

Relatório Provisório da Pesquisa: Impactos da Pandemia na Ocupação dos Jovens no Mercado de Trabalho brasileiro

1. Introdução

Este relatório documenta, passo a passo, o desenvolvimento da presente pesquisa, que objetiva, em linhas gerais, mensurar o impacto do choque econômico no contexto da pandemia de COVID-19 sobre a ocupação juvenil no mercado de trabalho. Mais propriamente, dentro do referencial teórico do Modelo de Search & Matching de Diamond-Mortensen-Pissarides (DMP), estimar as probabilidades condicionadas (Cadeias de Markov) de transição entre emprego e desemprego, com foco na integração dos jovens no setor laboral, e complementação com a mensuração de heterogeneidades no impacto dado diferentes características individuais (como sexo, raça, escolaridade, localidade, experiência), mas sobretudo, aplicação do referido modelo para regressão da eficiência de matching para o público-alvo.

Para este fim, utilizou-se microdados da PNAD Contínua Trimestral (2019-2022) e dados de admissões formais disponibilizados pelo CAGED como proxy para o estoque de vagas.

2. Construção do Painel Longitudinal da PNADC

2.1. Obtenção dos Dados

Foram utilizados microdados da PNAD Contínua Trimestral, para os anos de 2019 a 2022, coletados a partir do módulo *PNAD_SOCIAL* do STATA, desenvolvido pelo projeto Data Zoom da PUC Rio, que permite o emparelhamento dos indivíduos consultados pelo IBGE a partir da técnica proposta por Ribas e Soares (2008). Assim, às variáveis padrão são adicionados novos elementos de identificação *hous_id* e *ind_id* para identificação do domicílio e do indivíduo, respectivamente, caso sejam sorteados novamente na nova amostra. Estas informações serão úteis para a observação do mesmo indivíduo em dois ou mais trimestres consecutivos e registro de sua evolução no mercado de trabalho, para posterior geração das matrizes de transição e regressões contra as suas características e os períodos de análise, especialmente o que antecede e o em que se vive a pandemia.

O módulo aplicado às planilhas trimestrais da PNADC, baixadas diretamente do portal do IBGE (<https://www.ibge.gov.br/estatisticas/sociais/trabalho/9171-pesquisa-nacional-por-amostra-de-domicilios-continua-mensal.html?=&t=microdados>), retornou quatro arquivos de

dados .dta para cada ano de análise. Estes foram, em seguida, importados para o R, onde foram unidos em um mesmo banco de dados, via empilhamento, e feito o filtro das variáveis relevantes, conforme o script *importação_filtro.R*, tratado a seguir.

2.2. Filtragem dos Dados

No processo de filtro dos dados da PNADC já limpos e emparelhados pelo módulo anterior, prosseguimos para a seleção apenas das variáveis relevantes para o método desta pesquisa. Seguindo o convencionado pelo IBGE e estudos da OIT sobre mercado de trabalho, consideramos como Jovens apenas indivíduos na faixa etária de 14 a 29 anos de idade, e filtramos, restringindo às observações cuja variável de código *V2009* (que indica a idade relatada na semana de referência) se situassem neste intervalo de idades.

Para identificação do indivíduo, além das variáveis de emparelhamento geradas pela ferramenta do Data Zoom (*ind_id* foi renomeada para *id_pessoa*), e da idade, selecionamos variáveis idiossincráticas que podem ser úteis para análise dos impactos heterogêneos entre os mesmos jovens, como:

- *V2007*: Sexo (se homem, 1; se mulher, 2)
- *V1022*: Situação do domicílio (se urbana, 1; se rural, 2)
- *V2010*: Cor ou raça (branca, 1; preta, 2; parda, 4; e Outras)

Seguindo a literatura que compõem o referencial teórico da pesquisa, consideramos a questão da acumulação de experiência, além da qualificação, na inclusão e manutenção no mercado de trabalho. Como não existe nenhuma forma direta de mensurar esta dimensão do capital humano, consideramos como proxy do que denominamos “experiência” a idade do indivíduo, mas propriamente construímos três faixas de idade no intervalo etário de estudo: Adolescente (14-17), Jovem (18-24) e Jovem adulto (25-29).

Para análise do estado ocupacional, dentro do arcabouço do modelo de Search & Matching, é conveniente trabalhar apenas com jovens dentro da força de trabalho, o que nos levou à seleção das seguintes variáveis ocupacionais:

- *VD4001*: Condição em relação a força de trabalho (na força de trabalho, 1; fora da força de trabalho, 2)
- *VD4002*: Condição de ocupação (ocupada, 1; desocupada, 2)

Assim, criamos uma variável indicadora, *estado_ocupacional*, que assume valor 1 para as observações em que se está dentro da força de trabalho e ocupado, ou 0, para aquelas em que se está desocupado, mas ainda assim está dentro da força de trabalho. Ou seja, observar apenas jovens que estão ativamente no mercado de trabalho, seja trabalhando, seja buscando emprego.

