    PrblmsFmlrsEs     0.697    0.124                     1.000    1.000
    PrblmsCmrDrmr     0.647    0.114                     1.000    1.000
    Preparacao        0.582    0.112                     1.000    1.000
    Medo              0.688    0.119                     1.000    1.000
    PensamentNgtv     0.598    0.111                     1.000    1.000
```

Figura 23. Continuação da saída da função *summary()* para o modelo de preocupações parentais de meninos e meninas.

A Figura 23 apresenta as informações das estimativas, erros padrão, valor z, p-valor, estimativas padronizadas e covariâncias para as meninas (grupo 2). As variâncias para o grupo 2 também são apresentadas .

As informações apresentadas anteriormente para os dois grupos são resumidas na Figura 24 e na Figura 25.

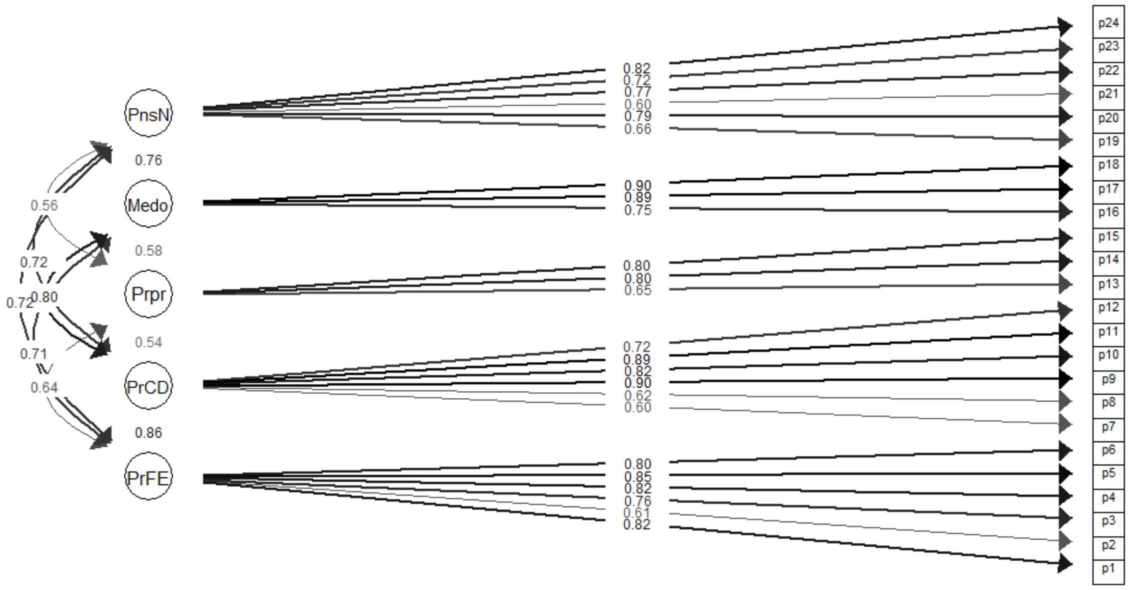


Figura 24. Modelo Fatorial de Preocupações Parentais em meninos (grupo 1). Apresenta as estimativas padronizadas e as correlações entre os fatores.

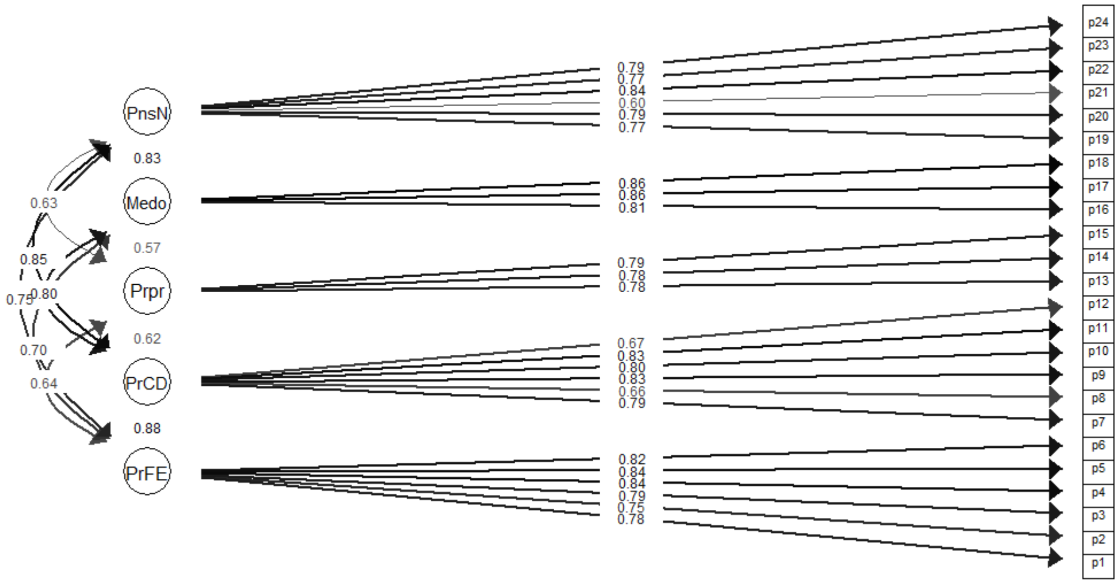


Figura 25. Modelo Fatorial de Preocupações Parentais em meninas (grupo 2).

Para prosseguir com a análise de invariância, deve-se olhar os índices de ajustamento do modelo em simultâneo para meninos e meninas e identificar se ele está bem ajustado ou não. Nesse caso olha-se como índices chaves o χ^2, CFI, RMSEA, PCFI, GFI. Para isso utiliza-se a função abaixo, cujo resultado é apresentado na Figura 26.

```
fitMeasures(fit, c("chisq","df","cfi","pcfi", "gfi", "rmsea"))
```

```
  chisq      df     cfi     gfi   rmsea
705.573 484.000   0.954   0.947   0.055
```

Figura 26. Saída da função *fitMeasures()*

Analisando os índices pode-se ver que a relação χ^2/gl = 1,45, olhando os valores de referência da Figura 4, é um valor considerado como bom ajustamento. Já o CFI (*Comparative Fit Index)* que é a razão entre o ajustamento do modelo em estudo e o ajustamento do pior modelo possível, indica ajustamento muito bom (cfi=0,954). O GFI (*Goodness-of- Fit Index),* que explica a proporção da covariância observada entre as variáveis observadas, explicada pelo modelo ajustado, indica que o modelo tem um bom ajustamento (gfi=0,947) e por último o RMSEA, um dos mais reconhecidos e importantes critérios, com um valor igual a 0,055, apontando também bom ajustamento.

Como o modelo parece estar corretamente ajustado, em simultâneo para meninos e meninas, pode-se prosseguir com as análises (BYRNE, 2001). Na análise de invariância entre os grupos, verifica-se se os pesos fatoriais e as correlações são invariantes entre meninos e meninas, ou seja, se a importância de cada fator é igual entre os grupos.

Deve-se proceder as análises, realizando restrições sobre parâmetros particulares, isto é, especificar os parâmetros como sendo invariantes (equivalentes em todos os grupos). Como foi comentado na seção 4.5 utilizar-se-á o pacote *semTools* 0.4 para facilitar a restrição dos parâmetros. A função *measurementInvariance ()* fornece a comparação entre 5 modelos, eles são:

Modelo 1: A mesma estrutura fatorial é imposta a todos os grupos.

Modelo 2: As cargas fatoriais são obrigadas a serem iguais em todos os grupos.

Modelo 3: As cargas fatoriais e interceptos (interceptos das variáveis observadas) são definidos iguais em todos os grupos.

Modelo 4: As cargas fatoriais, interceptos e variâncias residuais são definidos iguais em todos os grupos.

Modelo 5: As cargas fatoriais, interceptos, variâncias residuais e as medias (interceptos/médias das variáveis latentes) são definidos como iguais em todos os grupos.

Cada vez que um modelo mais restrito é definido, um teste da diferença entre os χ^2 dos modelos é avaliado, comparando o modelo atual com o anterior, e comparando o modelo atual com o modelo base (Modelo 1). O primeiro modelo que aparece nas comparações é o modelo que é assumido como correto, por exemplo, *Model 1 versus Model 2*, o Modelo 1 (modelo livre) é assumido como correto. Além disso, a diferença de CFI é também relatada (delta.cfi). Em seguida a função e a saída são apresentadas.

