

UFMG  
MQ<sup>2</sup> – Programa Intensivo em Metodologia

Pobreza no Brasil como variável latente  
*Uma introdução com dados da POF 2017-2018*

**Samuel Maia**

Programa de Pós-Graduação em Economia do Cedeplar

**Professor:** Dr. Danilo Medeiros

**Curso:** Teoria de Resposta ao Item

Julho de 2024

## 1. Introdução

A mensuração da pobreza possui mais de um século de história. Apesar disso, até o momento ela não alcançou consenso teórico ou metodológico. Aqui como em outros conceitos das ciências sociais, encontramos uma miríade de metodologias para medir a pobreza. Abaixo, meu foco estará numa dessas metodologias que se por um lado é recente, por outro ela se inspira numa tradição de pesquisa com mais de meio século. Falo da mensuração da pobreza segundo a *abordagem estatística*.

**Pergunta de pesquisa:** Para ser mais preciso, meu objetivo é apresentar um exercício de aplicação dessa abordagem usando dados brasileiros. Realizo três tipos de análise usando os dados da Pesquisa de Orçamentos Familiares (POF) da edição 2017-2018 do Brasil: análise de clusters, análise de componentes principais e análise fatorial. O objetivo final é testar um modelo fatorial de segunda ordem para a pobreza no Brasil com esses dados.

Estruturei o documento deste modo. Começo, na seção 2, apresentando a "cozinha de pesquisa", onde descrevo os caminhos tortuosos que me levaram aos resultados da análise. Na seção 3, apresento muito brevemente as principais abordagens de mensuração da pobreza, sendo a estatística uma delas. Na seção 4, trato do conceito e da estrutura da pobreza segundo essa abordagem. A seção 5 trata da operacionalização da proposta, isto é, dos modelos estatísticos e os dados necessários. Termina na seção 6, onde apresento e rapidamente discuto os resultados.

O arquivo do script utilizado para gerar todos os produtos do trabalho, sob o nome "Script do Trabalho Final POF-17-18", pode ser acessado no repositório deste projeto no GitHub, [aqui](#).

## 2. Cozinha de pesquisa: um preâmbulo

Este foi um exercício tão estimulante quanto frustrante – não necessariamente nessa ordem. Foi a primeira vez que fiz uso intensivo do R e que realizei uma análise com dados não simulados. Também por isso fiz uso intensivo de scripts prontos, desenvolvidos tanto pelo IBGE para tratamento dos dados da POF – para carregamento dos dados, merge de questionários e tratamento amostral –, quanto por Hector Nájera, um estudioso da pobreza e autor de um livro sobre a mensuração da pobreza empregando análise fatorial no R (Nájera, 2023). Foi desse trabalho de Nájera de onde tomei os *blueprints* para rodar os modelos. Por um lado, isso tornou a análise possível – não seria capaz de realizá-la sem esse suporte. Por outro, isso significa que ainda não adquiri a autonomia e controle no uso dos dados e do R necessários para conduzir pesquisas adequadamente.

Uma das partes mais estimulantes do exercício foi o de converter em binárias aquelas variáveis da POF originalmente categóricas multinômicas ou ordinais. Concretamente, tive de avaliar quando, por exemplo, determinada condição no domicílio (como o material de seus pisos) constituía ou não uma privação. Fazê-lo exigiu a mobilização um tanto criativa de diferentes critérios econômicos, sociológicos e também normativos.

Porém, devo confessar que tratar um conceito como a pobreza a partir de modelos de variáveis latentes ainda não é, de um ponto de vista mais teórico e conceitual, algo claro para mim. Veja: me parece ser muito intuitivo considerar que construtos mentais como inteligência ou ideologia são *não observáveis*, e por isso devem ser representados por variáveis latentes a serem capturadas a partir de seus *efeitos*. Mas esse não me parece ser o

caso de construtos como pobreza e privações. Para início de conversa, eles me parecem *ser observáveis*. Além disso, ainda me parece ser mais interessante tratá-los como efeitos cujas *causas* queremos atacar. De qualquer maneira, com o exercício vi como pode ser produtivo e iluminador deixar, ainda que temporariamente, questões como essas de lado em nome de explorações empíricas. Afinal, os resultados dessas explorações podem eventualmente modificar essas mesmas considerações teóricas e conceituais.

De fato, o grosso do exercício consistiu nisso. Digo, no uso do R para tratamento da base de dados, das variáveis, das propriedades amostrais da POF e, principalmente, do desenvolvimento, teste e modificação dos modelos. Como era esperado, uma das principais dificuldades surgiu da lide com bugs no código. Outros desafios dizem respeito mais à lógica de pesquisa. Por exemplo, grande parte dos resultados dos primeiros dias de análise descritiva e fatorial tiveram de ser jogados fora. Até então, não tinha realizado qualquer teste de adequação dos dados para análise fatorial. Quis então fazer este exercício e o resultado do teste KMO foi desesperador: "a adequação da amostra é medíocre". O motivo? Ora, eu estava tratando os dados sem levar em conta as propriedades amostrais da Pesquisa, o que tornava esses resultados, de saída, estatisticamente inválidos. Então refiz as análises, tentando ainda que precariamente dar conta dos pesos e estratificação da amostra. O resultado foi que, agora sim, ela se saiu bem nos testes de adequação.

Por fim, temos o processo quase incessante de desenho, testagem e refinamento dos modelos. Como os resultados do exercício mostram, a conclusão foi um tanto decepcionante: as variáveis escolhidas e os modelos desenhados não dão suporte a uma caracterização e representação da pobreza como variável latente explicativa de diferentes dimensões de privações captadas na POF. Na medida de minhas limitações, tentei não brigar com os fatos. Claro, dada a natureza exploratória e precária de meu uso dessas ferramentas, esses resultados não devem ser levados a sério. Muito provavelmente são artefatos estatísticos grosseiros que vão diminuir à medida que adquiro experiência de pesquisa; ou pelo menos assim eu espero.

### 3. Abordagens principais

A teoria da mensuração de Bradburn, Cartwright e Fuller (2017) é um bom ponto de partida. Segundo ela, são dois os requisitos para que uma medida científica seja confiável. O primeiro requisito é que ela possua três elementos: uma *caracterização* de um conceito, isto é, uma especificação de seu significado e abrangência; uma *representação* formal, isto é, algo como uma escala apropriada ou um modelo de sua estrutura ou propriedades; e um conjunto de *procedimentos* que especifique como os dados relevantes devem ser coletados e processados. O segundo requisito é que esses elementos sejam coerentes e se apoiem de forma mútua: a representação é adequada à caracterização; esta é sensível aos dados relevantes; e estes por sua vez capturam aquelas informações exigidas pela caracterização e representação; e assim por diante.

Assim como em outros conceitos nas ciências sociais, o conceito de pobreza é ele próprio objeto de intensa disputa. Nos termos de Bradburn et al., existe uma extensa literatura em torno de sua caracterização. E mesmo naqueles casos de concordância quanto a esse elemento, há uma plethora de disputas em torno da melhor maneira de representá-lo.

De maneira muito geral, as principais abordagens na caracterização e mensuração da pobreza são as seguintes:

- **Subsistência:** pobreza é privação de necessidades existenciais, especialmente re-

lacionadas à alimentação. O foco dessa abordagem tem sido estabelecer, a partir de critérios nutricionais, um mínimo calórico e nutricional cujo custo estabelece a linha de pobreza (cf. Rocha, 2006). Em geral, aqui os estudos pioneiros são aqueles de Seebom Rowntree (1901) sobre as condições de vida dos trabalhadores no Reino-Unido do início do século XX.

