

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM ESTATÍSTICA

Inácio Torres Cardoso

**Ajuste de análise fatorial exploratória com
método dos mínimos resíduos a dados de
testes neurocognitivos**

Goiânia

2025



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC no 1240/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei no 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo dos Trabalhos de Conclusão dos Cursos de Graduação disponibilizado no RI/UFG é de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o(s) autor(a)(es)(as) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG)

Nome(s) completo(s) do(a)(s) autor(a)(es)(as): Inácio Torres Cardoso.

Título do trabalho: Ajuste de análise fatorial exploratória com método dos mínimos resíduos a dados de testes neurocognitivos.

2. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador) Concorda com a liberação total do documento [X] SIM [] NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante: a) consulta ao(à)(s) autor(a)(es)(as) e ao(à) orientador(a); b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo do TCCG. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro.

Obs.: Este termo deve ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Renato Rodrigues Silva, Professor do Magistério Superior**, em 02/12/2025, às 21:52, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Inácio Torres Cardoso, Discente**, em 03/12/2025, às 18:07, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5831481** e o código CRC **3D8DFE99**.

Referência: Processo nº 23070.060796/2025-98

SEI nº 5831481

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM ESTATÍSTICA

Inácio Torres Cardoso

**Ajuste de análise fatorial exploratória com método dos
mínimos resíduos a dados de testes neurocognitivos**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Estatística da Universidade Federal de Goiás para aprovação no componente curricular TCC, como parte das exigências para a obtenção do título de bacharel em Estatística.
Orientador: Prof. Dr. Renato Rodrigues Silva

Goiânia

2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Cardoso, Inácio Torres

Ajuste de análise fatorial exploratória com método dos mínimos resíduos a dados de testes neurocognitivos [manuscrito] / Inácio Torres Cardoso. - 2025.

35 f.

Orientador: Prof. Renato Rodrigues Silva.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Estatística, Goiânia, 2025.

Bibliografia. Apêndice.

Inclui gráfico, tabelas.

1. Fatores latentes. 2. Rotação varimax. 3. Análise paralela. 4. Memória de trabalho. 5. Memória verbal. I. Silva, Renato Rodrigues, orient. II. Título.

CDU 519.22



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Aos vinte e oito dias do mês de novembro do ano de 2025 iniciou-se a sessão pública de defesa do Trabalho de Conclusão de Curso (TCC) intitulado “Ajuste de análise fatorial exploratória com método dos mínimos resíduos a dados de testes neurocognitivos”, de autoria de Inácio Torres Cardoso, do curso de Estatística, do Instituto de Matemática e Estatística da UFG. Os trabalhos foram instalados pelo Prof. Dr. Renato Rodrigues Silva com a participação dos demais membros da Banca Examinadora: Joelmir Divino Carlos Feliciano (IME/UFG) e Tallyta Carolyne Martins da Silva (IME/UFG). Após a apresentação, a banca examinadora realizou a arguição do estudante. Posteriormente, de forma reservada, a Banca Examinadora atribuiu a nota final de 9,3, tendo sido o TCC considerado aprovado.

Proclamados os resultados, os trabalhos foram encerrados e, para constar, lavrou-se a presente ata que segue assinada pelos Membros da Banca Examinadora.



Documento assinado eletronicamente por **Joelmir Divino Carlos Feliciano, Professor do Magistério Superior**, em 02/12/2025, às 18:22, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Tallyta Carolyne Martins Da Silva, Professor do Magistério Superior**, em 02/12/2025, às 20:25, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Renato Rodrigues Silva, Professor do Magistério Superior**, em 02/12/2025, às 21:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5808281** e o código CRC **DC21CAD7**.

Agradecimentos

Agradeço muito a todos que me apoiaram incondicionalmente durante minha graduação, que me deram forças e apoio nos momentos em que mais precisei.

Agradeço aos meus amigos de graduação, que estiveram lá para descontraír o ambiente e que certamente levo comigo no futuro profissional e pessoal. Agradeço à minha família, por todo o carinho e amor que alguém poderia ter. Sem eles, eu certamente não teria perseverado até o fim.

Finalmente, agradeço ao professor orientador, Dr. Renato Rodrigues Silva, pela confiança no meu trabalho e pelos conselhos que se mostraram imprescindíveis para a realização desta monografia. Um abraço especial ao meu amigo Felipe Melo Sartes, que não mediu esforços para me ajudar e tirar dúvidas sobre o desenvolvimento de um trabalho deste nível. Ambos fizeram as dificuldades serem superadas com muito mais facilidade.

Resumo

Neste presente trabalho, foi aplicada uma análise fatorial exploratória (AFE) utilizando o método dos mínimos resíduos (Minres) em dados provenientes do NeuroCognitive Performance Test (NCPT), uma bateria de testes neurocognitivos. O estudo teve como objetivo central investigar a aplicabilidade do método Minres como uma alternativa robusta aos métodos de extração tradicionais, como o de componentes principais e o de máxima verossimilhança. A pesquisa utilizou dados da bateria 26 do NCPT, composta por 11 subtestes que avaliam diferentes processos cognitivos. O procedimento analítico iniciou-se com a verificação da adequação dos dados para a AFE por meio do índice Kaiser-Meyer-Olkin (KMO), que resultou em boa adequação (0,82). A definição do número de fatores a serem retidos foi feita através da análise paralela, que indicou a existência de três fatores. A extração das cargas fatoriais foi então realizada com o algoritmo Minres, seguida de uma rotação ortogonal do tipo varimax para facilitar a interpretação dos fatores. Os resultados revelaram uma estrutura de três fatores. O fator 1 mostrou um contraste entre subtestes de memória de trabalho e inteligência fluida, o fator 2 foi predominantemente carregado por subtestes de memória verbal, enquanto o fator 3 foi dominado pelo subteste de raciocínio aritmético. A qualidade do ajuste do modelo foi atestada pelo Root Mean Square Error of Approximation, ou RMSEA (0,037), e o Tucker-Lewis Index, ou TLI (0,969). Apesar do sucesso na aplicação do método, o estudo possui algumas limitações, como as baixas comunalidades encontradas em alguns subtestes e a possível influência do tipo de rotação utilizada. Conclui-se que o método dos mínimos resíduos mostrou-se plenamente aplicável e viável para a análise fatorial exploratória de dados neurocognitivos.

Palavras-chave: Fatores latentes, Rotação varimax, Análise paralela, Memória de trabalho, Memória verbal.

Abstract

In this study, an exploratory factor analysis (EFA) was applied using the minimum residuals method (Minres) to data from the NeuroCognitive Performance Test (NCPT), a battery of neurocognitive tests. The central objective of the study was to investigate the applicability of the Minres method as a robust alternative to traditional extraction methods, such as principal components and maximum likelihood. The research used data from the NCPT battery 26, comprising 11 subtests that assess different cognitive processes. The analytical procedure began by verifying the suitability of the data for EFA using the Kaiser-Meyer-Olkin (KMO) index, which indicated good suitability (0.82). The determination of the number of factors to retain was conducted through parallel analysis, which indicated the existence of three factors. The extraction of factor loadings was then performed using the Minres algorithm, followed by an orthogonal varimax rotation to facilitate the interpretation of the factors. The results revealed a three-factor structure. Factor 1 showed a contrast between working memory and fluid intelligence subtests, Factor 2 was predominantly loaded by verbal memory subtests, and Factor 3 was dominated by the arithmetic reasoning subtest. The model's fit quality was confirmed by the Root Mean Square Error of Approximation, or RMSEA (0.037), and the Tucker-Lewis Index, or TLI (0.969), indices. Despite the successful application of the method, the study has some limitations, such as the low communalities found in some subtests and the possible influence of the type of rotation used. It is concluded that the minimum residuals method proved to be fully applicable and viable for exploratory factor analysis of neurocognitive data.

