



SÉRIE DE NOTAS TÉCNICAS EM ECONOMIA DA UFG
NT N. 016

NOTA TÉCNICA EM ECONOMIA
n. 016

Publicação cujo objetivo é auxiliar na elaboração de aulas e de pesquisas do Curso de Ciências Econômicas da UFG, divulgando demonstrações técnicas e metodológicas ou rotinas computacionais voltadas para a economia.

As opiniões contidas nesta publicação não representam o ponto de vista do Curso ou da FACE/UFG, sendo de inteira responsabilidade de seu(s) autor(es).

Reprodução permitida, desde que citada a fonte.

FACE – Faculdade de Administração, Ciências Contábeis e Ciências Econômicas
Curso de Ciências Econômicas
Direção FACE
Prof^{ra}. Andrea Freire de Lucena
Vice-Direção FACE
Prof^{ra}. Daiana Paula Pimenta
Coordenação do Curso de Ciências Econômicas
Prof. Everton Sotto Tibiriçá Rosa
Coordenação do Programa de Pós- Graduação em Economia
Prof. Felipe Queiroz Silva

Endereço

Campus Samambaia, Prédio da FACE – Rodovia
Goiânia/Nova Veneza, km. 0, CEP 74690-900,
Goiânia – GO. Tel. (62) 3521 – 1390

URL

<http://www.eco.face.ufg.br>

Alocação da Poupança das
Famílias Brasileiras:
Microdados da POF 2017–
2018 e Script em R para
Modelagem Composicional
com Regressão Dirichlet

Antonio Leonardo de Azevedo Carvalho
PPGECON/FACE/UFG
Tatiane Ferreira do Nascimento Melo
PPGECON e IME/UFG

FACE/UFG
Goiânia – Junho de 2025
Versão 1.0

CIÊNCIAS
ECONÔMICAS

FACE
FACULDADE DE
ADMINISTRAÇÃO, CIÊNCIAS CONTÁBEIS E
CIÊNCIAS ECONÔMICAS



Dados Internacionais de Catalogação na Publicação (CIP)
GPT/BC/UFG

Carvalho, Antonio Leonardo. Alocação da Poupança das Famílias Brasileiras: Microdados da POF 2017–2018 e Script em R para Modelagem Composicional com Regressão Dirichlet / Tatiane Ferreira do Nascimento Melo. - 2025.

41 f.: (Série de Notas Técnicas em Economia, 016)

Universidade Federal de Goiás, Goiânia, 2025. Bibliografia.

Anexos. 1. Software R. 2. POF. 3. Dados. I. Título

Alocação da Poupança das Famílias Brasileiras: Microdados da POF 2017–2018 e Script em R para Modelagem Composicional com Regressão Dirichlet

Antonio Leonardo de Azevedo Carvalho¹

PPGECON/FACE/UFG

Tatiane Ferreira do Nascimento Melo²

PPGECON e IME/UFG

Resumo: Esta nota técnica apresenta uma proposta de análise da alocação da poupança das famílias brasileiras entre diferentes categorias de ativos, agrupadas segundo o grau de liquidez, com base nos microdados da Pesquisa de Orçamentos Familiares (POF 2017-2018). As proporções da poupança destinadas a ativos financeiros, bens duráveis e previdência/imóveis são modeladas por meio de regressão Dirichlet — técnica apropriada para o tratamento de dados composicionais. Apesar do objetivo final ser a aplicação de um modelo estatístico, o estudo se destaca pela construção detalhada e criteriosa de uma base analítica composta por famílias homogêneas (casais com dois filhos), bem como pelo desenvolvimento de uma metodologia rigorosa e replicável para a identificação e quantificação da variação líquida dos ativos, a partir das diferentes bases de dados da POF 2017-2018. A nota documenta todas as etapas do processamento dos dados, assim como as decisões teóricas e técnicas incorporadas no script em R, permitindo a obtenção de resultados para diferentes perfis ocupacionais — com destaque para o perfil de poupança de empregados do setor privado. A proposta busca contribuir tanto para o avanço metodológico na mensuração da formação de riqueza quanto para a disseminação de boas práticas no uso de microdados e da linguagem de programação R.

Palavras chave: POF, acumulação de ativos, microdados, regressão Dirichlet, dados composicionais, modelagem estatística.

¹ antonio.leonardo@discente.ufg.br

² tmelo@ufg.br

Abstract: *This technical note presents a proposal for analyzing the allocation of Brazilian households' savings across different categories of assets, grouped by their degree of liquidity, based on microdata from the Household Budget Survey (POF 2017–2018). The proportions of savings allocated to financial assets, durable goods, and pension/real estate holdings are modeled using Dirichlet regression — a method well-suited for handling compositional data. Although the ultimate goal is the application of a statistical model, the study stands out for the detailed and careful construction of an analytical dataset composed of homogeneous families (couples with two children), as well as for the development of a rigorous and replicable methodology for identifying and quantifying the net variation in assets, based on the various datasets of the POF 2017–2018. The note documents all stages of data processing, along with the theoretical and technical decisions embedded in the R script, enabling the generation of results for different occupational profiles — with a particular focus on the savings behavior of private sector employees. The proposal aims to contribute both to the methodological advancement in measuring wealth formation and to the dissemination of best practices in the use of microdata and the R programming language.*

Key words: POF, wealth accumulation, microdata, Dirichlet regression, compositional data, statistical modeling.

1. Introdução

A compreensão da dinâmica de acumulação de ativos pelas famílias é um tema central na análise econômica, especialmente em contextos marcados por desigualdade de renda e acesso heterogêneo a instrumentos de poupança. A Pesquisa de Orçamentos Familiares (POF 2017–2018), conduzida pelo IBGE, constitui uma base de dados valiosa para investigar esse processo. No entanto, sua utilização requer uma etapa preliminar cuidadosa de transformação dos microdados em estruturas compatíveis com a análise da composição e da variação de riqueza.

Esta nota técnica apresenta uma proposta metodológica que combina modelagem estatística com uma estruturação rigorosa da base de dados. A abordagem tem início na seleção de famílias homogêneas — casais com dois filhos — e na construção de uma base analítica que permite mensurar, para cada unidade familiar, a variação líquida em diferentes categorias de ativos ao longo do período de referência da pesquisa. As proporções da poupança líquida

alocadas em ativos financeiros, bens duráveis e previdência/imóveis são então, modeladas como composições, respeitando a natureza relativa dos dados.

Para esse fim, utiliza-se a regressão Dirichlet com parametrização comum, conforme proposta por Hijazi (2003) e implementada no R por meio do pacote *DirichletReg*, descrito em Maier (2014). Essa modelagem é particularmente apropriada para dados composicionais, que possuem natureza aditiva e restrições de positividade e soma unitária, permitindo realizar inferências sobre como variáveis explicativas — como a idade da pessoa de referência — influenciam a alocação relativa da poupança entre as diferentes categorias de ativos.

Mais do que simplesmente aplicar um modelo estatístico, esta nota destaca-se pelo desenvolvimento de uma metodologia replicável para identificar e quantificar a variação dos ativos familiares. Essa metodologia foi implementada por meio de scripts modulares em linguagem R, que integram os diversos arquivos da POF e transformam seus registros em medidas econômicas relevantes, como o vetor composicional da poupança líquida. Além disso, a abordagem possibilita a obtenção de resultados específicos para diferentes perfis ocupacionais da pessoa de referência, ampliando significativamente o potencial analítico do estudo.

Portanto, esta nota técnica tem como objetivos principais: (i) demonstrar a construção de uma base de dados composicional a partir dos microdados da POF; (ii) apresentar uma aplicação prática da regressão Dirichlet para analisar a alocação da poupança familiar; e (iii) disponibilizar um roteiro metodológico claro e adaptável, que possa contribuir para pesquisas futuras nas áreas de economia do consumo, análise de riqueza e uso avançado de microdados públicos.

2. Arquivos e documentação da POF 2017–2018

A base analítica utilizada nesta Nota Técnica foi construída a partir dos microdados da POF 2017-2018, disponibilizados no portal oficial do IBGE:

<https://www.ibge.gov.br/estatisticas/sociais/saude/24786-pesquisa-de-orcamentos-familiares-2.html?=&t=microdados>

A estrutura dos microdados e informações relacionados necessários à replicação dos procedimentos estão distribuídos nas seis pastas comprimidas a seguir apresentadas, as quais podem ser acessadas e baixadas no citado endereço.

- Dados – atualizado em 13/07/2023;

- Documentação – atualizado em 13/07/2023;
- Questionários – atualizado em 23/04/2021;
- Tradutores das Tabelas – atualizado em 13/07/2023;
- Programas de Leitura – atualizado em 13/07/2023;
- Memória de Cálculo – atualizado em 29/09/2023; e

Foram extraídos e utilizados os seguintes arquivos de microdados constantes da pasta Dados:

- DOMICILIO.txt;
- MORADOR.txt;
- RENDIMENTO_TRABALHO.txt;
- OUTROS_RENDIMENTOS.txt;
- DESPESA_COLETIVA.txt; e
- DESPESA_INDIVIDUAL.txt.

Esses arquivos foram convertidos no ambiente R para o formato `.rds`, visando maior agilidade no processamento e na manipulação dos dados.

Destacam-se, para a correta interpretação dos códigos de variáveis e produtos dois arquivos auxiliares contidos na pasta Documentação:

- Cadastro de Produtos.xls – contém a lista completa de bens, serviços e ativos da POF, com codificação e descrição detalhada; e
- Dicionários de Variáveis.xls – fornece a estrutura e codificação dos campos presentes em cada base de microdados.

Esses documentos permitiram identificar com precisão as categorias de aplicação e resgate de ativos (como poupança, previdência, veículos, imóveis etc.), garantindo a consistência na consolidação dos indicadores utilizados nos modelos de regressão composicional.

A construção da base analítica final exigiu, portanto, a integração criteriosa entre arquivos diversos e a leitura cuidadosa da documentação oficial. A sistematização desses elementos constituiu etapa central do trabalho, assegurando a reprodutibilidade, a transparência

metodológica e a aderência técnica às boas práticas de uso de dados complexos em estudos empíricos. No Quadro 1 constam os elementos que compuseram os grupos de poupança.

