



Universidade de Brasília
Programa de Pós-Graduação em Psicologia Clínica e Cultura

Análise Fatorial Exploratória do Questionário Socioeconômico do ENEM 2023

Artur Carvalho de Medeiros - 251119392

Brasília
2025

Lista de Tabelas

1	Descrição das Variáveis do Questionário Socioeconômico do ENEM 2023 . . .	11
2	Resultados dos Testes de Fatorabilidade da Matriz de Correlações	15
3	Autovalores e Variância Explicada para os Componentes Iniciais	16
4	Índices de Ajuste para os Modelos Fatoriais de 3 e 4 Fatores	17
5	Índices de Ajuste do Modelo Fatorial Final	18
6	Matriz de Cargas Fatoriais	20
7	Variância Explicada por Fator para a Solução Final de 4 Fatores	21
8	Matriz de Correlação entre os Fatores	22
9	Comunalidades (h^2) dos Itens	23
10	Alfa de Cronbach dos fatores	24
11	Estatísticas Descritivas dos Escores Fatoriais	25

Lista de Figuras

1	Gráfico de Colunas dos Itens do Questionário - ENEM 2023	13
2	Matriz de Correlação dos Itens do Questionário - ENEM 2023	14
3	Análise Paralela e Screeplot	17
4	Diagrama de Fatores	19
5	Boxplot dos fatores	25

Sumário

1 Introdução	8
1.1 Considerações iniciais	8
2 Objetivos	9
2.1 Objetivo Geral	9
2.2 Objetivos Específicos	9
3 Metodologia	10
3.1 Amostra	10
3.2 Instrumento e Variáveis	10
3.3 Procedimentos e Análise de Dados	12
4 Resultados	13
4.1 Análise exploratória	13
4.2 Matriz de correlação	14
4.3 Fatorabilidade da matriz	15
4.4 Extração dos fatores	16
4.5 Resultados do modelo	18
4.5.1 Ajuste do modelo	18
4.5.2 Diagrama dos fatores	18
4.5.3 Matriz de cargas fatoriais	20
4.5.4 Variabilidade	21
4.5.5 Comunalidade e eficiência	22
4.5.6 Análise de Fidedignidade	24
4.6 Análise Descritiva dos Escores	24
5 Discussão	26
Referências	28

1 Introdução

1.1 Considerações iniciais

A busca por uma classificação fidedigna do nível socio-econômico do cidadão brasileiro é uma questão fundamental encontrada pelas instituições públicas do Brasil. Essa classificação é um indicador crucial para a estrutura econômica do país, permitindo a confecção de planos econômicos, políticas públicas, entre outras ações. Nesse contexto, a partir dos microdados do Exame Nacional do Ensino Médio (ENEM), o principal exame de acesso ao ensino superior do país, divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), é possível elaborar a construção de um indicador de Nível Socioeconômico (NSE) para os participantes, por meio da Análise Fatorial Exploratória.

Muitos dos itens do questionário do ENEM são análogos aos utilizados em importantes instrumentos de mercado, como o Critério Brasil da ABEP (Associação Brasileira de Empresas de Pesquisa, 2024). Este critério utiliza a posse de bens para classificar os domicílios segundo sua capacidade de consumo.

Contudo, a aplicação de um modelo baseado em uma lista fixa de bens pode ser excessivamente rígida, como aponta a literatura (SOARES; SOUZA; PEREIRA, 2004). Para um universo tão diverso como o do ENEM, torna-se relevante investigar, sob a ótica da psicometria, quais itens de fato contribuem para uma medida consistente de NSE.

Adicionalmente, é fundamental reconhecer que o construto de NSE não se resume à posse de bens. A literatura teórica, fundamentada em autores como Bourdieu (BOURDIEU, 1986) que aponta que dimensões como a renda familiar e o capital cultural dos pais são pilares centrais do indicador. O questionário do ENEM tem a vantagem de captar informações sobre todas essas dimensões, oferecendo uma base de dados rica para uma análise multidimensional.

Diante do exposto, o presente trabalho se propõe a investigar a estrutura fatorial subjacente aos itens do questionário socioeconômico do ENEM 2023. Utilizando a Análise Fatorial Exploratória (AFE), busca-se identificar e interpretar as dimensões latentes que compõem o construto de Nível Socioeconômico nesta população, contribuindo para a discussão sobre sua validade e complexidade.

2 Objetivos

2.1 Objetivo Geral

Investigar a estrutura dimensional do Questionário Socioeconômico do ENEM 2023 por meio da Análise Fatorial Exploratória.

2.2 Objetivos Específicos

- Verificar a possibilidade de distinguir fatores a partir da matriz de correlações dos itens do questionário;
- Extrair e reter o número de fatores latentes ideais para representar a estrutura dos dados;
- Interpretar a natureza de cada um dos fatores retidos com base nas cargas fatoriais dos itens;
- Analisar o padrão de correlação entre os fatores identificados;
- Verificar a fidedignidade da medida.

3 Metodologia

3.1 Amostra

Os dados presentes no estudo foram extraídos do banco de Microdados do Exame Nacional do Ensino Médio (ENEM) de 2023, disponibilizado publicamente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A população total de participantes do exame em 2023 foi de 3.933.955 indivíduos. Para a Análise Fatorial Exploratória, foram considerados todos os participantes que forneceram respostas ao Questionário Socioeconômico (QSE).

3.2 Instrumento e Variáveis

O instrumento utilizado foi o Questionário Socioeconômico (QSE) aplicado aos inscritos no ENEM 2023. Para investigar a estrutura latente do construto de Nível Socioeconômico (NSE), todos os itens foram incluídos na análise. A Tabela 1 apresenta uma descrição resumida das perguntas e das opções de resposta.