```
measurementInvariance(modelo, data = dados, group = "Sexo",strict=TRUE)
```

Observando a Figura 27, pode-se ver a comparação entre o Modelo 1 e o Modelo 2 (*Model 1 versus Model 2*) mostrando os graus de liberdade (delta.df =19), a diferença entre os χ^2(delta.chisq = 27,731) e o p-valor (delta.p.value =0,089). Com p-valor=0,089 não rejeita-se a hipótese nula de que o modelo com as cargas fatoriais fixas se ajusta tão bem aos dois grupos quanto o modelo com os pesos fatoriais livres. Assim, está demonstrada a invariância das cargas fatoriais nos dois grupos.

```
Model 1: configural invariance:
    chisq        df     pvalue       cfi      rmsea        bic
  705.573   484.000      0.000     0.954      0.055  17160.811

Model 2: weak invariance (equal loadings):
    chisq        df     pvalue       cfi      rmsea        bic
  733.304   503.000      0.000     0.952      0.055  17080.044

[Model 1 versus model 2]
  delta.chisq       delta.df delta.p.value      delta.cfi
       27.731         19.000         0.089          0.002

Model 3: strong invariance (equal loadings + intercepts):
    chisq        df     pvalue       cfi      rmsea        bic
  756.770   522.000      0.000     0.952      0.055  16995.012

[Model 1 versus model 3]
  delta.chisq       delta.df delta.p.value      delta.cfi
       51.197         38.000         0.075          0.003

[Model 2 versus model 3]
  delta.chisq       delta.df delta.p.value      delta.cfi
       23.466         19.000         0.217          0.001

Model 4: strict invariance (equal loadings + intercepts + residuals):
    chisq        df     pvalue       cfi      rmsea        bic
  816.209   546.000      0.000     0.944      0.057  16917.400

[Model 1 versus model 4]
  delta.chisq       delta.df delta.p.value      delta.cfi
      110.636         62.000         0.000          0.010

[Model 3 versus model 4]
  delta.chisq       delta.df delta.p.value      delta.cfi
       59.438         24.000         0.000          0.007

Model 5: equal loadings + intercepts + residuals + means:
    chisq        df     pvalue       cfi      rmsea        bic
  816.910   551.000      0.000     0.945      0.057  16889.549

[Model 1 versus model 5]
  delta.chisq       delta.df delta.p.value      delta.cfi
      111.337         67.000         0.001          0.009

[Model 4 versus model 5]
  delta.chisq       delta.df delta.p.value      delta.cfi
        0.701          5.000         0.983         -0.001
```

Figura 27. Saída da função *measurementInvariance()*

A comparação entre ajustamento do Modelo 1 e ajustamento do Modelo 3 se dá na parte da saída indicada como *Model 1 versus Model 3*, que é a comparação entre o modelo livre com o ajustamento do modelo com cargas fatoriais e interceptos fixos. Obtém-se, nesta comparação, que delta.df=38, delta.chisq=51,197 e delta.p.value =0,075. Também não se rejeita a hipótese nula nessa comparação, ou seja, a hipótese de que as qualidade de ajustamento do modelo livre e do modelo com cargas fatoriais e interceptos fixos não diferem significativamente. Já a comparação do Modelo 1 com o Modelo 4 e do Modelo 1 com o Modelo 5, compara o modelo livre com o modelo de cargas fatoriais, interceptos e variâncias residuais fixos e com o modelo de cargas fatoriais, interceptos, variâncias residuais e medias fixas. Sendo delta.df=62, delta.chisq=110,636 e delta.p.value=0,000 para a primeira comparação e delta.df=67, delta.chisq=111,337 e delta.p.value=0,009 para a segunda comparação. Pode-se concluir, então, que a qualidade do ajustamento desses dois modelos é significativamente diferente. Porém essa hipótese é considerada muito restritiva e geralmente é ignorada no estudo da invariância (MARÔCO,2010). A invariância para meninos e meninas fica assim demonstrada.

Quando o Modelo 2 é considerado o correto em comparação com o Modelo 3 (*Model 2 versus Model 3*) o delta.p.value =0,217 indica ausência de diferenças significativas entre a qualidade do ajustamento do modelo com pesos fatoriais fixos e o modelo com pesos fatoriais e interceptos fixos. As outras duas comparações, *Model 3 versus Model 4* (delta.p.value =0,000) e *Model 4 versus Model 5* (delta.p.value=0,983) são muito restritivos e de uma forma geral a análise de invariância não requer a invariância dos erros. A estratégia de análise vai dos modelos mais simples (livres) para os modelos mais complexos (todos os parâmetros fixos).

5.4 Análise de um Instrumento de Medida

Nesta seção é realizada a análise do instrumento de medida utilizado na tese de doutorado "A Prática do Marketing de Relacionamento e a Retenção de Clientes: um estudo aplicado em um ambiente de serviços" de Gabriel Sperandio Milan. A orientação a relacionamentos de longo prazo ultrapassa toda a organização, influenciando as interações com os clientes, antes, durante e depois da venda. Por isso é fundamental que esse relacionamento seja entendido com uma competência que estabelece um diferencial para as empresas em um ambiente cada vez mais competitivo (MILAN,2006). É abordada a estratégia de retenção de clientes por meio da prática do marketing de relacionamento no ambiente de serviços, tendo um caráter exploratório por se tratar de um estudo que busca o melhor ajuste em relação aos dados, e não o teste de apenas um modelo.

Na construção do modelo teórico é necessário o levantamento de informações que constituirão o embasamento da pesquisa. Baseando-se nessas informações, o pesquisador constrói o modelo teórico que será testado e confirmará ou não a relação entre as variáveis do modelo. Segundo Milan (2006), muitos autores foram consultados em relação à construção dos constructos teóricos a fim de medir a retenção de clientes.

Os dados foram obtidos em uma empresa do Rio Grande do sul, mais especificamente localizados na região Nordeste do Estado e na Grande Porto Alegre pela representatividade em sua carteira de clientes, tanto em relação ao número de clientes, quanto ao volume de negócios gerados. A amostra foi então definida por conveniência (amostragem não probabilística) baseando-se no julgamento do pesquisador. É importante salientar que a utilização de uma técnica de amostragem não probabilística pode acabar gerando um viés na amostra em relação à representatividade das características e percepções dos clientes. Entretanto, segundo Milan (2006), mesmo com estas ressalvas a técnica de amostragem adotada é adequada à pesquisa, impressão esta compartilhada com a empresa.

O questionário utilizado foi elaborado e aplicado pelo autor, com o cuidado de garantir que as perguntas feitas construíssem exatamente os construtos que deveriam ser analisados. Foi realizada também uma validação do instrumento de coleta de dados e pré-teste do mesmo. Segundo Milan (2006) os fatores que envolvem a prática do marketing na retenção de clientes são cinco: Satisfação, Valor, Reputação, Confiança e Fidelidade. O objetivo foi testar se o modelo de medida proposto para medir a retenção de clientes se ajusta bem aos dados. Na Figura 29, pode-se ver a especificação do modelo que se deseja testar.

Examinando o diagrama, percebem-se algumas características evidentes. Observam-se os cinco fatores latentes, indicados pelos círculos (Fidelidade, Confiança, Reputação, Valor e Satisfação). Os cinco fatores são intercorrelacionados, indicados pelas setas bidirecionais. São 30 variáveis observadas, indicados pelos retângulos (p1-p30), onde cada variável observada carrega um e apenas um fator. O fator Satisfação é medido pelas variáveis observadas de 1 até 6, Valor é medido pelas variáveis de 7 até 10, Reputação é medido pelas variáveis de 11 até 15, Confiança e Fidelidade são medidos pelas variáveis de 16 até 24 e 25 até 30, respectivamente.

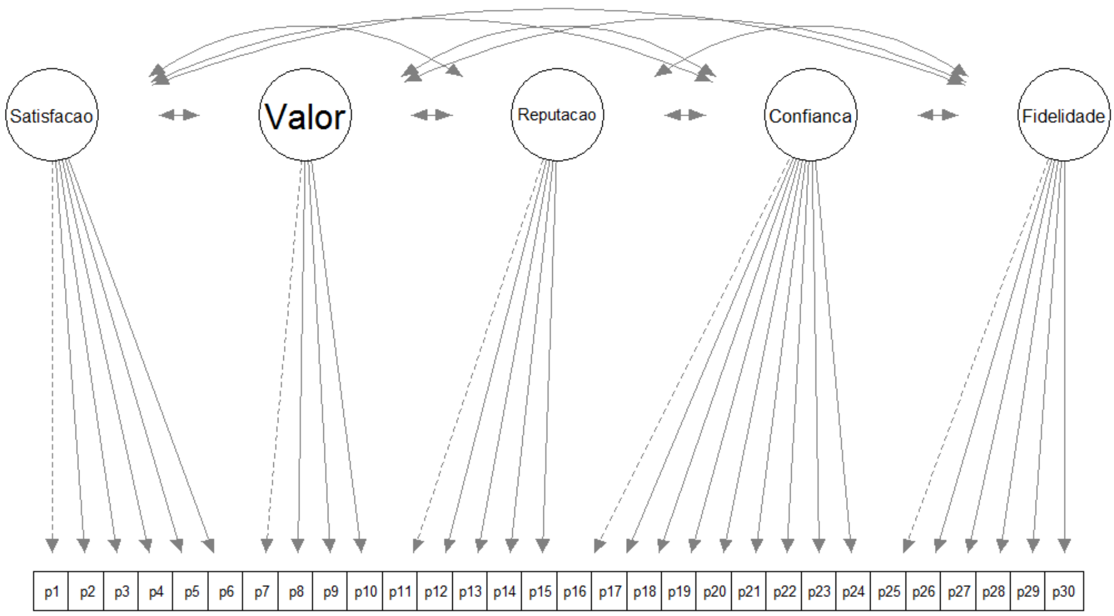


Figura 28. Diagrama de caminho do modelo de retenção de clientes.

A cada variável observada é associado um termo de erro, que é o omitido no diagrama de caminhos. Esses erros podem ser de medida ou erro derivado de alguma característica específica particular da variável observada, que não é aleatório. Para a identificação do modelo, a primeira de cada variável observada em cada conjunto de variáveis pertencente a um fator de análise fatorial confirmatória tem como carga fatorial definida o valor 1,0 (linha pontilhada), e os erros também são fixos.

Analisando o modelo, pode-se ver que há 60 pesos de regressão, 35 que foram fixados (peso de regressão de cada fator e 30 termos de erro) e 25 livres para serem estimados. Há também 10 covariâncias (entre os fatores) e 35 variâncias (5 dos fatores e 30 dos erros). Ao todo são 105 parâmetros, dos quais 70 foram estimados. Utiliza-se a fórmula [p(p+1)/2], para obter os momentos amostrais, obtendo-se 465 momentos distintos da amostra e 70 parâmetros para serem estimados, deixando o modelo com 395 graus de liberdade, ou seja, tem-se um modelo *overidentified*.

Depois de ativar o banco de dados no *software* R prossegue-se com a definição do modelo, conforme apresentado a seguir.