Quadro 1 – Dados selecionados da POF 2017-2018

BASE DE DADOS	QUADRO/CÓDIGO DO PRODUTO (V9001 ¹)
AF	
Instrumentos Financeiros	
OUTROS_RENDIMENTOS.rds	5600101, 5600201, 5600301, 5600401 5700101, 5700201, 5700301, 5700401
Título de Capitalização	
DESPEZA_INDIVIDUAL.rds	4803301 e 4803302
OUTROS_RENDIMENTOS.rds	5504401
Conta Vinculada do FGTS	
RENDIMENTO_TRABALHO.rds	V8500 ²
OUTROS_RENDIMENTOS.rds	5501001
ANF_D	
Ativos não Financeiros – Bens Duráveis	
DESPEZA_INDIVIDUAL.rds	Quadros 34, 35, 37, 38, 39, 43, 44, exceto códigos 3401701, 4400101, 4400202, 3901702, 3901705, 3901801, 3901903, 3802401, 4400201, 3901701, 3901704, 3901707 e 3901902
DESPEZA_COLETIVA.rds	Quadros 9, 15, 16, 17 e 18, exceto códigos 901401, 901501, 901601, 910101, 910201, 910301, 910401, 910501, 910502, 910503, 910601, 910701 e 910801
OUTROS_RENDIMENTOS.rds	Códigos de 5502101 a 5502634, exceto 5502301, 5502401, 5502618, 5502630, 5502633 e 5502634
Ativos não Financeiros – Bens Duráveis – Veículos	
DESPEZA_INDIVIDUAL.rds	Quadros 33 e 51
OUTROS_RENDIMENTOS.rds	55021 a 55026, exceto 55023 e 55024
ANF_I	
Ativos não Financeiros – Bens Duráveis – Imóveis	
DESPEZA_INDIVIDUAL.rds	4700101, 4700201, 4700202, 4702001, 4702101, 4702701, 4702801 e 4702901
DESPEZA_COLETIVA.rds	Quadro 11 e códigos 1200101 e 1200201
OUTROS_RENDIMENTOS.rds	5502101 e 5502201
Conta da Previdência Pública	
RENDIMENTO_TRABALHO.rds	V531112 ²
DESPEZA_INDIVIDUAL.rds	4800501
Conta da Previdência Privada Aberta e Fechada	
DESPEZA_INDIVIDUAL.rds	4800601, 4800602, 4800603 e 4800604
OUTROS_RENDIMENTOS.rds	5503301
Empréstimos concedidos/pagos e tomados/recebidos⁴	
DESPEZA_INDIVIDUAL.rds	4800101, 4800102 e 4802401
OUTROS_RENDIMENTOS.rds	5501601

Notas:

- Os dois primeiros dígitos do código do produto correspondem ao número do Quadro.
- Após a aplicação do fator de anualização, calcula-se o valor por meio da expressão $V8500_DEFLA * V9011 * FATOR_ANUALIZACAO * 0,08$.
- “Valor em reais (R\$) da dedução com previdência pública que incidiu sobre o rendimento recebido pelo informante no período de referência da pesquisa. Nos casos de valores ignorados (não determinados), este campo está preenchido com 9999999. Nos casos em que não ocorreu dedução ou quando não se aplica, o campo está em branco”.
- Apropriados aos grupos ANF_D e ANF_I, proporcionalmente à variação de cada grupo.

3. Carregamento das bibliotecas e definição do diretório de trabalho

Inicialmente, carregam-se no *script* as bibliotecas (ou “pacotes”) necessárias para realizar as operações de leitura, manipulação, modelagem e visualização dos dados. Essas bibliotecas não fazem parte do núcleo base da linguagem R, mas são indispensáveis para o desenvolvimento da análise realizada na dissertação ou no trabalho a que se destine. A seguir, descreve-se a função de cada biblioteca utilizada no *script*.

3.1. Pacote **dplyr**

O pacote `dplyr` é voltado à manipulação eficiente de dados no formato `data.frame` ou `tibble`. Ele fornece uma sintaxe clara e intuitiva para tarefas comuns de tratamento de dados. Essa biblioteca é fundamental para organizar os microdados brutos da POF 2017–2018 em estruturas adequadas para análise estatística e interpretação econômica.

Principais funções utilizadas:

- `filter()`: filtra linhas com base em condições lógicas;
- `select()`: seleciona colunas específicas;
- `mutate()`: cria ou modifica colunas;
- `group_by()`: agrupa os dados por uma ou mais variáveis;
- `summarise()`: gera estatísticas resumidas para os grupos;
- `left_join()`: une duas tabelas com base em uma chave comum;
- `arrange()`: ordena os dados por uma ou mais variáveis; e
- `bind_rows()`: empilha tabelas de mesma estrutura verticalmente.

3.2. Pacote **magrittr**

O pacote `magrittr` introduz o operador `%>%`, conhecido como pipe, que facilita o encadeamento de funções em uma sequência lógica e legível. Seu uso é particularmente útil em análises de dados, onde diversas transformações sucessivas precisam ser aplicadas a grandes bases de microdados da POF 2017–2018.

O operador permite que o resultado de uma função seja passado diretamente como argumento para a próxima, evitando múltiplas atribuições intermediárias e refletindo de maneira mais clara o raciocínio do pesquisador. Sem o operador pipe, seria necessário criar objetos intermediários a cada passo.

Uma aplicação concreta desse código está na construção da variável *Aplic_PrevPriv*, que representa os depósitos em previdência privada realizados por cada unidade de consumo:

```
PrevPriv_Deposito <- readRDS("DESPESA_INDIVIDUAL.rds") %>%  
  filter(V8000 != 9999999.99) %>%  
  mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC),  
         Valor = V8000_DEFLA * V9011 * FATOR_ANUALIZACAO) %>%  
  relocate(Celula) %>%  
  filter(V9001 == 4800601 | V9001 == 4800602 |  
         V9001 == 4800603 | V9001 == 4800604) %>%  
  subset(select = c(Celula, Valor))
```

Neste trecho:

1. `readRDS()` carrega a base `.rds`;
2. `filter(V8000 != 9999999)` remove observações com valores inválidos;
3. `mutate()` cria duas novas variáveis: a identificação da célula (`Celula`) e o valor anualizado da aplicação (`Valor`);
4. `relocate(Celula)` move a coluna "Celula" para o início;
5. `filter(V9001 == ...)` seleciona os códigos que representam as aplicações em previdência privada; e
6. `subset()` retém apenas as colunas necessárias, neste caso `Celula` e `Valor`.

3.3. Pacote **readxl**

O pacote `readxl` permite a leitura direta de arquivos do Excel com extensão `.xlsx`. Sua principal função é `read_excel()`, que importa planilhas sem a necessidade de conversão prévia para outros formatos, como `.csv`.

Embora seu uso não esteja evidenciado na parte inicial do *script*, essa biblioteca é carregada para possibilitar a eventual leitura de tabelas auxiliares criadas no Excel, como classificações de categorias de despesa, filtros, ou outros parâmetros externos.

3.4. Pacote **DirichletReg**

O pacote **DirichletReg** implementa métodos para estimação de modelos de regressão Dirichlet, que são apropriados para análise de dados composicionais — isto é, vetores de proporções que são não-negativos e somam exatamente 1.

Principais funções:

- `DR_data()`: converte vetores de proporções em objetos apropriados para regressão composicional;

- `DirichletReg()`: ajusta o modelo com base na distribuição Dirichlet multivariada.

Este pacote é fundamental nesta nota, pois a proposta metodológica envolve modelar a forma como diferentes domicílios distribuem sua renda entre consumo (por categoria) e poupança — caracterizando, assim, uma estrutura de dados composicional.

3.5. Pacote **ggplot2**

O **ggplot2** é um dos pacotes mais utilizados para construção de gráficos em R. Sua gramática gráfica permite a geração de visualizações altamente customizáveis com estrutura clara e modular.

Principais funções utilizadas:

- `ggplot()`: inicia o gráfico;

- `aes()`: define os mapeamentos estéticos (eixos, cores, grupos);

- `geom_bar()`, `geom_line()`: especificam o tipo de gráfico;

- `labs()`: adiciona rótulos e títulos; e

- `theme_minimal()`: aplica um estilo visual limpo.

A biblioteca é usada para representar graficamente os resultados da modelagem, facilitando a interpretação dos efeitos estimados.

3.6. Pacote **clipr**

O pacote `clipr` permite interações com a área de transferência do sistema operacional. A função `write_clip()` envia objetos diretamente do R para a área de transferência, permitindo colá-los facilmente em documentos Word, planilhas Excel ou outros editores.

Essa funcionalidade é especialmente útil nesta nota técnica, ao possibilitar a transferência ágil de tabelas geradas no R para documentos e planilhas, mantendo o formato e a estrutura dos dados.

3.7. Diretório de trabalho

O diretório de trabalho desta análise foi definido com o comando `setwd("C:/Users/Leonardo/Documents/Minhas consultas")`, que especifica o local onde os arquivos de entrada e saída serão lidos e salvos durante a execução dos *scripts*. A função `setwd()` — abreviação de *set working directory* — recebe como argumento o caminho da pasta onde estão armazenados os arquivos `.rds` utilizados ao longo do código. Sua função é garantir que os comandos subsequentes possam localizar corretamente os arquivos necessários, sem que seja preciso escrever o caminho completo a cada leitura de arquivo, o que evita erros do tipo “arquivo não encontrado”, tornando o *script* mais limpo e replicável dentro do ambiente de trabalho local do pesquisador.

No caso em questão, trata-se da pasta "Minhas consultas" no diretório pessoal do computador do pesquisador, de modo que, para replicar a análise em outro computador, será necessário ajustar o caminho para o diretório onde os arquivos `.rds` estiverem salvos no novo sistema.

A definição antecipada do diretório permite carregar as bases da POF 2017–2018 – que foram previamente convertidas para `.rds` – conforme a seguir exposto – com comandos diretos como `readRDS("DOMICILIO.rds")`.

4. Instrumentos financeiros

O seguinte trecho de código carrega e processa os dados da base `OUTROS_RENDIMENTOS.rds`, gerando a variável identificadora `Celula` para cada unidade de consumo (UC):

```

outros_rend_Cel <- readRDS("OUTROS_RENDIMENTOS.rds") %>%

  mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>%

  relocate(Celula)

```

A partir dessa base, calcula-se a variação líquida dos ativos financeiros de cada UC, com base nas categorias específicas declaradas na POF 2017–2018. As modalidades consideradas incluem: caderneta de poupança, fundos de aplicação, ações, além de outros depósitos e aplicações.

A base utilizada é o arquivo `OUTROS_RENDIMENTOS.rds`, que agrega informações dos quadros 56 e 57 da pesquisa. O comando acima realiza três operações sucessivas:

- Carrega a base de dados `OUTROS_RENDIMENTOS.rds` com a função `readRDS()`;
- Cria a variável `Celula`, que serve como identificador único da unidade de consumo e é construída pela concatenação das variáveis `COD_UPA`, `NUM_DOM` e `NUM_UC`, que representam, respectivamente, o número da unidade de pesquisa, o domicílio e a unidade de consumo dentro do domicílio; e
- Reorganiza as colunas com `relocate(Celula)` para posicionar a nova variável `Celula` como a primeira da base, o que facilita sua visualização e o uso como chave de junção; a variável `Celula`, por seu caráter de identificador, será utilizada em todos os blocos de programação seguintes.