- **Medidas axiomáticas:** mesmo que haja diferentes maneiras de caracterizá-la, as medidas de pobreza precisam satisfazer determinadas propriedades ou axiomas. Muitos desses axiomas são princípios normativos (por exemplo, a medida de pobreza deve ser sensível a transferência de recursos de pessoas pobres para não pobres e vice-versa, mesmo que o número de pobres se mantenha o mesmo). O economista Amartya Sen é um dos pioneiros dessa abordagem (Sen, 1976).
- **Medidas administrativas:** basicamente, falamos critérios governamentais para a concessão de benefícios sociais. A renda familiar per capita que separa beneficiários e não beneficiários do Programa Bolsa Família é um exemplo. Apesar dos esforços para que essas medidas obedeçam alguns critérios teóricos, empíricos e normativos, em grande parte elas respondem a determinantes mais mundanos das políticas públicas em geral como, por exemplo, a capacidade orçamentária.
- **Privação relativa:** a pobreza é caracterizada como a exclusão do padrão de vida de uma sociedade devido a falta de recursos. Aqui, o principal nome é o do sociólogo Peter Townsend que, além dessa caracterização, defendeu que para além de critérios normativos, medidas de pobreza precisam ser avaliadas a partir de ditames empíricos, tais como se basearem em hipóteses passíveis de teste (Townsend, 1979). Mais recentemente, essa proposta tem sido implementada na chamada abordagem estatística da mensuração da pobreza.
- **Abordagem estatística:** a ênfase é desenvolver e avaliar medidas de pobreza a partir da teoria da mensuração desenvolvida originalmente na psicometria onde, principalmente, medidas são concebidas como modelos estatísticos. Isso significa que as medidas de pobreza estão sujeitas a testes estatísticos, o que é geralmente interpretado por seus entusiastas como a ênfase townsendiana no desenvolvimento e teste de hipóteses (cf. Gordon, 2006; Nájera, 2023).

Levando isso em conta, o principal objetivo deste trabalho é apresentar uma aplicação da abordagem estatística de mensuração da pobreza a partir de dados brasileiros.

## 4. Caracterização e representação

Apesar das controvérsias em torno do próprio conceito, aceita-se que uma noção intimamente relacionada à da pobreza é a noção de *privação* ou *carência* (Medeiros, 2012). Mas os consensos param por aí. Por exemplo, dentre a série de escolhas teóricas e metodológicas necessárias para a construção de uma medida de pobreza, uma das mais importantes diz respeito a se estamos falando de um conceito ou fenômeno uni- ou multidimensional. Neste exercício, trato as medidas de pobreza da perspectiva multidimensional.

Segundo a apresentação de Nájera (2023), nas medidas de pobreza multidimensional a partir da abordagem estatística, a pobreza é concebida como um *construto*, isto é, um objeto abstrato que se manifesta em observáveis. Esses observáveis são precisamente

as privações que resultam da pobreza. E são essas privações que são capturadas pelos indicadores.

Os indicadores de privação estariam estatisticamente correlacionados pois eles têm uma causa comum, a pobreza. Isso significa que se a pobreza fosse eliminada, não haveria intercorrelação das variáveis, afinal, o fator comum responsável por suas mudanças teria desaparecido. Por isso esses são os modelos de *fator comum* para medir a pobreza. Neles, as manifestações observáveis são funções de duas coisas. Primeiro, do fator latente e, segundo, de outros fatores e de erro aleatório. Isso será apresentado formalmente na próxima seção.

## 5. Operacionalização

### 5.1. Modelos estatísticos

Nos modelos multidimensionais, cada variável observável ( $x_{ij}$ ) é produto de uma dimensão latente ( $\eta_j$ ) e de um fator de ordem superior ( $\zeta_h$ ). Aqueles são as diferentes dimensões da pobreza, enquanto este é a pobreza *simpliciter*. Como dito, além do efeito da variável latente, os modelos incluem erros aleatórios, como vemos a seguir:

$$x_{ij} = \lambda_{ij}\eta_j + \epsilon_{ij} \quad (1)$$

$$\eta_j = \gamma_j\zeta + \xi \quad (2)$$

O  $\lambda_{ij}$  e o  $\gamma_j$  são as cargas fatoriais e capturam a relação entre as variáveis latentes e as variáveis observáveis, além da relação entre as dimensões e a variável latente superior.

A [Figura 1](#) fornece uma representação de um modelo como esse. Indicadores são agrupados em algumas dimensões ( $\eta_j$ ) e as cargas fatoriais ( $\lambda_{ij}$ ) denotam a relação de cada mensuração com as dimensões em questão. A relação do fator de ordem superior (pobreza geral) com cada dimensão é dado pelas cargas fatoriais ( $\gamma_j$ ). Tanto  $\lambda_{ij}$  quanto  $\gamma_j$  são parâmetros do modelo e denotam a força da associação entre a variável latente e as variáveis observáveis.

Como veremos, é possível que o modelo mais adequado aos dados disponíveis prescindia dos fatores inferiores – as dimensões – ou mesmo do fator de ordem superior – da pobreza. Esta é precisamente uma das características distintivas da abordagem estatística: a estrutura conceitual da pobreza é ela própria sujeita a análise empírica.

### 5.2. Dados: fontes e preparação

Nesta seção apresento a principal fonte dos dados desse exercício, a POF 2017-2018, assim como os principais procedimentos adotados para seu tratamento.

#### A POF 2017-2018

A Pesquisa de Orçamentos Familiares 2017-2018 é a sexta pesquisa domiciliar do IBGE dedicada aos orçamentos e condições de vida de domicílios brasileiros.<sup>1</sup> Além da composi-

<sup>1</sup>A primeira edição, de 1974-1975, foi sob o nome de Estudo Nacional de Despesa Familiar. A ela se seguiram pesquisas de maior abrangência e já o nome corrente: as POFs de 1987-1988, 1995-1996, 2002-2003 e 2008-2009.

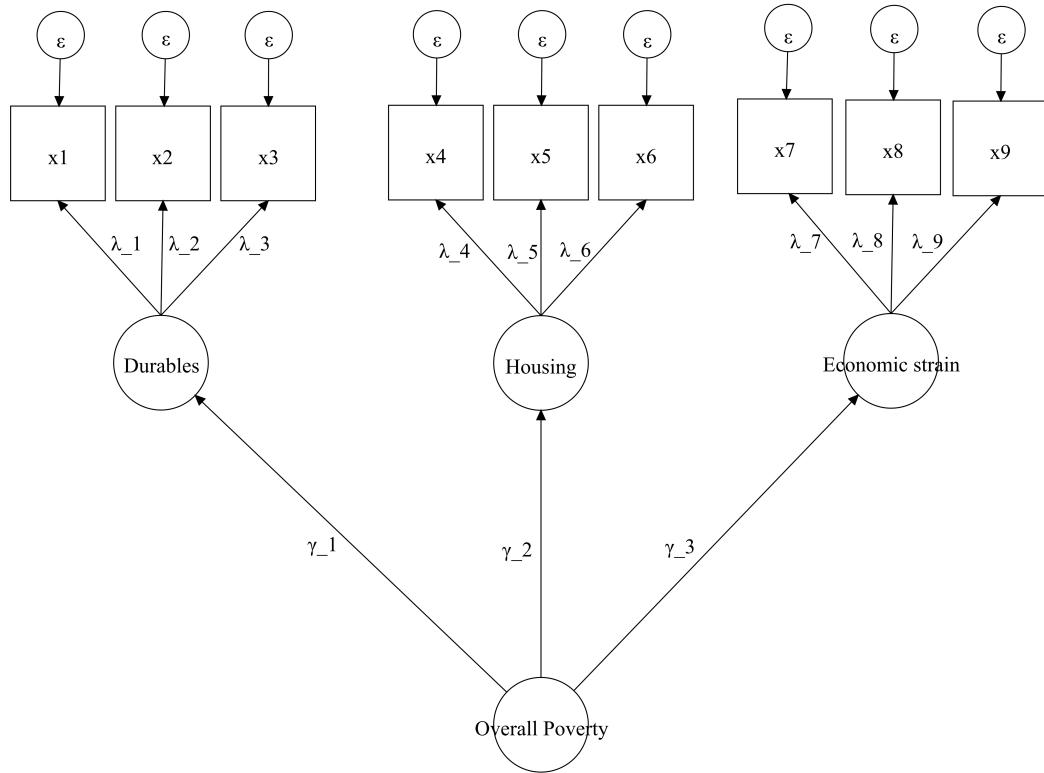


Figura 1: Diagrama produzido por Guio et al. (2009) para representar um modelo de mensuração da pobreza com um fator superior

ção dos gastos das famílias e da estrutura dos domicílios, a POF mais recente investigou o perfil nutricional e aspectos subjetivos das famílias do Brasil.