Key-words: Latent factors, Varimax rotation, Parallel analysis, Working memory, Verbal memory.

Sumário

Introdução	10
2 Fundamentação Teórica	12
2.1 Modelo Fatorial Ortogonal	12
2.1.1 Pressuposições do Modelo e Resultados Importantes	12
2.1.2 Modelo Fatorial Ortogonal com Variáveis Padronizadas	14
2.2 O Método dos Mínimos Resíduos (Minres)	14
2.2.1 Aplicação com Método de Gauss-Seidel	15
2.2.2 O Algoritmo Minres	17
2.3 Validações Finais do Modelo	17
2.3.1 Índice Kaiser-Meyer-Olkin (KMO)	17
2.3.2 Análise Paralela	18
2.3.3 Rotação dos Fatores	19
2.3.4 Índices de Qualidade de Ajuste	19
3 Material e Métodos	21
3.1 NeuroCognitive Performance Test (NCPT) e Alterações no Banco de Dados	21
3.2 Análise Fatorial Exploratória	21
4 Resultados e Discussão	24
4.1 Resultados	24
4.1.1 Fator 1 (Teste dos Cinco Dígitos)	25
4.1.2 Fator 2 (Teste de Aprendizagem Auditivo-Verbal de Rey)	25
4.1.3 Fator 3 (Raciocínio Aritmético)	26
4.2 Discussão	26
Conclusão	28
Referências	29
Apêndices	31
APÊNDICE A Código R de Manipulação do Dataset da Bateria 26	32
APÊNDICE B Código R da Análise Paralela e Fatorial	34

Introdução

A análise fatorial é uma técnica estatística multivariada que estabelece um ou mais fatores que expliquem a variabilidade das variáveis observadas (Rogers, 2022). A análise fatorial tem aplicação em muitas áreas, como, por exemplo, na neuropsicologia para testes neurocognitivos (Rentergem *et al.*, 2020; Doraiswamy *et al.*, 2022; Morrison *et al.*, 2015). A análise fatorial busca reduzir o conjunto de variáveis observadas a um número menor de fatores latentes. Essas novas dimensões permitem explicar as correlações entre as variáveis sem perda substancial das informações essenciais (Harman, 1967). No entanto, a análise fatorial desempenha um papel específico de identificar construtos latentes (Rogers, 2022). A análise fatorial pode ser dividida em análise fatorial confirmatória (AFC) e análise fatorial exploratória (AFE). Em uma análise fatorial confirmatória, objetiva-se confirmar empiricamente as relações entre variáveis observadas e/ou latentes, a partir de uma teoria subjacente. Em uma análise fatorial exploratória, a investigação dos construtos responsáveis pela correlação das variáveis é feita sem a pressuposição de um modelo teórico subjacente (Rogers, 2022; Brown, 2015).

Contudo, um ponto chave em uma análise fatorial é a extração dos fatores. A extração dos fatores consiste na estimação dos parâmetros de carga fatorial e de variâncias específicas. Existem diversos métodos de extração, como o método de componentes principais, o método da máxima verossimilhança, o método de mínimos resíduos, entre outros (Rogers, 2022). Segundo Rogers (2022) e Brown (2015), o método de componentes principais é equivocado para a aplicação de uma análise fatorial, muitas vezes vindo de uma confusão do que cada análise executa. A análise de componentes principais é usada para redução de dimensionalidade e busca a maior preservação possível da variância total. A análise fatorial tem como objetivo principal estimar os construtos latentes e a variância comum entre as variáveis observadas (Rogers, 2022; Hair *et al.*, 2009). Assim, ao se utilizar do método de componentes principais, o pesquisador não estaria executando uma análise fatorial típica, podendo superestimar as cargas fatoriais e a variância explicada, interferindo na elaboração dos construtos latentes (Hair *et al.*, 2009; Rogers, 2022).

Além disso, outros métodos de extração também apresentam limitações na técnica da análise fatorial. O método por estimação de máxima verossimilhança (EMV) pressupõe o conhecimento a respeito da distribuição da amostra aleatória. No caso do modelo da análise fatorial, pressupõe-se que os dados tenham distribuição normal multivariada. Ainda, utiliza-se o teste de razão de verossimilhança para testar o número de fatores (Harman, 1967). Isso traz limitações no contexto onde as variáveis são qualitativas ordinais, por exemplo trazer parâmetros viesados e estatísticas qui-quadrados enganosas, principalmente em casos onde elas apresentam assimetria (Rogers, 2022; Xia; Yang, 2019). Em contrapartida, o método de mínimos resíduos visa minimizar a soma dos quadrados dos resíduos (SQR) fora das diagonais da matriz de correlação das variáveis (Rogers, 2022; Harman; Jones, 1966; Harman, 1967). Dessa maneira, a

estimação das cargas fatoriais não se fundamenta em pressupostos da distribuição normal, ou seja, é um método mais robusto à violação da pressuposição de distribuição normal multivariada (Rogers, 2022).

Neste contexto, o presente estudo é a aplicação de uma análise fatorial exploratória em um conjunto de dados do NeuroCognitive Performance Test (NCPT), composto de baterias com subtestes neurocognitivos (Jaffe *et al.*, 2022). A extração foi feita por meio do método de mínimos resíduos, buscando evitar gargalos provenientes do método de componentes principais ou da máxima verossimilhança.

Os objetivos gerais são estudar os fundamentos teóricos do método dos mínimos resíduos e investigar a aplicabilidade do método a um conjunto de dados de uma bateria de subtestes neurocognitivos do NeuroCognitive Performance Test (NCPT).

Já os objetivos específicos consistem em:

- Determinar o número de fatores do modelo;
- Realizar a extração dos fatores pelo método de mínimos resíduos;
- Fazer a rotação dos fatores comuns;
- Avaliar a adequabilidade do modelo.

2 Fundamentação Teórica

2.1 Modelo Fatorial Ortogonal

Seja um vetor aleatório \mathbf{X} de dimensão $p \times 1$. O modelo fatorial ortogonal é descrito pelas seguintes equações:

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \epsilon_1, \quad (1)$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \epsilon_2, \quad (2)$$

$$\vdots \quad (3)$$

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \epsilon_p. \quad (4)$$

Em notação matricial, o modelo fatorial ortogonal pode ser reescrito da seguinte forma:

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}, \quad (5)$$

em que $\boldsymbol{\mu}$ é o vetor aleatório de médias das p variáveis observáveis, \mathbf{L} é a matriz de carga fatoriais de dimensão $p \times m$, \mathbf{F} é o vetor de variáveis latentes, denominadas de fatores comuns, com dimensão $m \times 1$; $\boldsymbol{\epsilon}$ é o vetor aleatório de fatores específicos de dimensão $p \times 1$ (Johnson; Wichern, 2007).