4.1. Aplicações e resgates em caderneta de poupança

No trecho abaixo, filtra-se a base para identificar, separadamente, os registros de aplicações (`V9001 == 5700101`) e resgates (`V9001 == 5600101`) em cadernetas de poupança. A variável `V9001` identifica o tipo de transação conforme os códigos da POF. O critério adicional `V8500 != 9999999` elimina registros com valores faltantes, pois o valor `9999999` é utilizado pelo IBGE para indicar ausência de declaração válida.

```

Cel_APoup <- outros_rend_Cel %>% filter(V9001 == 5700101 & V8500 != 9999999)
%>% rename(APoup = V8500_DEFLA) %>% select(Celula, APoup)

Cel_RPoup <- outros_rend_Cel %>% filter(V9001 == 5600101 & V8500 != 9999999)
%>% rename(RPoup = V8500_DEFLA) %>% select(Celula, RPoup)

CelPoup <- merge(Cel_APoup, Cel_RPoup, by = "Celula", all = T)

CelPoup[is.na(CelPoup)] <- 0

Poup <- mutate(CelPoup, VariacaoPoup = APoup - RPoup)

```

As colunas resultantes `V8500_DEFLA` (valores monetários deflacionados) são renomeadas para `APoup` e `RPoup`, e apenas a chave `Celula` e o valor relevante são mantidos com `select()`. As duas tabelas resultantes são unidas por `merge()`, com `all = T` para manter registros que tenham somente aplicação ou somente resgate. Em seguida, NAs são convertidos para zero com `is.na()<-0`, e calcula-se a variação líquida da poupança (`VariacaoPoup`) como a diferença entre o valor aplicado e o valor resgatado.

4.2. Fundos de aplicação financeira

Aplica-se o mesmo procedimento para os **fundos de aplicação financeira**, identificados pelos códigos 5700201 (aplicações) e 5600201 (resgates) na variável `V9001`, conforme a codificação da POF. Assim como no caso da poupança, são excluídos os registros com valor 9999999 na variável `V8500`, que indica ausência de informação declarada. Os valores deflacionados das aplicações e resgates são renomeados como `AFundos` e `RFundos`, respectivamente, e os dados são combinados por `Celula`. Valores ausentes são substituídos por zero para evitar distorções na análise. Ao final, calcula-se a variável `VariacaoFundos`, que expressa a **variação líquida individual da unidade consumidora (UC)** em fundos de aplicação financeira. Veja o código abaixo:

```
Cel_AFundos <- outros_rend_Cel %>% filter(V9001 == 5700201 & V8500 !=
9999999) %>% rename(AFundos = V8500_DEFLA) %>% select(Celula, AFundos)

Cel_RFundos <- outros_rend_Cel %>% filter(V9001 == 5600201 & V8500 !=
9999999) %>% rename(RFundos = V8500_DEFLA) %>% select(Celula, RFundos)

CelFundos <- merge(Cel_AFundos, Cel_RFundos, by = "Celula", all = T)

CelFundos[is.na(CelFundos)] <- 0

Fundos <- mutate(CelFundos, VariacaoFundos = AFundos - RFundos)
```

4.3. Ações

Segue-se a mesma lógica para os **investimentos em ações**. Os filtros e renomeações aplicados são análogos aos utilizados para poupança e fundos, considerando agora os códigos 5700301 (compra de ações) e 5600301 (venda de ações), conforme especificado na variável `V9001` da POF. Novamente, excluem-se registros com `V8500 == 9999999`, que indicam ausência de declaração válida. Após o tratamento, os dados são integrados por `Celula`, valores ausentes são preenchidos com zero e calcula-se a variável `VariacaoAcoes`, que representa a **variação líquida individual da unidade consumidora (UC)** no mercado de ações. Veja,

```

    Cel_AAcoes <- outros_rend_Cel %>% filter(V9001 == 5700301 & V8500 !=
9999999) %>%

    rename(AAcoes = V8500_DEFLA) %>% select(Celula, AAcoes)
Cel_RAcoes <- outros_rend_Cel %>% filter(V9001 == 5600301 & V8500 !=
9999999) %>%

    rename(RAcoes = V8500_DEFLA) %>% select(Celula, RAcoes)
Cel_Acoes <- merge(Cel_AAcoes, Cel_RAcoes, by = "Celula", all = T)
Cel_Acoes[is.na(Cel_Acoes)] <- 0
Acoes <- mutate(Cel_Acoes, VariacaoAcoes = AAcoes - RAcoes)

```

4.4. Outros depósitos e aplicações

Neste caso, as operações referem-se a modalidades genéricas de depósito ou aplicação que não se enquadram nas categorias anteriores, com os códigos 5700401 e 5600401.

```

Cel_AOutros <- outros_rend_Cel %>% filter(V9001 == 5700401 & V8500 !=
9999999) %>% rename(AOutros = V8500_DEFLA) %>% select(Celula, AOutros)
Cel_ROutros <- outros_rend_Cel %>% filter(V9001 == 5600401 & V8500 !=
9999999) %>% rename(ROutros = V8500_DEFLA) %>% select(Celula, ROutros)
Cel_Outros <- merge(Cel_AOutros, Cel_ROutros, by = "Celula", all = T)
Cel_Outros[is.na(Cel_Outros)] <- 0
Outros <- mutate(Cel_Outros, VariacaoOutros = AOutros - ROutros)

```

4.5. Consolidação dos instrumentos financeiros

Os quatro tipos de ativo financeiro são unidos em uma única base chamada `Aplic_Juntos`, contendo a variação líquida de cada um deles para cada UC. Novamente, NAs são convertidos para zero para permitir soma correta.

```

Aplic_JuntosPF <- merge(Poup, Fundos, by = "Celula", all = T)
Aplic_JuntosAO <- merge(Acoes, Outros, by = "Celula", all = T)
Aplic_Juntos <- merge(Aplic_JuntosPF, Aplic_JuntosAO, by = "Celula", all =
T) %>%

subset(select = c(Celula, VariacaoPoup, VariacaoFundos, VariacaoAcoes,
VariacaoOutros))

Aplic_Juntos[is.na(Aplic_Juntos)] <- 0

```

A variável `Variacao_IF` consolida a soma da variação líquida de todos os instrumentos financeiros formais. O agrupamento por `Celula` é necessário para garantir que o resultado seja por UC, mesmo que ocorram múltiplos registros por modalidade.

```

Ativos_IF <- Aplic_Juntos %>%

mutate(Soma = VariacaoPoup + VariacaoFundos + VariacaoAcoes +
VariacaoOutros) %>%

group_by(Celula) %>% summarise(Variacao_IF = sum(Soma))

```

4.6. Limpeza da memória

O uso da função `rm()` elimina os objetos intermediários criados ao longo do processo, liberando memória e mantendo o ambiente de trabalho limpo e organizado. Essa função é empregada, como poderá ser observado, ao final de cada bloco de *scripts* apresentado neste trabalho.

```

rm(outros_rend_Cel, Cel_APoup, Cel_RPoup, CelPoup, Poup, Cel_AFundos,
Cel_RFundos, CelFundos, Fundos, Cel_AAcoes, Cel_RAcoes, CelAcoes, Acoes,
Cel_Outros, Cel_ROutros, CelOutros, Outros, Aplic_JuntosPF,
Aplic_JuntosAO, Aplic_Juntos)

```

4.7. Títulos de capitalização

Este bloco trata da variação líquida associada aos títulos de capitalização adquiridos e/ou resgatados pelas famílias brasileiras durante o período de referência da POF 2017–2018. A operação se baseia em duas fontes de informação: a base de despesas individuais (quadro 48 da POF) para identificar as aquisições, e a base de outros rendimentos (quadro 55) para captar os resgates.

4.7.1. Aquisição de Títulos de Capitalização

Esse trecho realiza o cálculo do valor anualizado gasto com aquisição de títulos de capitalização:

```
TitCap_Aquisicao <- readRDS("DESPESA_INDIVIDUAL.rds") %>% filter(V8000 !=
9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC),
Valor = V8000_DEFLA * V9011 * FATOR_ANUALIZACAO) %>%
relocate(Celula) %>%
filter(V9001 == 4803301 | V9001 == 4803302) %>%
subset(select = c(Celula, Valor))
TitCap_Aquisicao[is.na(TitCap_Aquisicao)] <- 0
TitCap_A <- group_by(TitCap_Aquisicao, Celula) %>% summarise(Aplic_TitCap =
sum(Valor))
```

- A base carregada é `DESPESA_INDIVIDUAL.rds`, que reúne gastos efetuados diretamente por indivíduos dentro da unidade de consumo;

- O filtro `V8000 != 9999999.99` remove registros com valor não informado (código de ausência definido pelo IBGE);

- A chave `Celula` é construída por concatenação dos identificadores da UC (`COD_UPA`, `NUM_DOM`, `NUM_UC`), garantindo unicidade e rastreabilidade;

- A variável `Valor` é calculada como o valor deflacionado da despesa (`V8000_DEFLA`) multiplicado pelo número de parcelas (`V9011`) e pelo fator de anualização, que permite transformar gastos parciais em uma estimativa do montante anual;

- O filtro `V9001==4803301|4803302` seleciona apenas os itens da despesa classificados como aquisição de títulos de capitalização, conforme a codificação do IBGE no quadro 48; e

- A função `group_by()` seguida de `summarise()` agrega as aquisições ao nível da UC, obtendo a variável `Aplic_TitCap`.

4.7.2. Resgate de Títulos de Capitalização

Agora, são processados os valores de resgates de títulos de capitalização:

```
TitCap_Resgate <- readRDS("OUTROS_RENDIMENTOS.rds") %>% filter(V8500 !=  
9999999) %>%  
  
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>%  
  
relocate(Celula) %>%  
  
filter(V9001 == 5504401) %>%  
  
group_by(Celula) %>%  
  
summarise(Resg_TitCap = sum(V8500_DEFLA))
```

- A base utilizada é `OUTROS_RENDIMENTOS.rds`, que contém rendimentos de natureza não salarial ou patrimonial;

- O código `V9001 == 5504401` identifica especificamente os resgates de títulos de capitalização no quadro 55 da POF;

- Valores ausentes são filtrados via `V8500 != 9999999`, que é o código reservado para respostas inválidas; e

- O valor deflacionado (`V8500_DEFLA`) é somado para cada UC, após agrupamento por `Celula`.