Concluída esta etapa, de definição do conjunto de dados sobre os quais queremos proceder nossas análises dos impactos do choque econômico desencadeado pela pandemia de COVID-19 sobre a transição dos jovens no mercado de trabalho, passamos para a estimação dessas transições.

3. Construção da Matriz de Markov

Para a confecção das matrizes de transição a lá Markov, principiamos o emparelhamento de observações consecutivas do estado ocupacional de um mesmo indivíduo. Para isso, inicialmente ordenamos os dados por código de identificação individual, ano e trimestre, a fim de organizamos observações do mesmo indivíduo uma abaixo da outra. Em seguida, agrupamos todas as observações pelo mesmo id, criamos variável *estado_t1* e armazenamos nela o estado ocupacional da observação posterior. Analiticamente isso significa,

$$E_{t1t} = E_{t+1}$$

Depois, aplicamos um filtro para garantir que estejamos observando apenas registros consecutivos, ou seja, trimestre após trimestre, para que seja possível capturar as transições no espaço de um trimestre.

Ainda, fizemos a definição dos períodos de análise em que estão localizadas as observações, criando uma variável *período*, que assume a categoria “Pré-pandemia” para observações no ano de 2019, “Durante pandemia”, entre 2020 e 2021, e “Pós-pandemia, em 2022.

A partir desse conjunto de informações, foi possível observar as transições, mensurar o número delas, e por fim, chegar à Matriz de Markov, calculando os valores relativos, ou seja, as probabilidades de transição.

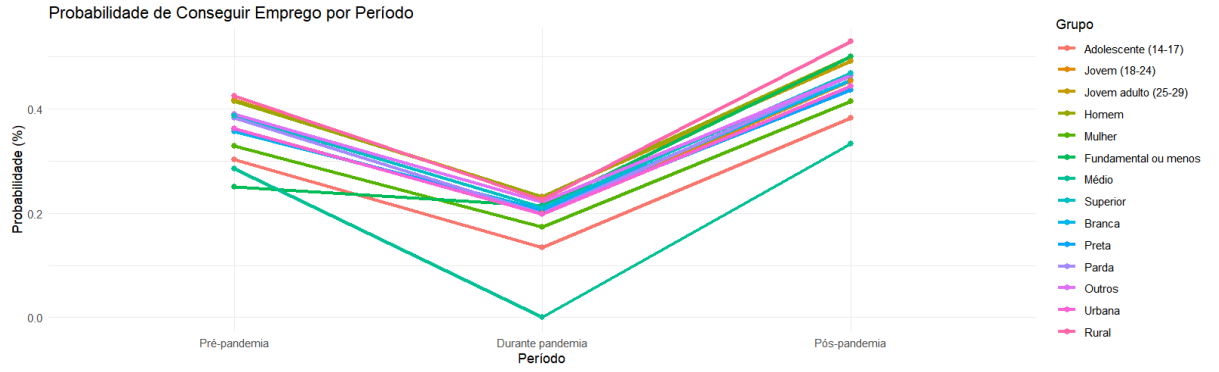
Pré-pandemia			Durante pandemia		Pós-pandemia	
	Desemprego	Emprego	Desemprego	Emprego	Desemprego	Emprego
Desemprego	0,628	0,372	0,798	0,202	0,542	0,458
Emprego	0,068	0,932	0,049	0,951	0,052	0,948

Construímos matrizes de transição para diversos períodos e características dos jovens, e geramos gráficos de barras para comparação, como estes:



As matrizes de transição foram armazenadas em uma planilha (*resultados/Matrizes de Transição.csv*), e permitem-nos responder algumas perguntas, como se: a persistência no desemprego aumentou; a manutenção do emprego caiu; houve retomada no pós-pandemia; se jovens negros e pardos têm menor probabilidade de conseguir emprego após desemprego; jovens com ensino superior têm maior probabilidade de manter emprego; como a idade/experiência influencia a estabilidade ou mobilidade ocupacional; jovens nas áreas urbanas mantêm o emprego com mais frequência; o interior apresenta maior persistência no desemprego; existe recuperação diferente p[ós-pandemia conforme a localidade.

Pode-se construir uma evolução temporal da mesma transição para diferentes grupos, como por exemplo a probabilidade de conseguir emprego mudou para as mulheres e homens entre os períodos.