A pesquisa foi realizada de junho de 2017 a julho de 2018 com abrangência nacional, sob quatro períodos de referência: 7 dias, 90 dias e 12 meses. A pesquisa é de natureza *cross-section* e amostral. As entrevistas foram feitas a partir de "domicílios particulares permanentes", de onde se identificam a unidade básica de pesquisa, a *unidade de consumo*. Essa unidade pode ser ou um único morador ou um conjunto de moradores que compartilham da mesma fonte de alimentação ou despesas com moradia (IBGE, 2019).

### Preparação das variáveis

Neste exercício, fiz uso dos microdados da pesquisa disponíveis para download na página do IBGE.<sup>2</sup> A POF 2017-2018 compreende sete questionários. Aqui foram usadas variáveis de dois deles: a POF 1 e a POF6. O primeiro é dedicado a informações gerais do domicílio. O segundo apresenta informações subjetivas sobre as condições de vida dos moradores, coletadas a partir de entrevista com pessoa responsável pela unidade de consumo ou por alguém assim considerada pelos demais moradores.

Para o carregamento dos dados no R, utilizei o programa de leitura fornecido pelo próprio IBGE, contido no arquivo "Leitura dos Microdados - R" (IBGE, 2023a). Esse arquivo contém um script para o carregamento de todos os questionários na forma de dataframes.

Para a seleção das variáveis, tomei primeiro como referência a metodologia adotada

<sup>2</sup>A página pode ser acessada [aqui](#).

para a medida oficial da pobreza multidimensional do México, segundo o *Consejo Nacional de Evaluación de la Política de Desarrollo Social* (KJ08, 2023). Fiz isso pois é a partir dessa medida que Nájera apresenta scripts para a geração de medidas estatísticas de pobreza. Assim, eu deveria "apenas" aplicá-los com as variáveis da POF análogas da medida mexicana. Contudo, no final acabei incluindo variáveis da Pesquisa relacionadas à qualidade de vida dos moradores entrevistados.

Para o exercício, considere 36 variáveis no total. Dessas, 33 estão originalmente presentes nos questionários da Pesquisa, e 3 foram produzidas pela combinação de variáveis originais. Todas as 36 variáveis procuram medir a presença ou ausência de uma privação. Por isso, todas elas são binárias. A lista de todas pode ser encontrada no [Apêndice A](#). Para seu significado, consulte o arquivo "[Dicionário de Privações da POF 17 e 18](#)" no repositório deste projeto.

O que poderia parecer um exercício trivial – a conversão de variáveis sobre condições de vida em binárias que indicam a presença ou ausência de privação – na verdade não é um exercício simples. Essa é uma das etapas onde precisamos fazer juízos que podem ser acusados de subjetivos ou mesmo de arbitrários. Por exemplo, tome o caso da variável da POF V0204 que trata do material que predomina no piso do domicílio. As opções são: (1) cerâmica, lajota ou pedra; (2) madeira apropriada para construção (aparelhada); (3) cimento; (4) terra; (5) outro material. Para o exercício, precisamos escolher uma ou mais dessas opções como indicadores de um tipo de carência ou privação. É razoável supor que não seria polêmico considerar um piso de terra como um indicador de uma carência relevante para a pobreza, e que pisos de cerâmica e madeira aparelhado, não.<sup>3</sup> Mas e o caso dos pisos de cimento? Para alguns, *prima facie*, se trata de um indicador de privação; para outros não. Diga-se, na construção de uma medida de pobreza, temos de realizar uma enormidade de juízos como esse.

### Merge dos questionários *domicílio e condições de vida e inferência estatística*

Parte importante do tratamento dos dados foi a combinação dos microdados provenientes dos questionários sobre o domicílio e condições de vida. Aqui, o principal objetivo é que dados de questionários distintos capturem uma mesma unidade de consumo. Para isso, tomei como base os procedimentos segundo o script, disponibilizado pelo próprio IBGE, "Programa para geração das estimativas da tabela 2" (IBGE, 2023b), também disponível no repositório do projeto. Também extraí desse script os procedimentos de inferência estatística levando em conta o desenho amostral da Pesquisa.

### Pacotes *psych* e *lavaan*

Nas análises estatísticas no R, fiz uso dos pacotes *psych* e *lavaan*. O pacote *psych* foi empregado para as análises fatoriais exploratórias e o pacote *lavaan* (Rosseel, 2012) foi utilizado nas confirmatórias.

---

<sup>3</sup> Vemos aqui um caso de como o terceiro elemento de uma medida, os procedimentos para processamento dos dados – nesse caso, dos itens – podem (e é aconselhável que sejam) guiados pela caracterização da pobreza. Por exemplo, se temos em mente formas mais extremas de pobreza, é possível que pisos de cimento não sejam privações relevantes. De fato, parte da literatura se dedica a desenvolver metodologias para realizarmos juízos como esse da maneira menos arbitrária possível (cf. Deeming, 2020).



## 6. Resultados

Esta é a seção com os resultados do exercício. Faço primeiro uma breve apresentação das prevalências dos diferentes tipos das privações no Brasil (6.1). Em seguida, vemos os resultados da análise fatorial (6.2): análises de clusters, de componentes principais e confirmatória. Termino o exercício com uma brevíssima retomada dos resultados (6.3).

### 6.1. Prevalências de privações no Brasil por tipo

Faço agora uma apresentação muito breve da prevalência dos diferentes tipos de privação no Brasil. Uma olhada na [Figura 2](#) me leva a destacar alguns pontos. Primeiro, mais de 70% das unidades de consumo relataram que, nos últimos três meses, ficaram sem dinheiro para ter uma alimentação saudável ou variada (variável *a\_ns*). Relacionado a isso, vemos que aquelas privações relacionadas às necessidades alimentares possuem a maior incidência no geral. Em segundo lugar, temos as privações relacionadas aos serviços de água, escoadouro, formas de cozinhar alimentos, de destino de lixo ou presença de iluminação na rua. Aqui, mais de 40% estão privados em pelo menos um aspecto (*s\_ct*). Entre os indicadores subjetivos, mais de 40% relataram um padrão muito ruim em relação ao lazer (*q\_lz*) e à violência na área de residência (*vlc*), enquanto mais de 27% disseram ter um rendimento que permite chegar ao fim do mês apenas com muita dificuldade (*q\_rd*). Por outro lado, é baixa a incidência daqueles que vivem em domicílios onde não há banheiro (*m\_bn*) – 5% – ou energia elétrica (*s\_el*) – 1%. É ainda menor o número dos que moram em habitações precárias como cortiços (*m\_tp*) – menos de 1%.

### 6.2. Análises fatoriais

Nesta subseção apresento as análises fatoriais propriamente ditas. Primeiro, (a) apresento os resultados de uma análise de clusters com todas as 36 variáveis. Em seguida, (b) apresento três análises de componentes principais, de onde me baseei para estruturar o modelo de base. Termino com (c) a análise confirmatória, onde esse modelo e algumas variantes são postos a testes.

#### a) Primeira Análise de Clusters

Para início, rodei uma análise de cluster com as 36 variáveis. Dela foi produzido um gráfico de cotovelo, apresentado na [Figura 3](#), que sugeriu como número ideal de fatores intermediários para a medida de pobreza o de três.