Tabela 1: Descrição das Variáveis do Questionário Socioeconômico do ENEM 2023

Item	Descrição da Pergunta (Resumida)	Opções de Resposta (Resumidas)	Tipo
Q001	Escolaridade do pai ou responsável.	A-H	Ordinal
Q002	Escolaridade da mãe ou responsável.	A-H	Ordinal
Q003	Grupo de ocupação do pai ou responsável.	A-F (Grupo 1 a Grupo 5)	Ordinal
Q004	Grupo de ocupação da mãe ou responsável.	A-F (Grupo 1 a Grupo 5)	Ordinal
Q005	Número de pessoas que moram na residência.	1 a 20	Númerica
Q006	Renda mensal da família (soma de todos).	A-Q (Nenhuma a > R\$26.400,00)	Ordinal
Q007	Em sua residência tem empregado(a) doméstico(a)?	A-D (Não a ≥ 5 dias na semana)	Ordinal
Q008	Em sua residência tem banheiro?	A-E (Não a 4 ou mais)	Ordinal
Q009	Em sua residência tem quartos para dormir?	A-E (Não a 4 ou mais)	Ordinal
Q010	Em sua residência tem carro?	A-E (Não a 4 ou mais)	Ordinal
Q011	Em sua residência tem motocicleta?	A-E (Não a 4 ou mais)	Ordinal
Q012	Em sua residência tem geladeira?	A-E (Não a 4 ou mais)	Ordinal
Q013	Em sua residência tem freezer?	A-E (Não a 4 ou mais)	Ordinal
Q014	Em sua residência tem máquina de lavar roupa?	A-E (Não a 4 ou mais)	Ordinal
Q015	Em sua residência tem máquina de secar roupa?	A-E (Não a 4 ou mais)	Ordinal
Q016	Em sua residência tem forno micro-ondas?	A-E (Não a 4 ou mais)	Ordinal
Q017	Em sua residência tem máquina de lavar louça?	A-E (Não a 4 ou mais)	Ordinal
Q018	Em sua residência tem aspirador de pó?	A (Não), B (Sim)	Dicotômica
Q019	Em sua residência tem televisão em cores?	A-E (Não a 4 ou mais)	Ordinal
Q020	Em sua residência tem aparelho de DVD?	A (Não), B (Sim)	Dicotômica
Q021	Em sua residência tem TV por assinatura?	A (Não), B (Sim)	Dicotômica
Q022	Em sua residência tem telefone celular?	A-E (Não a 4 ou mais)	Ordinal
Q023	Em sua residência tem telefone fixo?	A (Não), B (Sim)	Dicotômica
Q024	Em sua residência tem computador?	A-E (Não a 4 ou mais)	Ordinal
Q025	Em sua residência tem acesso à internet?	A (Não), B (Sim)	Dicotômica

Para viabilizar a análise estatística, as variáveis categóricas alfanuméricas foram recondicionadas em variáveis numéricas. Valores mais altos foram atribuídos às categorias que indicam maior nível do indicador (e.g., para Q006, “A”=1 a “Q”=17). Para os itens de duas opções, a resposta “Não” foi codificada como 0 e a resposta “Sim” como 1. O item Q005, já numérico, foi mantido em sua escala original.

3.3 Procedimentos e Análise de Dados

O tratamento e a análise dos dados foram conduzidos no software estatístico *R*. Para a manipulação dos dados, utilizou-se o conjunto de pacotes *tidyverse* (WICKHAM; WICKHAM, 2017), e as análises fatoriais e as análises fatoriais foram executadas com os pacotes *psych* (REVELLE, 2011) e *GPArotation* (BERNAARDS; JENNRICH; GILBERT, 2015).

Utilizou-se a Análise Fatorial Exploratória (AFE) para encontrar a estrutura subjacente dos 25 itens do questionário. O plano de análise seguiu as seguintes etapas:

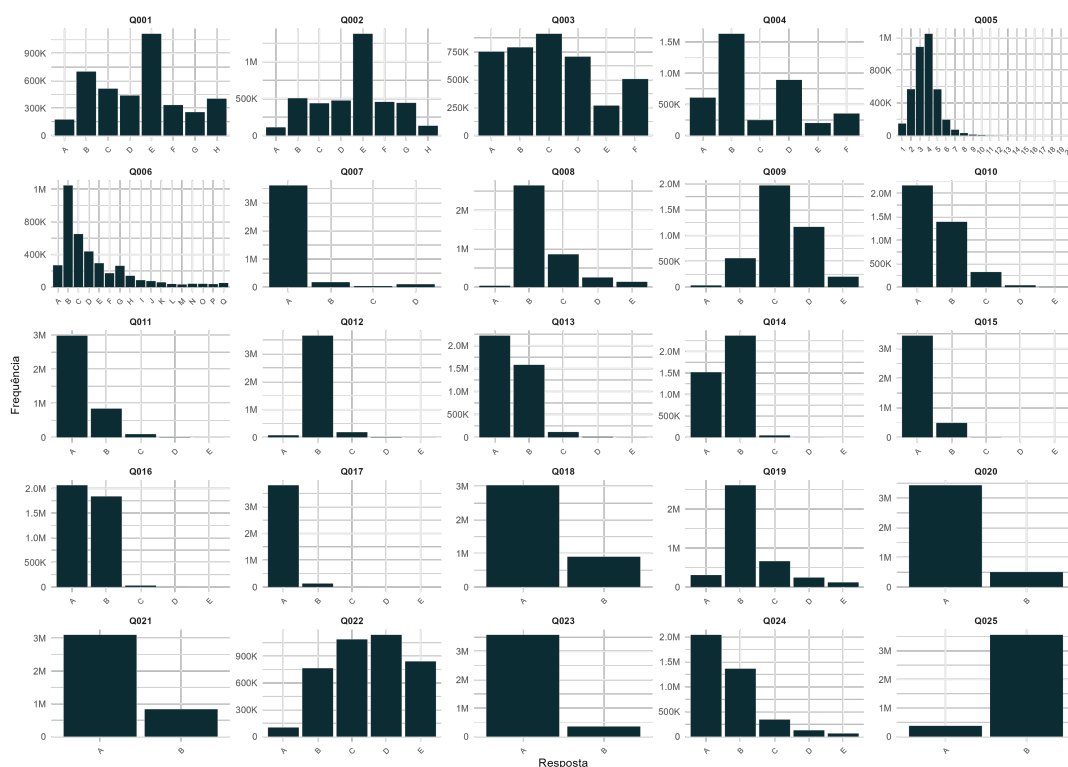
- **Análise de Fatorabilidade:** A adequação da matriz de correlações para a fatorização foi avaliada pelo teste de esfericidade de Bartlett (ARSHAM; LOVRIC, 2011), que testa a hipótese de que a matriz de correlação é uma matriz identidade, e pela medida de adequação da amostra de Kaiser-Meyer-Olkin (KMO) (SHRESTHA, 2021);
- **Extração e Retenção de Fatores:** A extração dos fatores foi realizada pelo método Minimum Residual (minres), que é robusto a violações da normalidade dos dados. A decisão sobre o número de fatores a serem retidos foi baseada em critérios de ajuste do modelo, como o índice Tucker-Lewis (TLI) e a Raiz do Erro Quadrático Médio de Aproximação (RMSEA), e na interpretabilidade teórica da solução fatorial;
- **Rotação e Interpretação:** Foi aplicada a rotação oblíqua Oblimin, pois esperava-se teoricamente que as dimensões do NSE fossem correlacionadas entre si. A interpretação e nomeação dos fatores foram realizadas a partir da análise da matriz de cargas fatoriais (pattern matrix), considerando-se como significativas as cargas fatoriais acima de 0,30.

4 Resultados

4.1 Análise exploratória

A primeira análise a ser realizada corresponde a uma verificação visual das respostas dos indivíduos. A figura 1 apresenta gráficos de colunas dos quantitativos de respostas para cada item.

Figura 1: Gráfico de Colunas dos Itens do Questionário - ENEM 2023



Fonte: ENEM 2023 - INEP

Primeiramente, observa-se um grande número de indivíduos com posse de bens básicos, como em Q012 (geladeira) e em Q019 (televisão em cores). Esses itens possuem baixa variância e, consequentemente, um poder limitado de discriminação entre diferentes estratos socioeconômicos na população do ENEM.

Em contrapartida, um segundo grupo de itens apresenta a maioria das respostas indicando a ausência do bem ou serviço. Itens como Q017 (máquina de lavar louça) e Q009 (empregado doméstico) se enquadram neste perfil, funcionando como bons indicadores para identificar participantes nos estratos socioeconômicos mais elevados, mas com pouca diferenciação nas faixas média e baixa.

Por fim, um terceiro grupo de itens, apresenta uma ampla variabilidade nas respostas.

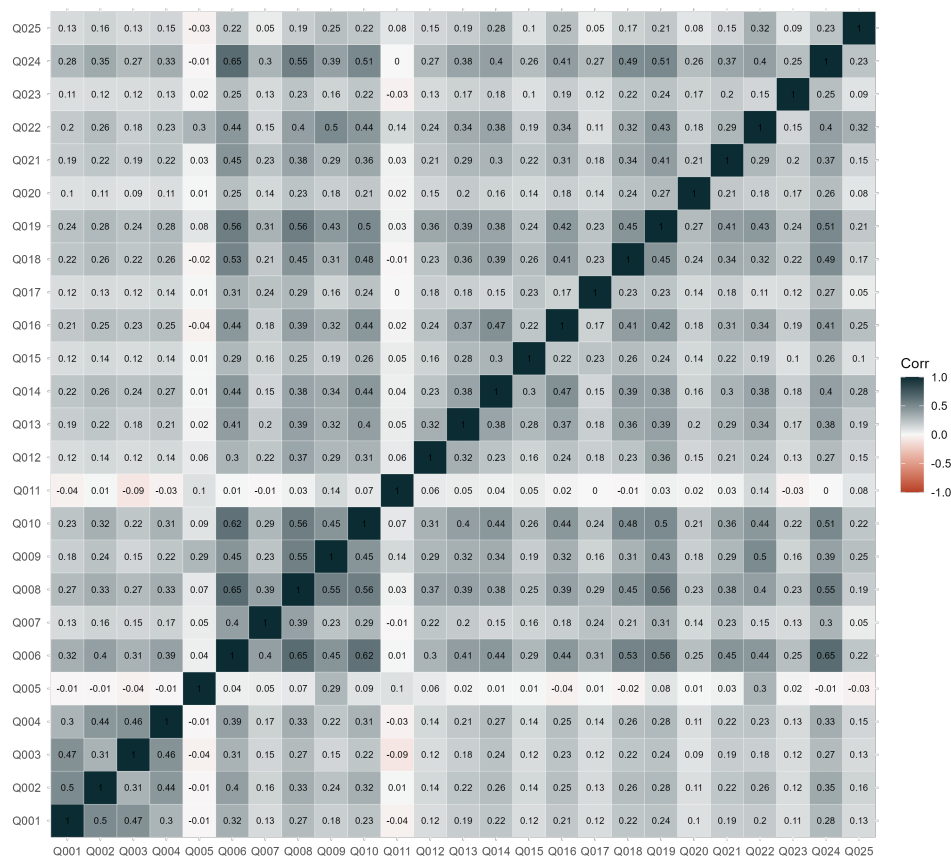
As variáveis Q001 e Q002, referentes à escolaridade dos pais, mostram um comportamento similar, com um pico de respostas na alternativa “E”(Ensino Médio completo), mas com uma distribuição substancial pelas demais categorias. De forma análoga, Q006 (renda familiar) também exibe respostas distribuídas por um vasto espectro de faixas. Essa variabilidade é psicometricamente desejável, indicando que estes itens são potencialmente os mais adequados para a extração de fatores latentes que expliquem as diferenças de NSE entre os estudantes.

