



Universidade de Brasília

IE – Departamento de Estatística

Estágio Supervisionado 2

Avaliação de Políticas Públicas Aplicada ao Programa Bolsa Família

Camila Neves Souto

Laís Raiane Miguel Amaral

Relatório do Projeto Final

Orientador: Prof. Donald Matthew Pianto

Brasília

Junho de 2011

Camila Neves Souto
Laís Raiane Miguel Amaral

Avaliação de políticas públicas aplicada ao Programa Bolsa Família

Projeto final apresentado à disciplina Estágio Supervisionado II, do curso de graduação em Estatística, Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília, para conclusão do curso de Estatística, bacharelado.

Orientador: Donald Matthew Pianto

Brasília

2011

Dedicamos aos nossos pais, irmãs, amigos e professores que de alguma maneira contribuíram para a conclusão desse trabalho.

Agradecimentos

Agradecemos primeiramente a Deus, principal responsável pela nossa conclusão desse curso, às nossas famílias pela compreensão, incentivo e inúmeros esforços feitos para tornar possível o nosso desejo e aos amigos que sempre foram solícitos em sanar nossas eventuais dúvidas. Em especial, agradecemos ao professor Donald M. Pianto, pela paciência, persistência, disponibilidade e atenção que nos foi dispensada.

Seria uma grave injustiça pretender dar juízos sobre aquilo por que não se interessou.

Luigi Giussani

Sumário

Introdução	1
1 Análise do banco de dados.....	4
2 Técnicas de avaliação de políticas públicas.....	6
2.1 O Modelo de Causalidade de Rubin (RCM): os possíveis resultados, o mecanismo de atribuição e interações.....	6
2.1.1 Resultados Potenciais.....	7
2.1.2 O mecanismo de atribuição	8
2.1.3 Interações e Efeitos Equilíbrio Geral	10
2.2 Estimadores e Hipóteses	10
2.2.1 Efeito Médio do Tratamento.....	10
2.2.2 Quantis da distribuição	11
2.2.3 Testando.....	12
2.3 Viés de Seleção.....	12
2.4 Experimentos Aleatórios	13
2.5 Outros Métodos de Controle do Viés de Seleção: Seleção em Observáveis	14
2.5.1 Hipóteses de Identificação	15
2.5.2 Matching (Emparelhamento).....	15
2.5.3 Escore de Propensão	16
2.6 Outros Métodos de Controle do Viés de Seleção: Seleção de Não Observáveis	17
2.6.1 Diferença em diferença.....	17
2.6.2 Regressão Descontínua	17
2.7 Análise de sobrevivência.....	20
3 Resultados	21
3.1 Análise descritiva e formação de grupos controle.....	21
3.2 Variáveis de caracterização do domicílio e indivíduo de famílias com renda per capita entre R\$ 50,00 e R\$ 55,00 e entre R\$ 55,00 e R\$ 60,00.	24
3.3 Variáveis de caracterização do domicílio e indivíduo de famílias com renda per capita entre R\$ 90,00 e R\$ 95,00 e entre R\$ 95,00 e R\$ 100,00.....	29
3.4 Avaliação do tempo de emprego (desemprego) formal.....	33
4 Conclusão	39
5 Trabalhos Futuros.....	40
6 Bibliografia	41

Resumo

Políticas públicas tem se tornado instrumentos importantes em toda América Latina em busca da extinção da fome e pobreza nos países subdesenvolvidos. No Brasil, isso é notado, pelos inúmeros programas de transferência de renda existentes, além daqueles que promovem a distribuição de alimentos ou outros utensílios necessários à sobrevivência com dignidade. O programa brasileiro de maior cobertura e eficiência na redução da miséria é o Programa Bolsa Família, que beneficia famílias na linha da pobreza e extrema pobreza, privilegiando aquelas que têm maior número de filhos na infância e adolescência. O objetivo desse estudo foi promover uma avaliação do Programa Bolsa Família a partir da análise do impacto que a participação no programa provoca no tempo de emprego (desemprego) formal. A motivação surgiu das frequentes discussões a respeito do comodismo e desencorajamento em busca de emprego que o programa pode causar nos beneficiários. Sendo assim, os indivíduos foram separados em grupos semelhantes, onde a diferença relevante entre eles era apenas a probabilidade de participação no PBF, a fim de minimizar o efeito que covariáveis poderiam causar no tempo de emprego (desemprego). Para este estudo, não foi encontrada influência mensurável da participação do PBF no tempo de emprego (desemprego) formal dos beneficiários.

Palavras-chave: Bolsa Família, tempo de emprego (desemprego), avaliação, renda

Introdução

O efeito causado por políticas ou programas sociais impacta na economia e em outros aspectos sociais. Por conta disso, pesquisas em análise estatística e econometria são utilizadas para estudar tais efeitos.

O principal problema a ser investigado nesses estudos é a avaliação dos efeitos causados pela exposição das unidades ao programa ou tratamento.

Em estudos econômicos, as unidades são tipicamente agentes econômicos, tais como indivíduos, firmas, estados, municípios ou países. Porém, em outras disciplinas onde métodos de avaliação são aplicados, as unidades podem ser animais, lotes ou terrenos. Podem-se exemplificar os tratamentos como programas educacionais, pesquisas com medicamentos, exposições ambientais, leis ou regulamentos.

Além disso, esses estudos são focados na observação de unidades expostas e não expostas ao tratamento, ou seja, a avaliação é feita a partir da comparação dos dois resultados para uma mesma unidade. Entretanto, aqui temos um problema: na maioria das vezes só conseguimos analisar um dos resultados, pois uma mesma unidade só pode ser exposta a um nível do tratamento.

Como tentativa de solucionar esse problema, separa-se os indivíduos de determinado programa em dois grupos homogêneos: o primeiro é aquele que receberá o tratamento, enquanto o segundo não o receberá. Por exemplo, para efetuar o teste de um determinado medicamento, formamos o grupo que receberá o fator testado e outro, chamado grupo de controle, que não receberá tal fator.

No Brasil, tem-se o Cadastro Único (CadÚnico) para Programas Sociais, que é um instrumento que identifica e caracteriza famílias com renda mensal de até meio salário mínimo por pessoa ou de três salários mínimos no total, controlado pelo Ministério do Desenvolvimento Social e Combate à Fome. Com isso, o Cadastro Único possibilita conhecer a realidade socioeconômica dessas famílias, a partir das informações que caracterizam o domicílio, as formas de acesso a serviços públicos essenciais e cada componente da família.

O CadÚnico é regulamentado pelo Decreto nº 6.135/07 e coordenado pelo Ministério do Desenvolvimento Social e Combate à Fome (MDS) e deve ser obrigatoriamente utilizado para seleção e integração de beneficiários em programas sociais do Governo Federal. Atualmente, o Cadastro Único é composto por mais de 19 milhões de famílias.