#### b) Três Análises de Componentes Principais

Foram realizadas três análises de componentes principais. A primeira considerou todas as 36 variáveis de privação. A segunda contou com 17 variáveis. A terceira com 10.

**Primeira ACP:** Foram três os componentes principais, RC1, RC2 e RC3, com os seguintes resultados (para os resultados completos, cf. [Apêndice B](#)). Em RC1, as variáveis com maiores cargas foram *s\_cb* (0.66), *s\_pv* (0.74), *s\_ct* (0.77) e *lur* (0.72), sugerindo que essas variáveis têm uma forte associação com o primeiro fator, provavelmente relacionado a uma dimensão de serviços no domicílio. O componente RC2 foi dominado pelas variáveis *a\_pr* (0.79), *a\_fm* (0.80) e *a\_sc* (0.75), o que sugere que ele captura algo como uma



dimensão relacionada à segurança alimentar dos moradores. No componente RC3, as variáveis *gt* (0.45), *fdu* (0.51) e *mdt* (0.44) apresentaram as maiores cargas, o que sugere uma dimensão relacionada à qualidade de vida.

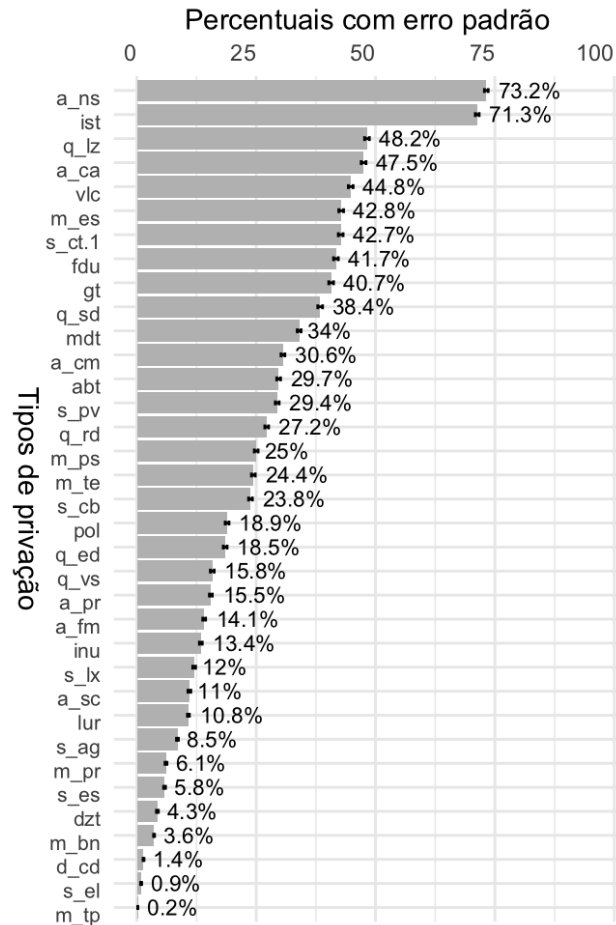


Figura 2: Prevalências de privações no Brasil por tipo. Para o significado das variáveis, cf. "[Dicionário de Privações da POF 17 e 18](#)"

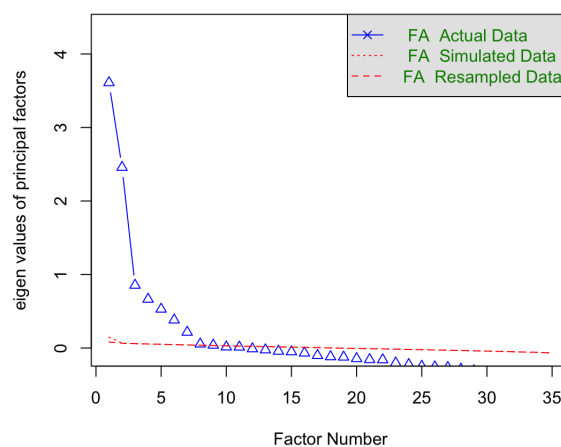
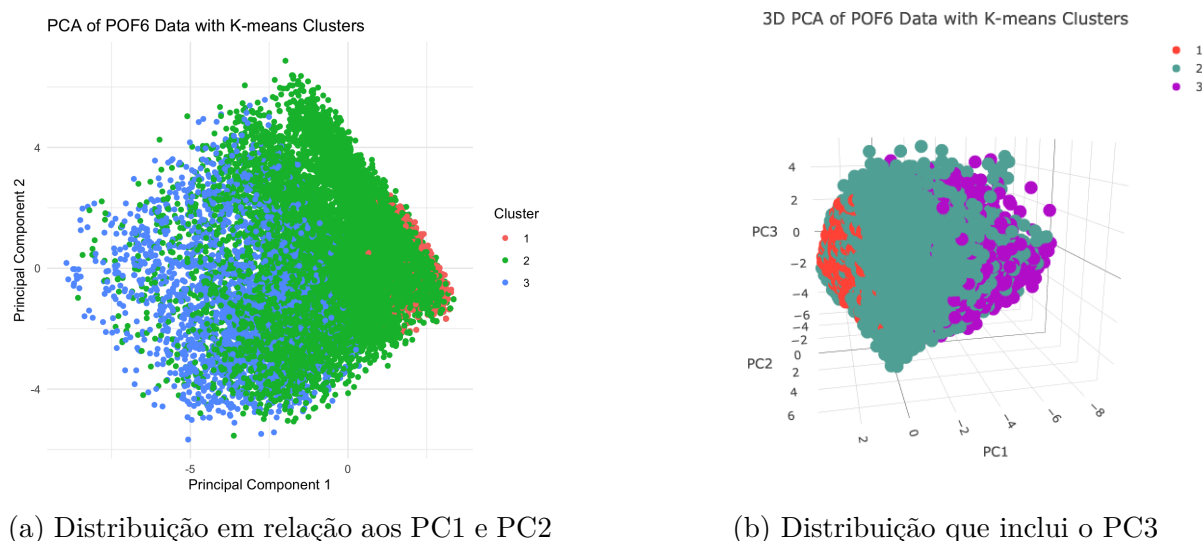


Figura 3: Gráfico de cotovelo para as 36 variáveis

Os três componentes principais juntos explicam apenas 27% da variância total dos dados. Especificamente, o componente RC1 explicou 12% da variância, o RC2 9% e o RC3 6%. A complexidade média dos itens foi de 1.2. O Root Mean Square of the Residuals (RMSR) foi 0.05, sugerindo que o ajuste do modelo aos dados é bom. O valor do  $\chi^2$  empírico foi 73430.79. O ajuste baseado nos valores fora da diagonal foi 0.85. Para fins meramente ilustrativos, a [Figura 4](#) apresenta uma representação em cluster dos resultados.

Esses resultados me motivaram a reduzir o número de variáveis.

Figura 4: Componentes principais da primeira ACP, com 36 variáveis



(a) Distribuição em relação aos PC1 e PC2

(b) Distribuição que inclui o PC3

**Segunda ACP:** Nesta segunda análise, foram abandonadas as variáveis com carga baixa nos diferentes componentes (h2) ou com baixa singularidade (u2). Adotei como critério cargas menores que 0,3 em todos os componentes ou comunalidades abaixo de 0,4. A lista das variáveis restantes pode ser encontrada, junto com todos os resultados dessa segunda ACP, no [Apêndice C](#).

Aqui eu apresento os resultados da ACP com rotação. Para isso fiz uso das funções **varimax** e **promax**. Aqui apresento apenas os resultados com a primeira função. No componente RC1, as variáveis com maiores cargas foram  $s\_cb$  (0.68),  $s\_pv$  (0.78),  $s\_ct$  (0.81) e  $lur$  (0.72). O componente RC2 foi dominado pelas variáveis  $a\_pr$  (0.81),  $a\_fm$  (0.83) e  $a\_sc$  (0.78). Se nesses dois fatores, a melhora foi pequena em comparação com a primeira ACP, no terceiro, a melhora foi significativa. Em RC3, as variáveis  $gt$  (0.69),  $fdu$  (0.71) e  $mdt$  (0.66) apresentaram as maiores cargas.