Uma observação demográfica relevante também emerge da comparação entre Q003 e Q004 (ocupação do pai e da mãe). Nota-se uma concentração significativamente maior de respostas na categoria “B”(Do lar, desempregada, etc.) para as mães em comparação aos pais, revelando padrões de estrutura ocupacional familiar que podem influenciar a análise.

4.2 Matriz de correlação

Inicialmente foi verificada a matriz de correlações de Pearson dos 25 itens do questionário, presente na Figura 2.

Figura 2: Matriz de Correlação dos Itens do Questionário - ENEM 2023



É possível observar que a matriz não é homogênea, revelando grupos de correlações mais expressivas, o que sugere a presença de possíveis “clusters”, ou seja, fatores. Destaca-se um primeiro agrupamento entre os itens Q001 a Q004, que se referem principalmente à escolaridade e ocupação dos pais.

Um segundo grande agrupamento, com correlações que vão de moderada a forte, é visível entre itens relacionados à renda (Q006) e à posse de bens e infraestrutura domiciliar, como carro (Q010), computador (Q024) e máquina de lavar (Q014).

Em contrapartida, itens como Q005 (número de moradores) e Q011 (posse de motocicleta) apresentam as correlações mais fracas com a maioria dos demais itens, indicando que podem não pertencer às mesmas dimensões latentes.

A existência desses padrões correlacionais, com grupos de itens isolados, reforça a adequação de uma análise multidimensional, justificando a utilização da Análise Fatorial Exploratória para investigar formalmente essa estrutura. Os testes formais de fatorabilidade da matriz são apresentados a seguir.

4.3 Fatorabilidade da matriz

A adequação da matriz de correlações para a análise fatorial foi verificada por meio do Teste de Esfericidade de Bartlett (BTS) e do índice de Kaiser-Meyer-Olkin (KMO) (MATOS; RODRIGUES, 2019), com os resultados na Tabela 2.

Tabela 2: Resultados dos Testes de Fatorabilidade da Matriz de Correlações

Teste	Estatística	Valor	Classificação / Decisão
Kaiser-Meyer-Olkin (KMO)	KMO Geral	0,94	Excelente
	MSA por item	[0,70; 0,98]	Bom a Excelente
Teste de Esfericidade de Bartlett	χ^2	30.108.296	
	gl	300	Rejeita-se H_0
	p-valor	0,001	

Nota: χ^2 = Qui-quadrado; gl = graus de liberdade; MSA = Medida de Adequação da Amostra.

O teste KMO avalia a proporção da variância nos itens que pode ser atribuída a fatores comuns. O índice global obtido (KMO = 0,94) é classificado como “excelente”, indicando adequação dos dados para a análise fatorial.

Da mesma forma, as medidas de adequação da amostra (MSA) individuais de cada item também se mostraram adequadas, variando entre 0,70 e 0,98.

Além disso, o Teste de Esfericidade de Bartlett, rejeitou a hipótese de que a matriz de correlações é significativamente diferente de uma matriz identidade (p-valor 0,001).

Dessa forma, os resultados de ambos os testes aprovam a realização da fatorabilidade da matriz de correlações, validando a aplicação da Análise Fatorial Exploratória.

4.4 Extração dos fatores

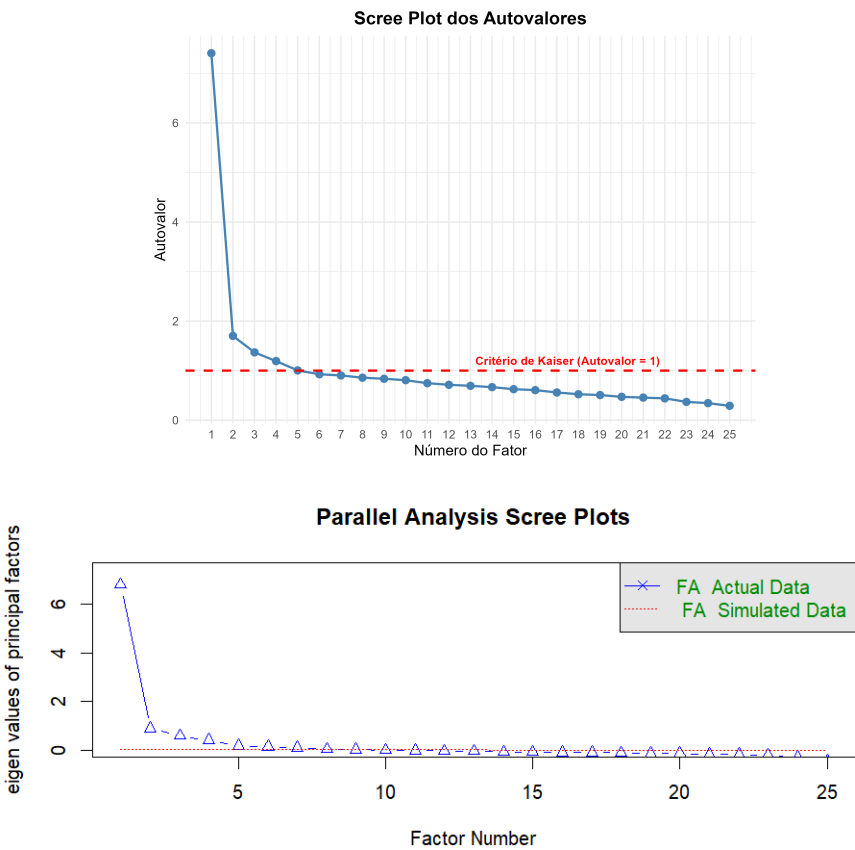
Para determinar o número de fatores a serem utilizados, foi realizada a análise dos autovalores, a qual indica utilizar fatores com autovalor maior que 1, segundo o Critério de Kaiser. A Tabela 3 apresenta os autovalores dos 6 primeiros componentes.