4.7.3. Cálculo da Variação Líquida e Consolidação

A etapa final consiste em:

- Unir as bases de aquisições e resgates com `merge()`, mantendo todas as observações (mesmo que haja só aquisição ou só resgate);

- Substituir valores ausentes por zero com `is.na() <- 0`, o que assegura que a subtração ocorra corretamente mesmo na ausência de uma das partes;

- Calcular a variação líquida da UC em relação aos títulos de capitalização: se positiva, indica mais aquisição do que resgate; se negativa, o contrário; e

- Selecionar apenas a `Celula` e a variável de interesse `Variacao_TitCap`.

```

TitCap_Juntos <- merge(TitCap_A, TitCap_Resgate, by = "Celula", all = T)
TitCap_Juntos[is.na(TitCap_Juntos)] <- 0
Ativos_TitCap <- mutate(TitCap_Juntos, Variacao_TitCap = Aplic_TitCap -
Resg_TitCap) %>% subset(select = c(Celula, Variacao_TitCap))
rm(TitCap_Aquisicao, TitCap_A, TitCap_Resgate, TitCap_Juntos)

```

4.8. Empréstimos concedidos e recebidos

Este bloco trata das operações de crédito não bancário realizadas pelas famílias: empréstimos que concederam a terceiros ou quitaram, e empréstimos que tomaram ou receberam o pagamento. O objetivo é calcular a variação líquida desse componente como item patrimonial relevante na alocação da poupança.

4.8.1. Saída de caixa por concessão ou quitação de empréstimos

Utiliza-se a base `DESPESA_INDIVIDUAL.rds` para captar:

- Código 4800101: empréstimos concedidos;
- Código 4800102: pagamento de empréstimos contraídos;
- Código 4802401: outras formas similares.

O valor é anualizado multiplicando-se os gastos deflacionados pelo número de parcelas e fator de ajuste. A soma por unidade de consumo (*Celula*) gera o total de saída financeira da família por esse motivo.

```

Emp_Aplic <- readRDS("DESPESA_INDIVIDUAL.rds") %>% filter(V8000 !=
9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC), Valor =
V8000_DEFLA*V9011*FATOR_ANUALIZACAO) %>% relocate(Celula) %>%
filter(V9001 == 4800101 | V9001 == 4800102 | V9001 == 4802401) %>%
subset(select = c(Celula, Valor))
Emp_Aplic[is.na(Emp_Aplic)] = 0
Emp_Aplic_A <- group_by(Emp_Aplic, Celula) %>%
summarise(Aplic_Emp = sum(Valor))

```

4.8.2. Entrada de caixa por empréstimos recebidos ou pagos por terceiros

A base `OUTROS_RENDIMENTOS.rds` fornece os registros de entrada por empréstimos tomados ou recebidos, filtrando-se o código 5501601. Após agrupar por `Celula`, obtém-se o valor total recebido por UC.

```
Emp_Resg <- readRDS("OUTROS_RENDIMENTOS.rds") %>% filter(V8500 != 9999999)
%>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>% relocate(Celula) %>%
filter (V9001 == 5501601) %>%
group_by(Celula) %>% summarise(Resg_Emp = sum(V8500_DEFLA))
```

4.8.3. Cálculo da variação líquida

A saída e a entrada de recursos são integradas em uma única tabela. Os valores ausentes são tratados como zero. A diferença entre o total concedido e o recebido fornece a variação líquida da posição de crédito da família.

```
Emp_Juntos <- merge(Emp_Aplic_A, Emp_Resg, by = "Celula", all = T)
Emp_Juntos [is.na (Emp_Juntos)] = 0
Ativos_Emp <- mutate(Emp_Juntos, Variacao_Emp = Aplic_Emp - Resg_Emp) %>%
subset(select = c(Celula, Variacao_Emp))
rm(Emp_Aplic, Emp_Aplic_A, Emp_Resg, Emp_Juntos)
```

5. Previdência social pública

Esta seção contabiliza os aportes realizados pelas famílias brasileiras à previdência pública. As contribuições são classificadas em dois grandes grupos: os contribuintes do setor público (Regimes Próprios de Previdência Social - RPPS, Tesouros Nacional e Regionais) e privado formal (Regime Geral de Previdência Social - RGPS), e os contribuintes individuais e facultativos. A soma de ambos compõe a aplicação líquida familiar destinada à previdência social obrigatória, considerada aqui como forma de alocação da poupança de longo prazo.

5.1. Contribuintes do setor formal (RGPS, RPPS, Tesouros)

Nesta etapa, utiliza-se a base `RENDIMENTO_TRABALHO.rds` (Quadro 53), focando na variável `V531112_DEFLA`, que representa o valor deflacionado das contribuições

previdenciárias formais. O valor é anualizado pela multiplicação com o número de parcelas (V9011), e agrupado por unidade de consumo (Celula).

```
rend_PrevPub <- readRDS("RENDIMENTO_TRABALHO.rds") %>% filter(V8500 !=
9999999) %>%
mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC), Valor = V531112_DEFLA*V9011)
%>% relocate(Celula, Valor) %>%
subset(select = c(Celula, Valor))
rend_PrevPub [is.na (rend_PrevPub)] = 0
Conta_PrevPub <- group_by(rend_PrevPub, Celula) %>%
summarise(Aplic_PrevPub = sum(Valor))
```

5.2. Contribuintes individuais e facultativos (RGPS)

A contribuição dos trabalhadores autônomos e facultativos é captada a partir da base DESPESA_INDIVIDUAL.rds (Quadro 48), com o código V9001 igual a 4800501. Multiplicam-se os valores deflacionados por V9011 e pelo fator de anualização para refletir o valor total aplicado ao longo de um ano. A agregação é feita por família.

```
rend_PrevPub_Outros <- readRDS("DESPESA_INDIVIDUAL.rds") %>% filter(V8000
!= 9999999.99) %>%
mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC), Valor =
V8000_DEFLA*V9011*FATOR_ANUALIZACAO) %>% relocate(Celula) %>%
filter(V9001 == 4800501) %>%
subset(select = c(Celula, Valor))
rend_PrevPub_Outros [is.na (rend_PrevPub_Outros)] = 0
Conta_PrevPub_Outros <- group_by(rend_PrevPub_Outros, Celula) %>%
summarise(Aplic_PrevPub_Outros = sum(Valor))
```

5.3. Cálculo da aplicação líquida em previdência pública

As duas fontes de contribuição previdenciária são integradas em uma mesma tabela, com valores ausentes substituídos por zero. A variável Variacao_PrevPub expressa a soma total anual de recursos destinados à previdência pública por família.

```
PrevPub_Juntos <- merge(Conta_PrevPub, Conta_PrevPub_Outros, by = "Celula",
all = T)
PrevPub_Juntos [is.na (PrevPub_Juntos)] = 0
Ativos_PrevPub <- mutate(PrevPub_Juntos, Variacao_PrevPub = Aplic_PrevPub +
Aplic_PrevPub_Outros) %>%
subset(select = c(Celula, Variacao_PrevPub))

rm(rend_PrevPub, Conta_PrevPub, rend_PrevPub_Outros, Conta_PrevPub_Outros,
PrevPub_Juntos)
```

6. Previdência privada aberta e fechada

A previdência privada constitui uma forma voluntária de poupança de longo prazo para a aposentadoria, podendo ser contratada tanto em regime aberto (oferecido por bancos e instituições financeiras) quanto em regime fechado (fundos de pensão, geralmente vinculados ao empregador). Ao contrário da previdência pública (RGPS/RPPS), trata-se de um ativo financeiro sob controle direto do contribuinte.

6.1 Aplicação

As aplicações em previdência privada referem-se aos aportes realizados pelos indivíduos em planos de previdência aberta ou fechada. Esses aportes são registrados como despesas na base `DESPESA_INDIVIDUAL.rds` e são identificados pelos códigos V9001 iguais a 4800601, 4800602, 4800603 e 4800604. Os valores monetários são ajustados com base em V8000_DEFLA, V9011 e FATOR_ANUALIZACAO, sendo então somados por unidade de consumo (Celula).

```
PrevPriv_Deposito <- readRDS("DESPESA_INDIVIDUAL.rds") %>% filter(V8000 !=
9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC), Valor =
V8000_DEFLA*V9011*FATOR_ANUALIZACAO) %>% relocate(Celula) %>%
filter(V9001 == 4800601 | V9001 == 4800602 | V9001 == 4800603 | V9001 ==
4800604) %>%
subset(select = c(Celula, Valor))
PrevPriv_Deposito [is.na (PrevPriv_Deposito)] = 0
PrevPriv_D <- group_by(PrevPriv_Deposito, Celula) %>%
summarise(Aplic_PrevPriv = sum(Valor))
```

6.2 Restituição ou Devolução

A restituição ou devolução de recursos dos planos de previdência privada, geralmente em decorrência de resgates, aposentadoria ou encerramento do plano, é informada na base `OUTROS_RENDIMENTOS.rds`, com código `V9001` igual a `5503301`. O montante é agregado por `Celula` para compor a entrada de caixa. A variação líquida do ativo previdência privada corresponde à diferença entre o total aplicado e o total restituído, caracterizando o saldo líquido que permanece investido.

```
PrevPriv_Restituicao <- readRDS("OUTROS_RENDIMENTOS.rds") %>% filter(V8500
!= 9999999) %>%
mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC)) %>% relocate(Celula) %>%
filter (V9001 == 5503301) %>%
group_by(Celula) %>% summarise(Rest_PrevPriv = sum(V8500_DEFLA))
```

6.3 Variação do saldo da Previdência Privada

```
PrevPriv_Juntos <- merge(PrevPriv_D, PrevPriv_Restituicao, by = "Celula",
all = T)
PrevPriv_Juntos [is.na (PrevPriv_Juntos)] = 0
Ativos_PrevPriv <- mutate(PrevPriv_Juntos, Variacao_PrevPriv =
Aplic_PrevPriv - Rest_PrevPriv) %>%
subset(select = c(Celula, Variacao_PrevPriv))
rm(PrevPriv_Deposito, PrevPriv_D, PrevPriv_Restituicao, PrevPriv_Juntos)
```

7. Fundo de Garantia do Tempo de Serviço (FGTS)

O FGTS constitui um dos ativos financeiros compulsórios mais relevantes para trabalhadores formais no Brasil. Por legislação, no início de cada mês, os empregadores devem depositar 8% do salário de cada funcionário em contas vinculadas na Caixa Econômica Federal. Estes valores pertencem ao empregado, que pode movimentá-los em situações específicas como demissão sem justa causa, aposentadoria, ou aquisição de imóvel.