Os três componentes principais juntos explicam 47% da variância total dos dados. Especificamente, o componente RC1 explica 21% da variância, o RC2 explica 17% e o RC3 explica 10%. A variância cumulativa explicada pelos dois primeiros componentes é 37%. Também uma melhora significativa em comparação com o conjunto maior de variáveis..

A complexidade média dos itens foi novamente de 1,2. O RMSR foi de 0,07. O valor do  $\chi^2$  empírico foi de 27770,97, com uma probabilidade inferior a 0, indicando que o modelo é estatisticamente significativo. O ajuste baseado nos valores fora da diagonal foi de 0,91. Apesar da melhora significativa com relação ao primeiro conjunto de dados, resolvi reduzir as variáveis um pouco mais e rodar uma nova ACP.

**Terceira ACP:** Nesta última ACP, foram abandonadas outras variáveis, cujas cargas fatoriais podem ser conferidas na [Tabela 1](#). Houve melhora significativa em todos os componentes. No componente RC1, as variáveis de maiores cargas foram novamente  $s\_cb$  (0,69),  $s\_pv$  (0,76),  $s\_lx$  (0,86) e  $lur$  (0,82). No RC2, as maiores foram também novamente das variáveis  $a\_pr$  (0,84),  $a\_fm$  (0,86) e  $a\_sc$  (0,84). No RC3, as maiores cargas também aumentaram:  $gt$  (0,73),  $fdu$  (0,73) e  $mdt$  (0,69). Os demais resultados podem ser visto no [Apêndice D](#).

Tabela 1: Terceiro ACP: cargas Fatoriais padronizadas e comunalidades

Variável	RC1	RC2	RC3	h2	u2
$a\_pr$	0.03	0.84	0.09	0.71	0.29
$a\_fm$	0.04	0.86	0.08	0.75	0.25
$a\_sc$	0.04	0.84	0.05	0.70	0.30
$s\_cb$	0.69	0.08	0.07	0.48	0.52
$s\_pv$	0.76	0.02	0.04	0.58	0.42
$s\_lx$	0.86	0.01	0.01	0.73	0.27
$lur$	0.82	0.00	0.00	0.67	0.33
$gt$	0.07	0.07	0.73	0.54	0.46
$fdu$	-0.11	0.03	0.73	0.54	0.46
$mdt$	0.14	0.09	0.69	0.50	0.50

Os três componentes principais juntos explicaram 62% da variância total dos dados. Especificamente, o componente RC1 explicou 25% da variância, o RC2 22% e o RC3 16%. A variância cumulativa explicada pelos dois primeiros componentes foi de 47%. Essa nova melhora significativa indicou que os componentes extraídos capturaram uma parte significativa da variância nos dados.

A complexidade média dos itens foi de 1. O RMSR foi de 0,09. O valor do  $\chi^2$  empírico foi de 15504,24 com uma probabilidade inferior a 0. O ajuste baseado nos valores fora da diagonal é 0,89. Esses resultados sugeriram que os três componentes eram suficientes.

Por isso, estabeleci o seguinte modelo da pobreza com fator superior,  $M_1$ , tomado como base para o exame nas análises confirmatórias. Nele, o fator superior é a *pobreza*; os fatores intermediários, *serviços*, *alimentação* e *moradia*, são as dimensões da pobreza; e seus indicadores são as variáveis  $s\_cb$ ,  $s\_pv$ ,  $s\_lx$ ,  $lur$ ,  $a\_pr$ ,  $a\_fm$ ,  $a\_sc$ ,  $gt$ ,  $fdu$ ,  $mdt$ . Sua estrutura completa está representada no diagrama da [Figura 5](#).

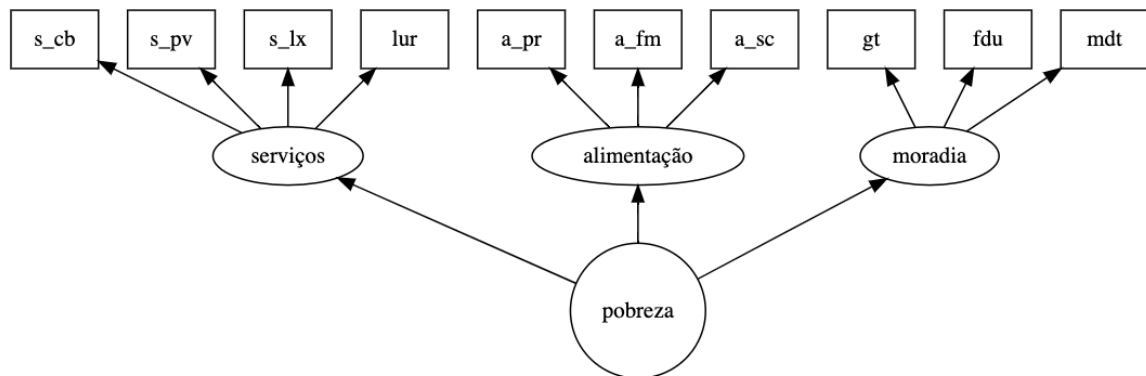


Figura 5: Representação do modelo base,  $M_1$

### c) Análises Fatoriais Confirmatórias

Finalmente, nesta seção apresento os resultados das análises fatoriais confirmatórias (AFCs) que realizei. No total, foram cinco AFCs, cada uma dos seguintes modelos:

- **Modelo  $M_1$ :** modelo base, com três fatores intermediários e um fator superior (*pobreza*)
- **Modelo  $M_2$ :** modelo com três fatores intermediários e com um fator superior, mas com menos variáveis no fator *moradia*
- **Modelo  $M_3$ :** modelo com três fatores intermediários e sem fator superior
- **Modelo  $M_4$ :** modelo com dois fatores médios (*serviços* e *alimentação*) e um fator superior
- **Modelo  $M_5$ :** modelo com dois fatores superiores (*serviços* e *alimentação*) e sem fator superior

Segue um resumo dos resultados das cinco AFCs (as análises completas podem ser conferidas no script deste exercício):

- **$M_1$ :** No modelo base, as cargas fatoriais do fator superior com os fatores intermediários foram, na melhor das hipóteses, razoáveis (baixas no fator *alimentação*, 0.1, e razoáveis nos fatores *serviços*, 0.88, e *moradia*, 1.19). Um resultado preocupante foi a variância negativa do fator superior com o fator *moradia*. Já para as cargas dos fatores intermediários com seus indicadores, os resultados foram mais positivos.

O modelo base apresentou um ajuste moderado, com índices CFI e TLI abaixo de 0,9 e um RMSEA de 0,089. O SRMR foi de 0,077, aceitável, mas ainda não ideal. Os demais resultados podem ser vistos no [Apêndice E](#). Apesar de não ter se saído bem, para fins ilustrativos adicionei a [Figura 6](#) com uma representação gráfica dos resultados.

- **$M_2$ :** Neste modelo, foram abandonadas dois indicadores do fator *moradia* com baixas cargas fatoriais (*gt* e *fdu*). Se isso levou a uma melhora apenas nas cargas do fator superior com o fator *moradia* (1.15), também levou a uma piora nas suas cargas com os fatores *alimentação* e *serviços*. A variância negativa no fator *moradia* com o fator superior também permaneceu.

Por outro lado, o modelo se saiu bem nos índices CFI e TLI, ambos acima de 0,9, e com um RMSEA de 0,042. O SRMR de 0,037 também indica um ajuste superior ao modelo anterior. Ainda assim, as baixas cargas fatoriais e a variância negativa sugeriram que abandonar o fator superior poderia melhorar os resultados. Isso significou que não encontrei, pelo menos neste exercício, suporte empírico para a hipótese inicial, isto é, de conceber a pobreza multidimensional como variável latente.