Um dos programas sociais que é composto de beneficiários inseridos no CadÚnico é o Programa Bolsa Família (PBF), programa de transferência direta de renda com condicionalidades, implementado desde outubro de 2003, que busca aliviar ou combater a pobreza, no qual são elegíveis famílias em situação de pobreza e de extrema pobreza. Em 2010, O PBF atendia a mais de 12 milhões de famílias, cada família podia receber de R\$ 22,00 a R\$ 200,00, dependendo da renda familiar por pessoa (limitada a R\$ 140,00), da quantidade e da idade dos filhos.

A seguir, nas tabelas 1 e 2, é apresentado como o valor recebido por cada família beneficiária é calculado, onde BVJ significa Benefício Variável Jovem:

Valor do Benefício Recebido por Famílias com Renda Familiar *Per Capita* Mensal até R\$ 70,00 – Brasil – 2010

Número de crianças e adolescentes de até 15 anos	Número de jovens de 16 e 17 anos	Tipo de benefício	Valor do Benefício
0	0	Básico	R\$ 68,00
1	0	Básico + 1 variável	R\$ 90,00
2	0	Básico + 2 variáveis	R\$ 112,00
3	0	Básico + 3 variáveis	R\$ 134,00
0	1	Básico + 1 BVJ	R\$ 101,00
1	1	Básico + 1 variável + 1 BVJ	R\$ 123,00
2	1	Básico + 2 variáveis + 1 BVJ	R\$ 145,00
3	1	Básico + 3 variáveis + 1 BVJ	R\$ 167,00
0	2	Básico + 2 BVJ	R\$ 134,00
1	2	Básico + 1 variável + 2 BVJ	R\$ 156,00
2	2	Básico + 2 variáveis + 2 BVJ	R\$ 178,00
3	2	Básico + 3 variáveis + 2 BVJ	R\$ 200,00

Tabela 1

Valor do Benefício Recebido por Famílias com Renda Familiar *Per Capita* Mensal entre R\$ 70,00 e R\$ 140,00 – Brasil – 2010

Número de crianças e adolescentes de até 15 anos	Número de jovens de 16 e 17 anos	Tipo de benefício	Valor do Benefício
0	0	Não recebe o benefício básico	-
1	0	1 variável	R\$ 22,00
2	0	2 variáveis	R\$ 44,00
3	0	3 variáveis	R\$ 66,00
0	1	1 BVJ	R\$ 33,00
1	1	1 variável + 1 BVJ	R\$ 55,00
2	1	2 variáveis + 1 BVJ	R\$ 77,00
3	1	3 variáveis + 1 BVJ	R\$ 99,00
0	2	2 BVJ	R\$ 66,00
1	2	1 variável + 2 BVJ	R\$ 88,00
2	2	2 variáveis + 2 BVJ	R\$ 110,00
3	2	3 variáveis + 2 BVJ	R\$ 132,00

Tabela 2

O cadastramento não implica na participação imediata das famílias no PBF. Com base nas informações do CadÚnico, o MDS seleciona as famílias que entrarão no Programa a cada mês. O Programa possui três vertentes: transferência de renda, condicionalidades e programas complementares. A transferência de renda é uma forma de alívio direto da pobreza. As condicionalidades garantem que as famílias mantenham seus filhos frequentando a escola e vacinados. Enquanto isso, os programas complementares buscam o desenvolvimento da família de forma que os beneficiários superem a situação de vulnerabilidade.

Tendo em vista o porte do Programa Bolsa Família, o impacto que ele traz à economia de alguns municípios brasileiros e as frequentes discussões a respeito do possível comodismo que o beneficiário pode desenvolver ao participar do PBF, o intuito deste projeto é verificar se existe relação entre a participação no programa e o tempo de emprego/desemprego formal dos beneficiários. Para isso, serão formados dois grupos compostos de indivíduos do Cadastro Único, sendo que ambos os grupos serão homogêneos em relação à renda familiar e outras covariáveis de interesse, mas terão diferentes probabilidades de participar no PBF. A partir desses grupos, será analisado se há alguma relação entre a participação no PBF e o tempo de desemprego dos indivíduos ou se o tempo em que cada indivíduo permanece empregado está relacionado com as demais covariáveis da população, como variáveis próprias de cada família: quantidade

de pessoas, valor da renda *per capita*, entre outros; e variáveis próprias do indivíduo: idade, sexo, etnia, escolaridade, entre outros.

A fim de realizar esse estudo, será utilizado o banco de dados fornecido pelo Ministério do Desenvolvimento Social e Combate à Fome, assim como informações da Relação Anual de Informações Sociais (RAIS), que é um instrumento de coleta de dados, do Ministério do Trabalho e Emprego, instituído pelo Decreto nº 76.900, de 23/12/75, que tem por objetivo:

- o suprimento às necessidades de controle da atividade trabalhista no País,
- o provimento de dados para a elaboração de estatísticas do trabalho,
- a disponibilização de informações do mercado de trabalho às entidades governamentais

Dessa forma, a população em questão será formada por indivíduos do CadÚnico, beneficiários ou não do PBF, que serão observados no âmbito do mercado formal. A base de dados obtida será composta de dados referentes à família de cada indivíduo, dados característicos de cada indivíduo, valor total recebido na participação no PBF e variáveis de emprego, como tempo em que o indivíduo permaneceu empregado/desempregado e a remuneração obtida.

O objetivo geral do projeto é desenvolver um estudo de avaliação de políticas públicas focado no Programa Bolsa Família, analisando os efeitos desse programa na população brasileira de baixa renda. E os objetivos específicos são formar grupos de tratamento e controle com uma mesma realidade sócio-econômica; e verificar a existência de relação entre a participação no PBF e o tempo de emprego formal e possível impacto negativo ou positivo dessa relação na população.

1 Análise do banco de dados

O banco de dados cedido pelo Ministério do Desenvolvimento Social possui variáveis de caracterização dos indivíduos das famílias inseridas no Cadastro Único Federal e dos domicílios aos quais eles pertencem, além das variáveis de participação na RAIS. Como o volume de dados do banco é grande, o mesmo foi disponibilizado dividido em 27 bases menores, uma para cada Unidade Federativa do Brasil (UF). As famílias presentes no banco têm renda familiar *per capita* abaixo de ½ salário mínimo mensal e podem ser selecionadas para participação em programas sociais do Governo Federal. A referência temporal do cadastro é dezembro de 2009, sendo que as variáveis

da RAIS foram incluídas para os anos de 2006, 2007, 2008 e 2009, ou seja, se um indivíduo com cadastro ativo em dezembro de 2009 é empregado no mercado formal nesse ano ou foi empregado em algum dos três anos anteriores, ele possui uma marcação na base, indicando como se deu essa participação.

Temos quatro grupos de variáveis importantes para o estudo: variáveis temporais, variáveis de domicílio, variáveis de indivíduo e variáveis da RAIS. As variáveis temporais indicam quando as famílias entraram no Cadastro Único e quando as beneficiárias entraram no Programa Bolsa Família. As variáveis de domicílio caracterizam as condições de moradia dessas famílias, evidenciando como é o abastecimento de água, tratamento do lixo, escoamento sanitário, tipo de domicílio, entre outros, além de variáveis que mostram a composição e renda familiar, participação ou não da família no Programa Bolsa Família e o quanto recebe pelo benefício. As variáveis de indivíduo trazem informações tais como escolaridade, idade, etnia, participação no mercado de trabalho e estado civil. Por fim, se tem as variáveis da RAIS, que indicam se o indivíduo esteve empregado formalmente, mês a mês, de janeiro de 2006 a janeiro de 2009.