7.1. Aplicação

A aplicação corresponde aos depósitos realizados na conta do FGTS durante o período de referência. Como a POF não apresenta diretamente esse valor, foi estimado multiplicando o rendimento bruto do trabalho por 0,08 (8%), apenas para indivíduos com vínculo empregatício formal (`V5304==1`). A base de dados utilizada foi a de rendimento do trabalho (Quadro 53), e o valor foi deflacionado e anualizado.

```
FGTS_Deposito <- readRDS("RENDIMENTO_TRABALHO.rds") %>% filter(V8500 !=
9999999) %>% filter(V5304 == 1) %>%
mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC), Valor =
V8500_DEFLA*V9011*FATOR_ANUALIZACAO, FGTS = 0.08*Valor) %>%
relocate(Celula, FGTS) %>%
group_by(Celula) %>% summarise(Aplic_FGTS = sum(FGTS))
```

7.2. Saque

O saque do FGTS corresponde às retiradas realizadas das contas vinculadas. Os dados foram obtidos a partir da base de outros rendimentos (Quadro 55), considerando os registros identificados pelo código V9001==5501001. A soma total de saques foi agregada por família com base na chave Celula.

```
FGTS_Saque <- readRDS("OUTROS_RENDIMENTOS.rds") %>% filter(V8500 !=
9999999) %>% mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC)) %>%
relocate(Celula)%>% filter (V9001 == 5501001) %>%
group_by(Celula) %>% summarise(Saque_FGTS = sum(V8500_DEFLA))
```

7.3 Variação do saldo do FGTS

A variação no saldo do FGTS é calculada pela diferença entre os depósitos estimados e os saques efetivos. Essa variação é considerada como um incremento (ou decréscimo) no ativo financeiro da família, compondo a estrutura de alocação da poupança analisada. Os valores ausentes são tratados como zero.

```
FGTS_Juntos <- merge(FGTS_Deposito, FGTS_Saque, by = "Celula", all = T)
FGTS_Juntos [is.na (FGTS_Juntos)] = 0
Ativos_FGTS <- mutate(FGTS_Juntos, Variacao_FGTS = Aplic_FGTS - Saque_FGTS)
%>% subset(select = c(Celula, Variacao_FGTS))
rm(FGTS_Deposito, FGTS_Saque, FGTS_Juntos)
```

8. Ativos não financeiros – duráveis

8.1 Demais duráveis

Este bloco se dedica à mensuração da variação patrimonial relacionada aos bens duráveis não imobiliários, como móveis, eletrodomésticos e demais itens com vida útil superior a um ano. Trata-se, portanto, de ativos não financeiros cuja aquisição, manutenção ou conserto representam aumento do patrimônio líquido da unidade de consumo (UC), ao passo que sua

venda representa redução do mesmo. Para tanto, o *script* manipula três bases distintas da POF 2017-2018: DESPESA_INDIVIDUAL.rds, DESPESA_COLETIVA.rds e OUTROS_RENDIMENTOS.rds.

A primeira etapa corresponde à identificação dos eventos de aquisição, manutenção e conserto de bens duráveis na base de despesas individuais. Esses eventos são extraídos dos Quadros 34, 35, 37, 38, 39, 43, 44 e 46, excetuando-se registros relacionados a itens de consumo imediato ou uso corrente, como aluguel de bens ou produtos perecíveis. Utiliza-se o valor deflacionado da despesa, ponderado pelo fator de anualização.

```
Duraveis_DI_Aumento <- readRDS("DESPESA_INDIVIDUAL.rds") %>%
filter(V8000 != 9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC),
Valor = V8000_DEFLA*FATOR_ANUALIZACAO) %>%
relocate(Celula) %>%
filter(QUADRO == 34 | QUADRO == 35 | QUADRO == 37 | QUADRO == 38 |
      QUADRO == 39 | QUADRO == 43 | QUADRO == 44 | QUADRO == 46) %>%
filter(V9001 != 3401701 & V9001 != 3502001 & V9001 != 3802401 &
      V9001 != 4400101 & V9001 != 4400102 & V9001 != 4400201 &
      V9001 != 4400202 & V9001 != 4400301 & V9001 != 3901701 &
      V9001 != 3901702 & V9001 != 3901703 & V9001 != 3901704 &
      V9001 != 3901705 & V9001 != 3901706 & V9001 != 3901707 &
      V9001 != 3901801 & V9001 != 3901901 & V9001 != 3901902 &
      V9001 != 3901903 & V9001 != 3901904) %>%
group_by(Celula) %>%
summarise(Aumento_DI = sum(Valor))
```

Paralelamente, realizam-se os mesmos procedimentos para as despesas coletivas, observando-se os quadros 9, 15 a 18. Também aqui são excluídos registros ligados a aluguel ou serviços que não representem aquisição de um ativo. O valor líquido é anualizado e agrupado por Celula:

```

Duraveis_DC_Aumento <- readRDS("DESPESA_COLETIVA.rds") %>%
filter(V8000 != 9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC),
Valor = V8000_DEFLA*FATOR_ANUALIZACAO) %>%
relocate(Celula) %>%
filter(QUADRO == 9 | QUADRO == 15 | QUADRO == 16 |
      QUADRO == 17 | QUADRO == 18) %>%
filter(V9001 != 901401 & V9001 != 901501 & V9001 != 901601 &
      V9001 != 910101 & V9001 != 910201 & V9001 != 910301 &
      V9001 != 910401 & V9001 != 910501 & V9001 != 910502 &
      V9001 != 910503 & V9001 != 910601 & V9001 != 910701 &
      V9001 != 910801) %>%
group_by(Celula) %>%
summarise(Aumento_DC = sum(Valor))

```

Em seguida, os totais apurados para despesas individuais e coletivas são agregados para compor a variável `Aumento_D`, que representa o total investido em bens duráveis no período. Por fim, identifica-se a alienação desses bens por meio da base `OUTROS_RENDIMENTOS.rds`, filtrando os códigos correspondentes a vendas de móveis, excluindo transações que não envolvam transferência de propriedade. A diferença entre aquisições e vendas compõe a variação líquida no ativo 'Demais Duráveis':

```

Duraveis_Juntos_Aumento <- merge(Duraveis_DI_Aumento, Duraveis_DC_Aumento,
by = "Celula", all = T)
Duraveis_Juntos_Aumento[is.na(Duraveis_Juntos_Aumento)] <- 0
Duraveis_Aumento <- mutate(Duraveis_Juntos_Aumento, Aumento_D = Aumento_DI
+ Aumento_DC)

Duraveis_Diminuicao <- readRDS("OUTROS_RENDIMENTOS.rds") %>%
filter(V8500 != 9999999) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>%
relocate(Celula) %>%
filter(V9001 >= 5502101 & V9001 <= 5502634) %>%
filter(!(V9001 %in% c(5502301, 5502401, 5502618, 5502630, 5502633,
5502634))) %>%
group_by(Celula) %>%
summarise(Diminuicao_D = sum(V8500_DEFLA))

Duraveis_Juntos <- merge(Duraveis_Aumento, Duraveis_Diminuicao, by =

```

```
"Celula", all = T)
Duraveis_Juntos[is.na(Duraveis_Juntos)] <- 0
Duraveis <- mutate(Duraveis_Juntos, Variacao_D = Aumento_D - Diminuicao_D)
Ativos_Duraveis <- subset(Duraveis, select = c(Celula, Variacao_D))
```

8.2. Veículos

Este segmento do *script* é dedicado ao cálculo da variação no estoque de ativos não financeiros representados por veículos. Para isso, são consideradas tanto as entradas (aquisições e melhorias) quanto as saídas (vendas), utilizando microdados da POF 2017–2018. As bases de dados utilizadas são `DESPESA_INDIVIDUAL.rds` e `OUTROS_RENDIMENTOS.rds`. O processo é dividido em três etapas: identificação dos aumentos no estoque de veículos, identificação das diminuições, e cálculo da variação líquida.

A primeira etapa consiste em identificar gastos relacionados à aquisição e conserto de veículos. Para isso, utiliza-se a base `DESPESA_INDIVIDUAL.rds`, considerando apenas os registros dos Quadros 33 e 51 da POF, os quais correspondem a despesas com transporte e manutenção de bens duráveis. Cada linha é filtrada para excluir valores faltantes (`V8000 != 9999999.99`) e é construída uma chave de identificação da unidade de consumo (UC) denominada `Celula`, a partir dos campos `COD_UPA`, `NUM_DOM` e `NUM_UC`. O valor da despesa é deflacionado e anualizado, utilizando o campo `V8000_DEFLA` multiplicado pelo `FATOR_ANUALIZACAO`.

```
Veiculos_Aumento <- readRDS("DESPESA_INDIVIDUAL.rds") %>%
filter(V8000 != 9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC),
Valor = V8000_DEFLA * FATOR_ANUALIZACAO) %>%
relocate(Celula) %>%
filter(QUADRO == 33 | QUADRO == 51) %>%
group_by(Celula) %>% summarise(Soma_A = sum(Valor))
```

A segunda etapa calcula a redução no estoque de veículos a partir da base `OUTROS_RENDIMENTOS.rds`, onde estão registradas transações como vendas. Mais uma vez, é criada a chave `Celula`, e um novo campo `codigo` é gerado ao truncar `V9001` (o código do rendimento declarado) para identificar apenas os rendimentos classificados entre os códigos 55023 e 55024, que correspondem à alienação de veículos.

```

Veiculos_Diminuicao <- readRDS("OUTROS_RENDIMENTOS.rds") %>%
filter(V8500 != 9999999) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>%
relocate(Celula) %>%
mutate(codigo = trunc(V9001/100)) %>%
filter(codigo == 55023 | codigo == 55024) %>%
group_by(Celula) %>% summarise(Soma_D = sum(V8500_DEFLA))

```

Na última etapa, as duas bases intermediárias são combinadas por meio da função `merge()`, unindo os valores de aquisição e venda por unidade de consumo. A função `mutate()` é então utilizada para calcular a variação líquida do ativo veículo. A estrutura é organizada na variável `Variacao_V`, que é a diferença entre os valores totais de aumento (`Soma_A`) e diminuição (`Soma_D`).

```

Veiculos_Juntos <- merge(Veiculos_Aumento, Veiculos_Diminuicao, by =
"Celula", all = T)
Veiculos_Juntos[is.na(Veiculos_Juntos)] = 0
Veiculos <- mutate(Veiculos_Juntos, Variacao_V = Soma_A - Soma_D)
Ativos_Veiculos <- subset(Veiculos, select = c(Celula, Variacao_V))
rm(Veiculos_Aumento, Veiculos_Diminuicao, Veiculos_Juntos, Veiculos)

```

Com isso, obtém-se uma planilha com a variação no saldo patrimonial relacionado aos veículos para cada unidade de consumo presente nos microdados da POF. Este procedimento permite incluir a dimensão dos ativos veiculares no cálculo da poupança das famílias.

