

Nota Técnica da Rotina de Cálculo do Índice de Desenvolvimento Familiar (IDF)

Núcleo de Avaliação e Gestão da Informação
Vitória Sesana

Janeiro de 2023

Sumário

1	Introdução	3
2	Índice de Desenvolvimento Social	4
2.1	O que é, para que serve e suas vantagens	4
2.2	Estrutura do IDF	5
3	Rotina de cálculo para o IDF	9
3.1	Dados de entrada e de saída	9
3.2	Estrutura do script de cálculo	11
3.3	Elaborando o script	15
3.4	Exemplo	17
4	Considerações Finais	19
5	Referências	20

1 Introdução

Interpretar adequadamente a realidade é essencial no processo de análise para tomadas de decisões. No entanto, dependendo da complexidade do cenário a ser observado, atingir esse objetivo se torna um trabalho árduo, uma vez que a temática e o volume dos dados interfere na maneira de obter ferramentas estratégicas e/ou operacionais.

Independente dos desafios, é indiscutível a utilização de dados que gerem informações a serem usadas para dispor um cenário próximo da realidade. Em paralelo, a criação de indicadores é uma alternativa válida, pois reúnem esses dados em uma única medida, facilitando a compreensão e comunicação dos resultados obtidos. Um outro benefício no uso dos indicadores seria obtê-los periodicamente no intuito de realizar comparações temporais e, dessa forma, verificar as alterações ocorridas em relação ao tempo.

Embora a ideia geral seja mensurar conceitos abstratos, há diferentes indicadores que são escolhidos de acordo com os dados disponíveis e a sua finalidade. O indicador analítico mensura elementos de uma forma objetiva e direta. Isso ocorre devido ao fato de serem unidimensionais e precisarem apenas de uma variável de cálculo. Um exemplo seria contar a quantidade de famílias que possuem crianças de 0 a 6 anos de idade, ou mulheres com renda menor que um salário mínimo. Essas informações se sustentam com apenas uma variável, idade e valor da renda, por exemplo. É claro que na prática há a possibilidade de precisar de uma variável auxiliar a fim de tratar ou deixar nos padrões desejados essas variáveis, mas, conceitualmente falando, elas não dependem de mais de uma dimensão.

Por outro lado, o indicador sintético estima conceitos multidimensionais, ou seja, contempla várias variáveis de cálculo em sua composição. A pobreza é um exemplo de um conceito multidimensional, pois envolve esferas como saúde, renda, educação e outras áreas que podem ser consideradas para constituir-lo. Desse modo, ocorre a condensação dessas várias dimensões para sintetizar, em uma única medida, a pobreza.

Um indicador sintético popularmente conhecido é o Índice de Desenvolvimento Humano (IDH). Esse índice foi criado no início da década de 1990 pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) com o intuito de medir o desenvolvimento populacional por região. As dimensões que o IDH contempla são a saúde (*esperança de vida ao nascer*), educação (*média de anos de estudo e os anos de estudo esperados*) e renda (*PIB per capita*). Elas são utilizadas para definir seus valores e seus dados são coletados a partir do censo populacional e a nível municipal.

A aplicação dos valores obtidos a partir dos indicadores é visualizada na elaboração de ferramentas de gestão e seleção, tais como uma fila de prioridade ou painéis para análises. O Bolsa Capixaba¹ é um exemplo de programa que prioriza seus beneficiários por meio de um indicador sintético.

Entretanto, conseguir esses indicadores para aplicá-los demanda um esforço operacional, tanto por consequência da complexidade de calculá-los quanto pela necessidade de atualizá-los periodicamente. É apropriado, portanto, a utilização de ferramentas computacionais no seu processo de obtenção. Mais precisamente, a elaboração de uma rotina de cálculo fornece uma sequência de etapas pela qual se obtém

¹Programa de transferência de renda realizado pelo Governo Estadual do Espírito Santo. Saiba mais em: setades.es.gov.br/Acessar-o-beneficio-do-Bolsa-Capixaba

os indicadores, garantindo a replicação do cálculo e podendo ser utilizadas por outras equipes ou órgãos, tanto com o objetivo de validar os dados ou para reaplicação em outros contextos.

Considerando os pontos abordados, o propósito deste documento é apresentar a estrutura de uma rotina de cálculo para adquirir o Índice de Desenvolvimento Familiar (IDF). Sendo esta, uma das alternativas para avaliar o quão desenvolvida uma família está no contexto social em que vive. A discussão sobre esse índice será melhor desenvolvida adiante.

2 Índice de Desenvolvimento Social

2.1 O que é, para que serve e suas vantagens

O IDF é um indicador sintético que busca medir o nível de desenvolvimento social que uma família possui em relação a diversas necessidades básicas como, por exemplo, conhecimento, trabalho e situação habitacional. Seu valor varia de 0 a 1 e quanto mais próximo de 0 mais vulnerável a família é.