```
modelo=' Satisfacao =~ p1 + p2 + p3 + p4 + p5 + p6

Valor =~ p7 + p8 + p9 + p10

Reputacao =~ p11 + p12 + p13 + p14 + p15

Confianca=~ p16 + p17 + p18 + p19 + p20 + p21 + p22 + p23 + p24

Fidelidade=~ p25 + p26 + p27 + p28 + p29 + p30'
```

As Equações estruturais que constituem esse modelo podem ser vista abaixo. É interessante reforçar, que cada equação possui uma carga fatorial diferente em relação a variável latente.

p1 = Satisfação + e1	p11 = reputação + e11	p21 = confiança + e21
p2 = Satisfação + e2	p12 = reputação + e12	p22 = confiança + e22
p3 = Satisfação + e3	p13 = reputação r + e13	p23 = confiança + e23
p4 = Satisfação + e4	p14 = reputação + e14	p24 = confiança + e24
p5 = Satisfação + e5	p15 = reputação + e15	p25 = fidelidade + e25
p6 = Satisfação + e6	p16 = confiança + e16	p26 = fidelidade + e26
p7 = valor + e7	p17 = confiança + e17	p27 = fidelidade + e27
p8 = valor + e8	p18 = confiança + e18	p28 = fidelidade + e28
p9 = valor + e9	p19 = confiança + e19	p29 = fidelidade + e29
p10 = valor + e10	p20 = confiança + e20	p30 = fidelidade + e30

Para ajustar o modelo utiliza-se, novamente, a função *cfa()*:

```
fit = cfa(modelo,data = dados)
```

A função *summary ()*, já utilizado algumas vezes ao longo do trabalho, traz as estimativas do modelo que são apresentadas na Figura 29.

```
summary(fit, standardized = TRUE)
```

```
lavaan (0.5-15) converged normally after  87 iterations

  Number of observations                           263

  Estimator                                         ML
  Minimum Function Test Statistic             1598.909
  Degrees of freedom                               395
  P-value (Chi-square)                           0.000

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)   Std.lv  Std.all
Latent variables:
  Satisfacao =~
    p1                1.000                               1.168    0.973
    p2                0.929    0.022   42.675    0.000    1.085    0.962
    p3                0.873    0.034   26.039    0.000    1.019    0.870
    p4                0.628    0.034   18.579    0.000    0.733    0.770
    p5                0.786    0.029   26.834    0.000    0.918    0.877
    p6                0.825    0.024   34.601    0.000    0.964    0.930
  ValorProver =~
    p7                1.000                               1.335    0.831
    p8                0.707    0.042   16.740    0.000    0.944    0.832
    p9                0.775    0.039   19.727    0.000    1.034    0.917
    p10               0.874    0.042   20.700    0.000    1.167    0.942
  Reputacao =~
    p11               1.000                               0.598    0.823
    p12               1.437    0.082   17.596    0.000    0.860    0.865
    p13               0.858    0.052   16.450    0.000    0.514    0.829
    p14               1.582    0.101   15.636    0.000    0.947    0.802
    p15               1.545    0.095   16.208    0.000    0.924    0.821
  Confianca =~
    p16               1.000                               0.787    0.836
    p17               0.835    0.056   15.026    0.000    0.657    0.766
    p18               1.215    0.061   20.026    0.000    0.956    0.910
    p19               1.561    0.079   19.748    0.000    1.229    0.903
    p20               1.478    0.074   20.044    0.000    1.164    0.910
    p21               1.121    0.063   17.686    0.000    0.883    0.849
    p22               1.065    0.052   20.528    0.000    0.838    0.922
    p23               1.055    0.051   20.571    0.000    0.830    0.923
    p24               1.383    0.073   19.017    0.000    1.088    0.885
  Fidelidade =~
    p25               1.000                               1.161    0.889
    p26               0.954    0.049   19.607    0.000    1.107    0.856
    p27               0.886    0.084   10.592    0.000    1.029    0.583
    p28               1.046    0.059   17.646    0.000    1.214    0.812
    p29               0.855    0.059   14.423    0.000    0.993    0.723
    p30               1.154    0.058   19.811    0.000    1.339    0.861