9. Ativos não financeiros duráveis – Imóvel

Nesta seção, apura-se a variação patrimonial líquida associada a imóveis. Consideram-se aqui as aquisições, construções e reformas de imóveis — inclusive terrenos — como eventos que resultam no aumento do saldo dos ativos não financeiros. Por outro lado, a venda de imóveis representa uma redução desses ativos.

A apuração das aquisições foi realizada a partir das bases Despesas Individuais e Despesas Coletivas da POF 2017–2018. A base individual contempla gastos atribuíveis diretamente à unidade de consumo (UC), enquanto a base coletiva contempla despesas compartilhadas entre membros do domicílio.

A extração das despesas relacionadas à aquisição, construção ou reforma de imóveis foi feita com base nos códigos específicos da variável V9001. As linhas de código a seguir exemplificam o procedimento aplicado à base de despesas individuais.

```
Imoveis_DI <- readRDS("DESPESA_INDIVIDUAL.rds") %>%
filter(V8000 != 9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC),
Valor = V8000_DEFLA * FATOR_ANUALIZACAO) %>%
relocate(Celula) %>%
filter(V9001 == 4700101 | V9001 == 4700201 | V9001 == 4700202 |
      V9001 == 4702001 | V9001 == 4702101 | V9001 == 4702701 |
      V9001 == 4702801 | V9001 == 4702901) %>%
group_by(Celula) %>% summarise(Aumento_DI = sum(Valor))
```

O código realiza os seguintes procedimentos:

- Carrega os dados da base DESPESA_INDIVIDUAL.rds;
- Remove registros com valores inválidos (9999999.99);
- Cria a chave de identificação da UC (Celula) por concatenação dos campos COD_UPA, NUM_DOM e NUM_UC;
- Corrige monetariamente os valores por meio do deflator V8000_DEFLA e do fator de anualização (FATOR_ANUALIZACAO);
- Filtra os códigos V9001 compatíveis com aquisição ou construção de imóveis; e
- Agrega os valores por UC.

Para capturar valores provenientes da base coletiva, aplicou-se procedimento análogo, com ajustes nos filtros.

```
Imoveis_DC <- readRDS("DESPESA_COLETIVA.rds") %>%
filter(V8000 != 9999999.99) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC),
Valor = V8000_DEFLA * FATOR_ANUALIZACAO) %>%
relocate(Celula) %>%
filter(QUADRO == 11 | V9001 == 1200101 | V9001 == 1200201) %>%
group_by(Celula) %>% summarise(Aumento_DC = sum(Valor))
```

Após a consolidação dos dados individuais e coletivos, os valores foram agregados para formar o total de acréscimo patrimonial relacionado a imóveis.

```

Imoveis_Juntos_Aumento <- merge(Imoveis_DI, Imoveis_DC, by = "Celula", all
= T)
Imoveis_Juntos_Aumento[is.na(Imoveis_Juntos_Aumento)] <- 0
Imoveis_Aumento <- mutate(Imoveis_Juntos_Aumento, Aumento_I = Aumento_DI +
Aumento_DC)

```

Por fim, as vendas de imóveis, que representam decréscimo patrimonial, foram extraídas da base `OUTROS_RENDIMENTOS.rds`, por meio dos códigos específicos da variável `V9001`:

```

Imoveis_Diminuicao <- readRDS("OUTROS_RENDIMENTOS.rds") %>%
filter(V8500 != 9999999) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>%
relocate(Celula) %>%
filter(V9001 == 5502101 | V9001 == 5502201) %>%
group_by(Celula) %>% summarise(Diminuicao_I = sum(V8500_DEFLA))

```

A variação líquida no saldo dos ativos associados a imóveis foi calculada como a diferença entre o total de aquisições (ou construções/reformas) e o total de vendas.

```

Imoveis_Juntos <- merge(Imoveis_Aumento, Imoveis_Diminuicao, by = "Celula",
all = T)
Imoveis_Juntos[is.na(Imoveis_Juntos)] <- 0
Imoveis <- mutate(Imoveis_Juntos, Variacao_I = Aumento_I - Diminuicao_I)
Ativos_Imoveis <- subset(Imoveis, select = c(Celula, Variacao_I))

rm(Imoveis_DI, Imoveis_DC, Imoveis_Juntos_Aumento, Imoveis_Aumento,
Imoveis_Diminuicao, Imoveis_Juntos, Imoveis)

```

10. Apropriação dos recursos de empréstimos entre bens duráveis, veículos e imóveis

Para compreender adequadamente como os recursos obtidos por meio de empréstimos são utilizados pelas famílias brasileiras, é necessário estimar a apropriação desses valores em diferentes tipos de ativos não financeiros. O trecho de código a seguir realiza esse procedimento com base nos ativos já computados anteriormente: bens duráveis (D), veículos (V) e imóveis (I).

Inicialmente, são unidas as bases referentes às variações nos saldos de duráveis, veículos e imóveis, além do total recebido por meio de empréstimos (objeto `Ativos_Emp`). O uso do

operador `merge` permite fazer a junção por célula, preservando os dados mesmo quando não há correspondência completa entre os subconjuntos.

Após a junção dos dados, a estrutura resultante recebe os valores ausentes preenchidos com zero, o que evita problemas nos cálculos posteriores. Em seguida, o *script* calcula a soma das variações dos ativos duráveis e de veículos (`Soma_D_V`), bem como a soma total considerando também os imóveis (`Soma_D_V_I`).

O ponto central do procedimento está na criação das proporções: a variável `Per_D_V` representa a proporção do total de empréstimos que se considera apropriada para duráveis e veículos, enquanto `Per_I` representa a fração residual, atribuída aos imóveis.

A lógica empregada respeita o sinal das variações: apenas quando duráveis/veículos e imóveis apontam para a mesma direção (ambos positivos ou ambos negativos), distribui-se proporcionalmente. Se duráveis/veículos estão positivos e imóveis negativos, assume-se que 100% do empréstimo foi aplicado nos primeiros; na situação inversa, assume-se que nada foi aplicado nos duráveis/veículos.

Por fim, essas proporções são aplicadas ao total de empréstimos (`Variacao_Emp`) para obter duas variáveis: `Variacao_Emp_DV`, correspondente à parcela apropriada para duráveis e veículos, e `Variacao_Emp_I`, apropriada para imóveis. Apenas as células com variação total positiva em ativos não financeiros são mantidas.

Esse tipo de alocação proporcional é particularmente útil para interpretar como as famílias tratam os recursos obtidos por financiamento — especialmente em contextos onde não há marcação explícita da destinação dos valores. A abordagem adotada no *script* oferece, assim, uma inferência prática, coerente com os dados observados.

```
Proporcao_D_V <- merge(Ativos_Duraveis, Ativos_Veiculos, by = "Celula", all = T)
Proporcao_D_V_I <- merge(Proporcao_D_V, Ativos_Imoveis, by = "Celula", all = T)
Proporcao_D_V_I_Emp <- merge(Proporcao_D_V_I, Ativos_Emp, by = "Celula", all = T)
Proporcao_D_V_I_Emp[is.na(Proporcao_D_V_I_Emp)] = 0

Ativos_Emp_P <- mutate(Proporcao_D_V_I_Emp,
                        Soma_D_V = round(Variacao_D + Variacao_V, 2),
                        Soma_D_V_I = round(Soma_D_V + Variacao_I, 2),
```

```

        Per_D_V = ifelse((Soma_D_V>=0 & Variacao_I>=0) |
(Soma_D_V<=0 & Variacao_I<=0), Soma_D_V/Soma_D_V_I,
        ifelse((Soma_D_V>0 & Variacao_I<0), 1,
0)),
        Per_I = 1-Per_D_V,
        Variacao_Emp_DV = Variacao_Emp*Per_D_V,
        Variacao_Emp_I = Variacao_Emp*Per_I) %>%
filter(Soma_D_V_I>0)

Ativos_Emp_DV <- subset(Ativos_Emp_P, select = c(Celula, Variacao_Emp_DV))
Ativos_Emp_I <- subset(Ativos_Emp_P, select = c(Celula, Variacao_Emp_I))

```

11. Seleção das famílias

Para fins de análise comparável e controlada da alocação da poupança, foi adotado um critério rigoroso de seleção das famílias. A amostra é composta exclusivamente por famílias com quatro membros: um casal heterossexual com dois filhos em comum, sendo a “Pessoa de Referência” (PR) um dos cônjuges, com idade entre 35 e 55 anos completos.

O primeiro bloco do código seleciona, a partir da base de moradores da POF (arquivo MORADOR.rds), as células familiares cuja PR (identificada pelo código de informante igual a 1) possui idade dentro da faixa definida e renda monetária per capita positiva.

```

Morador_Ref <- readRDS("MORADOR.rds") %>%
mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC), Renda = RENDA_DISP_PC*12*4)
%>%
relocate(Celula) %>% filter(COD_INFORMANTE == 1 & RENDA_MONET_PC > 0 &
RENDA_DISP_PC >= 5724/4 & RENDA_DISP_PC<= 23850/4) %>%
subset(select = c(Celula, V0403, Renda)) %>%
rename(Idade = V0403) %>% filter(Idade >= 35 & Idade <= 55)

```

Adicionalmente, impõem-se limites para a renda disponível per capita anualizada, filtrando-se os casos cuja renda total disponível familiar (estimada como 4 vezes a renda per capita mensal) esteja entre R\$ 5.724,00 e R\$ 23.850,00. Este intervalo elimina os extremos da distribuição e garante maior comparabilidade entre os grupos.

```

Morador_Ref <- readRDS("MORADOR.rds") %>%
mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC), Renda = RENDA_DISP_PC*12*4)
%>%
relocate(Celula) %>% filter(COD_INFORMANTE == 1 & RENDA_MONET_PC > 0 &
REDA_DISP_PC >= 5724/4 & REDA_DISP_PC<= 23850/4) %>%
subset(select = c(Celula, V0403, Renda)) %>%
rename(Idade = V0403) %>% filter(Idade >= 35 & Idade <= 55)

```

Em seguida, a seleção refina o escopo da amostra ao identificar aquelas células que contêm exatamente três outros membros além da PR: o cônjuge de sexo diferente (V0306 igual a 2) e dois filhos em comum (V0306 igual a 4). A função `group_by()` seguida de `summarise()` permite contar quantos membros compõem cada célula, restringindo a análise apenas às famílias que atendem à estrutura desejada (três membros além da PR). Finalmente, as bases são cruzadas e apenas as células que satisfazem ambas as condições são mantidas.

```

Morador <- readRDS("MORADOR.rds") %>%
mutate(Celula = paste(COD_UPA,NUM_DOM,NUM_UC)) %>%
relocate(Celula) %>% filter(V0306 == 1 | V0306 == 2 | V0306 == 4) %>%
subset(select = c(Celula, COD_INFORMANTE))
Morador_Q <- group_by(Morador,Celula) %>% summarise(Q = n()) %>% filter(Q
== 3)
Familia <- merge(Morador_Ref, Morador_Q, by = "Celula", all = T) %>%
filter(Q == 3, Idade != "NA")

```

12. Seleção da ocupação da pessoa de referência

A base de dados da POF permite a seleção de um conjunto amplo de ocupações, o que potencializa a replicação do estudo para distintos perfis profissionais. A seleção da ocupação da pessoa de referência (PR) é feita a partir da base de Rendimento do Trabalho, limitando-se ao primeiro informante do domicílio (identificado por `COD_INFORMANTE == 1`), com ocupação principal declarada (`SUB_QUADRO == 1`) e rendimentos monetários positivos (`V8500 != 9999999`). As categorias ocupacionais são diferenciadas por meio das variáveis `V5302` e `V5303`, conforme o tipo de vínculo empregatício ou forma de inserção no mercado de trabalho.