2.1.1 Pressuposições do Modelo e Resultados Importantes

As pressuposições do modelo fatorial ortogonal são:

- A esperança do vetor de fatores comuns \mathbf{F} é um vetor de zeros:

$$E(\mathbf{F}) = \mathbf{0},$$

- A matriz de covariâncias do vetor de fatores comuns \mathbf{F} é a matriz identidade (fatores não correlacionados e com variância unitária):

$$\text{Cov}(\mathbf{F}) = \mathbf{I},$$

- A esperança do vetor de fatores específicos $\boldsymbol{\epsilon}$ é um vetor de zeros:

$$E(\boldsymbol{\epsilon}) = \mathbf{0},$$

- A matriz de covariâncias do vetor de fatores específicos ϵ é uma matriz diagonal Ψ (erros não correlacionados):

$$\text{Cov}(\epsilon) = \Psi,$$

- Assume-se que o vetor de fatores comuns \mathbf{F} e o vetor de fatores específicos ϵ não são correlacionados:

$$\text{Cov}(\epsilon, \mathbf{F}) = \mathbf{0}.$$

Assumindo estas condições, se chega a uma estrutura da covariância do vetor aleatório \mathbf{X} , representado por Σ , para o modelo fatorial ortogonal (Johnson; Wichern, 2007). O resultado é indicado pelas seguintes manipulações, partindo da definição da covariância:

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'], \quad (6)$$

$$= E[(\mathbf{LF} + \epsilon)(\mathbf{LF} + \epsilon)'], \quad (7)$$

$$= E[(\mathbf{LF} + \epsilon)(\mathbf{F}'\mathbf{L}' + \epsilon')], \quad (8)$$

$$= E[\mathbf{LFF}'\mathbf{L}' + \mathbf{LF}\epsilon' + \epsilon\mathbf{F}'\mathbf{L}' + \epsilon\epsilon'], \quad (9)$$

$$= \mathbf{LE}[\mathbf{FF}']\mathbf{L}' + \mathbf{LE}[\mathbf{F}\epsilon'] + E[\epsilon\mathbf{F}'\mathbf{L}'] + E[\epsilon\epsilon'], \quad (10)$$

$$= \mathbf{LIL}' + \mathbf{L}(\mathbf{0}) + (\mathbf{0})\mathbf{L}' + \Psi, \quad (11)$$

$$= \mathbf{LL}' + \Psi. \quad (12)$$

Pela independência do vetor de fatores específicos ϵ e o vetor de fatores comuns \mathbf{F} , também se chega ao seguinte resultado para a covariância entre o vetor das variáveis \mathbf{X} e \mathbf{F} :

$$\text{Cov}(\mathbf{X}, \mathbf{F}) = E[(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}'], \quad (13)$$

$$= E[(\mathbf{LF} + \epsilon)\mathbf{F}'], \quad (14)$$

$$= E[\mathbf{LFF}' + \epsilon\mathbf{F}'], \quad (15)$$

$$= \mathbf{LE}[\mathbf{FF}'] + E[\epsilon\mathbf{F}'], \quad (16)$$

$$= \mathbf{LI} + \mathbf{0}, \quad (17)$$

$$= \mathbf{L}. \quad (18)$$

Estes resultados são a base para a interpretação dos coeficientes da matriz de cargas fatoriais \mathbf{L} . De acordo com (18), nota-se que l_{ij} é o valor da covariância da variável aleatória X_i e o fator comum F_j .

A variância de uma variável aleatória X_i pode ser decomposta em variância devido aos fatores comuns e a variância específica.

$$\sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i = h_i^2 + \psi_i, \quad (19)$$

em que σ_{ii} é a variância da i -ésima variável (o i -ésimo elemento na diagonal da matriz de covariâncias Σ), l_{ij}^2 é o quadrado da carga fatorial da i -ésima variável no j -ésimo fator, h_i^2 é a comunalidade da i -ésima variável e ψ_i é a variância específica da i -ésima variável (Johnson; Wichern, 2007).

2.1.2 Modelo Fatorial Ortogonal com Variáveis Padronizadas

Conforme visto na seção anterior, a estrutura de covariância para o modelo fatorial ortogonal é dada por (12). Para variáveis padronizadas, representadas por $Z_i = \frac{X_i - \mu_i}{\sigma_i}$, sendo i o valor para a i -ésima variável, a matriz de covariâncias Σ torna-se a matriz de correlações ρ . Assim, o modelo é expresso como:

$$\rho = \mathbf{L}\mathbf{L}' + \Psi. \quad (20)$$

Nesse caso, a comunalidade pode ser interpretada como a proporção da variância da variável X_i devido aos fatores comuns, ou seja,

$$\rho_{ii} = 1 = h_i^2 + \psi_i = \sum_{j=1}^m l_{ij}^2 + \psi_i. \quad (21)$$

Para ajustar o modelo fatorial ortogonal aos dados, utiliza-se a matriz de correlação amostral:

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{12} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \dots & 1 \end{bmatrix}. \quad (22)$$

2.2 O Método dos Mínimos Resíduos (Minres)

O método dos mínimos resíduos (Minres), proposto por Harman e Jones (1966), oferece uma solução direta para o problema de estimação das cargas fatoriais.

Inicialmente, considere o estimador correlação amostral reduzida, \mathbf{R}^+ , dado por:

$$\mathbf{R}^+ = \begin{bmatrix} \hat{h}_1^2 & r_{12} & \dots & r_{1p} \\ r_{12} & \hat{h}_2^2 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \dots & \hat{h}_p^2 \end{bmatrix}. \quad (23)$$

De acordo com Harman e Jones (1966), o teorema fundamental da análise fatorial afirma que:

$$\mathbf{R}^+ = \mathbf{L}\mathbf{L}'. \quad (24)$$

A minimização será feita baseada nos elementos fora da diagonal principal da matriz \mathbf{R} .

No entanto, tem-se a seguinte relação:

$$\mathbf{R} - \mathbf{I} = \mathbf{R}^+ - \mathbf{H}, \quad (25)$$

em que \mathbf{H} é a matriz diagonal das comunalidades:

$$\mathbf{H} = \mathbf{I} - \mathbf{\Psi} = \text{diag}(\mathbf{L}\mathbf{L}'). \quad (26)$$

Dessa forma, a condição para solução Minres pode ser precisamente:

$$\min_{\mathbf{L}} \|([\mathbf{R} - \mathbf{I}] - [\mathbf{L}\mathbf{L}' - \text{diag}(\mathbf{L}\mathbf{L}')])\|, \quad (27)$$

sendo $\|\cdot\|$ a norma de Frobenius (Harman, 1967).

Ou seja, o método dos mínimos resíduos consiste em minimizar a distância entre os elementos fora da diagonal da matriz \mathbf{R} e os elementos fora da matriz $\mathbf{L}\mathbf{L}'$.

Algebricamente, pode-se reescrever (27) da seguinte forma:

$$f(\mathbf{L}) = \sum_{k=j+1}^p \sum_{j=1}^{p-1} (r_{jk} - \sum_{p=1}^m l_{jp}l_{kp})^2, \quad (28)$$

onde r_{jk} é o elemento do j -ésima variável e k -ésimo fator da matriz de correlação amostral \mathbf{R} .

Após verificar estes resultados, pode se afirmar que o objetivo do método dos mínimos resíduos é minimizar a função $f(\mathbf{L})$ para uma específica função de m , variando os valores das cargas fatoriais (Harman, 1967). Como sub-produto, é assumido que a comunalidade satisfaz a seguinte condição:

$$h_i^2 = \sum_{i=1}^m l_{jp} \leq 1, \quad (29)$$

para $j = 1, 2, \dots, p$.