4. Modelagem Econométrica da Inclusão Juvenil no Mercado de Trabalho

Seguindo a teoria de fricções de mercado de trabalho, segundo a qual o matching não é perfeito e depende de atributos dos trabalhadores, regredimos a probabilidade de um jovem conseguir emprego, condicional às suas características, de forma a capturar os efeitos de: pandemia, escolaridade, sexo, raça, localidade, experiência (idade), e suas interações.

Para isso utilizamos como variável dependente a transição $0 \rightarrow 1$, ou seja, a incorporação ao mercado de trabalho, tal que a especificação do modelo de Regressão logística ficou da seguinte forma:

$$P(Y_{it+1} = 1|X_{it}) = \text{logit}^{-1}(X)_{it}'$$

Onde:

- Y_{it+1} : variável dependente *transição_emprego* = 1 se ocorreu transição de desempregado para empregado.
- X_{it} : vetor de características individuais (sexo, raça, escolaridade, etc.) no tempo t.

A variável binária *mulher* foi criada manualmente, e as outras, foram tratadas como fator e geradas em dummies internas, como por exemplo, *racaPreta*, *racaParda*, *racaOutros* (referência: Branca). Assim, o coeficiente de cada dummy representa o efeito marginal em log-odds daquela categoria em relação à base de referência.

De outra forma, pode-se dizer que, o coeficiente da variável binária *mulher* é o efeito de ser mulher comparado a homem, coeficiente de *periodoDurante pandemia*, efeito da pandemia comparado ao pré-pandemia, e assim por diante.

```
Call:
glm(formula = transicao_emprego ~ mulher + raca + escolaridade +
    localidade + experiencia + periodo, family = binomial, data = microdados_transicao)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.24177	0.41896	-7.738	1.01e-14	***
mulher	0.14207	0.01479	9.603	< 2e-16	***
racaPreta	0.41484	0.02531	16.387	< 2e-16	***
racaParda	0.35210	0.01639	21.484	< 2e-16	***
racaOutros	0.32443	0.07496	4.328	1.51e-05	***
escolaridadeMédio	-0.31671	0.71735	-0.442	0.65885	
escolaridadeSuperior	0.43245	0.41548	1.041	0.29794	
localidadeCapital	0.23352	0.01834	12.732	< 2e-16	***
experienciaJovem (18-24)	-0.16797	0.05389	-3.117	0.00183	**
experienciaJovem adulto (25-29)	-0.61899	0.05441	-11.376	< 2e-16	***
periodoDurante pandemia	-0.45900	0.01671	-27.470	< 2e-16	***
periodoPós-pandemia	-0.08904	0.01875	-4.748	2.05e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 157300 on 380098 degrees of freedom
Residual deviance: 154718 on 380087 degrees of freedom
(128275 observations deleted due to missingness)

Estes resultados nos permitem fazer algumas inferências:

- Raças não-brancas tem consistentemente mais chances de conseguir emprego.
- Mulheres, que historicamente tinham menos chances, estão com 15% a mais que os homens.
- O maior dinamismo urbano permite aos jovens dessas regiões uma vantagem de 26% sobre os das áreas rurais.
- Jovens adultos (25-29) têm grande dificuldade de inserção, e por outro lado, quanto mais novo menor a desvantagem.
- Período da pandemia teve efeito fortemente negativo (-37%), com recuperação parcial no pós (-9%);
- Ter ensino superior ou ao menos o Médio, nesse período, não altera significativamente as chances de transição comparado ao fundamental.

Essas assertivas se alinham bem as expectativas do arcabouço teórica da pesquisa para uma economia com empregos de péssima qualidade, baixa remuneração, em que há altas taxas de rotatividade e informalidade. Por outro lado, no pós-pandemia, a expansão de setores mais abertos à reinserção feminina, justifiquem essa melhora em relação aos homens.

No entanto, é preciso capturar, e esse é um bom desafio de pesquisa, os efeitos heterogêneos da pandemia sobre a transição ocupacional, mais especificamente, perguntar:

"Como o efeito de ser mulher, preto, pardo etc. muda no pós-pandemia em relação ao pré-pandemia?".

Neste caso utilizamos uma nova especificação de regressão logística, em que há interações entre variáveis individuais e o período P , tal qual:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X + \beta_2 P + \beta_3 (X \times P) + \dots$$

Resultado:

```
Call:
glm(formula = transicao_emprego ~ mulher * periodo + raca * periodo +
    escolaridade * periodo + localidade * periodo + experiencia *
    periodo, family = binomial, data = microdados_transicao)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.194151	1.012716	-4.141	3.45e-05	***
mulher	0.156468	0.022998	6.804	1.02e-11	***
periodoDurante pandemia	0.772659	1.174237	0.658	0.51053	
periodoPós-pandemia	1.360137	1.251743	1.087	0.27722	
racaPreta	0.488409	0.039848	12.257	< 2e-16	***
racaParda	0.440727	0.025773	17.100	< 2e-16	***
racaOutros	0.241821	0.125283	1.930	0.05358	.
escolaridadeMédio	1.282138	1.245271	1.030	0.30320	
escolaridadeSuperior	1.298107	1.009366	1.286	0.19842	
localidadeCapital	0.256429	0.028311	9.058	< 2e-16	***
experienciaJovem (18-24)	-0.158676	0.079773	-1.989	0.04669	*
experienciaJovem adulto (25-29)	-0.625733	0.080669	-7.757	8.71e-15	***
mulher:periodoDurante pandemia	-0.065084	0.033991	-1.915	0.05552	.
mulher:periodoPós-pandemia	0.035537	0.038145	0.932	0.35153	
periodoDurante pandemia:racaPreta	-0.121462	0.058684	-2.070	0.03847	*
periodoPós-pandemia:racaPreta	-0.123575	0.064371	-1.920	0.05489	.
periodoDurante pandemia:racaParda	-0.168657	0.037615	-4.484	7.34e-06	***
periodoPós-pandemia:racaParda	-0.121130	0.042399	-2.857	0.00428	**
periodoDurante pandemia:racaOutros	0.088755	0.179200	0.495	0.62040	
periodoPós-pandemia:racaOutros	0.177559	0.185663	0.956	0.33890	
periodoDurante pandemia:escolaridadeMédio	-10.359337	37.355906	-0.277	0.78154	
periodoPós-pandemia:escolaridadeMédio	-1.659437	1.768927	-0.938	0.34819	
periodoDurante pandemia:escolaridadeSuperior	-1.082615	1.167612	-0.927	0.35382	
periodoPós-pandemia:escolaridadeSuperior	-1.308413	1.244458	-1.051	0.29308	
periodoDurante pandemia:localidadeCapital	-0.025127	0.042193	-0.596	0.55149	
periodoPós-pandemia:localidadeCapital	-0.060047	0.047177	-1.273	0.20309	
periodoDurante pandemia:experienciaJovem (18-24)	-0.020861	0.125920	-0.166	0.86842	
periodoPós-pandemia:experienciaJovem (18-24)	-0.007341	0.136404	-0.054	0.95708	
periodoDurante pandemia:experienciaJovem adulto (25-29)	0.057411	0.127066	0.452	0.65140	
periodoPós-pandemia:experienciaJovem adulto (25-29)	-0.056524	0.137875	-0.410	0.68183	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Com estas estimativas pouquíssimas afirmações podemos fazer sobre mudanças desencadeadas pela pandemia, com algum respaldo estatístico. Os grupos mais vulneráveis durante e após a pandemia são:

Grupo	Durante a pandemia	Pós-pandemia
Mulheres	Penalização leve (6%)	Recuperação total

Pardos	Penalização forte (15%)	Penalização persiste (11%)
Pretos	Penalização moderada (11%)	Sem penalização