Em posse do banco de dados, iniciou-se o tratamento de algumas variáveis e seleção dos indivíduos que de fato são público do estudo em questão.

O primeiro passo foi selecionar apenas os indivíduos que podem participar do mercado formal, escolhendo, então, aqueles que têm idade superior a 17 anos e inferior a 66 anos.

O segundo passo foi, a partir das variáveis da RAIS, construir mais cinco variáveis que indicam o estado inicial do indivíduo em relação ao emprego (desemprego), ou seja, se em janeiro de 2006 o indivíduo estava empregado ou não, o tempo que permaneceu no estado inicial e o tempo no estado seguinte, e se cada um desses estados é censurado ou não. A censura no primeiro estado ocorre quando ele permanece os 48 meses do estudo num mesmo estado. Já a censura no segundo estado ocorre quando, antes do estudo terminar, ele não passou do segundo para o terceiro estado, não sendo possível calcular o tempo total do segundo estado. Com essas novas variáveis, foram sintetizadas as informações das 48 colunas que indicavam a participação do indivíduo na RAIS mês a mês de 2006 a 2009, o que possibilitou a exclusão dessas colunas da base de dados, tornando a base menor e mais fácil de trabalhar.

O terceiro passo foi calcular quantas pessoas há em cada família e calcular a renda familiar per capita a partir dessa informação e da renda total familiar.

O quarto passo foi selecionar as famílias que entraram no Programa Bolsa Família em data anterior a janeiro de 2006, isto é, apenas aquelas que durante todo o período de análise receberam o benefício, pois a entrada de famílias no período de 2006 a 2009 poderia causar viés no estudo, porque famílias que recebiam o benefício há dois meses, há 40 meses ou durante os quatro anos, seriam tratadas da mesma maneira quanto a variável recebimento do benefício.

O quinto passo foi retirar os indivíduos que estiveram desempregados ou empregados durante os 48 meses, o que apesar de ser uma perda de informação, foi um passo que possibilitou a diminuição do volume de dados, permitindo a execução do trabalho no software estatístico R. No capítulo 6, a falta dessas informações será comentada.

2 Técnicas de avaliação de políticas públicas

Para avaliar o impacto do Programa Bolsa Família, em relação à variável emprego, sobre os beneficiários, será apresentado o modelo de Causalidade de Rubin, estimadores e hipóteses, experimentos aleatórios, hipóteses de identificação: ignorabilidade e sobreposição, seleção sobre observáveis e não-observáveis tomando como base o artigo *Recent developments in the Econometrics of Program Evaluation*, publicado por GUIDO W. Imbens e Jeffrey M. Wooldridge (2009).

2.1 O Modelo de Causalidade de Rubin (RCM): os possíveis resultados, o mecanismo de atribuição e interações

Inicialmente, é necessário entender o conceito de causalidade dentro da avaliação de impacto.

Temos duas definições importantes nesse sentido:

- Causa: quando um resultado deriva da ocorrência de um evento;
- Determinação: quando uma variável influencia o resultado de outra variável.

Formalmente, o efeito causal ou efeito do tratamento é definido como:

$$Y_i(s) - Y_i(s'), \quad s \neq s' \text{ e } s, s' \in S(\text{conjunto finito de tratamentos})$$

A formulação de modelos de causalidade é baseada em análises contra-factuais, onde contra-factuais são possíveis resultados em diferentes estados hipotéticos da natureza. A análise contra-factual busca contrastar resultados em diferentes estados, onde só a ocorrência do evento em questão diferencia estes estados.

O principal problema da Avaliação de Impacto consiste no fato de que cada indivíduo é observado com somente um dos possíveis tratamentos. Além disso, há o problema de seleção, quando somente parte dos indivíduos é observada sob determinado tratamento, pois os resultados só são observados entre os que escolheram ou foram escolhidos para receber o tratamento, e, com isso, o resultado representativo de um grupo tratado pode não ser representativo da população como um todo. Conseqüentemente, sem a aplicação de restrições ou suposições, a construção empírica de um contra-factual é impossível.

Quando são estudados modelos estatísticos de causalidade, geralmente associa-se ao modelo causal de Rubin, que está interessado em estudar o efeito de uma causa e não a causa dos efeitos. Por exemplo: o estudo é focado no efeito que a participação no PBF causa no tempo de emprego (desemprego) do beneficiário e não no estudo das causas que levam o tempo de emprego (desemprego) dos beneficiários ser maior ou menor do que não beneficiários.

Suponha que se deseja analisar um programa de capacitação profissional, onde de uma população de N indivíduos, $i = 1, \dots, N$, alguns são participantes do programa e outros não. Dos indivíduos que não participam, têm aqueles que apenas escolheram não participar e aqueles que não possuem as características necessárias para entrada no programa, ou seja, não são elegíveis.

Para indicar se o indivíduo participa ou não do programa, usa-se o indicador W_i :

$$\begin{cases} W_i = 0, & \text{se o indivíduo não participa do programa} \\ W_i = 1, & \text{se o indivíduo participa do programa} \end{cases}$$

W é usado para denotar o vetor N -dimensional, com o i -ésimo elemento igual a W_i , e N_0 e N_1 para denotar o número de unidades controles e tratadas, respectivamente.

Para cada unidade também é observado um vetor coluna K -dimensional de covariáveis ou variáveis pré-tratadas, X_i , com X denotando a matriz $N \times K$ com a i -ésima linha igual a X_i' .

2.1.1 Resultados Potenciais

O primeiro elemento do RCM é a noção de resultados potenciais. Para cada indivíduo temos dois resultados potenciais possíveis, caso ele participe ou não do programa, pois ele não pode realizar as duas ações. Dessa forma, um será o resultado realizado e o outro contra-factual:

$$Y_i = Y_i(W_i) = Y_i(0)(1 - W_i) + Y_i(1)W_i$$

$$= \begin{cases} Y_i(0), & \text{se } W_i = 0, \\ Y_i(1), & \text{se } W_i = 1. \end{cases}$$

A diferença entre os resultados potenciais $(Y_i(0), Y_i(1))$ e o resultado realizado Y_i é o principal ponto da análise econométrica e estatística moderna dos efeitos do tratamento.

A análise baseada em resultados potenciais apresenta algumas vantagens em relação àquela baseada em resultados realizados. A primeira vantagem está no fato de que os resultados potenciais permitem a definição de resultados causais antes de especificar o mecanismo de atribuição e hipóteses sobre a distribuição. A segunda vantagem consiste no elo que os resultados potenciais mantém entre a análise dos efeitos causais e as manipulações explícitas, pois, considerando os dois resultados potenciais, o pesquisador é forçado a examinar atentamente os tipos de experimentos que podem revelar as causas dos efeitos. Como exemplo, considere os efeitos causados por etnia e gênero nos resultados potenciais. É complicada a interpretação dos resultados pela etnia. Eles seriam resultados da discriminação dos empregados ou seriam resultados das diferenças entre os examinados que provavelmente sofreram discriminação no estágio anterior da vida? A terceira vantagem é que a partir do conceito de resultados potenciais podemos separar o modelo dos resultados potenciais do modelo do mecanismo de atribuição. A quarta vantagem é que esses resultados nos permitem formular hipóteses probabilísticas em termos de variáveis potencialmente observáveis, ao invés de termos não observados. A quinta e última vantagem é que possibilita ver com clareza de onde vem a incerteza nos estimadores, pois até mesmo se observarmos toda a população, os efeitos causais serão incertos porque para cada unidade apenas um dos dois possíveis resultados é observado.