2.2.1 Aplicação com Método de Gauss-Seidel

Uma forma de permitir a aplicação do método Minres computacionalmente, de forma eficiente, baseia-se no processo de Gauss-Seidel (Harman, 1967). Também conhecido como

"método dos deslocamentos sucessivos", consiste em realizar pequenas modificações na variável alvo e depois substituí-la por estes novos valores, iterativamente.

Considere, para qualquer linha j da matriz L , um incremento ϵ_p , com $p = 1, 2, \dots, m$:

$$l_{j1} + \epsilon_1, l_{j2} + \epsilon_2, \dots, l_{jm} + \epsilon_m. \quad (30)$$

Dessa forma, uma nova carga fatorial b_{jp} pode ser escrita por:

$$b_{jp} = l_{j1} + \epsilon_p. \quad (31)$$

Então a correlação reproduzida de uma variável fixa j , com qualquer variável k , é:

$$r_{jk}^+ = \sum_{p=1}^m l_{kp} b_{jp}, \quad (32)$$

e a soma de quadrados das correlações residuais com esta variável é dada por:

$$f_j = \sum_{k=1, k \neq j}^p (r_{jk} - \sum_{p=1}^m l_{kp} b_{jp})^2. \quad (33)$$

Ao separar a carga fatorial original da mudança incremental, a expressão em (33) torna-se

$$f_j = \sum_{k=1, k \neq j}^p (r_{jk}^* - \sum_{p=1}^m l_{kp} \epsilon_p)^2, \quad (34)$$

em que r_{jk}^* são correlações residuais originais, ou seja:

$$r_{jk}^* = r_{jk} - \sum_{p=1}^m l_{kp} l_{jp}, \quad (35)$$

onde $k = 1, 2, \dots, p$ e $k \neq j$.

Para determinar os valores de ϵ que minimizam a função objetivo f , primeiro é considerada a derivada parcial com respeito à ϵ_q , como mostrado a seguir:

$$\frac{\partial f}{\partial \epsilon_q} = -2 \sum_{k=1, k \neq j}^p (r_{jk}^* - \sum_{p=1}^m l_{kp} \epsilon_p) (-l_{kq}). \quad (36)$$

Igualando (36) a zero, tem-se que:

$$\sum_{p=1}^m \left(\sum_{k=1, k \neq j}^p l_{kp} l_{kq} \right) \epsilon_p = \sum_{k=1, k \neq j}^p r_{jk}^* l_{kq}. \quad (37)$$

onde $q = 1, 2, \dots, m$.

Este procedimento pode ser escrito em formato matricial, como descrito a seguir:

$$\epsilon_j \mathbf{L}'_{j(j)} (\mathbf{L})_{j(j)} = \mathbf{r}_j^0 \mathbf{L}, \quad (38)$$

onde $\epsilon_j = (\epsilon_1, \epsilon_2, \dots, \epsilon_m)$ é o vetor linha incremental de mudança das cargas fatoriais para a variável j , $\mathbf{L}_{j(j)}$ é a matriz de cargas fatoriais com elementos na linha j substituídos por zeros e \mathbf{r}_j^0 é o vetor linha das correlações residuais da variável j com outras variáveis, e zero para o próprio resíduo.

Então, a solução para os deslocamentos para carga fatorial é dada por:

$$\epsilon_j = \mathbf{r}_j^0 \mathbf{L} (\mathbf{L}'_{j(j)} (\mathbf{L})_{j(j)})^{-1}. \quad (39)$$

2.2.2 O Algoritmo Minres

O processo iterativo para encontrar as cargas fatoriais pelo método Minres pode ser resumido no seguinte algoritmo:

1. Iniciar com uma estimativa para a matriz de cargas fatoriais arbitrária \mathbf{R} .
2. Obtenha ϵ por meio de (39). Designe a matriz de cargas fatoriais com a primeira linha substituída por (30).
3. Compute um ciclo de iterações L_1, L_2, \dots, L_p que serão determinadas por $\epsilon_1, \epsilon_2, \dots, \epsilon_p$.
4. Aplique teste de convergência após computar L_{cp} , sendo c o número de ciclos.
5. Repetir os passos 3 e 4 para todos os elementos de \mathbf{L} até que convirja.

2.3 Validações Finais do Modelo

A seguir, são detalhados os procedimentos e critérios empregados para garantir a adequação e a robustez da análise fatorial exploratória, com o uso do modelo fatorial ortogonal e aplicação do algoritmo dos mínimos resíduos.

2.3.1 Índice Kaiser-Meyer-Olkin (KMO)

Para que possamos proceder para a extração com o método dos mínimos resíduos, primeiro é possível avaliar a adequação da matriz de correlação amostral \mathbf{R} para a análise fatorial. O

índice Kaiser-Meyer-Olkin (KMO) é uma medida que compara as magnitudes dos coeficientes de correlação amostrais observados com as magnitudes dos coeficientes de cargas fatoriais parciais (Cureton; D'Agostino, 1993).

A lógica do índice Kaiser-Meyer-Olkin é que, se as variáveis compartilham fatores comuns, as correlações amostrais parciais entre pares de variáveis (após a remoção do efeito das demais) devem ser pequenas. Um valor elevado do KMO indica que a matriz de correlações amostrais é apropriada para a fatorização. A estatística é calculada como:

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum l_{ij}^2}, \quad (40)$$

onde r_{ij} é o elemento de matriz de correlação amostral R entre a variáveis i e o fator j , e l_{ij} é a carga fatorial da i -ésima variável e j -ésimo fator.

Os valores do KMO variam de 0 a 1. Valores próximos de 1 indicam que a análise fatorial exploratória é adequada, enquanto valores abaixo de 0,5 indicam que não é recomendada. Conforme a diretriz de Kaiser e Rice (1974), valores de KMO acima de 0,6 são considerados aceitáveis para prosseguir com a AFE.

2.3.2 Análise Paralela

A decisão sobre o número de fatores é um importante passo da aplicação de uma análise fatorial exploratória. Um dos métodos mais utilizados para essa determinação é a análise paralela, proposta originalmente por Horn (1965). A análise paralela é um procedimento baseado em autovalores (ou raízes latentes) que orienta a seleção do número apropriado de fatores na análise fatorial exploratória. É considerado um método de retenção fatorial com melhor acurácia em comparação com critérios tradicionais, como o critério de Kaiser (autovalor > 1), o gráfico *scree* e a porcentagem de variância explicada, que têm sido consistentemente criticados por superestimar o número de fatores (Rogers, 2022).

O método funciona comparando os autovalores da matriz de correlação dos dados observados com os autovalores de uma matriz de correlação gerada a partir de dados aleatórios, não correlacionados, com as mesmas dimensões (mesmo número de variáveis e de observações). O procedimento baseia-se em um gráfico *scree* comparativo que plota os autovalores obtidos a partir dos dados amostrais contra os autovalores estimados. Os dados aleatórios devem "paralelizar" aspectos dos dados reais da pesquisa, como o tamanho da amostra (Horn, 1965).

O critério de retenção é indicado pelo ponto onde as duas linhas se cruzam. A seleção de fatores é orientada pelo número de autovalores reais que são maiores do que os autovalores correspondentes gerados pelos dados aleatórios. Se um fator real explica menos variância do que o fator correspondente obtido a partir de números aleatórios, ele não deve ser incluído na

análise fatorial. A sugestão de um fator que explica menos variância do que um fator gerado aleatoriamente possivelmente se trata de um erro amostral (Horn, 1965).