Covariances:
  Satisfacao ~~
    ValorProver       1.416    0.145    9.766    0.000    0.908    0.908
    Reputacao         0.625    0.065    9.634    0.000    0.894    0.894
    Confianca         0.763    0.081    9.405    0.000    0.830    0.830
    Fidelidade        1.132    0.117    9.655    0.000    0.835    0.835
  ValorProver ~~
    Reputacao         0.690    0.078    8.848    0.000    0.864    0.864
    Confianca         0.843    0.097    8.651    0.000    0.802    0.802
    Fidelidade        1.315    0.145    9.065    0.000    0.848    0.848
  Reputacao ~~
    Confianca         0.453    0.049    9.319    0.000    0.961    0.961
    Fidelidade        0.626    0.067    9.282    0.000    0.902    0.902
  Confianca ~~
    Fidelidade        0.816    0.087    9.353    0.000    0.893    0.893
```

Figura 29. Saída da função *summary()* para o modelo de retenção de clientes.

Variances:

p1	0.077	0.011	0.077	0.053
p2	0.096	0.011	0.096	0.075
p3	0.334	0.031	0.334	0.243
p4	0.369	0.033	0.369	0.408
p5	0.252	0.024	0.252	0.230
p6	0.145	0.015	0.145	0.135
p7	0.796	0.077	0.796	0.309
p8	0.395	0.038	0.395	0.307
p9	0.203	0.023	0.203	0.159
p10	0.173	0.024	0.173	0.113
p11	0.171	0.016	0.171	0.323
p12	0.249	0.025	0.249	0.252
p13	0.120	0.011	0.120	0.313
p14	0.498	0.047	0.498	0.357
p15	0.413	0.039	0.413	0.326
p16	0.266	0.025	0.266	0.301
p17	0.304	0.027	0.304	0.414
p18	0.190	0.019	0.190	0.172
p19	0.341	0.033	0.341	0.184
p20	0.280	0.028	0.280	0.171
p21	0.303	0.028	0.303	0.280
p22	0.124	0.013	0.124	0.150
p23	0.120	0.012	0.120	0.149
p24	0.329	0.031	0.329	0.217
p25	0.358	0.040	0.358	0.210
p26	0.446	0.046	0.446	0.267
p27	2.054	0.184	2.054	0.660
p28	0.762	0.075	0.762	0.341
p29	0.903	0.084	0.903	0.478
p30	0.629	0.065	0.629	0.260
Satisfacao	1.364	0.126	1.000	1.000
ValorProver	1.782	0.216	1.000	1.000
Reputacao	0.358	0.044	1.000	1.000
Confianca	0.620	0.074	1.000	1.000
Fidelidade	1.348	0.148	1.000	1.000

Figura 30. Continuação da saída da função *summary()* para o modelo de retenção de clientes

Na Figura 29, observa-se que a amostra é constituída por 263 casos, o valor do χ^2 é igual a 1598,909 e o modelo possui 395 graus de liberdade. Como foi apresentado na seção 3.2, o recomendado pelos autores é que o modelo tenha de 5 a 10 casos por parâmetro o que não acontece, pois existem 70 parâmetros a serem estimados. O método usado na estimação é o Método da Máxima Verossimilhança.

Precisa-se verificar a significância das estimativas dos parâmetros. Na Figura 29 e 30 pode-se ver as estimativas, erros padrão, valor z e p-valor para avaliar a significância dos parâmetros. Outra maneira de ver as estimativas é diretamente no diagrama de caminhos, apresentado na Figura 31.

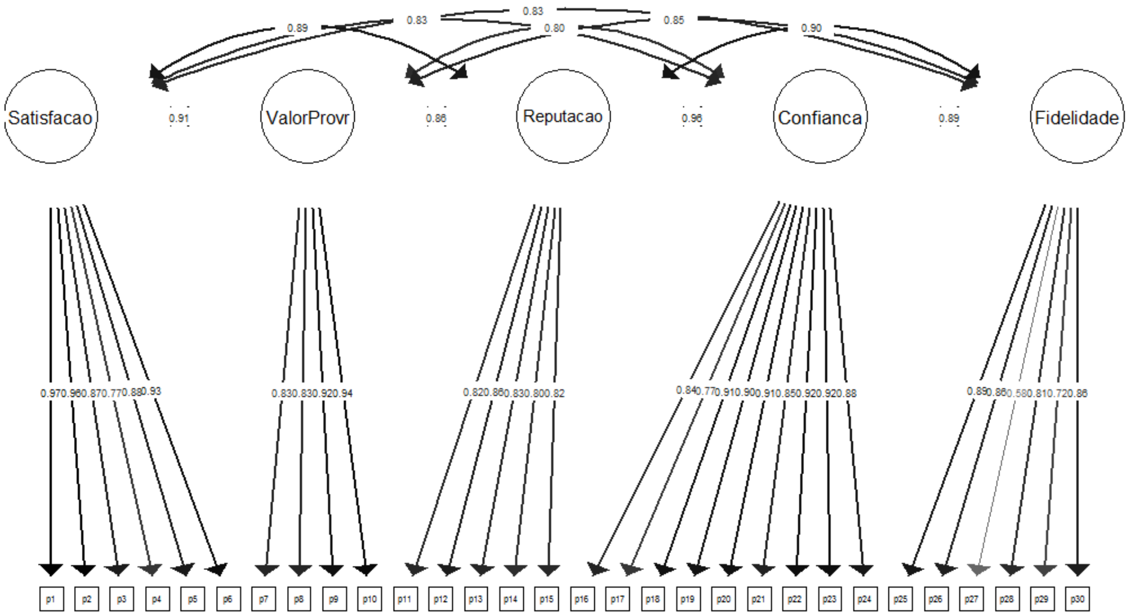


Figura 31. Diagrama de caminhos para o modelo de retenção de clientes com as estimativas padronizadas e correlações.

A plausibilidade do modelo está garantida, pois não existem dados discrepantes, correlações >1 e variâncias negativas apresentadas na Figura 29 e 30. Quanto à significância das estimativas, a estatística apresentada é a estatística z e p-valor para estatística z também é apresentado. Baseado no nível de significância 0,05, tem-se que as estimativas são todas razoáveis e estatisticamente significativas.

Deve-se agora olhar para as estatísticas de ajuste do modelo (*Goodness-of-Fit)* que são apresentadas pelo pacote *Lavaan.* Como visto anteriormente existem muitos critérios e índices de ajuste do modelo. O valor do χ^2 é igual a 1598,909 e o modelo possui 395 graus de liberdades e p-valor < 0,001. Como comentado anteriormente pelo fato do teste χ^2 ser sensível ao tamanho da amostra olha-se a relação χ^2/gl, que nesse caso é igual a 4,04 e, baseando se nos valores de referência da Figura 4, tem-se um ajustamento razoável. As estatísticas de ajuste são apresentadas na Figura 32.

```
fitMeasures (fit)
```

```
          fmin           chisq              df          pvalue  baseline.chisq
         3.040        1598.909         395.000           0.000       10656.107
   baseline.df baseline.pvalue             cfi             tli            nnfi
       435.000           0.000           0.882           0.870           0.870
           rfi             nfi            pnfi             ifi             rni
         0.835           0.850           0.772           0.883           0.882
          logl unrestricted.logl          npar             aic             bic
     -7613.296       -6813.842          70.000       15366.593       15616.643
        ntotal            bic2           rmsea  rmsea.ci.lower  rmsea.ci.upper
       263.000       15394.710           0.108           0.102           0.113
  rmsea.pvalue             rmr      rmr_nomean            srmr     srmr_nomean
         0.000           0.068           0.068           0.046           0.046
         cn_05           cn_01             gfi            agfi            pgfi
        73.759          77.209           0.690           0.635           0.586
           mfi            ecvi
         0.101           6.612
```

Figura 32. Índices de ajustamento do modelo

Analisando os principais índices tem-se, RMR = 0,068, GFI = 0,690, AGFI = 0,635, PGFI = 0,586, NFI=0,850, CFI=0,882, TLI =0,870 e RMSEA =0,108. Esses índices indicam que há um ajustamento ruim do modelo, e, portanto ele deve ser especificado novamente. Utiliza-se o auxílio dos Índices de Modificação para mudar o modelo conforme os valores que diminuem o χ^2. Como preferencialmente os Índices de Modificação devem ser maiores que 11, solicita-se que apenas os índices maiores que esse valor sejam exibidos. Os resultados são apresentados na Figura 33.

```
MI=modificationIndices(fit)