2.1.2 O mecanismo de atribuição

A segunda ferramenta do RCM é o mecanismo de atribuição, que consiste na probabilidade condicional de receber o tratamento, como uma função dos resultados potenciais e das covariáveis observadas. Temos três classes de mecanismos de atribuição.

A primeira classe do mecanismo de atribuição é a de experimentos aleatórios. Em experimentos aleatórios, a probabilidade do mecanismo de atribuição não varia com os resultados potenciais e é uma conhecida função de covariáveis. Na prática, há poucos experimentos na economia com aleatorização.

A segunda classe dos mecanismos de atribuição mantém a restrição de que condicional no valor das características a probabilidade de atribuição não depende dos resultados potenciais, ou

$$W_i \parallel (Y_i(0), Y_i(1)) | X_i$$

onde, $A \parallel B | C$ denota a independência condicional de A e B dado C. Entretanto, a probabilidade de atribuição não é mais assumida como uma conhecida função de covariáveis. De acordo com Rubin (1990), refere-se a esse mecanismo de atribuição como hipótese de não-confundimento.

A terceira classe dos mecanismos de atribuição contém todos os mecanismos de atribuição remanescentes, com alguma dependência dos resultados potenciais.

2.1.3 Interações e Efeitos Equilíbrio Geral

Na maior parte dessa literatura, é assumido que o tratamento recebido por uma unidade não afeta o resultado de outra unidade. Consideramos que somente o nível do tratamento aplicado ao indivíduo afeta o seu resultado. Na literatura estatística isso é conhecido como a Suposição de Valor Estável da Unidade do Tratamento (SUTVA, Stable-Unit-Treatment-Value-Assumption - Rubin (1978)). Neste projeto, as definições serão vistas sob esse pressuposto.

Essa suposição pode ser violada quando existem outras versões não representadas de tratamento ou quando há interação entre os indivíduos tratados e controles. Entretanto, desde que o efeito indireto causado a uma unidade devido à exposição a algumas unidades que receberam o tratamento seja muito menor que o efeito do tratamento sobre a primeira unidade, a interação não é um grande problema.

Como tentativa de solucionar o problema da interação, existem duas alternativas: redefinir a unidade de interesse ou modelar diretamente as interações.

2.2 Estimadores e Hipóteses

Na literatura atual sobre métodos de avaliação de políticas públicas, se tornou relevante a acomodação da heterogeneidade nos efeitos dos tratamentos, divergindo dos estudos iniciais que consideravam que os efeitos dos tratamentos eram constantes, o que implicava que o efeito de várias políticas pudesse ser explicado por um único parâmetro.

Devido à heterogeneidade nos efeitos dos tratamentos, na literatura atual não é possível resumir os efeitos de todas as políticas de interesse em termos de poucas medidas estatísticas.

Pesquisadores, no dia-a-dia, estudam estimativas dos efeitos em poucas políticas focalizadas. Estimativas de efeito médio são frequentemente utilizadas, seja de toda população ou de alguma subpopulação.

2.2.1 Efeito Médio do Tratamento

A literatura econométrica tem se focado em efeitos médios do tratamento. Os dois efeitos médios mais proeminentes são definidos sobre populações adjacentes. Um popular parâmetro a ser estimado do efeito médio dos tratamentos é o Efeito Médio do Tratamento Populacional (*Population Average Treatment Effect* – PATE), que consiste na esperança do efeito causal dos dois níveis possíveis para a mesma unidade:

$$\tau_{\text{PATE}} = E[Y_i(1) - Y_i(0)].$$

A quantidade acima é relevante quando a política sob consideração pode expor todas as unidades da população ao tratamento ou nenhuma delas.

Outro parâmetro popular é o Efeito Médio do Tratamento Populacional sobre os Tratados (*Population Average effect on the Treated – PATT*), que é o efeito médio sobre a subpopulação de unidades tratadas:

$$\tau_{\text{PATT}} = E[Y_i(1) - Y_i(0) | W_i = 1].$$

Em muitos estudos observacionais, o τ_{PATT} é uma estimativa mais interessante do que a estimativa geral da média, τ_{PATE} , como em casos onde não faz sentido calcular o possível resultado mesmo com o indivíduo inscrito no programa ou quando quem não está inscrito nunca poderá participar do programa.

Podemos também primeiramente focar no efeito médio causal condicionado a covariáveis da amostra:

$$\tau_{\text{CATE}} = \frac{1}{N} \sum_{i=1}^N E[Y_i(1) - Y_i(0) | X_i],$$

E, similarmente, a média sobre a subamostra de unidades tratadas:

$$\tau_{\text{CATT}} = \frac{1}{N_1} \sum_{i|W_i=1}^N E[Y_i(1) - Y_i(0) | X_i].$$

Se o efeito do tratamento for constante, todas as estimativas serão iguais. Entretanto, se há heterogeneidade no efeito dos tratamentos, as diferenças entre τ_{PATE} e τ_{CATE} , τ_{PATT} e τ_{CATT} serão sutis.

2.2.2 Quantis da distribuição

Uma classe alternativa da estimativa de parâmetros consiste nos quantis dos efeitos dos tratamentos, introduzido por Kjell Doksum (1974) e Erich L. Lehmann (1974) é definido:

$$\tau_q = F_{Y(1)}^{-1}(q) - F_{Y(0)}^{-1}(q),$$

como o efeito do tratamento no q -ésimo quantil. Primeiramente, note que esses efeitos nos quantis são definidos como diferenças entre quantis de duas distribuições marginais dos resultados potenciais e não como quantil do efeito do nível da unidade.

$$\tilde{\tau}_q = F_{Y(1)-Y(0)}^{-1}(q).$$

Em geral, o quantil da diferença difere da diferença entre quantis a menos que exista uma perfeita correlação entre resultados potenciais $Y_i(0)$ e $Y_i(1)$.

2.2.3 Testando

A bibliografia em teste de hipóteses em programa de avaliação é limitada. A maior parte dos testes aplicados é focada na hipótese nula de que o efeito médio de interesse é zero. Porque muitos estimadores comumente utilizados para o efeito médio dos tratamentos são assintoticamente normalmente distribuídos com viés assintótico zero, disso segue que os intervalos de confiança padrões podem ser usados para testar tais hipóteses.

Uma questão de interesse é se há algum efeito do programa, ou seja, se a distribuição de $Y_i(1)$ difere da de $Y_i(0)$. Isso é equivalente não só a média, mas todos os momentos são iguais nos dois tratamentos.