2.3.3 Rotação dos Fatores

Após a extração inicial dos fatores, a matriz de cargas fatoriais frequentemente não apresenta uma estrutura clara que facilite a interpretação. A rotação dos fatores é um procedimento que busca transformar a matriz de cargas fatoriais em uma solução mais simples e teoricamente mais significativa, sem alterar as propriedades matemáticas fundamentais do modelo, como as comunalidades e a variância total explicada. O objetivo é alcançar uma estrutura onde cada variável tenha uma carga alta em apenas um fator e cargas fatoriais baixas nos demais.

A rotação pode ser ortogonal, mantendo os fatores não correlacionados, ou oblíqua, permitindo que os fatores sejam correlacionados. A rotação ortogonal do tipo varimax, desenvolvida por Kaiser (1958), é uma das mais utilizadas. O critério varimax visa simplificar as colunas da matriz de cargas fatoriais, maximizando a variância dos quadrados das cargas em cada fator. Isso é alcançado ao se maximizar a seguinte função:

$$V = \sum_{j=1}^m \left[n \sum_{i=1}^n (\ell_{ij}^2 / h_i^2)^2 - \left(\sum_{i=1}^n \ell_{ij}^2 / h_i^2 \right)^2 \right], \quad (41)$$

onde ℓ_{ij} é a carga da variável i no fator j , e h_i^2 é a comunalidade da variável i . A maximização dessa função resulta em cargas que tendem a ser próximas de 1 ou 0, facilitando a atribuição de cada variável a um único fator.

2.3.4 Índices de Qualidade de Ajuste

Para avaliar o quão bem o modelo fatorial estimado se ajusta aos dados observados, diversos índices de qualidade de ajuste podem ser utilizados. No presente estudo, utilizaremos o Root Mean Square Error of Approximation (RMSEA) e o Tucker-Lewis Index (TLI). A seguir, explicamos suas definições e a forma com a qual são aplicados:

- O Root Mean Square Error of Approximation (RMSEA) é uma medida de erro de aproximação na população. Ele avalia o quão bem o modelo, com seus parâmetros estimados, se ajustaria à matriz de covariância da população. Valores de RMSEA menores que 0,05 indicam um bom ajuste, enquanto valores acima de 0,10 sugerem um ajuste inadequado. A fórmula é dada por:

$$\text{RMSEA} = \sqrt{\frac{\chi^2 - df}{df(N - 1)}}, \quad (42)$$

em que χ^2 é o valor da estatística qui-quadrado, df são os graus de liberdade e N é o tamanho da amostra (Hair *et al.*, 2009).

- O Tucker-Lewis Index (TLI) é um índices de ajuste comparativo. Ele compara o ajuste do modelo proposto a um modelo de base, geralmente o modelo nulo (ou de independência), no qual as variáveis não são correlacionadas. Os valores do TLI podem ultrapassar os limites de 0 e 1 (Hair *et al.*, 2009). Valores próximos de 1, preferencialmente acima de 0,95, indicam um bom ajuste do modelo (Shook-Sa; Chen; Zhou, 2017). A fórmula é dada por:

$$TLI = \frac{(\chi_{nulo}^2/df_{nulo}) - (\chi_{modelo}^2/df_{modelo})}{(\chi_{nulo}^2/df_{nulo}) - 1}, \quad (43)$$

onde χ_{nulo}^2 e df_{nulo} são os valores da estatística qui-quadrado e dos graus de liberdade para o modelo, e χ_{modelo}^2 e df_{modelo} os valores da estatística qui-quadrado e graus de liberdade para o modelo proposto (Hair *et al.*, 2009).

3 Material e Métodos

3.1 NeuroCognitive Performance Test (NCPT) e Alterações no Banco de Dados

O conjunto de dados é composto por resultados dos testes cognitivos informatizados Neuro-Cognitive Performance Test (NCPT). Esse conjunto de dados foi descrito primeiramente por Jaffe *et al.* (2022). O NCPT é composto por oito baterias de testes e foi aplicado a aproximadamente 750.000 adultos (Jaffe *et al.*, 2022). Neste trabalho, utilizou-se apenas a bateria número 26 (Jaffe *et al.*, 2022). A bateria 26 foi aplicada à 1504 indivíduos adultos e é composta pelos subtestes 36, 39, 40, 29, 28, 33, 30, 27, 32, 38 e 37. A descrição dos subtestes é feita a seguir: aprendizagem de lista verbal, ou memória verbal (36), teste de trilha parte A, ou velocidade de processamento (39), teste de trilha parte B, ou flexibilidade cognitiva (40), raciocínio aritmético (29), teste de memória de curto prazo e de trabalho (28), span de memória reverso, ou memória de trabalho (33), matriz progressiva de Raven (inteligência fluida) (30), atenção visual dividida (27), teste da função executiva tomada de decisão (32), teste de codificação de símbolos de dígitos, ou memória de trabalho e velocidade de processamento (38) e lista verbal atrasada, ou memória verbal (37). Descrições mais detalhadas sobre os subtestes podem ser encontradas na tabela 1. Além das pontuações dos testes, Jaffe *et al.* (2022) coletou variáveis sociodemográficas (idade, sexo, nível educacional, país de origem) e a hora do dia que foi iniciado o teste. Dessa forma, o banco de dados contém 16 variáveis e 416.118 linhas. Após a importação dos dados, aplicou-se a técnica de pivoteamento. Como resultado do pivoteamento, obteve-se um conjunto de dados em que cada coluna representa a pontuação de um dos 11 subtestes da bateria 26. A preparação dos dados foi realizada utilizando-se o software estatístico R (R Core Team, 2021) e os pacotes *sparklyr* e *dplyr* (Santos, 2021).

3.2 Análise Fatorial Exploratória

A análise fatorial exploratória foi realizada baseada nas pontuações dos subtestes da bateria número 26. O procedimento de análise é descrito a seguir.

Matrizes de correlações foram estimadas para estudar a relação entre cada subteste observado. O índice Kaiser-Meyer-Olkin (KMO) foi utilizado para verificar a adequação dos dados a uma AFE (Cureton; D'Agostino, 1993). Valores de KMO acima de 0,6 foram considerados aceitáveis para prosseguir com a análise fatorial (Kaiser; Rice, 1974). A escolha de números de fatores foi feita a partir do método da análise paralela (Timmerman; Lorenzo-Seva, 2011).

A extração dos fatores, ou seja, a estimação dos parâmetros de carga fatoriais e variâncias específicas será feita por meio do método dos mínimos resíduos. O método dos resíduos mínimos pode ser englobado em uma classe de métodos mais gerais denominados métodos dos mínimos quadrados (Harman; Jones, 1966). Rotação ortogonal dos fatores será aplicada para facilitar a interpretação das cargas fatoriais. A rotação do tipo varimax foi adotada nesse trabalho (Kaiser, 1958).

A adequação do modelo foi avaliada por meio dos índices de qualidade de ajuste Root Mean Square Error of Approximation (RMSEA) e Tucker-Lewis Index (TLI), conforme Shook-Sa, Chen e Zhou (2017). Valores próximos de 1, preferencialmente acima de 0,95, indicam boa adequação do modelo no caso do TLI, enquanto para o RMSEA são indicados valores menores que 0,05. Todos os cálculos e estimativas foram feitos com o auxílio do software estatístico R (R Core Team, 2021) e os pacotes *tidyverse* e *psych*.