subset(MI, mi > 11)
```

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
1	Satisfacao	=~	p7	64.206	1.191	1.391	0.866	0.866
2	Satisfacao	=~	p9	19.762	-0.409	-0.478	-0.424	-0.424
3	Valor	=~	p2	12.252	-0.176	-0.235	-0.209	-0.209
4	Valor	=~	p11	12.192	-0.171	-0.229	-0.315	-0.315
5	Valor	=~	p25	11.078	-0.252	-0.337	-0.258	-0.258
6	Valor	=~	p28	14.720	0.378	0.504	0.337	0.337
7	Valor	=~	p29	14.829	0.393	0.525	0.382	0.382
8	Reputacao	=~	p7	28.570	1.234	0.738	0.460	0.460
9	Reputacao	=~	p8	15.845	-0.648	-0.388	-0.342	-0.342
10	Reputacao	=~	p9	12.252	-0.480	-0.287	-0.255	-0.255
11	Reputacao	=~	p25	25.902	1.234	0.738	0.565	0.565
12	Reputacao	=~	p28	11.445	-1.030	-0.616	-0.412	-0.412
13	Confianca	=~	p7	22.969	0.661	0.520	0.324	0.324
14	Confianca	=~	p8	18.087	-0.414	-0.326	-0.287	-0.287
15	Confianca	=~	p25	39.048	1.003	0.790	0.605	0.605
16	Confianca	=~	p28	21.940	-0.958	-0.754	-0.504	-0.504
17	Confianca	=~	p29	17.264	-0.875	-0.688	-0.501	-0.501
18	Fidelidade	=~	p7	19.174	0.501	0.582	0.363	0.363
19	p1	~~	p2	66.966	0.087	0.087	0.064	0.064
20	p1	~~	p4	23.877	-0.066	-0.066	-0.058	-0.058
21	p1	~~	p7	13.043	0.072	0.072	0.037	0.037
22	p1	~~	p17	19.396	-0.052	-0.052	-0.051	-0.051
23	p2	~~	p5	21.344	-0.056	-0.056	-0.047	-0.047
24	p2	~~	p6	16.893	-0.041	-0.041	-0.035	-0.035
25	p2	~~	p18	23.362	0.049	0.049	0.041	0.041
26	p3	~~	p17	11.749	0.071	0.071	0.070	0.070
27	p3	~~	p23	11.237	-0.046	-0.046	-0.044	-0.044
28	p4	~~	p17	19.610	0.094	0.094	0.116	0.116
29	p4	~~	p18	11.519	-0.060	-0.060	-0.060	-0.060
30	p5	~~	p6	32.345	0.076	0.076	0.070	0.070
31	p5	~~	p16	12.855	0.061	0.061	0.062	0.062
32	p5	~~	p28	11.383	0.100	0.100	0.064	0.064
33	p5	~~	p30	11.058	0.092	0.092	0.056	0.056
34	p6	~~	p18	11.480	-0.039	-0.039	-0.036	-0.036
35	p7	~~	p9	35.100	-0.190	-0.190	-0.105	-0.105
36	p8	~~	p9	62.691	0.179	0.179	0.140	0.140
37	p8	~~	p10	12.860	-0.085	-0.085	-0.060	-0.060
38	p8	~~	p25	13.609	-0.101	-0.101	-0.068	-0.068
39	p8	~~	p27	19.313	-0.260	-0.260	-0.130	-0.130
40	p8	~~	p28	17.511	0.157	0.157	0.093	0.093
41	p8	~~	p29	11.546	0.136	0.136	0.087	0.087
42	p9	~~	p10	11.293	0.077	0.077	0.055	0.055
43	p9	~~	p16	18.559	0.072	0.072	0.068	0.068
44	p9	~~	p22	11.488	-0.040	-0.040	-0.039	-0.039
45	p10	~~	p27	14.552	0.174	0.174	0.080	0.080
46	p11	~~	p13	24.504	0.049	0.049	0.109	0.109
47	p11	~~	p17	20.550	0.067	0.067	0.108	0.108
48	p13	~~	p17	33.106	0.072	0.072	0.135	0.135
49	p13	~~	p18	14.060	-0.039	-0.039	-0.060	-0.060
50	p13	~~	p23	12.256	0.029	0.029	0.052	0.052
51	p13	~~	p26	12.236	0.057	0.057	0.071	0.071
52	p15	~~	p26	13.726	-0.111	-0.111	-0.076	-0.076
53	p15	~~	p30	14.973	0.138	0.138	0.079	0.079
54	p16	~~	p17	27.854	0.097	0.097	0.120	0.120
55	p17	~~	p18	15.177	-0.063	-0.063	-0.070	-0.070
56	p17	~~	p22	12.554	-0.047	-0.047	-0.060	-0.060
57	p17	~~	p24	22.939	0.100	0.100	0.095	0.095
58	p19	~~	p20	90.872	0.209	0.209	0.120	0.120
59	p19	~~	p22	12.972	-0.053	-0.053	-0.043	-0.043
60	p19	~~	p24	13.258	-0.085	-0.085	-0.051	-0.051
61	p20	~~	p22	16.369	-0.055	-0.055	-0.047	-0.047
62	p20	~~	p23	20.685	-0.061	-0.061	-0.053	-0.053
63	p22	~~	p23	25.670	0.046	0.046	0.056	0.056
64	p24	~~	p26	23.017	0.128	0.128	0.081	0.081
65	p25	~~	p26	34.097	0.193	0.193	0.115	0.115
66	p25	~~	p28	19.145	-0.179	-0.179	-0.091	-0.091
67	p25	~~	p29	31.617	-0.236	-0.236	-0.132	-0.132
68	p26	~~	p28	16.399	-0.175	-0.175	-0.090	-0.090
69	p28	~~	p29	81.145	0.511	0.511	0.248	0.248
70	p28	~~	p30	32.112	0.292	0.292	0.125	0.125

Figura 33. Índices de Modificação

Na Figura 33, o valor destacado na linha 58 indica um Índice de Modificação de 90,872, o maior entre os índices encontrados, significa que se colocarmos covariância entre o erro das questões p20 e p19 baixaria pelo menos o valor do χ^2. A estimativa desse novo parâmetro incorporado ao modelo seria de aproximadamente 0,209. O segundo maior dentre esses valores é o apresentado na linha 69, com um índice de 81,145 entre as questões p28 e p29 e a estimativa desse novo parâmetro incorporado ao modelo seria de aproximadamente 0,511. Os valores para p1 e p2 são 66,966 e 0,087. Os Índices de Modificação foram retirados um a uma para garantir que com a mudança de um índice o segundo mais relevante não mudaria.

Qualquer modificação exige que as mudanças sejam plausíveis, ou seja, façam sentido teórico. Como o trabalho mede a retenção de clientes através de cinco fatores que são "parecidos" as questões podem ao mesmo tempo medir um ou outro fator. Portanto modifica-se o modelo, no caso, optou-se por modificar apenas os 3 maiores valores do Índice de Modificação, e depois se avalia o novo modelo. O novo modelo está definido abaixo.

```
Modelo2=' Satisfacao =~ p1 + p2 + p3 + p4 + p5 + p6

Valor =~ p7 + p8 + p9 + p10

Reputacao =~ p11 + p12 + p13 + p14 + p15

Confianca=~ p16 + p17 + p18 + p19 + p20 + p21 + p22 + p23 + p24

Fidelidade=~ p25 + p26 + p27 + p28 + p29 + p30

p19~~p20

p28~~p29

p1~~p2'
```

Com o novo modelo ajustado as mudanças que se obteve nos índices de ajuste foram uma queda no χ^2 (1356,315), os graus de liberdade (392). A relação entre os graus de liberdade e o χ^2 ainda está alto. Alguns índices como o CFI (de 0,882 para 0,906), o GFI (de 0,690 para 0,735), o TLI (0,870 para 0,895) e o RMSEA (0,108 para 0,097) apesarem de terem melhorado ainda não apresentam um ajustamento bom o suficiente. Ver Figura 34.

fmin	chisq	df	pvalue	baseline.chisq
2.579	1356.315	392.000	0.000	10656.107
baseline.df	baseline.pvalue	cfi	tli	nnfi
435.000	0.000	0.906	0.895	0.895
rfi	nfi	pnfi	ifi	rni
0.859	0.873	0.786	0.906	0.906
logl	unrestricted.logl	npar	aic	bic
-7492.000	-6813.842	73.000	15129.999	15390.766
ntotal	bic2	rmsea	rmsea.ci.lower	rmsea.ci.upper
263.000	15159.321	0.097	0.091	0.102
rmsea.pvalue	rmr	rmr_nomean	srmr	srmr_nomean
0.000	0.069	0.069	0.045	0.045
cn_05	cn_01	gfi	agfi	pgfi
86.157	90.210	0.735	0.686	0.620
mfi	ecvi			
0.160	5.712			

Figura 34. Índices de ajustamento do modelo 2.

Recorrendo novamente aos Índices de Modificação na Figura 35.

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
1	Satisfacao	=~	p7	62.321	1.368	1.559	0.971	0.971
2	Satisfacao	=~	p9	19.881	-0.485	-0.552	-0.489	-0.489
3	Valor	=~	p11	12.357	-0.171	-0.228	-0.314	-0.314
4	Valor	=~	p16	12.170	0.157	0.209	0.222	0.222
5	Valor	=~	p25	16.157	-0.280	-0.373	-0.286	-0.286
6	Reputacao	=~	p7	28.721	1.227	0.736	0.459	0.459
7	Reputacao	=~	p8	15.516	-0.633	-0.380	-0.335	-0.335
8	Reputacao	=~	p9	12.947	-0.487	-0.292	-0.259	-0.259
9	Confianca	=~	p7	23.285	0.660	0.521	0.325	0.325
10	Confianca	=~	p8	17.367	-0.400	-0.316	-0.278	-0.278
11	Confianca	=~	p9	11.209	-0.265	-0.210	-0.186	-0.186
12	Confianca	=~	p25	14.560	0.711	0.561	0.429	0.429
13	Fidelidade	=~	p7	23.051	0.501	0.595	0.370	0.370
14	p1	~~	p4	17.956	-0.053	-0.053	-0.046	-0.046
15	p1	~~	p7	11.763	0.064	0.064	0.033	0.033
16	p2	~~	p4	15.727	0.051	0.051	0.047	0.047
17	p2	~~	p18	11.833	0.033	0.033	0.028	0.028
18	p3	~~	p23	13.303	-0.049	-0.049	-0.047	-0.047
19	p4	~~	p17	15.349	0.082	0.082	0.101	0.101
20	p5	~~	p6	12.107	0.045	0.045	0.042	0.042
21	p5	~~	p30	12.866	0.099	0.099	0.061	0.061
22	p7	~~	p9	36.279	-0.193	-0.193	-0.106	-0.106
23	p8	~~	p9	60.874	0.175	0.175	0.137	0.137
24	p8	~~	p10	13.124	-0.086	-0.086	-0.061	-0.061
25	p8	~~	p27	16.650	-0.240	-0.240	-0.120	-0.120
26	p9	~~	p10	11.073	0.076	0.076	0.054	0.054
27	p9	~~	p16	17.858	0.070	0.070	0.066	0.066
28	p9	~~	p22	11.045	-0.038	-0.038	-0.037	-0.037
29	p10	~~	p27	15.600	0.181	0.181	0.083	0.083
30	p11	~~	p13	23.037	0.047	0.047	0.104	0.104
31	p11	~~	p17	18.563	0.063	0.063	0.101	0.101
32	p13	~~	p17	31.178	0.069	0.069	0.130	0.130
33	p13	~~	p18	13.623	-0.039	-0.039	-0.059	-0.059
34	p15	~~	p26	12.840	-0.103	-0.103	-0.071	-0.071
35	p15	~~	p30	16.708	0.151	0.151	0.086	0.086
36	p16	~~	p17	26.101	0.093	0.093	0.115	0.115
37	p17	~~	p18	16.115	-0.066	-0.066	-0.073	-0.073
38	p17	~~	p22	19.927	-0.057	-0.057	-0.074	-0.074
39	p17	~~	p24	20.231	0.092	0.092	0.087	0.087
40	p18	~~	p19	11.128	0.051	0.051	0.036	0.036
41	p20	~~	p23	12.257	-0.038	-0.038	-0.033	-0.033
42	p22	~~	p23	17.284	0.037	0.037	0.045	0.045
43	p24	~~	p26	22.597	0.121	0.121	0.076	0.076
44	p25	~~	p26	15.825	0.129	0.129	0.077	0.077
45	p28	~~	p30	28.583	0.240	0.240	0.103	0.103

Figura 35. Índices de modificação do modelo 2

Na Figura 35, existem 9 valores destacados, são apenas os Índices de Modificação mais altos das covariâncias. Com os índices de modificação podem se alterar quando há uma mudança no modelo, irá se verificar o Índice de Modificação a cada relação alterada para garantir que essas são as relações que devem ser remodeladas. Isso será feito até o modelo apresentar um bom ajustamento, sempre se atenta à plausibilidade do modelo. Após a inserção da correlação entre os erros das variáveis observadas p8 e p9, o novo índice mais alto apresentado é a relação p9 e p10 e o modelo ainda não está com um ajustamento suficientemente bom. Após a modificação do modelo com a inserção da correlação entre p9 e p10, tem-se a modificação p13 e p17, p16 e p17, p28 e p30, p29 e p30, p17 e p24, p24 e p26, p1 e p4 respectivamente, podem-se ver que alguns índices foram mais relevantes que outros conforme as mudanças.

O modelo final é apresentado logo abaixo.