Há alguns motivos para que o IDF seja escolhido em virtude dos demais indicadores de desenvolvimento social, como o IDH. Um dos primeiros pontos a serem considerados é que o IDF não precisa do censo populacional para ser calculado. A diferença se dá pelo tempo de atualização dos dados, pois a coleta do censo de um país geralmente ocorre a cada dez anos. No caso do IDF utiliza-se uma base que não precisa ter todos os dados da população, apenas do grupo de interesse.

Além disso, o IDF mensura o desenvolvimento social a partir de 6 dimensões: *ausência de vulnerabilidade, acesso ao conhecimento, acesso ao trabalho, disponibilidade de recurso, desenvolvimento infantil e condições habitacionais*. Para efeito de comparação, o IDH apresenta apenas 3 dimensões.

Um outro benefício do IDF é que este possui a família como unidade de análise, dado pelo fato de cada família possuir um código que é atribuído para cada um de seus integrantes. Desse modo, pode-se analisar de forma mais detalhada a sociedade, tanto em nível familiar quanto individual, podendo até obter dados relacionados a grupos sociais e demográficos. Em comparação, o IDH produz apenas informações a nível regional, de forma per capita, dificultando em uma interpretação mais aprofundada e diversificada dos resultados.

Levando em consideração essas características, o IDF demonstrou ser vantajoso em atender a demanda de mensurar o nível de vulnerabilidade das famílias e grupos sociais. Consequentemente, há uma otimização no processo de entendimento do cenário social, a fim de captar problemas existentes e medir o desempenho social ao longo do tempo, sem perder informações úteis, isto é, correspondendo com uma visão mais realista da sociedade.

Os dados utilizados para obter o IDF são registrados no Cadastro Único para Programas Federais. O indicador é calculado por uma série de etapas que envolvem esse banco de dados, mas para conseguir realizar os cálculos é fundamental compreender antes como o índice é estruturado.

2.2 Estrutura do IDF

O índice pretende contemplar as principais dimensões que impactam direta ou indiretamente no desenvolvimento social familiar. São elas:

- **Ausência de vulnerabilidade** procura medir se algum membro ou integrante da família possui características que necessitam de cuidados, como idosos e crianças, ou grupos étnicos como quilombolas, indígenas e imigrantes. A importância dessa dimensão se dá em razão dos custos e estilo de vida afetados por esses atributos, impactando em seu desempenho social.
- **Acesso ao conhecimento** pressupõe mensurar o nível de alfabetização e a escolaridade dos integrantes da família. Tendo em vista que a qualidade de ensino influencia na capacidade de qualificação no mercado de trabalho e dando oportunidade para a família satisfazer as suas necessidades.
- **Acesso ao trabalho** é voltada para as condições que a família possui em relação ao trabalho. Contabilizando a quantidade de pessoas na faixa etária para trabalhar e verificando a qualidade de trabalho dos integrantes.
- **Disponibilidade de recursos** geralmente é a dimensão que é mais associada ao desenvolvimento social da família. De fato ela é essencial para verificar a qualidade de vida de uma família, pois é com os recursos financeiros que possibilita a aquisição de uma boa alimentação, móveis para casa, cestas básicas, acesso a moradias com estrutura física de qualidade, por exemplo.
- **Desenvolvimento infantil** garantir o acesso a educação e saúde além de evitar o trabalho infantil são aspectos que influenciam na qualidade de vida das crianças e adolescentes.
- **Condição habitacional** verifica a situação do ambiente em que a família vive, a estrutura física como a disponibilidade de água, luz, saneamento básico. Também há a análise da quantidade de pessoas que são abrigadas dentro da residência.

Além das dimensões, o IDF possui componentes e indicadores na sua estrutura, elas irão definir melhor cada dimensão. As dimensões, componentes e indicadores são arranjadas de forma hierárquica, como mostrado na figura abaixo:

