**Metodologia.**

**Dados.**

No intuito de avaliar a trajetória dos trabalhadores por conta própria (TCP), foram acessados dados da Pesquisa Nacional por Amostragem de Domicílio Contínua (PNADc), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), entre o primeiro trimestre anos de 2023 e o último de 2024. A PNADc é realizada trimestralmente e acompanha domicílios ao longo de 5 trimestres consecutivos, sendo que todos os períodos possuem tem um grupo amostral iniciando e finalizando a pesquisa. Dessa forma, ao longo de 8 trimestres, 4 grupos amostrais completam a pesquisa (IBGE, 2025).

A variável objeto de análise foi a que descreve a ocupação do entrevistado na semana anterior. A variável é categórica e pode assumir nove alternativas, conforme a coluna “Antes do tratamento” do quadro 1. Para a finalidade da presente análise, foi realizado um tratamento que considerou apenas as seguintes categorias: Empregado do setor privado, Trabalhador doméstico, Trabalhador por conta própria, Empregador, Fora da Força de Trabalho, Desocupado. Ademais, algumas categorias iniciais foram agrupadas (ex.: empregado do setor privado e trabalhador doméstico), conforme ilustrado no quadro 1. Os tratamentos implicaram em uma amostra de 43.527.

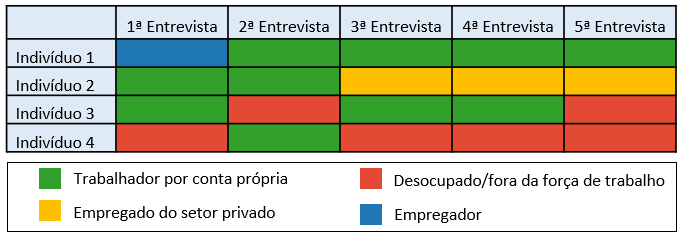
Quadro 1 – Tratamentos de grupos ocupacionais

|  |  |  |
| --- | --- | --- |
| **Antes do tratamento** | **Após o tratamento** | **Sigla** |
| Empregado do setor privado | Empregado do setor privado | ESP |
| Trabalhador doméstico |
| Trabalhador por conta própria | Trabalhador por conta própria | TCP |
| Empregador | Empregador | EMP |
| Fora da força de trabalho | Pessoa fora de trabalho | PFT |
| Desocupado |
| Trabalhador familiar não remunerado | Excluídos | |
| Empregado do setor público (inclusive empresas de economia mista) |
| Militar do exército, da marinha, da aeronáutica, da polícia militar ou do corpo de bombeiros militar |

Fonte: elaborado pelos autores.

**Análise de Sequência**

A técnica de análise empregada foi a análise de sequência. A análise de sequência é tipicamente usada, no campo sociológico, para estudar trajetórias de carreira e cursos de vida e suas transições. A fim de acompanhar estes indivíduos, foi realizada um tratamento para identificar cada entrevistado da base por meio da concatenação das seguintes variáveis: Unidade Primária de Amostragem (UPA), número de seleção do domicílio, grupo de amostra, estrato, dia, mês e ano de nascimento e sexo. A Análise de Sequência é realizada seguindo três etapas (Abbott e Tsay,2000): codificação de narrativas ou processos como sequências, medição de dissimilaridades em pares de sequências e redução de dados. Para a primeira etapa são necessários três elementos: unidades observacionais, que são representadas pelos indivíduos. No exemplo da figura 1, cada linha representa uma unidade observacional. O segundo elemento se chama pontos no tempo. No exemplo da figura 1, cada coluna representa um ponto no tempo. Por fim, o terceiro elemento se chama estado, que é a representação das ocupações definidas no quadro anterior. Na figura 1, as cores representam os estados (Liao et al., 2025).



Aa segunda etapa, medição de dissimilaridades, é realizada para cada par distinto de indivíduos. O objetivo é quantificar o quanto duas sequências são diferentes. Aplicamos a técnica Optimal Matching (OM) com valor 2 para substituições e 1,5 para inserções e exclusões de estados, o cálculo é realizado minimizando o custo de tornar uma sequência igual a outra. Para ilustrar esse processo, considere as seguintes trajetórias:

* Sequência A: TCP, TCP, ESPR, TCP, TCP;
* Sequência B: PFT, TCP, ESPR, ESPR, ESP;
* Sequência C: TCP, TCP, TCP, TCP, TCP.

Os únicos pontos no tempo em que não houve correspondência entre as sequências A e B são a primeira, quarta e quinta entrevista, totalizando um custo igual a 6 para gerar uma equivalência entre as sequências. Para tornar sequências A e C iguais, tem-se um custo igual a 2 e, portanto, são mais similares. O resultado desse processo é uma matriz de dissimilaridades de dimensões n x n, em que o n é o tamanho da amostra. Tanto a codificação, quanto a medição de dissimilaridades são executadas usando a biblioteca TraMineR em linguagem R.

A medição da dissimilaridade é especialmente importante para a última etapa da análise de sequência, denominada redução de dados. Técnicas de clusterização são empregadas com o objetivo de identificar padrões de trajetórias com similaridade (Liao et al., 2025). Primeiro é aplicada a técnica de redução de dimensionalidade com o Escalonamento Multidimensional Clássico (MDS) sobre a matriz de dissimilaridades, reduzindo-a para duas dimensões. E por fim, a clusterização é realizada através do K-means, definindo três clusters. Esta última etapa é executada usando funções nativas do R.