

# Nota Metodológica: Painel de Desigualdades

Esta nota detalha a metodologia utilizada para o cálculo dos indicadores de desigualdade no mercado de trabalho contido no site [FIPE-CDRT](#). Os arquivos e instruções de replicação se encontram no nosso repositório no [Github](#).

## Índices de Equilíbrio Racial e de Gênero

Os Índices de Equilíbrio são calculados baseados na divulgação trimestral mais recente da Pesquisa Nacional por Amostra de Domicílios Contínua e se baseia em [Fripo, França e Rodrigues, 2020](#). Os dois indicadores seguem metodologias muito semelhantes, por isso detalhamos a construção do Índice de Equilíbrio Racial (doravante *IER*) a título de exemplo. O equivalente de gênero segue trivialmente da adaptação desse.

O *IER* é uma medida da distribuição ocupacional por raça relativa à distribuição racial em uma dada população de referência. Note que nossa análise se concentra na comparação entre setores e entre UFs. Assim, para cada setor e UF, o *IER* de uma ocupação  $j$  deste setor é definido como <sup>1</sup>:

$$IER_j(p) = \left( \frac{b_j - p}{p} \right) \left( \frac{p}{1-p} \right)^{b_j}$$

onde  $b_j$  é a proporção de negros na ocupação  $j$  do setor e  $p$  é a proporção de negros na População Economicamente Ativa (pessoas com 14 anos ou mais de idade, ocupadas ou desocupadas) na UF. Para a classificação ocupacional utilizamos Classificação das Ocupações para as Pesquisas Domiciliares -- COD definida para o trabalho principal (variável V4010). Por questões de representatividade estatística, utilizamos a COD a nível de 2 dígitos. A definição de setor também se baseia no trabalho principal e segue a classificação da variável VD4010 da PNADC.

Finalmente, nós agregamos os *IER<sub>j</sub>* a nível setorial pela média ponderada utilizando a massa salarial de cada ocupação dentro de cada setor como peso. Utilizamos o rendimento mensal habitual do trabalho principal (VD4016) para construir a massa salarial das ocupações. Dessa forma, construímos a base de dados final com os *IERs* para cada par setor-UF no Brasil.

## Índice de Gini

O Índice de Gini da renda do trabalho é calculado utilizando as PNADCs Trimestrais dos últimos quatro anos, ou seja, dos últimos dezesseis trimestres. Calculamos o indicador para cada trimestre a nível estadual e nacional a partir da variável VD4019, que denota o rendimento mensal habitual de todos os trabalhos. Utilizamos as funções do pacote `survey` do software `R`.

## Gaps Salariais

Os gaps salariais foram calculados utilizando os microdados da Relação Anual de Informações Sociais (RAIS) de 2019. Consideramos apenas os vínculos ativos em 31/12 <sup>2</sup>. Para cada um dos 645 municípios do estado de São Paulo, estimamos a seguinte equação por Mínimos Quadrados Ordinários:

$$\log(\text{salario/hora})_i = \beta_0 + \beta_1 \cdot \text{Negro}_i + \beta_2 \cdot \text{Mulher}_i + \beta_3 \cdot \text{Idade}_i + \beta_4 \cdot \text{Idade}_i^2 + \beta_5 \cdot \text{Ocupacao}_i + \beta_6 \cdot \text{Setor}_i + u_i$$

onde *salário* é o valor da remuneração nominal em dezembro de 2019, *hora* denota a quantidade de horas contratadas, *Ocupacao* representa o código de ocupação da Classificação Brasileira de Ocupações (CBO, 2002) e *Setor* diz respeito às Classes da Classificação Nacional de Atividades Econômicas (CNAE 2.0).

Nós utilizamos os coeficientes estimados  $\hat{\beta}_1$  e  $\hat{\beta}_2$ , conjuntamente com seus p-valores, para representar as diferenças médias de rendimento entre negros e brancos e mulheres e homens, respectivamente, no mercado formal de cada município paulista. Consideramos estimativas com p-valor maior do que 0.10 como zero. Além disso, devido à existência de efeitos-fixos de alta dimensão relativos a ocupação e setor, não houve graus de liberdade necessários para estimar  $\hat{\beta}_1$  para o município de Lourdes. Nesse caso, Lourdes foi incluído na mesma categoria dos municípios com diferencial salarial positivo (*i.e.*,  $\hat{\beta}_1 > 0$   $\hat{\beta}_2 > 0$ ), representados pela cor cinza no mapa.

As demais categorias foram definidas baseadas no desvio-padrão de cada grupo de estimativas. Seja  $\sigma$  o desvio-padrão da distribuição de algum  $\hat{\beta}_i$ . Os grupos foram definidos como segue:

- Muito baixo:  $(-0.25\sigma, 0]$ .
- Baixo:  $(-0.5\sigma, -0.25\sigma]$ .
- Médio:  $(-\sigma, -0.5\sigma]$ .
- Alto:  $(-2\sigma, -\sigma]$ .
- Muito alto:  $(-\infty, -2\sigma]$ .

---

1. Para detalhes acerca da derivação da expressão do  $IER_j$ , ver [Frippo, França e Rodrigues, 2020](#). ↵

2. Além disso, excluímos as observações com salário ou horas contratadas iguais a 0 ou que não declararam raça ou ocupação. ↵