Um segundo conjunto de questões é a preocupação sobre a heterogeneidade dos efeitos dos tratamentos. Mesmo que o efeito da média seja zero, pode ser importante estabelecer se há implementação de uma intervenção somente com aqueles que esperam ser beneficiados, o que pode melhorar a média dos resultados.

2.3 Viés de Seleção

Quando os indivíduos são separados em dois grupos, um que recebe o tratamento e o outro de controle, a diferença entre os resultados para ambos, que é o interesse do pesquisador, pode ser atribuída tanto ao tratamento, como às diferenças pré-existentes (viés de seleção), que existem mesmo quando se toma o cuidado de criar grupos com características semelhantes.

Suponha um estudo que consiste na mensuração do impacto do uso de livros de literatura no aprendizado das escolas. Seja $Y_i(1)$ o escore médio do teste das crianças da escola i , caso faça uso de livros de literatura e $Y_i(0)$ o resultado do teste das crianças da mesma escola i , caso não faça uso de livros de literatura. Depois, defina Y_i como o resultado que de fato ocorreu na escola i . O interesse está na diferença $Y_i(1) - Y_i(0)$, que é o efeito do uso de livros de literatura na escola i .

Como não é possível observar os dois resultados para uma mesma escola, será observado o efeito médio esperado que o uso de livros de literatura provoca numa

população de escolas: $E[Y_i(1) - Y_i(0)]$.

Em uma larga amostra dessa população, essa diferença consistirá em

$$\begin{aligned} D &= E[Y_i(1)|\text{escola usa livros de literatura}] - \\ &- E[Y_i(0)|\text{escola não usa livros de literatura}] = \\ &= E[Y_i(1)|T] - E[Y_i(0)|C]. \end{aligned}$$

Subtraindo e adicionando $E[Y_i(0)|T]$, isto é, o resultado esperado sob não tratamento de uma escola do grupo dos tratados, uma quantidade que não pode ser observada, mas que pode ser definida, obtém-se:

$$D = E[Y_i(1)|T] - E[Y_i(0)|T] - E[Y_i(0)|C] + E[Y_i(0)|T]$$

$$D = E[Y_i(1) - Y_i(0)|T] + E[Y_i(0)|T] - E[Y_i(0)|C]$$

O primeiro termo, $E[Y_i(1) - Y_i(0)|T]$, é o efeito do tratamento que se deseja isolar. Para o exemplo dado, esse termo significa a diferença que os livros-texto têm, em média, no tratamento das escolas.

O segundo termo, $E[Y_i(0)|T] - E[Y_i(0)|C]$ é o viés de seleção. Ele capta a diferença dos resultados potenciais dos indivíduos não tratados entre as escolas tratadas e de comparação; escolas de tratamento podem apresentar diferentes escores médios mesmo se não tiverem sido tratadas. Isso acontece, por exemplo, quando as escolas que recebem livros de literatura são escolas onde os pais consideram educação uma prioridade, pois é natural que os filhos sejam mais encorajados a estudar e tenham resultados melhores. Nesse caso, $E[Y_i(0)|T]$ seria maior que $E[Y_i(0)|C]$. O viés também pode seguir por outra direção. Por exemplo, caso os livros de literatura sejam encaminhados para escolas em comunidades particularmente desfavorecidas, $E[Y_i(0)|T]$ será menor do que $E[Y_i(0)|C]$. Em geral, além de qualquer efeito dos livros didáticos, pode haver diferenças sistemáticas entre escolas com livros e sem.

Como $E[Y_i(0)|T]$ não é observado, é em geral impossível saber a magnitude do viés de seleção e em qual direção ele explica a diferença dos resultados dos grupos tratados e de comparação. Muitos estudos empíricos são focados em estudar situações onde pode ser assumida a inexistência do viés ou uma maneira de corrigi-lo.

2.4 Experimentos Aleatórios

Segundo Wooldridge (2009), uma forma de remover totalmente o viés de seleção é quando os indivíduos ou grupos de indivíduos são aleatoriamente atribuídos aos grupos de tratamento e controle. Em uma avaliação aleatorizada, uma amostra de N indivíduos é selecionada da população de interesse. Note que essa “população” não precisa ser uma amostra aleatória da população inteira, mas pode ser selecionada de acordo com os observáveis; portanto, será estudado o efeito do tratamento sobre uma particular subpopulação sobre a qual a amostra foi retirada. Essa amostra experimental é então dividida nos grupos de tratamento, N_T , e controle N_C .

O grupo de tratamento é exposto ao programa e o outro não e, então, o resultado Y é comparado para ambos os grupos. O efeito médio do tratamento pode ser estimado como a diferença entre as médias de Y entre os dois grupos,

$$\hat{D} = \hat{E}[Y_i|T] - \hat{E}[Y_i|C],$$

como o tratamento foi atribuído aleatoriamente, os grupos controle e tratamento diferem apenas quanto à exposição ao programa ou política pública. Se os dois grupos não tivessem recebido o tratamento, seus resultados seriam em média o mesmo, o que implica que o viés de seleção, $E[Y_i(0)|T] - E[Y_i(0)|C]$, é igual a zero. Se, em adição, o resultado potencial de um indivíduo não é relacionado com o de outro, isto é, se o SUTVA é obedecido, tem-se

$$E[Y_i|T] - E[Y_i|C] = E[Y_i(1) - Y_i(0)|T] = E[Y_i(1) - Y_i(0)],$$

o parâmetro causal de interesse do tratamento T .

Em contrapartida, a regressão para obter \hat{D} é

$$Y_i = \alpha + \beta T + \varepsilon_i,$$

onde T é uma variável binária que indica a atribuição ao grupo de tratamento. Essa última equação pode ser estimada pelo método dos mínimos quadrados ordinários e facilmente pode ser provado que $\hat{\beta}_{OLS} = \hat{E}(Y_i|T) - \hat{E}(Y_i|C)$, o que mostra que quando a randomização é corretamente implementada, fornece uma estimativa não viesada do impacto do programa na amostra sob estudo.

Embora esse procedimento seja o mais adequado do ponto de vista inferencial, nem sempre é possível sua implementação em estudos de avaliação de políticas públicas, e mesmo nas situações onde é possível ser aplicada, a aleatorização encontra problemas, devido à dificuldade de formar grupos de controle totalmente aleatórios. Na maioria das situações práticas a alocação das unidades aos grupos de tratamento e controle é feito por escolha ou indicação, persistindo o problema da auto-seleção, que implica em endogeneidade ou dependência entre a variável de seleção e os resultados potenciais, que resulta em em estimadores viesados.

2.5 Outros Métodos de Controle do Viés de Seleção: Seleção em Observáveis

Tendo em vista a dificuldade prática que algumas vezes ocorre na implementação da randomização, outros métodos podem ser utilizados para resolver o problema do viés de seleção. O objetivo de todos esses métodos é a formação de grupos de comparação sob as hipóteses de identificação. Essas hipóteses não são diretamente testáveis, e a validade de qualquer estudo depende do quão convincente são as hipóteses.