Tabela 1 – Descrições dos subtestes da bateria 26 por processo cognitivo e o tipo de escore

ID do Subteste	Nome do Subteste	Processos Cognitivos	Escore
27	Atenção visual dividida	Atenção Dividida	Número de respostas corretas em 12 tentativas.
28	Extensão de memória direita	Visão Espacial e Memória de Trabalho	Nível de amplitude máxima da sequência em que o participante acertou pelo menos uma tentativa.
29	Raciocínio Aritmético	Inteligência Fluida e Cristalizada	Número de respostas corretas em 45 segundos.
30	Matriz Progressiva de Raven	Inteligência Fluida	Número de respostas corretas menos o número de respostas incorretas em 45 segundos.
32	Go/no-go	Controle inibitório	Tempo de resposta média em tentativas corretas em milissegundos.
33	Span de memória reverso	Memória de trabalho	Nível de amplitude máxima da sequência em que o participante acertou pelo menos uma tentativa.
36	Aprendizagem de lista verbal	Memória verbal	Número total de palavras lembradas corretamente na terceira rodada.
37	Lista verbal atrasada	Memória verbal	Número total de palavras lembradas corretamente.
38	Teste de codificação de símbolos de dígitos	Memória de trabalho e velocidade de processamento	Número de respostas corretas em 90 segundos.
39	Teste de trilha parte A	Velocidade de processamento	Tempo gasto para clicar em todos os números, em segundos.
40	Teste de trilha parte B	Flexibilidade cognitiva	Tempo gasto para clicar em todos os círculos, em segundos.

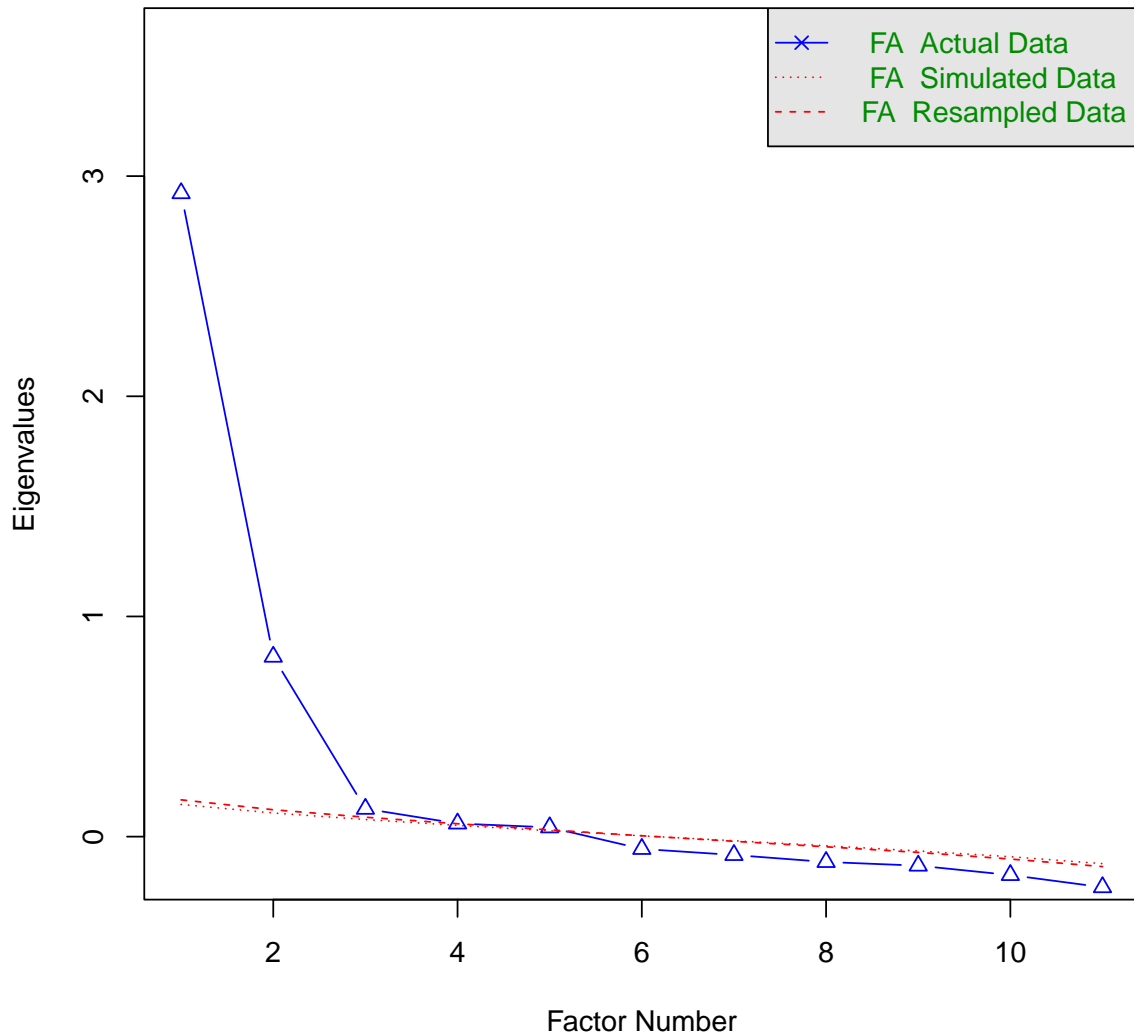
Fonte: Elaborado pelo autor.

4 Resultados e Discussão

4.1 Resultados

A estimativa do índice Kaiser-Meyer-Olkin (0,82) configura uma boa adequação para um estudo do tipo análise fatorial exploratória. A análise paralela indicou três fatores, pois, até o terceiro fator, os autovalores estimados da amostra original são maiores do que o autovalores obtidos por simulação, como mostrado na figura 1:

Figura 1 – Gráfico da Análise Paralela



Fonte: Elaborado pelo autor

O próximo passo realizado foi o ajuste do modelo fatorial ortogonal com três fatores usando o método de extração Minres, com rotação de fatores varimax. As cargas fatoriais associadas a

cada subteste, além da comunalidade (\hat{h}^2) e variância específica ($\hat{\psi}^2$), são mostrados a seguir na tabela 2:

Tabela 2 – Cargas Fatoriais Padronizadas para os fatores F1, F2 e F3, Comunalidades e Variância Específica por Subteste

Subtestes	F1	F2	F3	\hat{h}^2	$\hat{\psi}^2$
36 (Aprendizagem de lista verbal)	0.14	0.86	0.11	0.77	0.23
39 (Teste de trilha parte A)	-0.38	-0.07	-0.12	0.16	0.84
40 (Teste de trilha parte B)	-0.53	-0.18	-0.25	0.38	0.62
29 (Raciocínio Aritmético)	0.30	0.14	0.79	0.74	0.26
28 (Extensão de memória direta)	0.58	0.06	0.09	0.35	0.65
33 (Span de memória reverso)	0.53	0.10	0.10	0.30	0.70
30 (Matriz Progressiva de Raven)	0.35	0.12	0.30	0.23	0.77
27 (Atenção visual dividida)	0.41	0.13	0.08	0.19	0.81
32 (Go/no-go)	-0.41	-0.03	-0.18	0.20	0.80
38 (Teste de codificação de símbolos de dígitos)	0.66	0.15	0.29	0.55	0.45
37 (Lista verbal atrasada)	0.16	0.75	0.10	0.60	0.40

Fonte: Elaborado pelo autor

Os índices de qualidade de ajuste demonstraram boa adequação ao uso da análise fatorial (RMSEA = 0,037; TLI = 0,969). A partir dos resultados acima, análises acerca de cada fator, refletindo sobre as relações existentes entre os subtestes e suas comunalidades e variâncias específicas, podem ser feitas.