Tabela 3: Autovalores e Variância Explicada para os Componentes Iniciais

Componente	Autovalor	% da Variância	% Acumulada
1	7,430	29,72%	29,72%
2	1,686	6,75%	36,47%
3	1,383	5,53%	42,00%
4	1,140	4,56%	46,56%
5	1,008	4,03%	50,59%
6	0,927	3,71%	54,30%

Observa-se que 5 componentes apresentam autovalor superior a 1. O primeiro componente sozinho explica quase 30% da variância total, mas ainda observa-se uma queda expressiva até o quarto componente. Dessa forma, um modelo construído a partir de 3 ou 4 fatores seria o ideal. Para uma melhor verificação do número de fatores, foram realizadas a Análise Paralela e o Screeplot.

Figura 3: Análise Paralela e Screeplot



Fonte: ENEM 2023 - INEP

O Scree Plot e a Análise Paralela (Figura 3) mostram uma inflexão na curva após o quarto componente, embora uma solução de três fatores também pudesse ser considerada. Foi realizada a comparação dos dois modelos possíveis para verificar qual a melhor opção.

Tabela 4: Índices de Ajuste para os Modelos Fatoriais de 3 e 4 Fatores

CrITÉrio de Ajuste	Modelo com 3 Fatores	Modelo com 4 Fatores
Tucker-Lewis Index (TLI)	0,900	0,926
RMSEA	0,051	0,043
BIC	2.295.946	1.477.676

Nota: RMSEA = Raiz do Erro Quadrático Médio de Aproximação; BIC = Critério de Informação Bayesiano.

A partir da análise dos critérios de ajuste verificou-se que o modelo de 4 fatores é superior em todos os aspectos. Os índices de ajuste relativo (TLI = 0,928) e absoluto (RMSEA = 0,043) foram consideravelmente melhores. De forma decisiva, o valor do BIC para o modelo de 4 fatores foi drasticamente inferior, indicando que a melhoria no ajuste aos dados justifica largamente a inclusão do quarto fator.

Diante da convergência de evidências dos critérios gráficos (Scree Plot) e estatísticos (TLI, RMSEA e BIC), optou-se pelo modelo de 4 fatores como a solução mais adequada para representar a estrutura latente do Questionário Socioeconômico do ENEM 2023.

4.5 Resultados do modelo

Com base nos critérios de ajuste comparativos, foi selecionado o modelo de 4 fatores como a solução final. Com base na matriz de correlação ainda se optou por utilizar uma rotação fatorial oblíqua, a qual permite que os fatores estejam correlacionados entre si, já que os itens em geral possivelmente se correlacionam com itens de outros possíveis fatores. Essa rotação não impõe independência entre os fatores.

4.5.1 Ajuste do modelo

A Tabela 5 resume os principais índices de ajuste do modelo, que são considerados bons (TLI > 0,90; RMSEA < 0,05), validando a adequação do modelo à estrutura dos dados.

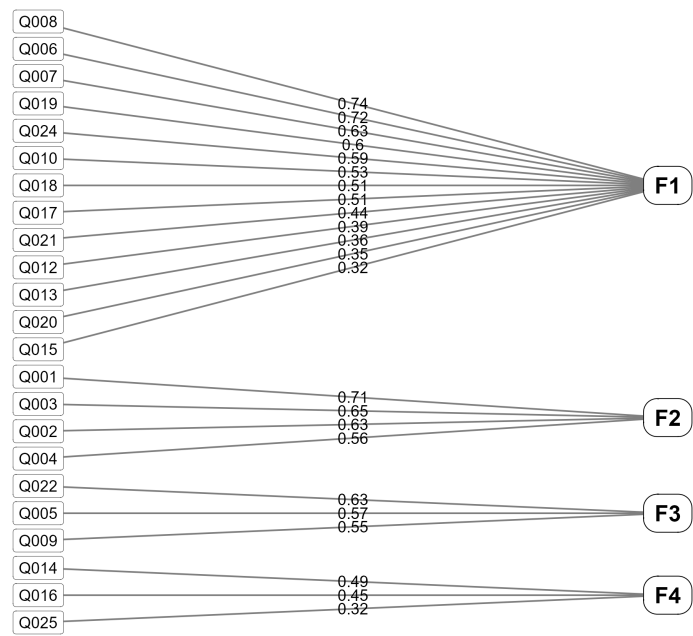
Tabela 5: Índices de Ajuste do Modelo Fatorial Final

Índice	Valor
Tucker-Lewis Index (TLI)	0,926
RMSEA	0,043
BIC	1.477.676

4.5.2 Diagrama dos fatores

A Figura 4 apresenta o diagrama dos fatores. Além disso, a matriz de cargas fatoriais com rotação Oblimin é apresentada na Tabela 6. A análise das cargas permitiu a interpretação de quatro fatores substanciais:

Figura 4: Diagrama de Fatores



A partir do diagrama é possível verificar que o fator 1 é o que apresenta maior força, com 13 itens que o mensuram, os demais fatores apresentam 3 itens.