- **$M_3$ :** O abandono do fator superior levou a uma melhora substancial nas cargas entre os fatores médios e seus indicadores, mas as cargas do fator *moradia* ainda se mostraram baixas. E agora todas as variâncias foram positivas. Note, ao contrário da intuição motivadora dos modelos de fator comum, a eliminação da pobreza não

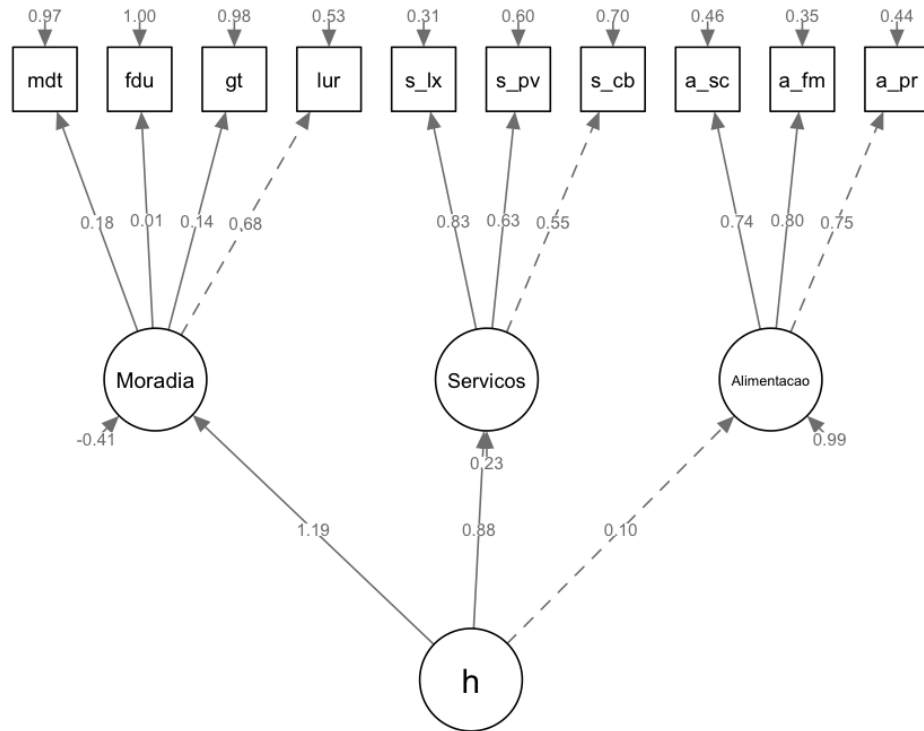


Figura 6: Representação dos resultados do modelo base,  $M_1$ , modelo com fator superior  $h$ , isto é, a pobreza

significou o desaparecimento da intercorrelação entre as variáveis: os três fatores se mostraram correlacionados, ainda que fracamente.

Entre os três primeiros, este modelo apresentou o melhor ajuste, com CFI e TLI de 0,998, um RMSEA de 0,017 e o SRMR de 0,012. Os resultados completos podem ser vistos no [Apêndice F](#). Uma representação de sua estrutura com o fator superior ausente e com as cargas fatoriais pode ser visto na [Figura 7](#).

- **$M_4$ :** E se o abandono o fator *moradia* melhorasse as coisas para o fator superior? Foi com esta a principal motivação do teste desse modelo. Mesmo assim, a variância com o fator *serviços* foi negativa (-0,28). Pelo menos ajuste foi bom, com CFI e TLI de 0,998 e um RMSEA de 0,019. O SRMR de 0,012 é igual ao do modelo  $M_3$ . Em suma, o fator superior *pobreza* continuou parecendo um elemento não adequado aos dados.
- **$M_5$ :** Finalmente, neste modelo sem o fator *moradia* e sem o fator superior, todas as variâncias foram positivas (inclusive superiores às do  $M_3$ ), superando assim também o modelo  $M_4$ . O ajuste foi praticamente idêntico ao  $M_3$ , com CFI e TLI de 0,998 e um RMSEA de 0,017. O SRMR de 0,012 também foi excelente. Desses pontos de vista, esse é o melhor dos cinco modelos.

### 6.3. Breve discussão

Em linhas muito gerais, estes foram os resultados deste exercício:

- *Modelos com fator superior* ( $M_1$  e  $M_2$ ): o  $M_2$  supera o  $M_1$  em termos de ajuste, mas ainda enfrenta problemas como baixas cargas fatoriais e variância negativa.
- *Modelos sem fator superior* ( $M_3$  e  $M_5$ ): o modelo  $M_3$  apresenta um excelente ajuste e nenhuma variância negativa, assim como o  $M_5$ . Ambos os modelos têm índices de ajuste favoráveis.
- *Comparação Geral*: os modelos sem fator superior ( $M_3$  e  $M_5$ ) apresentaram os melhores índices de ajuste e variância. O  $M_1$  teve o pior ajuste e variância.

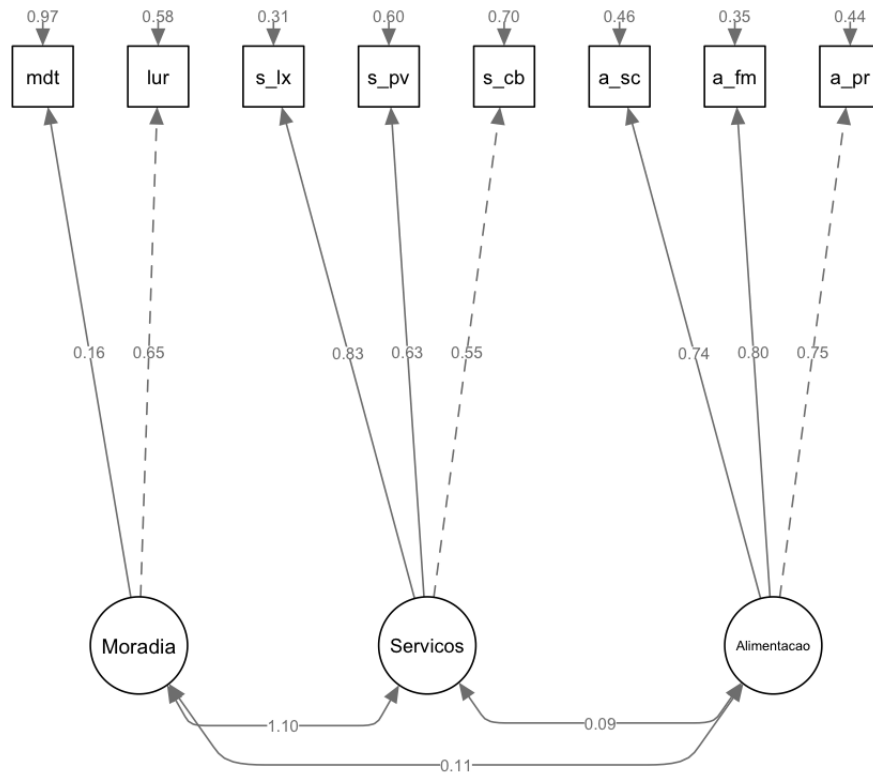


Figura 7: Representação do  $M_3$ , modelo sem o fator superior

Neste exercício, os fatores *moradia*, *serviços* e *alimentação* não pareceram ser explicados por um fator superior que, suponha-se, poderia ser caracterizado como a *pobreza*. Vale notar que, apesar de independentes, os três fatores apresentaram leves correlações entre si, o que poderia merecer investigação futura. Ainda assim, dado o baixo valor dessas correlações, a que se ter cautela, pois isso pode ser mero artefato estatístico.