2.5.1 Hipóteses de Identificação

Hipótese 1 (Ignorabilidade)

O pressuposto fundamental é a ignorabilidade, introduzida por Rosenbaum e Rubin (1983),

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) | X_i.$$

Esta hipótese diz que condicional em X_i , a alocação dos grupos de tratamento e controle é feita de forma aleatória, o que implica que $\{Y_i(0), Y_i(1)\}$ e W_i são independentes.

Esse pressuposto é frequentemente controverso à medida que assume que além das covariáveis X_i observadas não há características dos indivíduos associadas tanto com os resultados potenciais quanto com o tratamento. No entanto, é utilizado rotineiramente em análise de regressão múltipla. De fato, suponhamos que o efeito que o tratamento assume, τ , seja constante para cada sorteio i , $\tau = Y_i(1) - Y_i(0)$. Depois, assumamos que $Y_i(0) = \alpha + \beta'X_i + \varepsilon_i$, onde $\varepsilon_i = Y_i(0) - E[Y_i(0)|X_i]$ é o resíduo dos não observáveis que afetam a resposta na ausência do tratamento.

Após, com o resultado observado definido como $Y_i = (1 - W_i)Y_i(0) + W_iY_i(1)$, pode-se escrever:

$$Y_i = \alpha + \tau W_i + \beta'X_i + \varepsilon_i,$$

e a ignorabilidade será equivalente a independência de ε_i e de W_i , condicional a X_i .

Hipótese 2 (Sobreposição)

$$0 < pr(W_i = 1 | X_i = x) < 1, \text{ para todo } x.$$

À medida que a probabilidade acima implica que o suporte da distribuição condicional de X_i , dado $W_i = 1$, se sobrepõe completamente à distribuição condicional de X_i , dado $W_i = 0$, ela é chamada de hipótese de sobreposição.

Com uma amostra aleatória, $(W_i, X_i)_{i=1}^N$, é possível estimar o escore de propensão $e(x) = pr(W_i = 1 | X_i = x)$ que pode nos ajudar a determinar se a hipótese de sobreposição existe. É necessário, porém, ter cautela, pois modelos paramétricos comuns, como probit e logit, asseguram que todas as probabilidades estimadas estão estritamente entre 0 e 1, então a análise das probabilidades ajustadas por esses métodos pode ser prejudicada.

2.5.2 Matching (Emparelhamento)

Para estimar os resultados contra-factuais utiliza-se o emparelhamento, uma maneira de eliminar o viés de seleção, a noção deste método advém do fato de que, em

cada célula os grupos de tratamento e controle são homogêneos em relação à X , pois estes, por construção, possuem os mesmos valores das observáveis. Este emparelhamento é a base para a estimativa do contra-factual, pois nestas condições é plausível estimar o resultado contra-factual do grupo de tratamento pelo resultado do grupo de controle. É interessante notar que o emparelhamento só é possível em situações onde em cada nível da variável de X a probabilidade de observar não tratados seja diferente de zero. Esta é conhecida como condição de suporte comum.

2.5.3 Escore de Propensão

Escore de propensão é a probabilidade de um indivíduo participar do programa dado as variáveis de pré-tratamento X , sendo representada por $p(X_i = x)$, ou seja,

$$p(X_i = x) = P(W_i = 1 | X_i = x).$$

Rosenbaum e Rubin (1983) mostraram que o escore de propensão possui duas propriedades importantes:

- *Propriedade de balanceamento das variáveis pré-tratamento dado o escore de propensão:* condicional ao escore de propensão, a variável de seleção W e o vetor X são independentes, ou seja, $W \parallel X \mid p(X_i)$. Na prática $p(X_i)$ tem que ser estimado, e assim esta propriedade deve ser verificada para garantir o sucesso do emparelhamento por escore de propensão. Se esta propriedade for obedecida para um certo conjunto de dados então podemos concluir que para as unidades de controle e de tratamento que possuem o mesmo escore de propensão a distribuição de X é idêntica

- *Propriedade de não confundimento ou ignorabilidade condicional ao escore de propensão:* assumindo que a propriedade de ignorabilidade é válida condicional às observáveis X , então segue que esta propriedade é também válida, condicionando-se no escore de propensão. Ou seja, a variável de resultado é independente da variável de seleção, condicional ao escore de propensão. Formalmente : $\{Y(1), Y(0)\} \parallel W \mid p(X)$.

A metodologia de emparelhamento (*matching*) é baseada na escolha ideal de um grupo de controle a partir de uma amostra maior que a do grupo de tratamento. O grupo de controle é pareado com o de tratamento a partir das variáveis observadas ou pelo escore de propensão. Quanto mais próximos os escores, melhor o emparelhamento. Conforme Ravallion (2000), um bom grupo de controle vem do mesmo ambiente econômico, social e cultural que o grupo de tratamento.

2.6 Outros Métodos de Controle do Viés de Seleção: Seleção de Não Observáveis

Ao contrário da seção anterior onde as hipóteses de identificação eram bem definidas, cada método apresentado a seguir tem a sua particularidade, descartando total ou parcialmente a hipótese de ignorabilidade. São eles: diferença em diferença e regressão descontínua.

2.6.1 Diferença em diferença

Para a utilização do método diferença em diferença (*Difference in Difference* – DID) é necessário obter informações sobre o resultado em análise para os dois grupos, em dois períodos, antes e depois da implementação do programa. A ideia deste método é comparar os resultados dos beneficiários depois do programa com os seus resultados antes do programa, subtraindo de cada um deles o resultado dos não beneficiários depois e antes do programa, respectivamente. Com esse estimador, sob a hipótese de covariáveis não observáveis não variam de acordo com o tempo, o problema do viés de seleção é resolvido. Este método foi utilizado em Heckman, Ichimura, Smith e Todd (1988) e Ichimura e Todd (1997, 1998). É essencial verificar a homogeneidade das características dos grupos na situação antes do programa. Neste contexto longitudinal o estimador do impacto do programa será dado pelo DID:

$$\begin{aligned} DID &= E[Y_i(1)] - E[Y_i(0)] \\ &= (E[Y_i|G_i = 1, T_i = 1] - E[Y_i|G_i = 1, T_i = 0]) - \\ &\quad - (E[Y_i|G_i = 0, T_i = 1] - E[Y_i|G_i = 0, T_i = 0]) \end{aligned}$$

sendo $G_i \in \{0,1\}$, onde grupo 1 é o grupo tratado que é observado em um período $T_i \in \{0,1\}$.

2.6.2 Regressão Descontínua

Um caso interessante de controle de alguma variável ocorre quando a probabilidade de atribuição de um grupo de tratamento segue uma função descontínua de uma ou mais variáveis observáveis. Por exemplo, uma organização de microcrédito pode limitar elegibilidade a empréstimos a mulheres que moram em casas com menos de um acre de terra; estudantes que passam em provas se a nota for no mínimo 50%; ou tamanho de turma que não é permitido exceder mais de 25 estudantes. Se o impacto de alguma variável não observável correlacionada com a variável usada na atribuição do tratamento é suave, a seguinte suposição é razoável para um ε pequeno:

$$E[Y_1^C | T, X < X^- + \varepsilon, X > X^- - \varepsilon] = E[Y_1^C | C, X < \bar{X} + \varepsilon, X > \bar{X} - \varepsilon],$$

onde X é a variável subjacente e \bar{X} é o limite de atribuição. Essa suposição implica que sob um intervalo ε de \bar{X} , o viés de seleção é zero e é a base do “delineamento da estimativa da regressão descontínua”.