4.5.3 Matriz de cargas fatoriais

Tabela 6: Matriz de Cargas Fatoriais

Item e Descrição	Fator 1	Fator 2	Fator 3	Fator 4
<i>Fator 1: Poder Aquisitivo Domiciliar</i>				
(Q008) Quartos para dormir	0,741			
(Q006) Renda mensal da família	0,721			
(Q019) Televisão em cores	0,598			
(Q024) Computador	0,590			
(Q010) Carro	0,528			
(Q018) Aspirador de pó	0,512			
(Q017) Máquina de lavar louça	0,506			
(Q021) TV por assinatura	0,444			
(Q012) Geladeira	0,392			
(Q013) Freezer	0,356			
(Q020) Aparelho de DVD	0,346			
(Q015) Máquina de secar roupa	0,316			
<i>Fator 2: Capital Socioeducacional Familiar</i>				
(Q001) Escolaridade do pai		0,709		
(Q003) Ocupação do pai		0,654		
(Q002) Escolaridade da mãe		0,629		
(Q004) Ocupação da mãe		0,557		
<i>Fator 3: Composição e Funcionamento do Domicílio</i>				
(Q022) Telefone Celular			0,628	
(Q005) Nº de pessoas na residência			0,565	
(Q009) Empregado(a) doméstico(a)*	0,304		0,548	
<i>Fator 4: Infraestrutura Tecnológica Doméstica</i>				
(Q014) Máquina de lavar roupa				0,486
(Q016) Forno micro-ondas				0,446
(Q025) Acesso à internet				0,323

Nota: Cargas abaixo de 0,30 foram omitidas. *O item Q009 apresentou carga cruzada relevante, carregando nos Fatores 1 e 3.

Considerando os itens relacionados a cada fator, justifica-se escolha do que é mensurado pelo fator:

- Fator 1 - Poder Aquisitivo Domiciliar: Este fator, o mais forte, agrupa itens de renda, posse de bens de maior valor (computador, carro) e infraestrutura domiciliar (quartos, banheiros).
- Fator 2 - Capital Socioeducacional Familiar: Este fator é definido exclusivamente pelos quatro itens de escolaridade e ocupação dos pais.
- Fator 3 - Composição e Funcionamento do Domicílio: Este fator agrupa itens relacionados à estrutura social do domicílio, como o número de moradores e a presença de empregado doméstico e telefone celular.
- Fator 4 - Infraestrutura Tecnológica Doméstica: A extração de um quinto fator permitiu a emergência de um grupo específico de itens, composto por eletrodomésticos como freezer, máquina de lavar e micro-ondas, que antes se misturavam com outros bens no Fator 1.

4.5.4 Variabilidade

A qualidade de uma solução fatorial também é avaliada pela sua capacidade de explicar a variância total dos itens originais. A Tabela 7 resume a contribuição de cada um dos quatro fatores no modelo final.

Tabela 7: Variância Explicada por Fator para a Solução Final de 4 Fatores

Fator	SS loadings	% da Variância	% Acumulada
Fator 1	4,54	18,2%	18,2%
Fator 2	1,92	7,7%	25,8%
Fator 3	1,45	5,8%	31,6%
Fator 4	1,35	5,4%	37,0%

Nota: SS loadings representam a soma dos quadrados das cargas fatoriais, análogos aos autovalores após a rotação.

O modelo de 4 fatores explica cerca de 37% da variância comum entre os 25 itens. Observa-se que a estrutura é dominada pelo primeiro fator (“Poder Aquisitivo Domiciliar”), que sozinho é responsável por 18,2% da variância explicada. Os três outros fatores contribuem com porções menores da variância.

Tabela 8: Matriz de Correlação entre os Fatores

	Fator 1	Fator 2	Fator 3	Fator 4
Fator 1	1,00			
Fator 2	0,54	1,00		
Fator 3	0,47	0,24	1,00	
Fator 4	0,46	0,34	0,25	1,00

A Tabela 8 mostra que os quatro fatores são positivamente correlacionados, com correlação moderada entre o Poder Aquisitivo Domiciliar (Fator 1) e os demais fatores. Já as demais correlações entre os fatores se mantêm fracas, apesar de ainda positivas.

4.5.5 Comunalidade e eficiência

A comunalidade (h^2) de cada item indica a proporção da variância de uma variável que é explicada conjuntamente pelos fatores extraídos. A Tabela 9 apresenta estes valores em ordem decrescente.

Tabela 9: Comunalidades (h^2) dos Itens

Item	Comunalidade (h^2)	Especificidade
(Q006) Renda	0,694	0,306
(Q008) Quartos	0,638	0,362
(Q022) Telefone Celular	0,588	0,412
(Q009) Empregado(a)	0,547	0,453
(Q010) Carro	0,528	0,472
(Q024) Computador	0,531	0,469
(Q019) Televisão	0,513	0,487
(Q014) Máquina de lavar roupa	0,463	0,537
(Q001) Escolaridade do pai	0,453	0,547
(Q017) Máquina de lavar louça	0,437	0,563
(Q016) Forno micro-ondas	0,437	0,563
(Q002) Escolaridade da mãe	0,436	0,564
(Q018) Aspirador de pó	0,429	0,571
(Q003) Ocupação do pai	0,412	0,588
(Q004) Ocupação da mãe	0,379	0,621
(Q013) Freezer	0,337	0,623
(Q005) N° de pessoas	0,294	0,706
(Q021) TV por assinatura	0,281	0,719
(Q007) Banheiro	0,280	0,720
(Q012) Geladeira	0,200	0,800
(Q025) Acesso à internet	0,188	0,812
(Q015) Máquina de secar roupa	0,157	0,843
(Q020) Aparelho de DVD	0,122	0,878
(Q023) Telefone fixo	0,106	0,894
(Q011) Motocicleta	0,070	0,930