Claro, o fracasso desse modelo fatorial de mensuração da pobreza multidimensional não pode ser interpretado como razão para abandonarmos esse modo de caracterizar a representar esse conceito. Para além dos erros que provavelmente cometi na condução dos testes, vale a pena experimentar outras estruturas que incluam outras variáveis e, especialmente, outras dimensões de privação. Em especial, aquelas relacionados aos *recursos* dos moradores; afinal, a única variável aqui considerada é um pouco próxima dessa dimensão, a *q\_rd*, certamente é mais que imperfeita.

# Bibliografia

- Bradburn, N., Cartwright, N., & Fuller, J. (2017). A Theory of Measurement. Em L. McClimans (Ed.), *Measurement in Medicine: Philosophical Essays on Assessment and Evaluation* (pp. 73–88). Rowman & Littlefield.
- Deeming, C. (Ed.). (2020). *Minimum Income Standards and Reference Budgets: International and Comparative Policy Perspectives*. Policy Press.
- Gordon, D. (2006). The concept and measurement of poverty. Em C. Pantazis, D. Gordon & R. Levitas (Ed.), *Poverty and Social Exclusion in Britain* (pp. 29–70). The Policy Press.
- Guio, A.-C., Fusco, A., & Marlier, E. (2009). A European Union Approach to Material Deprivation Using EU-SILC and Eurobarometer Data. 19. <https://liser.elsevierpure.com/en/publications/a-european-union-approach-to-material-deprivation-using-eu-silc-a>
- IBGE. (2019). *Pesquisa de orçamentos familiares 2017-2018: primeiros resultados* [Coordenação de Trabalho e Rendimento].
- IBGE. (2023a). Leitura dos Microdados - R [Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/saude/24786-pesquisa-de-orcamentos-familiares-2.html?t=downloads>. Acesso em: 23 de julho de 2024].
- IBGE. (2023b). *Programa para geração das estimativas da tabela 2 da publicação: Pesquisa de Orçamentos Familiares 2017-2018* [Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/saude/24786-pesquisa-de-orcamentos-familiares-2.html?=&t=downloads>. Acesso em: 23 de julho de 2024].
- KJ08. (2023). Medición Multidimensional de la Pobreza 2020, basado no Consejo Nacional de Evaluación de la Política de Desarrollo Social [RPubs, 15 May 2023. Disponível em: <https://rpubs.com/KJ08/MMP>. Acesso em: 25 de julho de 2024].
- Medeiros, M. (2012). *Medidas de Desigualdade e Pobreza*. Editora UnB.
- Nájera, H., et al. (2023). *Multidimensional Poverty Measurement: A Statistical Approach with Applications*. GitHub. <https://github.com/hectornajera83/book>
- Rocha, S. (2006). *Pobreza no Brasil: afinal do que se trata?* FGV Editora.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling [Acesso em: 25 de julho de 2024]. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rowntree, B. S. (1901). *Poverty: A Study of Town Life*. Macmillan; Co.
- Sen, A. (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44(2), 219–231. <https://doi.org/10.2307/1912718>
- Townsend, P. (1979). *Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living*. Allen Lane; Penguin Books. <https://www.poverty.ac.uk/free-resources-books/poverty-united-kingdom>



# Apêndice A

## Variáveis de privação

Tabela A.1: Variáveis de privação e os códigos na POF correlatos

COD	Código da POF
a_ca	V6110
a_pr	V6112
a_cm	V6113
a_ns	V6109
a_fm	V6115
a_sc	V6115
m_tp	V0201
m_pr	V0202
m_te	V0203
m_ps	V0204
m_es	V61061
s_ag	V0207
s_es	V0212
s_el	V61052
s_cb	V02162
s_pv	V0220
s_ct	V0207, V0212, V02162, V02162, V0220
d_cd	V0217
m_bn	V0207
q_rd	V6101
q_vs	V61043
q_ed	V61044
q_sd	V61045
q_lz	V61046
lur	V61053
gt	V61063
fdu	V61064
mdt	V61065
ist	V61066
abt	V61067
pol	V61068
inu	V61069
dzt	V61070
vlc	V610611
s_lx	V0213

# Apêndice B

## Primeira Análise de Componentes Principais

Tabela B.1: Cargas Fatoriais Padronizadas e Comunalidades

Variável	RC1	RC2	RC3	h2	u2	Com
a_ca	0.07	0.54	0.15	0.3146	0.69	1.2
a_pr	0.07	0.79	0.04	0.6246	0.38	1.0
a_cm	0.10	0.73	0.12	0.5552	0.44	1.1
a_ns	0.09	0.40	0.22	0.2118	0.79	1.7
a_fm	0.08	0.80	0.04	0.6464	0.35	1.0
a_sc	0.08	0.75	0.01	0.5681	0.43	1.0
m_tp	-0.03	0.05	-0.01	0.0035	1.00	1.7
m_pr	0.43	0.02	0.07	0.1925	0.81	1.1
m_te	-0.24	-0.04	-0.06	0.0631	0.94	1.2
m_ps	0.36	0.12	0.09	0.1556	0.84	1.4
m_es	0.05	0.13	0.30	0.1103	0.89	1.4
d_cd	0.26	-0.05	-0.10	0.0828	0.92	1.4
m_bn	0.46	0.09	0.03	0.2183	0.78	1.1
s_ag	0.53	-0.01	0.00	0.2766	0.72	1.0
s_es	0.42	0.02	0.17	0.2010	0.80	1.3
s_el	0.30	0.06	0.02	0.0938	0.91	1.1
s_cb	0.66	0.06	-0.04	0.4435	0.56	1.0
s_pv	0.74	-0.03	0.02	0.5499	0.45	1.0
s_ct	0.77	-0.01	0.05	0.5999	0.40	1.0
q_rd	-0.02	0.38	0.21	0.1872	0.81	1.6
q_vs	0.02	0.24	0.33	0.1650	0.83	1.8
q_ed	0.04	0.07	0.35	0.1272	0.87	1.1
q_sd	-0.01	0.08	0.41	0.1781	0.82	1.1
q_lz	-0.06	0.10	0.40	0.1725	0.83	1.2
lur	0.72	-0.04	-0.07	0.5201	0.48	1.0
gt	0.16	0.11	0.45	0.2435	0.76	1.4
fdu	-0.03	0.07	0.51	0.2693	0.73	1.0
mdt	0.20	0.12	0.44	0.2450	0.76	1.6
ist	0.09	-0.01	0.42	0.1815	0.82	1.1
abt	-0.19	-0.01	0.34	0.1527	0.85	1.6
pol	-0.01	-0.04	0.40	0.1582	0.84	1.0
inu	0.04	0.00	0.38	0.1433	0.86	1.0
dzt	0.02	0.00	0.23	0.0552	0.94	1.0
vlc	-0.26	0.03	0.33	0.1796	0.82	1.9
s_lx	0.77	-0.02	-0.09	0.6015	0.40	1.0

Tabela B.2: Resumo dos Componentes Principais

	RC1	RC2	RC3
SS loadings	4.11	3.11	2.27
Proportion Var	0.12	0.09	0.06
Cumulative Var	0.12	0.21	0.27
Proportion Explained	0.43	0.33	0.24
Cumulative Proportion	0.43	0.76	1.00

Tabela B.3: Estatísticas do Modelo

	Valor
Complexidade média dos itens	1.2
RMSR	0.05
$\chi^2$ empírico	73430.79
Probabilidade	< 0.05
Ajuste baseado nos valores fora da diagonal	0.85