A ideia da regressão descontínua é estimar o efeito do tratamento usando indivíduos pouco abaixo do limite de atribuição como um controle para os que estão pouco acima.

Este delineamento tem se tornado popular entre pesquisadores que trabalham em programas de avaliação em países desenvolvidos e afirmam que esta regressão elimina o viés se as regras de atribuição são realmente implementadas.

A regressão descontínua é dividida em duas configurações gerais, a *Sharp e a Fuzzy Regression Discontinuity Design* (SRD e FRD).

2.6.2.1 *Sharp Regression Discontinuity Design*

Na regressão Sharp (afiado) a probabilidade de atribuição assume valor igual à zero ou um. Um se o valor da covariável for no mínimo igual a uma constante c , então este indivíduo está no grupo de tratamento (e a participação é obrigatória a esses indivíduos). E zero se todas as unidades com covariáveis têm valores inferiores a c , então estes indivíduos estão no grupo de controle (membros deste grupo não são elegíveis para o tratamento).

2.6.2.2 *Fuzzy Regression Discontinuity Design*

Na regressão Fuzzy (embaçado) a probabilidade de receber o tratamento não pula de zero para um no limite c . No lugar disso, ela só precisa de uma descontinuidade na probabilidade de atribuição do tratamento no limite c :

$$\lim_{x \downarrow c} P(W_i = 1 | X_i = x) \neq \lim_{x \uparrow c} P(W_i = 1 | X_i = x)$$

2.6.2.3 Ilustração

A seguir, a partir de dados fictícios, foram gerados gráficos para ilustrar a técnica de regressão linear descontínua.

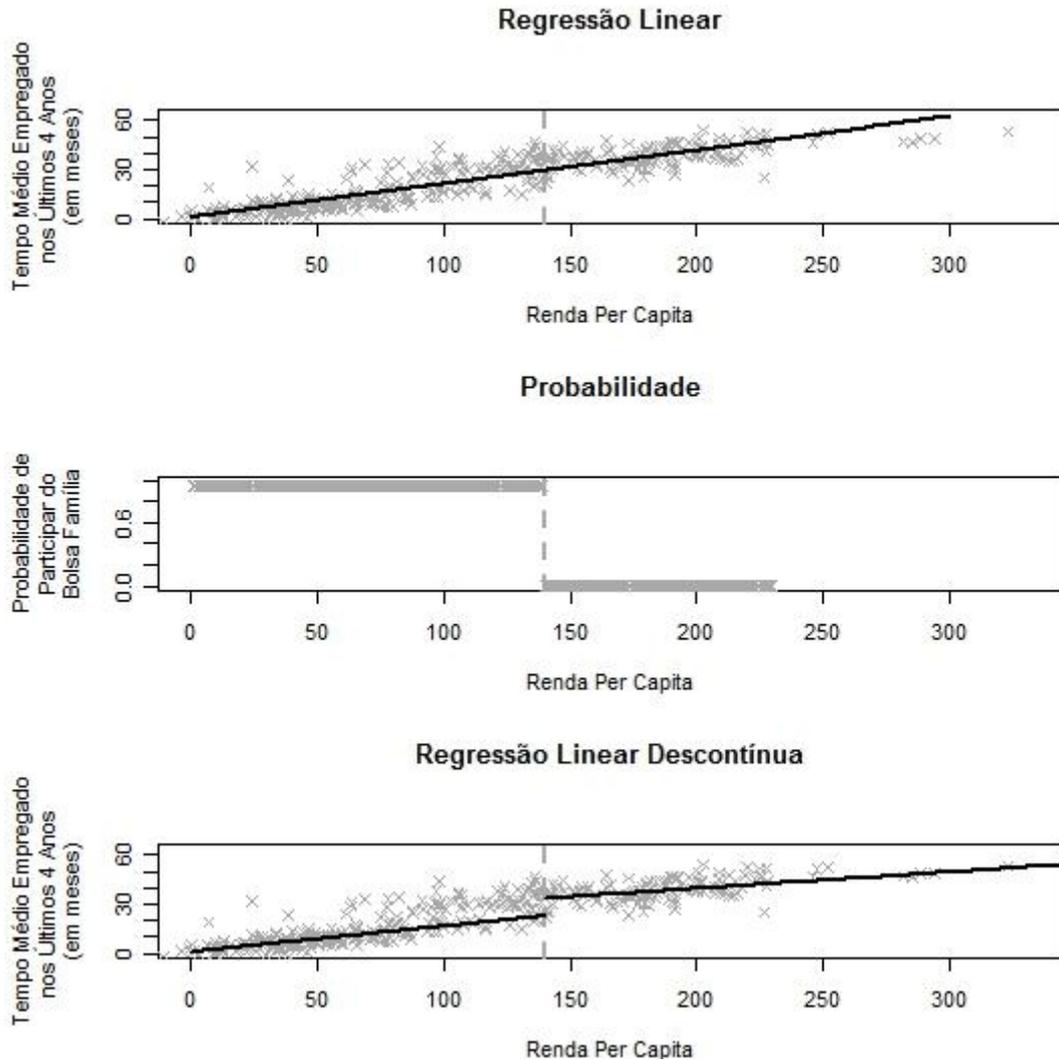


Figura 1

Para o primeiro gráfico da Figura 1, foi gerada a regressão linear simples para os dados fictícios. Neste gráfico é mostrado que ainda que o tempo médio de emprego para o primeiro grupo (*renda per capita* <140 – beneficiários) sendo menor que o segundo (*renda per capita* >140 – não beneficiários) não se pode afirmar que à posse desse benefício, o indivíduo ficará empregado por menos tempo, pois, isto poderia ser consequência de outras covariáveis, talvez capacidade do indivíduo, nível de escolaridade, entre outras.

O segundo gráfico da Figura 1 mostra a probabilidade do indivíduo participar do PBF de acordo com a *renda per capita*, como só são elegíveis pessoas com *renda per*

capita menor que R\$ 140,00, a probabilidade de indivíduos com renda maior que R\$ 140,00 participar do programa é nula, ao mesmo tempo a probabilidade de indivíduos com renda menor que R\$ 140,00 de participar do programa é próximo de 1 (pois o número de elegíveis é pouco maior que o número de vagas). Então se pode ter que entre esses dois grupos a única característica que difere entre eles é a participação no programa, representado pelo salto no gráfico da probabilidade de participar do programa.

No terceiro gráfico da Figura 1, foram geradas as retas pelo método de regressão descontínua. Os dados são os mesmos do primeiro gráfico, no entanto há um ponto de descontinuidade quando a renda *per capita* é igual a R\$ 140,00. Pela interpretação da regressão descontínua temos que este salto pode ter como causa o fato do indivíduo participar ou não do programa, pois há uma descontinuidade na probabilidade do indivíduo ser beneficiário ou não de acordo com a renda. Entretanto, se não houvesse salto, a certeza de que ser do grupo dos tratados ou do grupo de controle não tem diferença no resultado estudado seria concreta.