Observa-se que os itens teoricamente centrais para o construto de NSE, como Q006 (Renda, $h^2 = 0,694$), Q008 (Quartos, $h^2 = 0,638$) e Q009 (Empregado doméstico, $h^2 = 0,547$), apresentam as maiores comunalidades. Isso indica que mais de 50% da variância desses itens é capturada pela estrutura do modelo. Da mesma forma, os itens que compõem o fator "Capital Socioeducacional Familiar" (Q001 a Q004) também apresentam comunalidades moderadas.

Por outro lado, a análise das comunalidades também permite identificar os itens que menos se ajustam à solução fatorial. É o caso de itens como Q011 (Motocicleta, $h^2 = 0,070$), Q023 (Telefone fixo, $h^2 = 0,106$) e Q020 (Aparelho de DVD, $h^2 = 0,122$), os quais exibem

as comunalidades mais baixas. Isso significa que a maior parte da variância desses itens é específica de cada um ou representa erro de medida.

4.5.6 Análise de Fidedignidade

Além disso, observou-se também o Alfa de Cronbach para cada fator, apresentado na Tabela 10, a fim de verificar a fidedignidade de cada fator.

Tabela 10: Alfa de Cronbach dos fatores

Fator / Subescala	Nº de Itens	Alfa de Cronbach (α)
1. Poder Aquisitivo Domiciliar	12	0,726
2. Capital Socioeducacional Familiar	4	0,736
3. Composição e Funcionamento do Domicílio	3	0,447
4. Infraestrutura Tecnológica Doméstica	3	0,596

As escalas “Poder Aquisitivo Domiciliar” ($\alpha = 0,726$) e “Capital Socioeducacional Familiar” ($\alpha = 0,736$) apresentaram consistência “Aceitável”. Indicando que os itens que compõem estes fatores medem de forma coesa e fidedigna seus respectivos construtos, podendo ser considerados escalas psicometricamente adequadas.

Em contrapartida, as escalas com menor número de itens apresentaram fidedignidade abaixo do ideal. O fator “Infraestrutura Tecnológica Doméstica” ($\alpha = 0,596$) ainda é considerado “Questionável”. Já o fator o fator “Composição e Funcionamento do Domicílio” ($\alpha = 0,447$) mostrou uma consistência interna “Inaceitável”.

4.6 Análise Descritiva dos Escores

Após a extração dos fatores, foram calculados os escores fatoriais para cada indivíduo da amostra, permitindo uma análise descritiva da distribuição de cada dimensão do NSE. A Figura 5 apresenta os boxplots e a Tabela 11 resume as estatísticas descritivas para cada fator. Por construção, todos os escores são padronizados com média 0 e desvio padrão 1.

Figura 5: Boxplot dos fatores

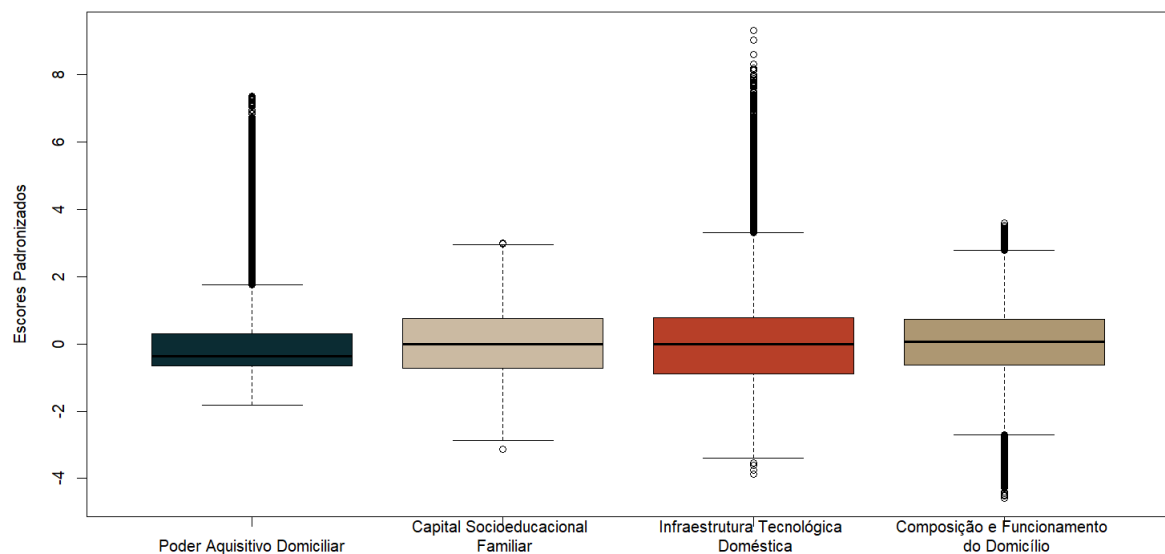


Tabela 11: Estatísticas Descritivas dos Escores Fatoriais

Estatística	Poder Aquisitivo Domiciliar	Capital Socioeducacional Familiar	Infraestrutura Tecnológica Doméstica	Composição e Funcionamento do Domicílio
Mínimo	-1,81	-3,12	-3,88	-4,58
1º Quartil	-0,66	-0,73	-0,90	-0,64
Média	0,00	0,00	0,00	0,00
Mediana	-0,37	-0,01	-0,00	0,05
3º Quartil	0,30	0,74	0,78	0,73
Máximo	7,35	3,00	9,30	3,60
Desvio Padrão	1,00	1,00	1,00	1,00