```
Modelo final=' Satisfacao =~ p1 + p2 + p3 + p4 + p5 + p6

Valor =~ p7 + p8 + p9 + p10

Reputacao =~ p11 + p12 + p13 + p14 + p15

Confianca=~ p16 + p17 + p18 + p19 + p20 + p21 + p22 + p23 + p24

Fidelidade=~ p25 + p26 + p27 + p28 + p29 + p30

p19~~p20

p28~~p29

p1~~p2

p8~~p9

p9~~p10

p13~~p17

p16~~p17

p28~~p30

p29~~p30

p17~~p24

p24~~p26

p1~~p4'
```

Com o novo modelo especificado os índices de ajuste sofreram uma melhora (Figura 36).

fmin	chisq	df	pvalue	baseline.chisq
2.034	1069.671	383.000	0.000	10656.107
baseline.df	baseline.pvalue	cfi	tli	nnfi
435.000	0.000	0.933	0.924	0.924
rfi	nfi	pnfi	ifi	rni
0.886	0.900	0.792	0.933	0.933
logl	unrestricted.logl	npar	aic	bic
-7348.677	-6813.842	82.000	14861.355	15154.271
ntotal	bic2	rmsea	rmsea.ci.lower	rmsea.ci.upper
263.000	14894.292	0.083	0.077	0.088
rmsea.pvalue	rmr	rmr_nomean	srmr	srmr_nomean
0.000	0.063	0.063	0.043	0.043
cn_05	cn_01	gfi	agfi	pgfi
106.634	111.718	0.791	0.746	0.651
mfi	ecvi			
0.271	4.691			

Figura 36. Índices de ajustamento do modelo final.

Com o novo modelo ajustado as mudanças que obteve-se nos índices de ajuste foi uma queda no χ^2 (1069,671) e os graus de liberdade (382). A relação entre os graus de liberdade e o χ^2 melhorou (χ^2/gl = 2,79). Os valores dos índices CFI (0,933), GFI (0,791), TLI (0,924) melhoraram, porém o valor do RMSEA (0,083) ainda não é satisfatório. Como ainda não chegou-se a um bom ajustamento do modelo, deve-se verificar a significância das estimativas dos parâmetros incluídos no modelo, ou seja, as covariâncias. Nas Figuras 37 a 39 todos os p-valores, inclusive das covariâncias entre os termos de erro são significativos.

```
  Number of observations                           263

  Estimator                                         ML
  Minimum Function Test Statistic             1069.671
  Degrees of freedom                               383
  P-value (Chi-square)                           0.000

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)   Std.lv  Std.all
Latent variables:
  Satisfacao =~
    p1                1.000                               1.147    0.955
    p2                0.919    0.019   48.680    0.000    1.054    0.935
    p3                0.893    0.036   24.928    0.000    1.024    0.874
    p4                0.655    0.037   17.511    0.000    0.751    0.789
    p5                0.812    0.031   26.507    0.000    0.932    0.890
    p6                0.853    0.025   33.679    0.000    0.978    0.944
  Valor =~
    p7                1.000                               1.408    0.877
    p8                0.627    0.039   16.150    0.000    0.883    0.779
    p9                0.676    0.035   19.106    0.000    0.953    0.854
    p10               0.795    0.037   21.538    0.000    1.120    0.903
  Reputacao =~
    p11               1.000                               0.595    0.819
    p12               1.438    0.083   17.325    0.000    0.856    0.861
    p13               0.851    0.053   16.205    0.000    0.507    0.824
    p14               1.585    0.102   15.466    0.000    0.943    0.799
    p15               1.552    0.096   16.087    0.000    0.924    0.820
  Confianca =~
    p16               1.000                               0.784    0.833
    p17               0.812    0.047   17.377    0.000    0.636    0.749
    p18               1.222    0.062   19.867    0.000    0.958    0.911
    p19               1.540    0.081   18.929    0.000    1.207    0.887
    p20               1.459    0.076   19.216    0.000    1.144    0.895
    p21               1.123    0.064   17.462    0.000    0.881    0.846
    p22               1.081    0.052   20.719    0.000    0.847    0.932
    p23               1.067    0.052   20.619    0.000    0.836    0.929
    p24               1.392    0.074   18.918    0.000    1.091    0.886
  Fidelidade =~
    p25               1.000                               1.207    0.924
    p26               0.941    0.042   22.320    0.000    1.136    0.877
    p27               0.863    0.079   10.986    0.000    1.042    0.591
    p28               0.914    0.059   15.489    0.000    1.103    0.738
    p29               0.730    0.059   12.316    0.000    0.881    0.641
    p30               1.053    0.056   18.874    0.000    1.270    0.816
```
Figura 37. Informações básicas do modelo e estimativas.