4.1.1 Fator 1 (Teste dos Cinco Dígitos)

No primeiro fator F1, o contraste das cargas fatoriais entre os subtestes que englobam processos cognitivos como memória de trabalho (28, 33 e 38) e inteligência fluida (29 e 30), com valores elevados positivos, contra controle inibitório (32), velocidade de processamento (39) e flexibilidade cognitiva (40), com valores elevados negativos. Na literatura que abrange as avaliações neuropsicológicas, o Teste de Cinco Dígitos (FDT) é o que mais se assemelha a esta oposição encontrada nos dados, visto que avalia velocidade de processamento, funções executivas e funcionamento atencional (Campos *et al.*, 2016).

4.1.2 Fator 2 (Teste de Aprendizagem Auditivo-Verbal de Rey)

Na análise do segundo fator F2, os subtestes 36 e 37, que apresentam a memória verbal como processo cognitivo, são os que possuem maior carga fatorial positivas (0,86 e 0,75, respectivamente). Altos valores para as comunalidades (77% e 60%, respectivamente) demonstram que grande parte da variabilidade destas variáveis é explicada na estrutura fatorial. O construto do fator se assemelha ao que é avaliado no Teste de Aprendizagem Auditivo-Verbal de Rey

(RAVLT). Este teste é considerado um dos mais eficazes para avaliação da aquisição, formação, conservação e evocação das informações verbais (Cotta *et al.*, 2012).

4.1.3 Fator 3 (Raciocínio Aritmético)

No terceiro fator F3, há a predominância do subteste 29, raciocínio aritmético, com carga fatorial alta e positiva (0,79). Apesar de ter uma carga moderadamente relevante no primeiro teste (0,30), o processo cognitivo do subteste em questão não é medido nos testes FDT ou RAVLT, relacionados com os outros fatores.

4.2 Discussão

Após analisar os resultados da pesquisa realizada a partir dos subtestes que compõem o conjunto de dados, o próximo passo lógico envolve comparar os resultados obtidos neste estudo com os demais trabalhos disponíveis na literatura. Santos *et al.* (2015) realizou uma análise fatorial exploratória com diferentes formas de rotação dos fatores (varimax e oblimin) e métodos de extração (fatoração do eixo principal, máxima verossimilhança e componentes principais). Os autores retiraram variáveis neurocognitivas que apresentaram baixa comunalidade, algo não realizado nesta presente pesquisa. Dentre os três fatores encontrados por estes autores, o primeiro indica processos cognitivos similares, como memória de trabalho, em uma denominação de função executiva geral, ou "GENEXEC", enquanto o segundo também apresenta similaridades, pois indica variáveis que apresentavam processos cognitivos relacionados à memória, mais especificamente nomeado como "MEM" (Santos *et al.*, 2015).

Com relação às baixas comunalidades encontradas em algumas variáveis, é notável frisar que os subtestes 32, 39 e 40 fazem parte de famílias de testes que mensuram também outros processos cognitivos. Os testes de trilha parte a e b, por exemplo, levam em conta flexibilidade cognitiva e velocidade de processamento, como exposto por Salthouse (2011), o que lhes configura com uma alta variância específica e com cargas fatoriais negativas para os três fatores. O subteste go/no-go também apresenta cargas fatoriais negativas e baixa comunalidade, o que indica a ausência de contemplação do controle inibitório, processo cognitivo principal do subteste (Littman; Takács, 2017).

Outra limitação do estudo realizado advém da escolha do Minres como método de extração, visto que, apesar de ser mais flexível quanto à distribuição das variáveis, a utilização de correlação de Pearson não segue a mesma flexibilização sobre as pressuposições. Além disso, a escolha de um diferente método de rotação poderia impactar na performance em um cenário favorável para a rotação oblimin, recomendada na literatura (Santos *et al.*, 2015). Possivelmente, o fator 3 do presente estudo pode se assemelhar ao teste de raciocínio de Baddeley, que já apresentou

correlação com outros testes de medição de conhecimento matemático e raciocínio aritmético (Baghaei; Khoshdel-Niyat; Tabatabaee-Yazdi, 2017).

Conclusão

No presente estudo, se investigou a aplicabilidade do método dos mínimos resíduos em uma bateria de subtestes neurocognitivos NCPT. Apesar de algumas limitações como a baixa comunalidade de algumas das variáveis e uma falta de interpretabilidade clara do terceiro fator, o método Minres se mostrou tão aplicável quanto as demais formas tradicionais de extração para problemas aplicáveis à análise fatorial exploratória. Futuras reedições de pesquisas com testes neurocognitivos podem aperfeiçoar os resultados ao tentar alternativas vistas na literatura, como remover variáveis com alta variância específica ou encontrar testes aptos a serem aplicados com determinados processos cognitivos em mente, de modo a melhorar a capacidade dos construtos latentes de explicar a variabilidade dos subtestes.

Referências

- BAGHAEI, P.; KHOSHDEL-NIYAT, F.; TABATABAEE-YAZDI, M. The Persian adaptation of Baddeley's 3-min grammatical reasoning test. **Psicologia: Reflexão e Crítica**, 2017. Research Open Access. Received: 4 May 2017; Accepted: 30 July 2017. Citado na página 27.
- BROWN, T. A. **Confirmatory Factor Analysis for Applied Research**. Second edition. New York, NY: The Guilford Press, 2015. (Methodology in the Social Sciences). ISBN 978-1-4625-1779-4. Citado na página 10.
- CAMPOS, M. C. *et al.* Confiabilidade do teste dos cinco dígitos em adultos brasileiros. **Jornal Brasileiro de Psiquiatria**, Instituto de Psiquiatria da Universidade Federal do Rio de Janeiro, v. 65, n. 2, p. 135–139, Apr 2016. ISSN 0047-2085. Disponível em: <<https://doi.org/10.1590/0047-2085000000114>>. Citado na página 25.
- COTTA, M. F. *et al.* O Teste de Aprendizagem Auditivo-Verbal de Rey (RAVLT) no diagnóstico diferencial do envelhecimento cognitivo normal e patológico. **Contextos Clínicos**, v. 5, n. 1, p. 10–25, janeiro-junho 2012. Citado na página 26.
- CURETON, E.; D'AGOSTINO, R. **Factor Analysis: An Applied Approach (1st ed.)**. [S.l.]: Psychology Press, 1993. Citado 2 vezes nas páginas 18 e 21.
- DORAISWAMY, P. M. *et al.* Validity of the web-based, self-directed, neurocognitive performance test in mild cognitive impairment. **Journal of Alzheimer's Disease**, v. 86, n. 3, p. 1131–1136, 2022. PMID: 35180109. Disponível em: <<https://journals.sagepub.com/doi/abs/10.3233/JAD-220015>>. Citado na página 10.
- HAIR, J. F. J. *et al.* **Análise Multivariada de Dados**. 6a. ed. [S.l.]: Bookman Companhia Editora, 2009. Tradução autorizada da 6ª edição da obra original em inglês "Multivariate Data Analysis", publicada por Pearson Education, Inc., sob o selo de Prentice Hall, em 2006. ISBN 978-85-7780-534-1. Citado 2 vezes nas páginas 10 e 20.
- HARMAN, H.; JONES, W. Factor analysis by minimizing residuals (minres). **Psychometrika**, v. 31, p. 351–368, 1966. Citado 4 vezes nas páginas 10, 14, 15 e 22.
- HARMAN, H. H. **Modern Factor Analysis**. Second, revised. Chicago and London: The University of Chicago Press, 1967. Citado 2 vezes nas páginas 10 e 15.
- HORN, J. L. A rationale and test for the number of factors in factor analysis. **Psychometrika**, v. 30, p. 179–185, 1965. Citado 2 vezes nas páginas 18 e 19.
- JAFFE, P. *et al.* A massive dataset of the neurocognitive performance test, a web-based cognitive assessment. **Sci Data**, 2022. Citado 2 vezes nas páginas 11 e 21.
- JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6th. ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2007. Citado 3 vezes nas páginas 12, 13 e 14.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v. 23, n. 3, p. 187–200, 1958. Citado 2 vezes nas páginas 19 e 22.