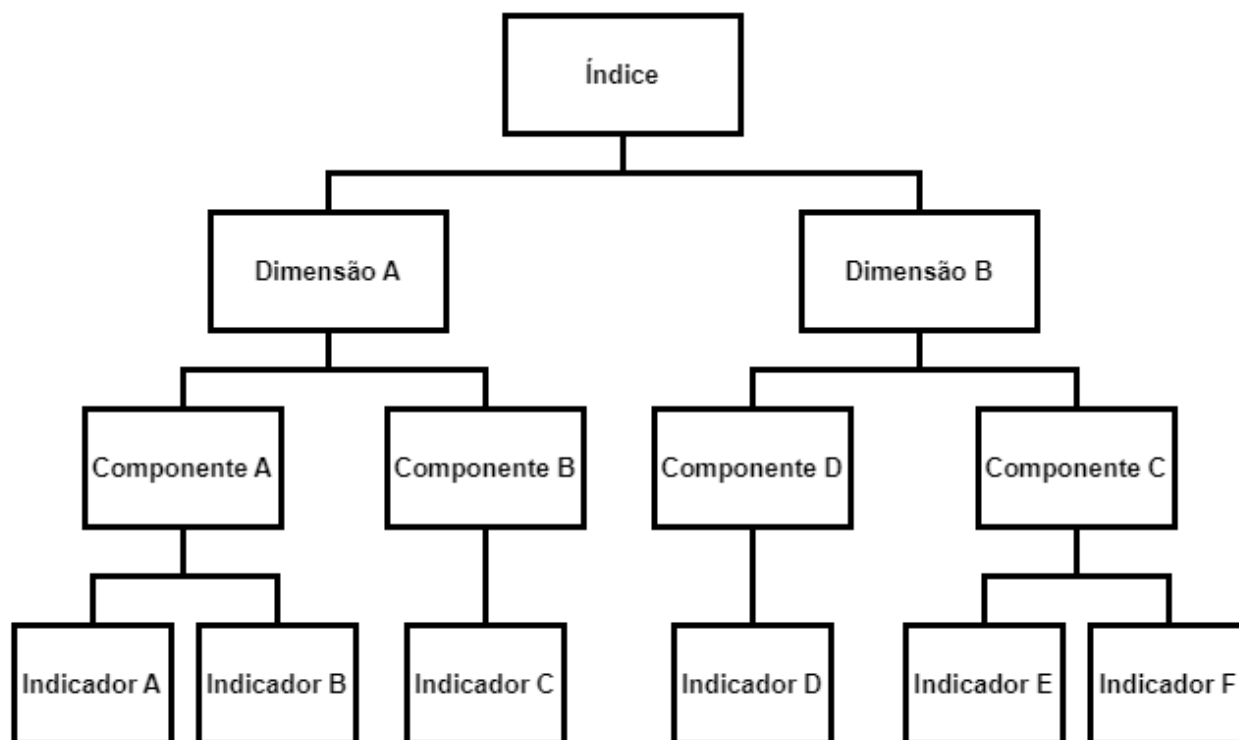


Figure 1: Exemplificação da estruturação do Índice de Desenvolvimento Familiar

A quantidade de indicadores e componentes irão variar de acordo com a sua dimensão. Além disso, o IDF não atribui pesos aos itens que o compõem. Isto é, a condição habitacional de uma família tem o mesmo impacto que a sua disponibilidade de renda, mesmo com a quantidade de componentes e indicadores sendo diferentes para essas dimensões.

Os indicadores são a unidade mínima para conseguir quantificar o IDF. Isso ocorre devido ao fato de atribuirmos um número de acordo com as características do grupo a ser analisado. Tal valor dependerá se o grupo atenda ou não uma condição. Nesta nota técnica, as condições foram seguidas de acordo com metodologia disponibilizada no Perfil da Pobreza do Espírito Santo, publicada pelo Instituto João Santos Neves²(IJSN).

De modo geral, caso a condição seja verdadeira, o grupo receberá a nota **um**, caso contrário receberá **zero**. Para ilustrar, imagine uma família que não possui crianças entre zero a seis anos. Esta família terá o valor um atribuído no indicador que contabiliza a ausência de crianças nessa faixa etária pelo fato de não ser afetada por esse tipo de vulnerabilidade.

Este primeiro momento é crucial para o processo de cálculo do IDF, pois é nele que há o contato direto entre os dados que possuímos e as informações que queremos mensurar.

O cálculo do IDF, das dimensões e dos componentes se dá pela média aritmética das unidades que estão abaixo de si. Por exemplo, os componentes possuem seu valor definido pelos indicadores. Portanto seu

²Confira todas as condições no anexo do Perfil da Pobreza no Espírito Santo. Disponível em: <http://ijsn.es.gov.br/artigos/6074-perfil-da-pobreza-no-espírito-santo-familias-inscritas-no-cadunico-2021>

cálculo se dá pela soma dos valores de seus indicadores dividido pela quantidade de indicadores existentes para aquele componente. O mesmo ocorre para o IDF e para as dimensões.

Ao todo, o IDF possui 6 dimensões, 27 componentes e 65 indicadores. A relação entre eles pode ser apresentada da seguinte maneira:

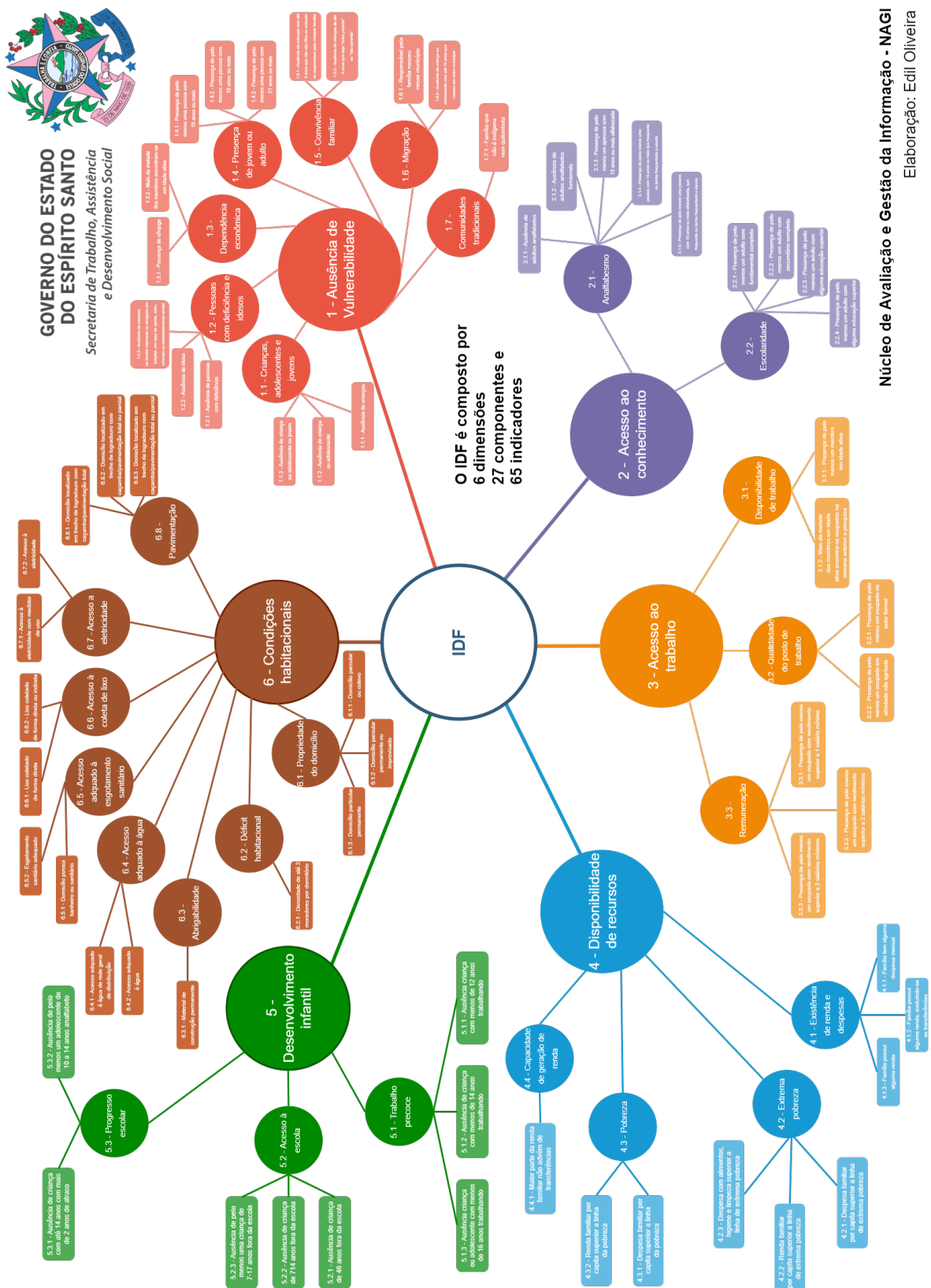


Figure 2: Diagrama de Árvore do Índice de Desenvolvimento Familiar (IDF)

A complexidade do IDF se dá principalmente pelo alto número de variáveis de cálculo que o compõem. Dessa forma, uma solução seria desenvolver uma rotina de cálculo para evitar erros manuais e obter o IDF de forma consistente.

3 Rotina de cálculo para o IDF

Uma rotina de cálculo tem como objetivo desenvolver um conjunto de ações para executar e obter algum objeto, como um arquivo, tabela ou documento. Em geral, sua elaboração utiliza ferramentas tecnológicas para estabelecer e executar as regras de cálculo necessárias para encontrar esses valores. Sendo amplamente utilizados para otimizar a precisão e a agilidade no processo de trabalho.

Sendo assim, desenvolver uma rotina de cálculo para o IDF é vantajoso por conta da alta quantidade de variáveis de cálculo a serem processadas e da necessidade de obter os resultados mais atuais possíveis.

Escolher um programa que seja capaz de efetuar todas as atividades necessárias para o cálculo do índice é o primeiro passo que deve ser feito. Levando em consideração os dados a serem processados, na disponibilidade do software e na flexibilidade de execução de atividades, a escolha da ferramenta a ser utilizada para o desenvolvimento da rotina foi a **linguagem de programação R**. Software capaz de obter, tratar, manipular, analisar e gerar dados. Adequado para atender a demanda em questão.

Além disso, o R é um software livre e possui uma comunidade brasileira bastante ativa, oferecendo suporte e soluções para problemas relacionadas à ferramenta. Sua disponibilização é gratuita podendo ser executada em sistemas operacionais Windows e Linux. A familiaridade que a equipe desenvolvedora desta nota técnica possui com esta linguagem também foi um fator para a sua escolha.

Depois da escolha do programa, a lógica para estruturar a rotina começa estabelecendo os dados de entrada, que irão sofrer uma sequência de manipulações para gerar o IDF. Portanto, deve-se definir bem os dados iniciais.

3.1 Dados de entrada e de saída

Para gerar o índice, foram criadas funções para os indicadores, componentes, dimensões, variáveis auxiliares e o próprio IDF. Cada função dependerá de argumentos que são fundamentais para executar os cálculos necessários para obter os valores desejados. Esses argumentos serão utilizados nas regras de cálculo de cada elemento e por isso devem ser definidos antes do cálculo.