```
Covariances:
  p19 ~~
    p20               0.207    0.029    7.127    0.000    0.207    0.577
  p28 ~~
    p29               0.650    0.083    7.864    0.000    0.650    0.610
  p1 ~~
    p2                0.081    0.015    5.411    0.000    0.081    0.570
  p8 ~~
    p9                0.208    0.029    7.290    0.000    0.208    0.503
  p9 ~~
    p10               0.130    0.023    5.654    0.000    0.130    0.422
  p13 ~~
    p17               0.071    0.012    5.740    0.000    0.071    0.360
  p16 ~~
    p17               0.094    0.018    5.223    0.000    0.094    0.321
  p28 ~~
    p30               0.425    0.070    6.084    0.000    0.425    0.468
  p29 ~~
    p30               0.315    0.068    4.615    0.000    0.315    0.332
  p17 ~~
    p24               0.077    0.018    4.195    0.000    0.077    0.241
  p24 ~~
    p26               0.108    0.026    4.252    0.000    0.108    0.305
  p1 ~~
    p4               -0.053    0.012   -4.386    0.000   -0.053   -0.252
  Satisfacao ~~
    Valor             1.556    0.152   10.242    0.000    0.963    0.963
    Reputacao         0.614    0.064    9.558    0.000    0.899    0.899
    Confianca         0.743    0.080    9.287    0.000    0.827    0.827
    Fidelidade        1.153    0.119    9.726    0.000    0.833    0.833
  Valor ~~
    Reputacao         0.753    0.082    9.179    0.000    0.898    0.898
    Confianca         0.927    0.103    8.997    0.000    0.840    0.840
    Fidelidade        1.447    0.154    9.427    0.000    0.852    0.852
  Reputacao ~~
    Confianca         0.450    0.048    9.287    0.000    0.965    0.965
    Fidelidade        0.660    0.069    9.496    0.000    0.918    0.918
  Confianca ~~
    Fidelidade        0.860    0.090    9.568    0.000    0.909    0.909
```

Figura 38. Continuação das estimativas do modelo.

Variances:				
p1	0.128	0.017	0.128	0.088
p2	0.160	0.018	0.160	0.126
p3	0.324	0.031	0.324	0.236
p4	0.343	0.032	0.343	0.378
p5	0.227	0.022	0.227	0.207
p6	0.116	0.013	0.116	0.108
p7	0.595	0.064	0.595	0.231
p8	0.506	0.048	0.506	0.393
p9	0.337	0.032	0.337	0.271
p10	0.282	0.033	0.282	0.184
p11	0.174	0.016	0.174	0.330
p12	0.256	0.025	0.256	0.259
p13	0.122	0.012	0.122	0.322
p14	0.504	0.047	0.504	0.362
p15	0.414	0.039	0.414	0.327
p16	0.272	0.025	0.272	0.307
p17	0.317	0.028	0.317	0.439
p18	0.187	0.019	0.187	0.170
p19	0.393	0.038	0.393	0.213
p20	0.326	0.032	0.326	0.199
p21	0.307	0.029	0.307	0.284
p22	0.109	0.011	0.109	0.132
p23	0.110	0.012	0.110	0.136
p24	0.325	0.031	0.325	0.214
p25	0.249	0.034	0.249	0.146
p26	0.389	0.042	0.389	0.232
p27	2.027	0.182	2.027	0.651
p28	1.019	0.096	1.019	0.456
p29	1.113	0.102	1.113	0.589
p30	0.809	0.080	0.809	0.334
Satisfacao	1.316	0.126	1.000	1.000
Valor	1.984	0.222	1.000	1.000
Reputacao	0.354	0.044	1.000	1.000
Confianca	0.614	0.074	1.000	1.000
Fidelidade	1.456	0.149	1.000	1.000

Figura 39. Continuação das estimativas do modelo.

Como foi visto na seção 5.2 pela limitação do cálculo da covariância dos resíduos padronizados no pacote *Lavaan*, utiliza-se os resíduos normalizados.

$cov

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13
p1	-0.021												
p2	-0.022	0.000											
p3	0.038	0.238	0.000										
p4	0.097	0.199	0.260	0.000									
p5	-0.022	-0.231	-0.288	-0.421	0.000								
p6	-0.092	-0.083	0.015	0.008	0.363	0.000							
p7	0.378	0.121	0.483	-0.056	-0.175	-0.056	0.000						
p8	-0.245	-0.240	0.114	-0.582	0.464	0.278	-0.390	0.000					
p9	-0.074	-0.171	-0.108	-0.026	0.413	0.060	-0.547	0.328	0.251				
p10	-0.042	-0.106	-0.485	-0.226	0.070	-0.124	-0.175	0.677	0.263	0.000			
p11	-0.307	-0.376	-0.719	0.200	-0.852	-0.464	-0.318	-1.173	-0.631	-0.432	0.000		
p12	0.345	0.736	0.433	0.978	0.170	-0.007	0.278	-0.239	0.704	0.858	0.218	0.000	
p13	-0.148	0.050	-0.356	0.555	-0.353	-0.030	-0.738	-0.868	-0.086	0.014	1.274	-0.443	0.159
p14	0.204	0.044	-0.758	-0.613	-0.087	-0.776	-0.309	-0.208	0.115	0.075	-0.073	-0.003	0.162
p15	0.502	0.636	0.221	-0.228	0.317	0.573	0.866	0.739	0.338	0.390	-0.453	0.060	-0.252
p16	0.423	0.511	0.903	0.474	0.800	0.591	0.749	0.688	1.326	0.689	0.006	0.302	0.473
p17	-0.120	0.045	0.953	1.703	0.249	0.143	0.652	0.209	0.898	0.756	1.883	0.475	0.497
p18	0.551	0.885	0.125	-0.458	-0.493	-0.339	0.173	-1.084	-0.431	-0.004	-0.125	0.297	-0.487
p19	0.614	0.652	0.208	0.182	-0.319	-0.058	0.471	-1.164	-0.070	0.365	0.057	0.293	0.224
p20	0.296	0.229	-0.040	0.459	-0.460	-0.014	0.408	-0.382	0.685	0.929	0.125	0.298	0.333
p21	-0.197	-0.101	-0.907	-0.508	-1.390	-0.840	0.061	-1.614	-0.604	-0.110	-0.106	-0.589	-0.014
p22	0.242	0.367	-0.307	-0.040	-0.734	-0.288	0.123	-1.369	-0.839	-0.117	-0.061	-0.335	-0.321
p23	0.473	0.503	-0.359	0.153	-0.052	0.234	0.037	-0.760	-0.040	0.451	0.634	-0.057	0.796
p24	-0.460	-0.009	-0.073	-0.057	-1.085	-0.742	-0.145	-1.023	-0.875	-0.335	0.401	-0.577	0.307
p25	-0.166	-0.120	-0.362	-0.472	-0.556	-0.249	-0.126	-1.354	-0.600	-0.448	-0.049	-0.153	0.460
p26	-0.384	-0.265	0.201	0.543	-0.560	-0.142	-0.016	-0.752	-0.406	-0.288	0.078	-0.647	0.670
p27	-0.296	-0.327	-0.440	0.206	-0.385	-0.446	0.111	-1.939	-0.207	0.855	-1.136	-0.047	-0.497
p28	0.903	0.701	0.753	0.527	1.667	1.333	1.325	2.396	2.132	1.930	-0.682	-0.144	0.041
p29	1.503	1.320	1.632	1.159	2.103	1.681	1.592	2.656	2.260	2.198	-0.785	-0.270	0.097
p30	0.421	0.438	0.586	-0.142	1.104	0.560	0.568	1.094	0.754	0.949	-0.764	0.195	-0.256

	p14	p15	p16	p17	p18	p19	p20	p21	p22	p23	p24	p25	p26
p1													
p2													
p3													
p4													
p5													
p6													
p7													
p8													
p9													
p10													
p11													
p12													
p13													
p14	0.000												
p15	0.680	0.000											
p16	-0.113	0.433	0.000										
p17	0.294	0.106	0.202	0.219									
p18	-0.299	-0.416	-0.342	-0.520	0.000								
p19	-0.057	-0.509	-0.043	-0.077	0.465	0.000							
p20	0.126	0.099	0.110	0.256	0.215	0.000	0.000						
p21	-0.229	-0.078	-0.164	0.079	0.058	0.391	0.646	0.000					
p22	0.038	-0.342	0.061	-0.470	-0.011	-0.330	-0.363	0.287	0.000				
p23	0.059	-0.355	-0.252	0.178	0.009	-0.075	-0.375	-0.450	0.283	0.000			
p24	-0.255	-0.136	0.213	0.375	-0.036	-0.367	-0.102	0.010	0.079	-0.132	-0.021		
p25	0.473	-0.321	0.359	0.025	0.142	0.263	0.276	-0.224	0.092	0.194	0.062	0.000	
p26	0.140	-0.897	0.148	0.905	-0.221	-0.134	-0.104	-0.798	-0.048	-0.162	-0.091	0.164	-0.042
p27	-1.387	-0.607	-0.105	-0.526	-0.543	0.242	0.499	0.142	-0.506	-0.174	-0.545	0.008	0.227