KAISER, H. F.; RICE, J. Little jiffy, mark iv. **Educational and Psychological Measurement**, v. 34, n. 1, p. 111–117, 1974. Disponível em: <<https://doi.org/10.1177/001316447403400115>>. Citado 2 vezes nas páginas 18 e 21.

LITTMAN, R.; TAKÁCS, Á. Do all inhibitions act alike? a study of go/no-go and stop-signal paradigms. **PLoS one**, Public Library of Science San Francisco, CA USA, v. 12, n. 10, p. e0186774, 2017. Citado na página 26.

MORRISON, G. E. *et al.* Reliability and validity of the neurocognitive performance test, a web-based neuropsychological assessment. **Frontiers in psychology**, Frontiers Media SA, v. 6, p. 1652, 2015. Citado na página 10.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Citado 2 vezes nas páginas 21 e 22.

RENTERGEM, J. A. Agelink van *et al.* The factor structure of cognitive functioning in cognitively healthy participants: a meta-analysis and meta-analysis of individual participant data. **Neuropsychology Review**, v. 30, p. 51–96, feb 2020. Citado na página 10.

ROGERS, P. Melhores práticas para sua análise fatorial exploratória: Tutorial no factor. **Revista de Administração Contemporânea**, Fundação Getulio Vargas, v. 26, n. 6, p. e210085, 2022. Citado 3 vezes nas páginas 10, 11 e 18.

SALTHOUSE, T. A. What cognitive abilities are involved in trail-making performance? **Intelligence**, Elsevier, v. 39, n. 4, p. 222–232, 2011. Citado na página 26.

SANTOS, D. D. Análise de dados de alta dimensão utilizando apache spark no r. **Universidade Federal Fluminense**, 2021. Citado na página 21.

SANTOS, N. C. *et al.* Exploring the Factor Structure of Neurocognitive Measures in Older Individuals. **PLoS ONE**, v. 10, n. 4, p. e0124229, Apr 2015. Publicado em 16 de abril de 2015. Citado na página 26.

SHOOK-SA, B. E.; CHEN, D.-G.; ZHOU, H. Using structural equation modeling to assess the links between tobacco smoke exposure, volatile organic compounds, and respiratory function for adolescents aged 6 to 18 in the united states. **International Journal of Environmental Research and Public Health**, v. 14, n. 10, 2017. ISSN 1660-4601. Disponível em: <<https://www.mdpi.com/1660-4601/14/10/1112>>. Citado 2 vezes nas páginas 20 e 22.

TIMMERMAN, M. E.; LORENZO-SEVA, U. Dimensionality assessment of ordered polytomous items with parallel analysis. **Psychological Methods**, p. 209–220, 2011. Citado na página 21.

XIA, Y.; YANG, Y. Rmsea, cfi, and tli in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. **Behavior Research Methods**, v. 51, p. 409–428, 2019. Citado na página 10.

Apêndices

APÊNDICE A – Código R de Manipulação do Dataset da Bateria 26

```
rm(list=ls())
library(sparklyr)
library(dplyr)

sc <- spark_connect(master = "local", version = "3.0.0")
battery26_spark = spark_read_csv(sc, "battery26", "battery26_df.csv")
battery_local = collect(battery26_spark)

#Quantos subtestes foram feitos por pessoa?

battery_local |>
  group_by(test_run_id) |>
  summarise(
    n_total_linhas = n(),
    .groups = "drop"
  ) -> subteste_por_pessoa

#Identificando usuários com dados duplicados do mesmo subteste

battery_local |>
  dplyr::summarise(
    n = dplyr::n(),
    .by = c(user_id, age, gender, education_level, country,
            test_run_id, battery_id, time_of_day, grand_index, specific
          ) |>
  dplyr::filter(n > 1L) -> dados_duplicados

#Arrumando os dados

battery_local |>
  #filter(user_id != 61751545) |>
```

```

pivot_wider(
  names_from = specific_subtest_id,
  values_from = raw_score
) -> battery_local_wider

battery_local_wider |>
  rename(sub_36 = '36',
         sub_39 = '39',
         sub_40 = '40',
         sub_29 = '29',
         sub_28 = '28',
         sub_33 = '33',
         sub_30 = '30',
         sub_27 = '27',
         sub_32 = '32',
         sub_38 = '38',
         sub_37 = '37') -> battery_local_wider_rn

#battery_local_wider_rn |>
#  select( all_of(
#           c('user_id',
#             'sub_29',
#             'sub_28',
#             'sub_30',
#             'sub_27',
#             'sub_32',
#             'sub_31')) -> battery_tidy

#rownames(battery_tidy) = battery_local_wider_rn$user_id

#Exportar os dados

write.csv(x=battery_local_wider_rn,
         file='battery_local_wider.csv')

```

APÊNDICE B – Código R da Análise Paralela e Fatorial

```
rm(list=ls())
library(tidyverse)
library(psych)
library(ggpubr)
library(GGally)

#Ler os dados
battery = read.csv('battery_local_wider.csv',
                  header = TRUE)

#Arrumando os dados
battery |>
  select( all_of(
    c(
      'sub_36' ,
      'sub_39' ,
      'sub_40' ,
      'sub_29' ,
      'sub_28' ,
      'sub_33' ,
      'sub_30' ,
      'sub_27' ,
      'sub_32' ,
      'sub_38' ,
      'sub_37' ))) -> battery_tidy

pdf('matriz.cor.pdf')
ggpairs(battery_tidy,
        columns = 1:11,axisLabels = "none",
        upper = list(continuous =
```

```

        wrap("cor", size = 2.00 )),
    lower = list(continuous = 'blank'),
    diag = list(continuous = "density"))+
  theme_void()
dev.off()

#KMO
KMO(battery_tidy)

#Análise Paralela
pdf('Analise_Paralela.pdf')
res_paralell = fa.parallel(
  battery_tidy,
  main = "",
  fm="minres",
  plot = TRUE,
  fa="fa",
  n.iter = 500,
  SMC=TRUE,
  ylabel="Eigenvalues",
  #xlab="Número de Fatores",
  quant=0.95)
dev.off()

#Analise Fatorial com todos os itens

model_all = fa(battery_tidy, 3, rotate='varimax')

model_all

```