3.1.1 Cadastro Único

A primeira seleção que deve ser feita é em relação às informações do público alvo. Isto é, coletar os dados das pessoas cujo objetivo é ter seu desempenho social mensurado.

Por esse motivo, o IDF utiliza a base de dados do Cadastro Único³, pois é onde há os registros necessários para o cálculo dos indicadores, servindo de base para o cálculo do índice.

Um exemplo é, se quisermos gerar o IDF para as famílias que residem no Espírito Santo, é no Cadastro Único que terá as informações das famílias e de seus integrantes mais vulneráveis nesta região.

Uma característica do Cadastro Único é o fato de possuir o código familiar, sendo usado como número de identificação. Este valor servirá como chave-primária para o processo de análise e manipulação dos dados. No final de todo o processo da geração do IDF, cada família receberá um único valor. Outro ponto é o fato de que cada família possui seu representante, sendo este o encarregado de prestar as informações de toda a família.

3.1.2 Parâmetros da função

Para se calcular o IDF utiliza-se alguns parâmetros que podem variar de acordo com a necessidade do gestor ou com o tempo. Dado que estes valores interferem nas classificações das famílias, optou-se por colocá-los como argumentos da função, de modo que seja facilmente alterado pelo pesquisador. São eles:

- Salário mínimo: importante parâmetro para verificar se o indivíduo recebe uma renda próxima daquela dita como necessariamente mínima para sobreviver. Este valor é definido por lei.
- Linha da Pobreza: valor monetário mensal definido para indicar se a pessoa vive em situação de pobreza. O método de seleção desse valor será definido arbitrariamente por quem irá utilizar o IDF.
- Linha da Extrema pobreza: Similar ao da linha da pobreza, porém a linha da extrema pobreza informa se os indivíduos estão ou não na situação de extrema pobreza de acordo com a sua renda. Uma observação é que a linha de extrema pobreza sempre é menor que a linha da pobreza. Assim como a linha da pobreza, o método de seleção desse valor será definido arbitrariamente por quem irá utilizar o IDF.

Leva-se em consideração que estes parâmetros devem estar alinhados de acordo com a data de referência (período em que os dados foram coletados ou disponibilizados) da base de dados. Já que elas fornecem um panorama de como o sujeito está em relação à sociedade naquele período. Fornecendo uma comparação entre como o indivíduo está no ambiente em que convive.

3.1.3 Variáveis Auxiliares

Há algumas variáveis que não estão no cadastro único, mas que podem ser obtidas por meio de cálculos derivados das informações que nele existem, podendo precisar se relacionar ou não com valores externos.

³O Cadastro Único é um sistema estruturado de acordo com o Decreto 9364 de 2001 e visa reunir as informações das famílias mais vulneráveis do Brasil. Saiba mais em: <https://www.gov.br/pt-br/servicos/inscrever-se-no-cadastro-unico-para-programas-sociais-do-governo-federal>

Um exemplo disso é a necessidade de calcular a idade de cada indivíduo. Para isso, temos a data de nascimento relacionada com a data de referência da base coletada. A partir desses dois valores consegue-se, portanto, achar a idade de cada pessoa cadastrada.

Essas e outras funções são chamadas de variáveis auxiliares e é interessante criar uma etapa de cálculo individual para cada uma. Tanto por conta da sua necessidade de ser utilizada inúmeras vezes durante o processo de execução da rotina, quanto pelo fato do seu cálculo ser tão complexo que o ideal seria separá-la da rotina principal.

3.1.4 INPC e valores de renda deflacionados

No banco de dados do Cadastro Único há o armazenamento de dados sobre a renda de cada indivíduo. No entanto, dependendo da data de coleta dessas informações, as interpretações dos valores podem sofrer distorções devido à inflação que ocorre ao longo dos meses. Já que R\$100,00 há 2 anos não representa igualmente o valor de R\$100,00 nos dias atuais.

Por isso utiliza-se o Índice Nacional de Preços ao Consumidor (INPC), disponibilizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no site do Sistema IBGE de Recuperação Automática (SIDRA). Com os valores do INPC consegue-se calcular os deflatores para modificar os valores de renda e, desse modo, obter a renda deflacionada dos indivíduos, ou seja, os valores correspondentes aos dias atuais ou da data que se deseja.

Esse é um exemplo de uma variável auxiliar. A partir dos valores de renda deflacionados consegue-se obter uma análise mais apurada e realista dos dados disponíveis, sendo fundamental para aplicar na rotina de cálculo do IDF.

3.2 Estrutura do script de cálculo

A rotina principal é realizada logo após estabelecer os dados de entrada e conceituar as variáveis auxiliares. Seria então necessário desenvolver cada cálculo para obter os indicadores, componentes, dimensões e, por fim, o IDF.

Cada elemento que constitui o IDF necessita de um cálculo. Reunir todas essas etapas em um único local resultaria em um script com muitas linhas de código, tornando-o complexo e pouco profissional.