p28	0.481	1.086	-0.076	0.262	-0.181	-0.034	0.364	0.206	-0.576	-0.044	-0.288	-0.229	-0.212
p29	1.254	1.081	0.717	0.055	-0.415	-0.104	0.156	-0.137	-0.359	-0.250	0.133	-0.589	0.161
p30	0.882	1.119	0.088	0.320	-0.025	0.042	-0.042	-0.498	0.084	0.159	0.225	-0.173	-0.185

	p27	p28	p29	p30
p1				
p2				
p3				
p4				
p5				
p6				
p7				
p8				
p9				
p10				
p11				
p12				
p13				
p14				
p15				
p16				
p17				
p18				
p19				
p20				
p21				
p22				
p23				
p24				
p25				
p26				
p27	0.000			
p28	-0.399	0.000		
p29	0.117	0.000	0.000	
p30	0.408	0.000	0.000	0.000

Figura 40. Matriz de covariância residual normalizada

A matriz residual apresentou apenas poucos casos maiores que 1,96 mas bem próximos a 1,96, não apresentando maiores problemas. Finalmente o modelo para medir a retenção de clientes foi definido com algumas diferenças do modelo original, covariâncias entre as questões do questionário, cinco variáveis latentes e 30 variáveis observadas. O modelo final pode ser visto na Figura 42. Algumas correlações não aparecem no modelo por conta da proximidade das variáveis observadas.

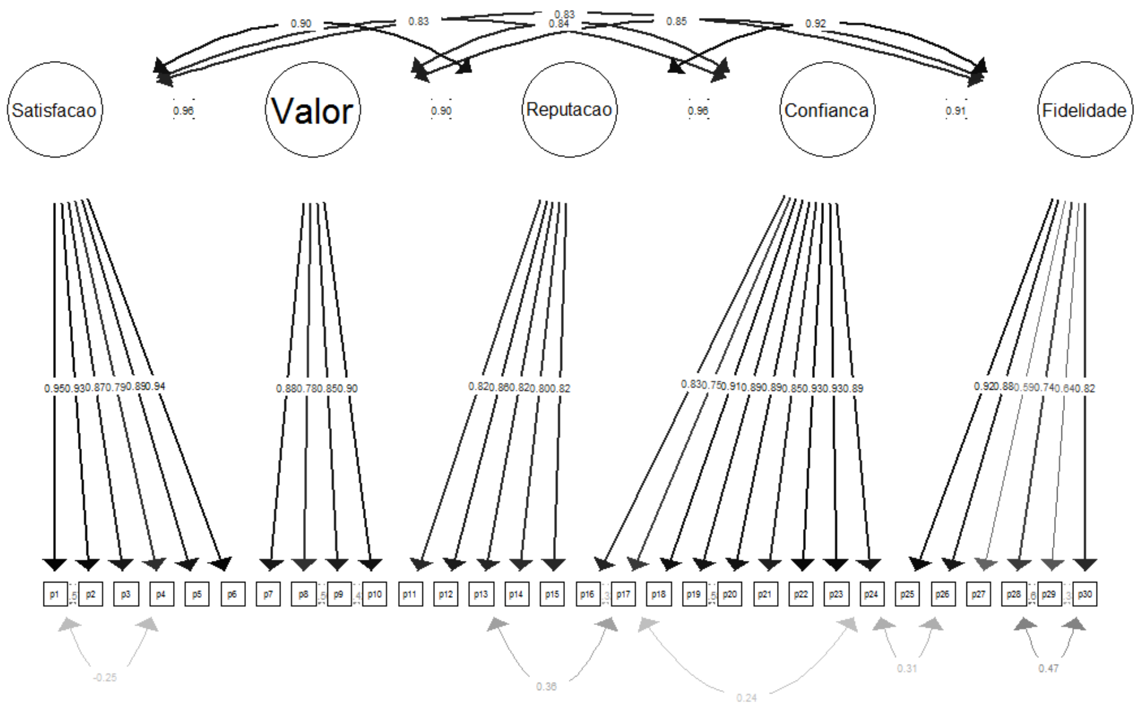


Figura 42. Diagrama de caminhos para o modelo final de retenção de clientes com as estimativas padronizadas e correlações

6. CONSIDERAÇÕES FINAIS

A Modelagem de Equações Estruturais (SEM) é um conjunto de técnicas estatísticas multivariadas que inclui análise de trajetória e análise fatorial integrando-as em modelos completos de regressão estrutural, estimando, simultaneamente, os parâmetros de uma série de equações de regressão linear que, embora separadas são interdependentes. O primeiro passo é a especificação do modelo, o qual deve possuir profundas raízes teóricas. A especificação inicial pode tomar a forma de um diagrama ou de uma sequência de equações. A técnica é de natureza confirmatória, em que o pesquisador tem o objetivo de avaliar modelos através dos índices de ajuste que verificam se a estrutura de variância e covariância da matriz de dados gerados pela amostra é consistente com a estrutura do modelo do pesquisador.

Existem diversos *softwares* que abordam a técnica de equações estruturais sendo em sua maioria comerciais. O objetivo principal deste trabalho foi desenvolver a técnica através do *software* livre R que é muito utilizado no meio estatístico. Para tanto foi utilizado o pacote *Lavaan* 0.5, que além de ser de programação intuitiva apresenta uma solução completa para a análise dos modelos, como estimativas dos parâmetros com suas significâncias estatísticas, índices de ajuste do modelo, matriz residual e índice de modificação.

Entre os objetivos específicos estava a exploração da forma prática das várias utilizações do uso de Equações Estruturais explicitando as etapas de utilização para cada uma das técnicas. Ao longo do trabalho foram analisados modelos de Equações Estruturais que abordam algumas técnicas como Regressão Multivariada, Análise Fatorial Confirmatória e Análise de Múltiplos Grupos.

No desenvolvimento deste assunto sentiu-se falta de um material de apoio devido à carência de recursos bibliográficos da Modelagem de Equações Estruturais no pacote *Lavaan*. Houve também dificuldades de encontrar referências bibliográficas que tivessem fundamentação teórica mais completa para algumas técnicas de apoio.

Algumas técnicas de Modelagem de Equações Estruturais não foram contempladas neste trabalho ficando como sugestão para outros estudos, como Modelo de Crescimento Latent (*growth curve models*) e Análise Fatorial de Segunda Ordem.

Pode-se também estender a modelagem para além do modelo de medida, analisando-se o modelo estrutural.

REFERÊNCIAS BIBLIOGRÁFICAS

ALGARVIO, S., LEAL, I., MAROCO, J., & MORENO,M. *Parental Concerns*: *Comparative Study between a group of Portuguese Parents and a group of Mozambican Parents. Internacional Journal of Developmental and Educational Psychology*, 1 (4), 199-208, 2008.

ARBUCKLE, J. L. AMOS 17 *users' guide.* Chicago, IL: SPSS, 2008.

ARBUCKLE, J. L. AMOS 18 *Reference Guide (Version 18) [Computer software]*. Chicago, IL: SPSS Inc., 2009.

BENTLER, P. M. *EQS – Structural Equations Program Manual*. University of California, Los Angeles, 1995a.

BENTLER, P. M., & WU, E. J. C. *EQS for Windows: User's Guide Encino*, CA: Multivariate Software, Inc., 1995b.

BROWNE, M., & CUDECK, R. *Single Sample Cross-validation Indices for Covariance Structures.* Multivariate Behavioral Research,24, 445-455, 1989.

BOLLEN, K. A. *Structural Equations with Latent Variables*. John Wiley & Sons, New York, 1989.

BYRNE, B. M. *Structural Equation Modeling With Amos: Basic Concepts, Applications, and Programming (Multivariate Applications Series).* Lawrence Erlbaum Associates, New Jersey. 2001.

HAIR, J. F., ANDERSON, R. E., THATHAN, R. L. & BLACK, W. C. Análise Multivariada de dados. 5 ed. Bookman, Porto Alegre, 2005.

JORESKOG, K. G. A general method for the analysis of covariance structures. Psychometrika, 34, 183-202 , 1970.

LEMKE, C., Modelos de Equações Estruturais com Ênfase em Analise Fatorial Confirmatória no *Software* AMOS. Universidade Federal do Rio grande do Sul, Instituto de Matematica, Departamento de Estatística. Monografia. Porto Alegre, 2005.

MAROCO, J. Análise de Equações Estruturais: Fundamentos teóricos, *Software* & Aplicações. Report number, Pêro Pinheiro, 2010.

MILAN, G. S. A prática do Marketing de Relacionamento e a Retenção de Clientes: Um Estudo Aplicado em um Ambiente de Serviços, Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Programa de Pós-Graduação em Engenharia de Produção. Tese. Porto Alegre, 2006.

OLSSON, U. H, TROYE, S. V., & HOWELL, R. D. *Theoretic fit and empirical fit: The performance of maximum likelihood versus generalized least squares estimation in structural equation models.* Multivariate Behavioral Research, 34(1), 31-59,1999.

ROSSEEL, Y., *Lavaan: An R Package for Structural Equation Modeling*. *Journal of Statistical Software*, 2012.

ROSSEEL, Y., *The Lavaan tutorial. Departament of Data Analysis,* Ghent University. Belgium, 2013.

SANTOS, R. B. Modelos de Equações Estruturais, Universidade Federal do Rio Grande do Sul, Instituto de Matemática, Departamento de Estatística. Monografia. Porto Alegre, 2002.

SILVA, J. F. Modelagem de equações estruturais : apresentação de uma metodologia, Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Programa de Pós-Graduação em Engenharia de Produção. Tese. Porto Alegre, 2006.