12.1. Servidor Público Estatutário

```
Renda_Trab <- readRDS("RENDIMENTO_TRABALHO.rds") %>%
filter(V8500 != 9999999 & SUB_QUADRO == 1 & COD_INFORMANTE == 1 &
V5302 == 4 & V5303 == 1) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>%
subset(select = c(Celula))
Renda_Trab [is.na (Renda_Trab)] = 0
```

12.2. Empregado do setor privado

```
Renda_Trab <- readRDS("RENDIMENTO_TRABALHO.rds") %>%
filter(V8500 != 9999999 & SUB_QUADRO == 1 & COD_INFORMANTE == 1 & V5302 ==
3) %>%
mutate(Celula = paste(COD_UPA, NUM_DOM, NUM_UC)) %>%
subset(select = c(Celula))
Renda_Trab [is.na (Renda_Trab)] = 0
```

Além dessas duas categorias exemplificativas, foram implementadas rotinas para outras ocupações disponíveis na POF, quais sejam:

- Trabalhador doméstico (V5302 == 1);
- Militar (V5302 == 2);
- Empregado público com vínculo CLT (V5302 == 4 & V5303 == 2);
- Empregador (V5302 == 5); e
- Conta própria (V5302 == 6).

Essas rotinas ampliam o escopo de análises possíveis, permitindo investigações comparativas adicionais — como as que contrastam a poupança de autônomos e empregadores com a de trabalhadores formais. Para cada ocupação, o código filtra as observações relevantes e identifica as famílias correspondentes por meio do código `Celula`.

Após a seleção da ocupação da Pessoa de Referência (PR) — por exemplo, a de servidor público estatutário ou empregado do setor privado —, procede-se à junção desta base com aquela que reúne os domicílios filtrados quanto à composição familiar e faixa etária da PR. Essa junção, feita pela variável identificadora `Celula`, resulta na base denominada `Trabalhador`, que contém as famílias com perfis compatíveis e cuja PR exerce a ocupação de interesse da análise.

A partir da base `Trabalhador`, realiza-se uma nova junção com a base `Aplic_Percentuais`, a qual contém, para cada família, os percentuais alocados nos três tipos de ativos: financeiros, não financeiros duráveis (inclusive veículos), e previdência + imóveis (considerados ativos de longo prazo). O objetivo dessa nova base —

Trabalhador_Var_Indep — é estruturar os dados com as variáveis independentes selecionadas, tais como idade da PR e renda monetária disponível anual da família.

```
Trabalhador <- merge(Familia, Renda_Trab, by = "Celula")

Trabalhador_Var_Indep <- merge(Trabalhador, Aplic_Percentuais, by =
"Celula") %>%
  subset(select = c(Per_Fin, Per_Dur, Per_Prev_I, Idade, Variacao, Renda))
%>%
mutate(Taxa_Poupanca = Variacao/Renda)

rm(Familia, Renda_Trab_Pub, Trabalhador, Morador, Morador_Q)
```

13. Modelo econométrico – regressão Dirichlet

A última etapa da análise proposta nesta nota técnica diz respeito à modelagem da alocação da poupança das famílias entre três componentes principais, agrupados segundo o grau de liquidez: ativos financeiros, bens duráveis/veículos e imóveis/previdência. Para isso, utiliza-se a regressão Dirichlet, técnica apropriada para dados composicionais.

Este tipo de abordagem, conforme discutido em Hijazi e Jernigan (2009), permite modelar vetores de proporções como função de covariáveis observadas, respeitando as restrições naturais da composição, como a soma dos componentes igual a um e a interdependência entre eles. Diferentemente de modelos multivariados convencionais — que podem gerar inferências inadequadas ao ignorar essa estrutura —, a **regressão Dirichlet** preserva as propriedades composicionais dos dados, oferecendo uma modelagem estatisticamente mais coerente.

No caso em questão, utiliza-se a função `DirichReg()` do pacote `DirichletReg` em R para ajustar um modelo do tipo:

```
head(Trabalhador_Var_Indep)

AL <- DR_data(Trabalhador_Var_Indep[,1:3])
AL[1:10, ]

Modelo <- DirichReg(AL ~ I(Idade^2)-1 | I(Idade^2) - 1 | I(Idade^2) - 1, data
= Trabalhador_Var_Indep)

summary(Modelo)
```

Aqui, `AL` é um objeto da classe `DR_data` que representa a matriz de proporções (ativos financeiros, duráveis/veículos, imóveis/previdência), e a variável explicativa comum nos três submodelos é a idade ao quadrado da pessoa de referência da família (PR). A retirada do intercepto (com `-1`) se justifica por características do modelo e da amostra, uma vez que o objetivo é modelar variações em função exclusiva da idade, sem termo constante.

Esse tipo de estrutura com submodelos separados por | reflete o caráter multivariado do modelo Dirichlet e permite, conforme Hijazi (2003), que diferentes relações com a covariável sejam capturadas para cada componente do vetor de resposta. O uso do termo $I(Idade^2)$ indica que se adota uma forma funcional quadrática, coerente com a hipótese de ciclo de vida, segundo a qual a taxa de poupança pode aumentar até certo ponto da idade adulta e depois declinar. Ademais, o autor propôs o uso de funções quadráticas para capturar efeitos não-lineares (da idade).

A execução do modelo retorna os coeficientes estimados, seus erros padrões e estatísticas de significância, permitindo avaliar a influência da idade sobre a alocação da poupança entre os componentes. Com base nesses resultados, torna-se possível simular a trajetória da composição da poupança ao longo da vida ativa e comparar diferentes grupos ocupacionais ou faixas de renda, como proposto em seções anteriores desta nota técnica.

14. Exemplo de aplicação do modelo – empregados do setor privado

A aplicação do modelo de regressão Dirichlet aos empregados do setor privado visa compreender como a idade da pessoa de referência influencia a alocação da poupança acumulada entre três componentes principais: ativos financeiros (Per_Fin), bens duráveis e veículos (Per_Dur), e previdência/imóveis (Per_Prev_I). Os resultados estimados para os empregados do setor privado foram os seguintes:

```
Call:
DirichReg(formula = AL ~ I(Idade^2) - 1 | I(Idade^2) - 1 | I(Idade^2) - 1,
data = Trabalhador_Var_Indep)

Standardized Residuals:
              Min          1Q      Median          3Q          Max
Per_Fin      -1.3068  -0.8009  -0.2555   0.3256   3.4144
Per_Dur      -2.5497  -0.6470   0.2496   1.0663   2.6332
Per_Prev_I   -1.6823  -0.7496  -0.2312   0.4190   3.6578

-----
Beta-Coefficients for variable no. 1: Per_Fin
      Estimate Std. Error z value Pr(>|z|)
I(Idade^2) 9.254e-05  2.198e-05   4.21 2.55e-05 ***
-----
Beta-Coefficients for variable no. 2: Per_Dur
      Estimate Std. Error z value Pr(>|z|)
I(Idade^2) 3.809e-04  2.172e-05  17.54  <2e-16 ***
-----
Beta-Coefficients for variable no. 3: Per_Prev_I
      Estimate Std. Error z value Pr(>|z|)
I(Idade^2) 2.452e-04  2.199e-05  11.15  <2e-16 ***
-----
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Log-likelihood: 503.5 on 3 df (64 BFGS + 2 NR Iterations)
AIC: -1001, BIC: -988.1
Number of Observations: 535
Link: Log
Parametrization: common
```

Não obstante esteja fora do escopo desta nota técnica a realização de análises, observa-se que os coeficientes estimados para as três equações são estatisticamente significativos (valores- $p < 0,001$), o que indica forte evidência de que a idade — mais precisamente sua idade ao quadrado — está associada à forma como a poupança é distribuída entre essas categorias. Os sinais positivos dos coeficientes sugerem uma tendência de aumento relativo na alocação para todos os componentes com o avanço da idade, ainda que, devido à natureza composicional dos dados, essa interpretação deva ser feita com cautela.

Os resíduos padronizados estão em faixas aceitáveis, sem indicar problemas relevantes de ajuste. Por fim, o valor do log-verossimilhança (503,5), bem como os critérios de informação AIC (-1001) e BIC (-988,1), apontam para um bom desempenho do modelo, especialmente considerando sua estrutura parcimoniosa.

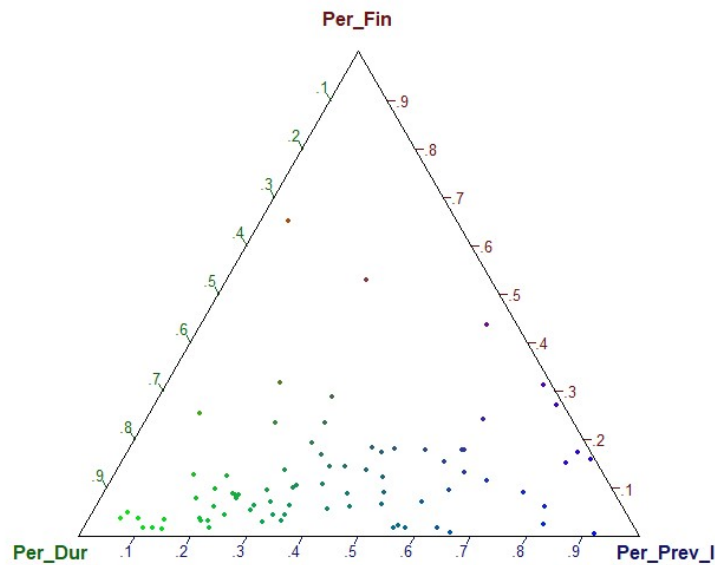
15. Visualização Gráfica

15.1. Diagrama Ternário

Como forma de explorar visualmente a estrutura dos dados composicionais, foi utilizado o diagrama ternário, recurso especialmente útil para variáveis que representam proporções de um todo. Esse tipo de gráfico permite visualizar a distribuição conjunta de três componentes — no caso, os percentuais de poupança alocados em ativos financeiros (`Per_Fin`), bens duráveis (`Per_Dur`) e previdência/imóveis (`Per_Prev_I`) — dentro de um simplex tridimensional, onde cada ponto representa uma observação cuja soma das partes é igual a 1.