A linguagem de ferramenta R permite gerar *funções*, técnica capaz de salvar e executar sequências de comandos, podendo coletar dados de entrada, processá-los e disponibilizar o resultado de maneira versátil. Essas funções são nomeadas e armazenadas em arquivos do tipo ‘**R**’.

Sua aplicação é vista no momento de compartilhar informações entre arquivos. Já que, mesmo que as funcionalidades fossem definidas em um local específico, elas conseguem ser chamadas e executadas quando e onde necessárias. Esse método, conhecido como *modularização*, possibilita fragmentar o código em vários códigos menores. Impactando na organização da rotina como um todo.

Utilizando dessa técnica para a elaboração da rotina, definimos um módulo para cada objeto (IDF, dimensão, componente, indicador, variáveis auxiliares). Dentro de cada módulo terá uma função que

irá receber os valores de entrada necessários para a geração daquele elemento e retornando uma tabela com seus resultados.

Com a modularização, a obtenção de cálculo é realizada de forma mais independente. Por exemplo, caso haja o interesse em apenas observar as condições habitacionais das famílias, bastaria apenas chamar a funcionalidade dessa dimensão. Informando os dados de entrada necessários para executá-la.

Dessa forma, evita-se a necessidade de ter todos os dados de entrada atribuídos, somente aqueles que serão utilizados pela funcionalidade. Caso a execução do cálculo fosse feita em um único arquivo, provavelmente ocorreria a necessidade de ter em mãos todas as informações necessárias para o cálculo do idf. Tornando complexa e desnecessária a coleta de todos os dados de entrada.

As validações internas para verificar a veracidade e a consistência da rotina do código são facilitadas com a modularização, pois há o mapeamento das funções e de seus comandos de forma separada e organizada.

É importante pontuar que as funcionalidades podem mencionar umas às outras dentro da sua própria rotina. Isso será bastante utilizado devido a relação existente entre cada elemento do IDF, observadas no tópico anterior. Por exemplo, os componentes são calculados pela média aritmética dos indicadores. Desse modo, no script para calcular determinado componente são requisitadas as funcionalidades dos indicadores que servirão de base para calculá-lo. O mesmo ocorre para as outras divisões e também para as variáveis auxiliares, já que elas servem de base para obter diversos elementos do IDF.

O fracionamento da rotina, apesar de ser vantajoso, resulta na criação de vários arquivos e funções. Por esse motivo, elaboramos uma nomenclatura para cada termos existente, a fim de padronizar os arquivos criados, o conteúdo das suas funções e como são gerados os resultados.

A princípio, vamos estruturar um sistema de numeração para os objetos para logo após definir uma padronização dos termos da rotina de cálculo.

Table 1: Sistematização da enumeração dos objetos da Rotina de Cálculo do IDF

Objeto	Sigla nº Objeto	Código	Intervalo nº Objeto	Total de Arquivos em .R
IDF	-	-	-	1
Dimensão	X	X.	1 a 6	6
Componente	Y	X.Y.	1 a 8	27
Indicador	Z	X.Y.Z.	1 a 5	65
Variável Auxiliar	-	-	-	5

O objetivo desse sistema é enumerar os principais elementos do IDF: dimensões, componentes e indicadores. Essa enumeração informará de qual ramificação o elemento faz parte. Isso porque vimos no tópico de estruturação que o IDF é constituído de forma hierárquica. Esse sistema será aplicado tanto nas funcionalidades quanto no nome dos arquivos criados para o armazenamento das funções. Também será aplicado para definir as colunas que serão geradas.

1. Sigla do nº Objeto

É a sigla referente a enumeração do objeto. X representa o número da dimensão, Y representa o número do componente e Z representa o número do indicador. Ela será utilizada para orientar a sintaxe dos termos.

2. Código

A codificação das dimensões, componentes e indicadores se deu pelo artigo realizada pelo Instituto João Santo Neves (IJSN) sobre o Perfil da Pobreza⁴

Essa codificação auxilia na interpretação do objeto a ser calculado e de qual ramificação ele faz parte, já que o IDF possui seus elementos organizados de forma hierárquica.

Exemplos:

- 3: É referente à dimensão 3, relacionada ao Acesso ao trabalho
- 6.4: Mostra o componente 4 da Dimensão 6, Acesso adequado à água
- 3.2: Informa sobre o componente 2 da Dimensão 3, qualidade do posto de trabalho
- 1.4.2: Nos mostra que é um indicador do componente 4 da dimensão 1.
- 5: Dimensão relacionada ao Desenvolvimento infantil.

3. Intervalo nº Objeto

É o intervalo de variação dos valores de X, Y e Z. Como há apenas 6 dimensões, então os valores de X variam de 1 a 6. Os componentes variam de acordo com a dimensão que está relacionada e cada dimensão pode ter quantidade de componentes diferentes. Por exemplo, a dimensão 1 possui 7 componentes, já a dimensão 3 possui apenas 3 componentes. O mesmo ocorre para os indicadores.

4. Total de Arquivos em .R

É a quantidade total de arquivos gerados em ‘R’ de cada objeto, ou seja, a quantidade de módulos existentes. Nota-se que para as quantidades de dimensões, componentes e indicadores coincidem com os valores vistos na figura 2 do tópico anterior.