# Apêndice C

## Segunda Análise de Componentes Principais

Tabela C.1: Cargas Fatoriais Padronizadas e Comunalidades

Variável	RC1	RC2	RC3	h2	u2	Com
a_ca	0.04	0.53	0.16	0.31	0.69	1.2
a_pr	0.04	0.81	0.05	0.66	0.34	1.0
a_cm	0.07	0.74	0.12	0.56	0.44	1.1
a_fm	0.04	0.83	0.04	0.69	0.31	1.0
a_sc	0.04	0.78	0.00	0.61	0.39	1.0
m_ps	0.36	0.10	0.28	0.22	0.78	2.0
s_ag	0.53	0.01	0.00	0.29	0.71	1.0
s_es	0.37	0.08	0.07	0.15	0.85	1.2
s_cb	0.68	0.07	0.03	0.47	0.53	1.0
s_pv	0.78	0.01	0.00	0.60	0.40	1.0
s_ct	0.81	0.04	0.05	0.67	0.33	1.0
hur	0.72	-0.01	-0.07	0.53	0.47	1.0
gt	0.10	0.10	0.69	0.50	0.50	1.1
fdu	-0.07	0.05	0.71	0.51	0.49	1.0
mdt	0.15	0.10	0.66	0.47	0.53	1.2
abt	-0.21	0.03	0.25	0.11	0.89	2.0
s_lx	0.78	0.00	-0.05	0.61	0.39	1.0

Tabela C.2: Resumo dos Componentes Principais

	RC1	RC2	RC3
SS loadings	3.52	2.82	1.62
Proportion Var	0.21	0.17	0.10
Cumulative Var	0.21	0.37	0.47
Proportion Explained	0.44	0.35	0.20
Cumulative Proportion	0.44	0.80	1.00

Tabela C.3: Estatísticas do Modelo

	Valor
Complexidade média dos itens	1.2
RMSR	0.07
$\chi^2$ empírico	27770.97
Probabilidade	< 0.05
Ajuste baseado nos valores fora da diagonal	0.91

## Apêndice D

# Terceira Análise de Componentes Principais

Tabela D.1: Resumo dos Componentes Principais

	RC1	RC2	RC3
SS loadings	2.49	2.16	1.55
Proportion Var	0.25	0.22	0.16
Cumulative Var	0.25	0.47	0.62
Proportion Explained	0.40	0.35	0.25
Cumulative Proportion	0.40	0.75	1.00

Tabela D.2: Estatísticas do Modelo

	Valor
Complexidade média dos itens	1
RMSR	0.09
$\chi^2$ empírico	15504.24
Probabilidade	< 0.05
Ajuste baseado nos valores fora da diagonal	0.89

# Apêndice E

## Análise Fatorial Confirmatória do $M_1$

Tabela E.1: Resultados da Primeira AFC

Variável	RC1	RC2	RC3	h2	u2	Cargas Padronizadas
a_pr	0.03	0.84	0.09	0.71	0.29	0.747
a_fm	0.04	0.86	0.08	0.75	0.25	0.804
a_sc	0.04	0.84	0.05	0.70	0.30	0.738
s_cb	0.69	0.08	0.07	0.48	0.52	0.549
s_pv	0.76	0.02	0.04	0.58	0.42	0.632
s_lx	0.86	0.01	0.01	0.73	0.27	0.829
lur	0.82	0.00	0.00	0.67	0.33	0.684
gt	0.07	0.07	0.73	0.54	0.46	0.138
fdu	-0.11	0.03	0.73	0.54	0.46	0.012
mdt	0.14	0.09	0.69	0.50	0.50	0.176

Tabela E.2: Estatísticas de Ajuste do Modelo

Estatística	Valor
$\chi^2$	5926.914
Graus de liberdade	32
P-valor ( $\chi^2$ )	0.000
CFI (Comparative Fit Index)	0.892
TLI (Tucker-Lewis Index)	0.848
RMSEA	0.089
Intervalo de confiança (90%) - Inferior	0.087
Intervalo de confiança (90%) - Superior	0.091
P-valor H0: RMSEA $\leq$ 0.05	0.000
P-valor H0: RMSEA $\geq$ 0.08	1.000
SRMR	0.077

Tabela E.3: Loglikelihood e Critérios de Informação

<b>Estatística</b>	<b>Valor</b>
Loglikelihood do modelo usuário (H0)	-305846.652
Loglikelihood do modelo irrestrito (H1)	-302883.195
Akaike (AIC)	611739.304
Bayesian (BIC)	611924.570
Sample-size adjusted Bayesian (SABIC)	611851.477

Tabela E.4: Variâncias dos Erros e Latentes

<b>Variável</b>	<b>Variância (Std.lv)</b>	<b>Variância (Std.all)</b>
a_pr	0.441	0.441
a_fm	0.353	0.353
a_sc	0.456	0.456
s_cb	0.699	0.699
s_pv	0.601	0.601
s_lx	0.312	0.312
lur	0.532	0.532
gt	0.981	0.981
fdu	1.000	1.000
mdt	0.969	0.969
Alimentacao	0.553	0.989
Servicos	0.069	0.229
Moradia	-0.190	-0.406
h	0.006	1.000



## Apêndice F

### Análise Fatorial Confirmatória do $M_3$

Tabela F.1: Estatísticas de Ajuste do Modelo

Estatística	Valor
$\chi^2$	724.865
Graus de liberdade	17
P-valor ( $\chi^2$ )	0.000
CFI (Comparative Fit Index)	0.986
TLI (Tucker-Lewis Index)	0.976
RMSEA	0.042
Intervalo de confiança (90%) - Inferior	0.040
Intervalo de confiança (90%) - Superior	0.045
P-valor H0: RMSEA $\leq$ 0.05	1.000
P-valor H0: RMSEA $\geq$ 0.08	0.000
SRMR	0.037

Tabela F.2: Loglikelihood e Critérios de Informação

Estatística	Valor
Loglikelihood do modelo usuário (H0)	-239983.650
Loglikelihood do modelo irrestrito (H1)	-239621.218
Akaike (AIC)	480005.301
Bayesian (BIC)	480158.347
Sample-size adjusted Bayesian (SABIC)	480097.965

Tabela F.3: Estimativas dos Parâmetros

Variável	Estimativa	Erro Padrão	Valor z	P(> z )	Std.lv	Std.all
<b>Latent Variables:</b>						
Alimentacao =						
a_pr	1.000	-	-	-	0.747	0.747
a_fm	1.076	0.011	97.509	0.000	0.804	0.804
a_sc	0.987	0.010	96.245	0.000	0.738	0.738
Servicos =						
s_cb	1.000	-	-	-	0.547	0.547
s_pv	1.151	0.017	68.195	0.000	0.630	0.630
s_lx	1.521	0.020	75.451	0.000	0.832	0.832
Moradia =						
lur	1.000	-	-	-	0.646	0.647
mdt	0.246	0.011	23.401	0.000	0.159	0.159

Tabela F.4: Covariâncias entre Variáveis Latentes

Variáveis	Estimativa	Erro Padrão	Valor z	P(> z )	Std.lv	Std.all
Alimentacao Servicos	0.037	0.003	10.712	0.000	0.090	0.090
Alimentacao Moradia	0.052	0.005	9.575	0.000	0.107	0.107
Servicos Moradia	0.390	0.006	60.304	0.000	1.104	1.104

Tabela F.5: Variâncias dos Erros e Latentes

Variável	Estimativa	Erro Padrão	Valor z	P(> z )	Std.lv	Std.all
.a_pr	0.441	0.006	72.121	0.000	0.441	0.441
.a_fm	0.353	0.006	56.969	0.000	0.353	0.353
.a_sc	0.456	0.006	74.439	0.000	0.456	0.456
.s_cb	0.701	0.007	96.863	0.000	0.701	0.701
.s_pv	0.603	0.007	90.352	0.000	0.603	0.603
.s_lx	0.307	0.006	48.782	0.000	0.307	0.307
.lur	0.582	0.021	27.402	0.000	0.582	0.582
.mdt	0.975	0.009	106.864	0.000	0.975	0.975
Alimentacao	0.559	0.009	58.928	0.000	1.000	1.000
Servicos	0.299	0.007	40.787	0.000	1.000	1.000
Moradia	0.418	0.022	19.101	0.000	1.000	1.000