Uma outra forma de tentar isolar os efeitos da pandemia sobre os diferentes grupos de jovens é adotando um modelo DIF-IN-DIF, ou seja, que isole Diferenças em Diferenças com Heterogeneidade.

```
Call:
glm(formula = transicao_emprego ~ pandemia_fase * (mulher + raca +
  escolaridade + localidade + experiencia), family = binomial,
  data = microdados_transicao)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.194151	1.012716	-4.141	3.45e-05	***
pandemia_faseDurante pandemia	0.772659	1.174237	0.658	0.51053	
pandemia_fasePós-pandemia	1.360137	1.251743	1.087	0.27722	
mulher	0.156468	0.022998	6.804	1.02e-11	***
racaPreta	0.488409	0.039848	12.257	< 2e-16	***
racaParda	0.440727	0.025773	17.100	< 2e-16	***
racaOutros	0.241821	0.125283	1.930	0.05358	.
escolaridadeMédio	1.282138	1.245271	1.030	0.30320	
escolaridadeSuperior	1.298107	1.009366	1.286	0.19842	
localidadeCapital	0.256429	0.028311	9.058	< 2e-16	***
experienciaJovem (18-24)	-0.158676	0.079773	-1.989	0.04669	*
experienciaJovem adulto (25-29)	-0.625733	0.080669	-7.757	8.71e-15	***
pandemia_faseDurante pandemia:mulher	-0.065084	0.033991	-1.915	0.05552	.
pandemia_fasePós-pandemia:mulher	0.035537	0.038145	0.932	0.35153	
pandemia_faseDurante pandemia:racaPreta	-0.121462	0.058684	-2.070	0.03847	*
pandemia_fasePós-pandemia:racaPreta	-0.123575	0.064371	-1.920	0.05489	.
pandemia_faseDurante pandemia:racaParda	-0.168657	0.037615	-4.484	7.34e-06	***
pandemia_fasePós-pandemia:racaParda	-0.121130	0.042399	-2.857	0.00428	**
pandemia_faseDurante pandemia:racaOutros	0.088755	0.179200	0.495	0.62040	
pandemia_fasePós-pandemia:racaOutros	0.177559	0.185663	0.956	0.33890	
pandemia_faseDurante pandemia:escolaridadeMédio	-10.359337	37.355906	-0.277	0.78154	
pandemia_fasePós-pandemia:escolaridadeMédio	-1.659437	1.768927	-0.938	0.34819	
pandemia_faseDurante pandemia:escolaridadeSuperior	-1.082615	1.167612	-0.927	0.35382	
pandemia_fasePós-pandemia:escolaridadeSuperior	-1.308413	1.244458	-1.051	0.29308	
pandemia_faseDurante pandemia:localidadeCapital	-0.025127	0.042193	-0.596	0.55149	
pandemia_fasePós-pandemia:localidadeCapital	-0.060047	0.047177	-1.273	0.20309	
pandemia_faseDurante pandemia:experienciaJovem (18-24)	-0.020861	0.125920	-0.166	0.86842	
pandemia_fasePós-pandemia:experienciaJovem (18-24)	-0.007341	0.136404	-0.054	0.95708	
pandemia_faseDurante pandemia:experienciaJovem adulto (25-29)	0.057411	0.127066	0.452	0.65140	
pandemia_fasePós-pandemia:experienciaJovem adulto (25-29)	-0.056524	0.137875	-0.410	0.68183	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Grupo	Durante Pandemia	Pós-pandemia
Mulheres	Penalizadas marginalmente (~6%)	Recuperadas (sem efeito no pós)
Pardos	Penalizados significativamente (~15%)	Penalização persiste (~11%)
Pretos	Penalizados (~11%)	Recuperação no pós
Outras Raças	Sem efeito claro	Sem efeito claro
Escolaridade	Sem interação significativa	—
Capital	Benefício mantido, sem mudança	—
Jovens (18–29)	Desvantagem consistente, mas sem mudança ao longo dos períodos	

Podemos concluir que, a pandemia penalizou especialmente os jovens pardos e, em menor grau, pretos e mulheres. Apenas as mulheres se recuperaram plenamente no pós-pandemia.

5. Construção da Matriz de Matching

Para estimação do modelo DMP, construímos um painel trimestral agregado por período e grupo, com:

- Estoques:

U : Número de jovens desocupados

E : Número de jovens ocupados

$N = U + E$: força de trabalho jovem

- Transições:

T_{UE} : Número de transições desempregado-empregado

T_{EU} : transições empregado-desempregado

- Indicadores:

$$f_t = T_{UE}/U$$

$$s_t = T_{EU}/E$$

$$u_t = U/(U + E)$$

A primeira versão do modelo estimado utilizou número de jovens empregados como proxy para vagas. Apesar de gerar um coeficiente elevado (elasticidade > 1), trata-se de um caso de endogeneidade estrutural, pois $f_t = T_{UE}/U$ e $\theta_t = E/U$ são construídos com a mesma base.

Também usamos os dados do IPEADATA para extrair o total de ocupados trimestrais no país. Apesar de teoricamente mais plausível, a elasticidade estimada foi negativa e não significativa, provavelmente pela rigidez do estoque de ocupados totais diante de variações específicas entre os jovens.

6. Estimação da Função Matching com os dados do CAGED

Como alternativa utilizamos dados de admissões formais mensais do Novo CAGED, agregados por trimestre, e construímos:

$$\theta_{CAGED,t} = \frac{Admissões\ formais}{U_t}$$

Estimamos então a equação: $\log(f_t) = \beta_0 + \beta_1 \log(\theta_t) + \epsilon_t$. E, chegamos aos seguintes resultados gerais:

- Com proxy interna (E/U): $\beta_1 \approx 1.33, R^2 > 0.6$ (mas enviesado)
- Com total de ocupados: $\beta_1 < 0$, não significativo
- Com CAGED: $\beta_1 = 0.30$, sinal correto, mas $p > 0.3$ (baixo poder estatístico)