Com a enumeração especificada, vamos padronizar a nomenclatura dos termos que serão criados, chamados e calculados durante toda a rotina. O objetivo de padronizar os termos é de deixar o acesso, a compreensão e a criação dos arquivos o mais intuitivo possível.

⁴Disponível nas páginas 91 a 96, podendo ser acessado em: <http://ijsn.es.gov.br/artigos/6074-perfil-da-pobreza-no-espirito-santo-familias-inscritas-no-cadunico-2021>.

Table 2: Sintaxes dos termos padronizados da Rotina de Cálculo do IDF

Objeto	Arquivo	Função	Coluna	Pasta
IDF	IDF_calculo	IDF	idf	IDF
Dimensão	DIM0X	DX	dX	DIM0X
Componente	DX_CY	DX_CY	dX_cY	DIM0X
Indicador	DX_CY_indicadorZ	DX_CY_IZ	dX_cY_iZ	DIM0X
Variável Auxiliar	auxiliar_nome	auxiliar_nome	aux_nome	AUXILIARES

1. Arquivo

Elaborar e armazenar os comandos produzidos na linguagem R resulta na geração de arquivos do tipo ‘R’. Cada arquivo conterá um script que irá conter a função para calcular algum objeto.

Por conta disso, foi atribuída uma sintaxe para nomear esses arquivos a fim de especificar o caminho de armazenamento das funcionalidades.

Exemplo:

- *D1_C3*: Arquivo que contém a função para calcular o componente 3 da dimensão 1. Relacionada a dependência econômica da dimensão sobre ausência de vulnerabilidade.

Vale ressaltar que, para as variáveis auxiliares, o ‘nome’ representa o nome do que a variável auxiliar está calculando, não sendo algo específico do IDF. São utilizadas apenas para dar suporte ao cálculo dos elementos do IDF.

Exemplo:

- *auxiliar_idade*: funcionalidade utilizada para calcular as idades de cada indivíduo.

2. Função

Dentro do arquivo ‘R’, nós salvamos um conjunto de comandos que serão executados para gerar valores do elemento. Essa variável terá uma função atribuída a ela e será necessário os dados de entradas fundamentais para a execução do cálculo.

Exemplo:

- *D6_C1*: é a função que calcula o componente 1 da dimensão 6.

3. Coluna

A execução de cada função resultará em uma tabela com a coluna dos valores que foram calculados. Essa coluna terá seu nome relacionado com o objeto que está representando.

Exemplo:

- *d2_c4_i1*: é a coluna gerada pela função *D2_C4_I1* e que contém os valores do indicador 1, do componente 4 da dimensão 2.

4. Pasta

Houve a necessidade de organizar todos os arquivos gerados em ‘R’ em pastas. Já que chamar as funcionalidades dependerá do local do arquivo onde estão armazenadas e há um grande número de arquivos gerados. Cada arquivo está localizado em alguma das 8 pastas abaixo:

- IDF, DIM01, DIM02, DIM03, DIM04, DIM05, DIM06 e AUXILIARES

As pastas *DIM01*, *DIM02*, *DIM03*, *DIM04*, *DIM05* e *DIM06* armazenam os arquivos relacionados aos elementos (componentes e indicadores) daquela dimensão, incluindo o módulo que geram a si próprias. A pasta *AUXILIARES* contém as funções para o cálculo das variáveis auxiliares e outras funções que otimizam o código. Já a pasta *IDF* contém a funcionalidade que resultará no índice de desenvolvimento familiar.

3.3 Elaborando o script

Como mencionado anteriormente, cada elemento do IDF terá um arquivo que conterà as regras de cálculo necessárias para obtê-los. Essas regras de cálculos são chamadas de scripts ou códigos e serão desenvolvidas utilizando como base a linguagem de programação R.

Já foi definida a padronização dos termos referentes a criação da rotina do cálculo do IDF e agora será mostrado como de fato foi organizado essa rotina.

3.3.1 Cabeçalho

Cada script conterà o cabeçalho abaixo, comentado, para identificar o elemento que módulo está sendo responsável pelo cálculo:

```
# NAGI - SETADES / ano
# Responsável: 'nome do responsável pela elaboração do código'

# Cálculo do Índice de Desenvolvimento Familiar
# Dimensão: X. 'nome da dimensão'
# Componente: X.Y. 'nome do componente'
# Indicador: X.Y.Z. 'nome do indicador'
```

Para o caso de variáveis auxiliares:

```
## NAGI - SETADES / ano
## Responsável: 'nome do responsável pela elaboração do código'

# Índice de Desenvolvimento Familiar
# Função Auxiliar: 'nome da função auxiliar'
```

3.3.2 Corpo do código

O corpo do código representa a função que foi definida de acordo com as manipulações necessárias para adquirir determinado elemento do IDF.

```
nome_da_função <- function(dados_de_entrada){

  comandos_e_regras_de_cálculo

  return(saída)

}
```

Com o esquema acima podemos perceber como é realizada a estruturação, no R, para a criação de uma função.