A análise da distribuição revela perfis distintos para cada fator. O Fator 1 (Poder Aquisitivo Domiciliar) apresenta uma clara assimetria à direita. Isso é confirmado na Tabela 11, onde a mediana (-0,37) é consideravelmente inferior à média (0,00) e está muito mais próxima do primeiro quartil (-0,66) do que do terceiro (0,30). O boxplot corrobora essa observação, mostrando uma cauda superior longa e um número expressivo de outliers positivos. Essa distribuição é um reflexo da estrutura socioeconômica brasileira, na qual uma grande massa de indivíduos se concentra nas faixas de renda mais baixas, enquanto a alta renda e a posse de múltiplos bens de alto valor são características de uma parcela menor da população.

O Fator 2 (Capital Socioeducacional Familiar), por sua vez, exibe uma distribuição notavelmente mais simétrica. A mediana (-0,01) é praticamente idêntica à média (0,00), e os quartis estão distribuídos de forma mais equilibrada ao redor do centro. Isso sugere que, na

população de participantes do ENEM, a dimensão de escolaridade e ocupação dos pais é o que apresenta maior similaridade com a distribuição normal, apresentando menos casos de valores extremos.

O Fator 3 (Infraestrutura Tecnológica Doméstica) e o Fator 4 (Composição e Funcionamento do Domicílio) apresentam distribuições relativamente simétricas em torno da média, como indicado pela proximidade entre suas médias e medianas. No entanto, o boxplot do Fator 3 revela a presença de outliers superiores, indicando a existência de um grupo de estudantes com acesso tecnológico muito acima da média, mesmo que a distribuição geral seja centrada.

Em suma, a análise descritiva dos escores revela como a desigualdade social se manifesta de formas diferentes em cada um desses fatores, de maneira mais acentuada na dimensão econômica e de forma mais gradual na dimensão de capital cultural.

5 Discussão

O presente estudo se propôs a realizar uma avaliação psicométrica do Questionário Socioeconômico (QSE) do ENEM 2023, com o objetivo de investigar sua estrutura latente para a mensuração do Nível Socioeconômico (NSE). A Análise Fatorial Exploratória indicou que um modelo de 4 fatores, explicando 37% da variância total dos itens.

A partir da estrutura fatorial foram definidas as quatro dimensões distintas do NSE. O primeiro fator, Poder Aquisitivo Domiciliar, agrupa itens de renda, posse de bens de maior valor e indicadores de infraestrutura domiciliar.

Enquanto o segundo fator, Capital Socioeducacional Familiar, utiliza itens a respeito da escolaridade e ocupação dos pais.

Já o terceiro, Composição e Funcionamento do Domicílio, está relacionado a estrutura social e aos serviços do domicílio, como o número de moradores, a presença de empregado doméstico e a posse de telefone celular.

Por fim, o quarto fator, Infraestrutura Tecnológica Doméstica se utiliza itens específicos de tecnologia e serviços domésticos, como máquina de lavar roupa, forno micro-ondas e acesso à internet.

A análise demonstrou ainda que estes quatro fatores são positivamente correlacionados, com destaque para a relação entre o "Poder Aquisitivo" e o "Capital Socioeducacional" ($r=0,54$), validando a concepção de que, embora distintas, estas dimensões não são independentes na constituição da posição social de um indivíduo.

Apesar disso, em relação a fidedignidade do fator, somente os 2 primeiros fatores foram considerados "Aceitáveis" através do Alfa de Cronbach, ao passo que a terceira foi classificada como "Questionável" e a quarta "Inaceitável".

Em conclusão, foi demonstrado que o construto de NSE pode ser bem representado por um modelo de quatro dimensões inter-relacionadas, mas distintas, com atenção especial em relação à fidedignidade do fator.

Referências

ARSHAM, H.; LOVRIC, M. Bartlett's test. *International encyclopedia of statistical science*, v. 2, n. 2, p. 20–23, 2011.

Associação Brasileira de Empresas de Pesquisa. *Critério Brasil 2024: atualização para classificação econômica*. 2024. <<https://www.abep.org/criterio-brasil>>. Acesso em: 13 jun. 2025.

BERNAARDS, C.; JENNRICH, R.; GILBERT, M. P. Package 'gparotation'. *Retrieved February*, v. 19, p. 2015, 2015.

BOURDIEU, P. The forms of capital. In: RICHARDSON, J. (Ed.). *Handbook of Theory and Research for the Sociology of Education*. New York: Greenwood, 1986. p. 241–258.

MATOS, D. A. S.; RODRIGUES, E. C. Análise fatorial. Escola Nacional de Administração Pública (Enap), 2019.

REVELLE, W. An overview of the psych package. Citeseer, 2011.

SHRESTHA, N. Factor analysis as a tool for survey analysis. *American journal of Applied Mathematics and statistics*, v. 9, n. 1, p. 4–11, 2021.

SOARES, T. M.; SOUZA, R. C.; PEREIRA, V. R. Métodos alternativos no critério brasil para construção de indicadores sócio-econômicos: Teoria da resposta ao item. *XXXVI Simpósio Brasileiro de Pesquisa Operacional*, v. 35, 2004.

WICKHAM, H.; WICKHAM, M. H. Package tidyverse. *Easily install and load the 'Tidyverse'*, 2017.