A geração do gráfico é realizada a partir do objeto transformado pela função `DR_data()`, utilizando o comando `plot()` do pacote `DirichletReg`, que possui métodos específicos para esse tipo de dado. O *script* a seguir implementa essa visualização:

```
plot(AL, cex = 0.5, a2d = list(colored = TRUE, c.grid = FALSE))
```



Esse comando elabora o diagrama ternário com pontos em tamanho reduzido (`cex = 0.5`) e com coloração automática por densidade (`colored = TRUE`), facilitando a percepção de aglomerações e padrões. A opção `c.grid = FALSE` remove a malha de grade auxiliar, proporcionando um gráfico mais limpo. Essa visualização é particularmente útil para identificar regiões de concentração de dados, relações de dominância entre componentes e possíveis agrupamentos naturais.

A leitura do gráfico permite compreender como as famílias analisadas distribuem sua poupança entre as categorias propostas, podendo ser uma etapa exploratória relevante antes da estimação do modelo econométrico.

15.2. Visualização gráfica da alocação da poupança

A partir dos resultados obtidos na regressão Dirichlet, esta subseção dedica-se à visualização gráfica dos padrões de alocação da poupança segundo a idade da pessoa de referência da família, bem como à representação da precisão associada às estimativas do modelo. O *script* segue a formatação a seguir.

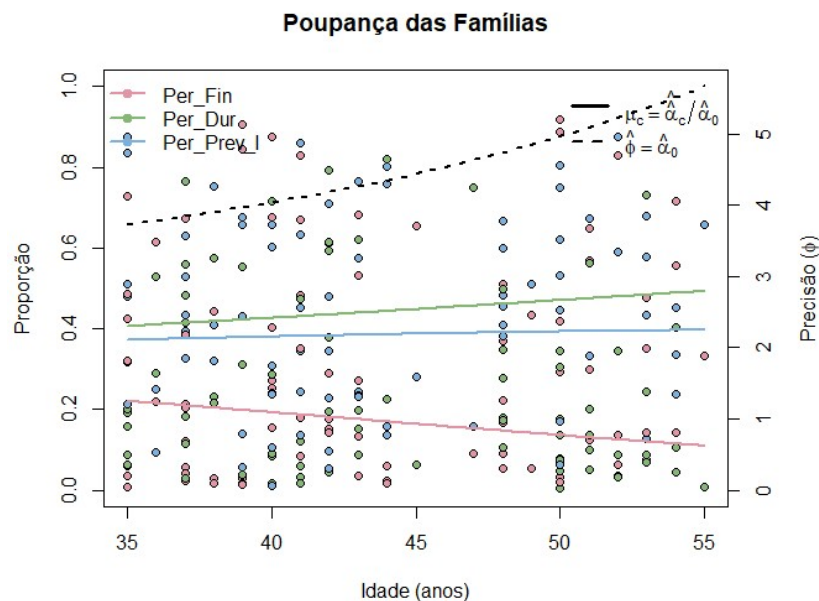
```

plot(rep(Trabalhador_Var_Indep$Idade, 3), as.numeric(AL), pch = 21, bg =
rep(c("#E495A5", "#86B875", "#7DB0DD"), each = 39), xlab = "Idade (anos)",
ylab = "Proporção", ylim = 0:1)
par(mar = c(4, 4, 4, 4) + 0.1)
plot(rep(Trabalhador_Var_Indep$Idade, 3), as.numeric(AL), pch = 21,
      bg = rep(c("#E495A5", "#86B875", "#7DB0DD"), each = 39),
      xlab = "Idade (anos)", ylab = "Proporção", ylim = 0:1,
      main = "Poupança das Famílias")

Xnew <- data.frame(Idade = seq(min(Trabalhador_Var_Indep$Idade),
max(Trabalhador_Var_Indep$Idade), length.out = 100))
for (i in 1:3) lines(cbind(Xnew, predict(Modelo, Xnew)[, i]), col =
c("#E495A5", "#86B875", "#7DB0DD")[i], lwd = 2)
legend("topleft", legend = c("Per_Fin", "Per_Dur", "Per_Prev_I"), lwd = 2,
col = c("#E495A5", "#86B875", "#7DB0DD"), pt.bg = c("#E495A5", "#86B875",
"#7DB0DD"), pch = 21, bty = "n")
par(new = TRUE)
plot(cbind(Xnew, predict(Modelo, Xnew, F, F, T)), lty = "24", type = "l",
ylim = c(0, max(predict(Modelo, Xnew, F, F, T))), axes = F, ann = F, lwd =
2)
axis(4)
mtext(expression(paste("Precisão (", phi, ")", sep = "")), 4, line = 3)
legend("topright", legend = c(expression(hat(mu[c] ==
hat(alpha)[c]/hat(alpha)[0])), expression(hat(phi) == hat(alpha)[0])), lty
= c(1, 2), lwd = c(3, 2), bty = "n")
AL <- Trabalhador_Var_Indep
AL$AL <- DR_data(Trabalhador_Var_Indep[, 1:3])
dd <- range(Trabalhador_Var_Indep$Idade)
X <- data.frame(Idade = seq(dd[1], dd[2], length.out = 200))
pp <- predict(DirichReg(AL ~ I(Idade^2) - 1 | I(Idade^2) - 1 | I(Idade^2) -
1, AL), X)

```

O gráfico principal exhibe, em um sistema de coordenadas cartesianas, os pontos empíricos correspondentes às proporções de poupança alocadas nas três categorias de ativos consideradas — ativos financeiros, bens duráveis e previdência/imóveis — em função da idade da pessoa de referência. Cada observação é representada por um ponto com cor específica, conforme o agrupamento da variável resposta: tons de rosa para ativos financeiros, verde para bens duráveis e azul para previdência/imóveis.



Sobre os pontos observados, são traçadas três curvas suavizadas, obtidas por meio da função `predict()` aplicada ao modelo ajustado. Essas curvas representam as médias estimadas da alocação da poupança para cada categoria ao longo da faixa etária da amostra, com base no termo quadrático da idade. A estrutura funcional quadrática é compatível com o modelo comum de regressão Dirichlet adotado.

Adicionalmente, sobre o mesmo plano, é traçada uma curva de precisão estimada, representada pelo parâmetro ϕ (phi), o qual resulta da soma dos parâmetros α do modelo Dirichlet. Essa função é plotada em um segundo eixo vertical (à direita), o que permite observar, para cada idade simulada, o nível de precisão da alocação predita. Um valor mais elevado de ϕ indica menor variância na distribuição das proporções entre as categorias de ativos, e, portanto, maior certeza associada à predição feita pelo modelo.

O *script* que gera essa visualização finaliza com a criação de um grid de valores para a idade, o qual é utilizado como base para o cálculo das predições e da precisão. Os comandos `par(new = TRUE)` e `axis(4)` permitem a sobreposição da curva de precisão sobre o gráfico anterior, sem comprometer a legibilidade dos dois conjuntos de informações. A legenda à direita explicita as expressões estatísticas envolvidas na estimação do modelo, favorecendo a leitura analítica do resultado.

Esse conjunto gráfico cumpre o importante papel de tornar acessível a interpretação dos resultados estimados, ilustrando como a idade da pessoa de referência afeta a distribuição da

poupança entre as diferentes formas de ativos, e qual a robustez estatística dessas estimativas ao longo da amostra analisada.

16. Considerações Finais

Esta nota técnica apresenta uma proposta de análise da alocação da poupança das famílias brasileiras entre diferentes categorias de ativos agrupados por grau de liquidez, com base nos microdados da POF 2017–2018. As proporções de poupança alocadas em ativos financeiros, bens duráveis e previdência/imóveis são modeladas por meio de regressão Dirichlet, técnica estatística apropriada para o tratamento de dados composicionais.

Para além da aplicação do modelo, o estudo se destaca pela construção criteriosa da base analítica, composta por famílias homogêneas — casais com dois filhos —, e pela elaboração de uma metodologia replicável para identificar e quantificar a variação líquida dos ativos com base em diversas bases da POF. O processo completo é documentado com apoio de *scripts* em R, permitindo que o usuário reproduza integralmente a abordagem adotada. A nota também exemplifica os resultados com base no perfil ocupacional dos empregados do setor privado.

Contudo, esta nota técnica não pretende esgotar o tema nem abordar todas as possibilidades de modelagem e aplicação dos dados. Seu objetivo é oferecer uma introdução prática e fundamentada, que contribua para o uso rigoroso dos microdados e para a compreensão da dinâmica de formação da riqueza. Usuários interessados em aprofundamentos adicionais são encorajados a consultar a bibliografia especializada e a documentação dos pacotes utilizados.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] HIJAZI, R. H. Analysis of Compositional Data Using Dirichlet Covariate Models. PhD thesis, American University, Washington, D.C, 2003.
- [2] HIJAZI, R. H. and JERNIGAN, R. W. (2009). Modeling compositional data using dirichlet regression models. *Journal of Applied Probability e Statistics*, 4(1):77–91.
- [3] Instituto Brasileiro de Geografia e Estatística (IBGE). Pesquisa de orçamentos familiares 2017-2018: Primeiros resultados/IBGE, Coordenação de Trabalho e Rendimento. Rio de Janeiro: IBGE, 2019.

[4] MAIER, M. J. DirichletReg: Dirichlet regression for compositional data in R. Research Report Series/ Department of Statistics and Mathematics 125, WU Vienna University of Economics and Business, Vienna, January 2014.

[5] MAIER, M. J. DirichletReg: Dirichlet Regression in R. Versão 0.7-1. Pacote de software. Disponível em: <<https://cran.r-project.org/web/packages/DirichletReg/index.html>>. Acesso em: 26 jun. 2024.

[6] R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>