3.3.3 Boas Práticas

Adotamos alguns bons hábitos, pré-estabelecidos, para a criação dos códigos e regras de cálculos:

- Não utilizar ponto nos nomes das funções e nome dos scripts;
- Comentar qual elemento está sendo chamado em cada **source** e com o encoding ‘*UTF-8*’, para evitar desformatação;
- Datas estão no formato ano-mês-dia ‘%Y-%m-%d’;
- Sempre nomear as bases que estão sendo manipuladas de **dado** e no final da manipulação enviar o output da função com o nome ‘saída’;
- Toda a manipulação feita em ‘**data.table**’, pois nesse pacote há funções que otimizam o cálculo de bases com grande volume de arquivos de maneira rápida;
- Comentar as etapas da manipulação com o ‘ctrl + shift + R’.

3.4 Exemplo

Abaixo está um exemplo de um dos scripts:

```
## NAGI - SETADES / 2022
## Responsável: Vitoria Sesana

# Cálculo do Índice de Desenvolvimento Familiar
# Dimensão: 1. Ausência de Vulnerabilidade
# Componente: 1.1. Criança, adolescentes e jovens
# Indicador: 1.1.1. Ausência de criança

D1_C1_I1 <- function(base){

  require(data.table)

  # chamando as colunas necessárias -----

  if (("aux_idade" %in% colnames(base)) == TRUE) {

    dado <- base[, .(d.cod_familiar_fam,
                     aux_idade)]

  } else {

    source("Script/AUXILIARES/auxiliar_idade.R")

    dado <- auxiliar_idade(base)

    dado <- dado[, .(d.cod_familiar_fam,
                     aux_idade)]
  }

  # marca criança 0 a 6 -----

  dado[, marca_idade_0_a_6 := fifelse(aux_idade >= 0 & aux_idade <= 6, 1,
                                       fifelse(is.na(aux_idade), NA_real_, 0))]

  # calculando o indicador -----

  dado <- dado[, .(total_pessoas_0_a_6 =
```

```

        sum(marca_idade_0_a_6, na.rm = TRUE)),
    by = c("d.cod_familiar_fam")]

dado[, d1_c1_i1 := fifelse(total_pessoas_0_a_6 > 0, 0, 1)]

# saída -----

saida <- dado[, .(d.cod_familiar_fam,
                  d1_c1_i1)]

return(saida)
}

```

Podemos perceber que este é uma funcionalidade que gerará um indicador que representa se a família possui ou não uma criança de 0 a 6 anos. Ela é o primeiro indicador do componente 1 da dimensão 1, representando a Ausência de Vulnerabilidade.

Esse indicador necessita da variável auxiliar para calcular idade para ser obtido, devido ao fato de verificar se a família possui ou não crianças de 0 a 6 anos de idade e a base do Cadastro Único não coleta a idade dos indivíduos, mas sim a sua data de nascimento.

4 Considerações Finais

A importância da nota técnica se dá pela preservação do conhecimento produzido e da lógica utilizada para a elaboração da rotina de cálculo. Oferecendo uma base para as pessoas que vierem fazer parte desse trabalho no futuro. Auxiliando, desse modo, em dar continuidade ao trabalho e possibilitar a sugestão de alterações para otimizar os meios de obter o IDF.

Além disso, a nota também orienta aqueles que desejam reproduzir estudos e análises do IDF em outros contextos sociais. A disponibilidade do código e o fato da ferramenta utilizada ser um software livre favorecem para que isso ocorra.

A rotina de cálculo tenta atender a demanda de conseguir, de maneira ágil, periódica e consistente, os valores do IDF. A estrutura da rotina que foi elaborada também torna possível um estudo mais aprofundado desse índice, já que possibilita análises das dimensões, componentes e indicadores, além do próprio índice. Com as informações obtidas, consegue-se ter uma noção do quão desenvolvido um grupo social está. Possibilitando análises comparativas temporais ou regionais e de diferentes grupos sociais.

5 Referências

Perfil da Pobreza no Espírito Santo: Famílias Inscritas no CadÚnico 2021. Site: Instituto João Santo Neves e acessado em fevereiro de 2023. Disponível em: <http://ijsn.es.gov.br/artigos/6074-perfil-da-pobreza-no-espírito-santo-famílias-inscritas-no-cadunico-2021>

Texto para Discussão (TD) 986: O índice de desenvolvimento da família (IDF).

Site: Repositório do conhecimento do IPEA e acessado em fevereiro de 2023. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/2946>

Índice Nacional de Preços ao Consumidor - INPC. Site: SIDRA, acessado em fevereiro de 2023 Disponível em: <https://sidra.ibge.gov.br/pesquisa/snipc/inpc/quadros/brasil/janeiro-2023>

Cadastro Único. Site: Governo Federal e acessado em fevereiro de 2023. Disponível em: <https://www.gov.br/pt-br/servicos/inscrever-se-no-cadastro-unico-para-programas-sociais-do-governo-federal>