2.7 Análise de sobrevivência

Como visto capítulo 1, na análise do banco de dados foram encontrados dados censurados na variável de interesse para o estudo: tempo de emprego (desemprego) formal dos indivíduos. Com isso, apesar de não ser comum no estudo de avaliação de políticas públicas, é possível que, devido à presença de dados censurados, a análise dos dados seja feita a partir de gráficos de sobrevivência. É necessário, então, definir alguns conceitos da análise de sobrevivência.

A análise de sobrevivência consiste no estudo de indivíduos onde um evento bem definido (falha) ocorre depois de algum tempo (tempo de falha). Os dados de sobrevivência são caracterizados pelos tempos de falha e pelas censuras, onde censuras são as informações incompletas. No estudo do impacto do PBF no tempo de desemprego, a falha é o indivíduo passar do estado de desemprego para o estado de emprego e a censura ocorre quando um indivíduo não sai do estado de desemprego até o final do estudo, não sendo possível quantificar o tempo em que permaneceu nesse estado.

3 Resultados

3.1 Análise descritiva e formação de grupos controle

Com a base de dados tratada, é possível iniciar o processo de avaliação de que a participação no PBF produz algum impacto no tempo de desemprego formal dos indivíduos de 18 a 65 anos selecionados para o estudo.

Primeiramente, para formar os grupos a serem comparados, foram calculadas as probabilidades dos indivíduos participarem do PBF em 2009, de acordo com as faixas de renda *per capita* das famílias as quais pertencem, a fim de encontrar um corte de renda em que os limites a direita e a esquerda da probabilidade de participação no PBF sejam diferentes, representando assim um salto nessa probabilidade. A Figura 2 mostra os gráficos com essas probabilidades e a tabela 3 indica esses valores.

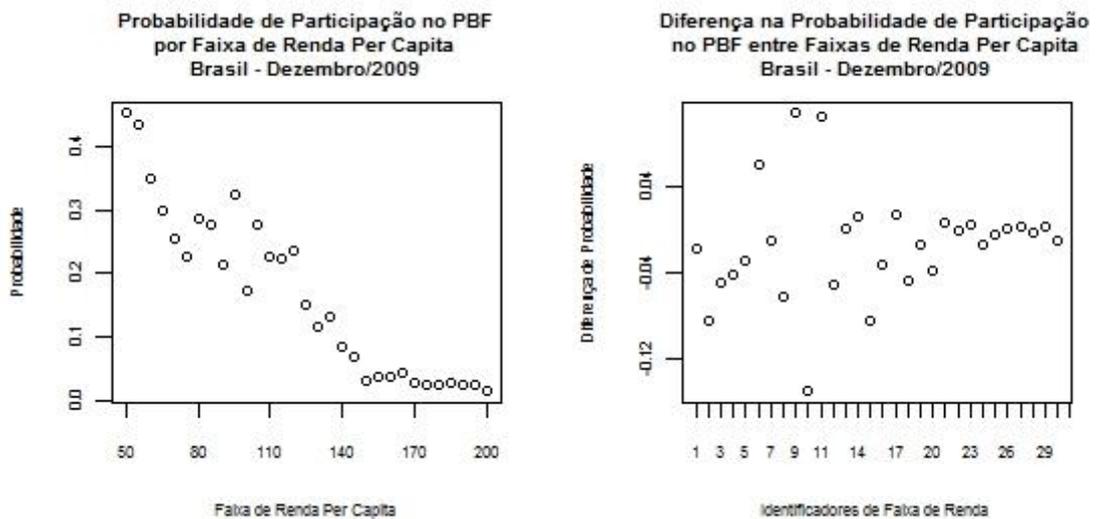


Figura 2

Probabilidade de participação no PBF por Faixa de Renda - CadÚnico - Brasil - Dez/2009

ID	Faixa de renda em R\$	Probabilidade de participação no PBF	Salto de probabilidade (%)
1	195 - 200	0,01461092	-
2	190 - 195	0,02613403	1,152311
3	185 - 190	0,02399337	-0,214066
4	180 - 185	0,02746656	0,347319
5	175 - 180	0,02517101	-0,229555
6	170 - 175	0,02399492	-0,117609
7	165 - 170	0,02975843	0,576351
8	160 - 165	0,04297348	1,321505
9	155 - 160	0,03821467	-0,475881
10	150 - 155	0,03931781	0,110314
11	145 - 150	0,0320661	-0,725171
12	140 - 145	0,07087776	3,881166
13	135 - 140	0,08462711	1,374935
14	130 - 135	0,13180593	4,717882
15	125 - 130	0,11696414	-1,484179
16	120 - 125	0,15055703	3,359289
17	115 - 120	0,23627535	8,571832
18	110 - 115	0,22456222	-1,171313
19	105 - 110	0,22440731	-0,015491
20	100 - 105	0,27565292	5,124561
21	95 - 100	0,17107186	-10,458106
22	90 - 95	0,32330928	15,223742
23	85 - 90	0,21347673	-10,983255
24	80 - 85	0,27638649	6,290976
25	75 - 80	0,28686504	1,047855
26	70 - 75	0,22596973	-6,089531
27	65 - 70	0,25654345	3,057372
28	60 - 65	0,29959928	4,305583
29	55 - 60	0,3495457	4,994642
30	50 - 55	0,43430572	8,476002
31	45 - 50	0,45294347	1,863775

Tabela 3

No primeiro gráfico é descrita a relação entre a probabilidade de participação no PBF e a faixa de renda *per capita* dos indivíduos. A partir do mesmo, nota-se que a probabilidade diminui à medida que a renda *per capita* se eleva, destacando-se saltos de probabilidade em algumas faixas de renda. O segundo gráfico e a tabela mostram as diferenças nas probabilidades de participação no PBF entre as faixas de renda, ou seja, explicam os saltos do primeiro gráfico, das faixas com menor renda para as faixas com

maior renda. As faixas com maior salto de renda são aquelas que estão entre R\$ 85,00 e R\$ 100,00 *per capita*, seguidas das faixas entre R\$ 115,00 e R\$ 125,00 e entre R\$ 50,00 e R\$ 60,00. O comportamento das faixas que estão próximas a R\$ 100,00 *per capita* não é natural, pois não segue o padrão de diminuição das probabilidades que o primeiro gráfico apresentou à medida que a renda se elevava. Pode ser que esse comportamento se deva ao fato de que em 2006 o benefício era concedido às famílias com renda *per capita* de até R\$ 100,00, como o estudo refere-se a beneficiários que desde 2006 participam do PBF, esse pode ser o motivo do comportamento atípico. Já as outras faixas de destaque são o salto de probabilidade de aproximadamente 0,08 entre R\$ 115,00 e R\$ 125,00 e entre R\$ 50,00 e R\$ 60,00, que são faixas próximas as linhas de pobreza e extrema pobreza.

Com o intuito de verificar o impacto da participação do PBF sobre os beneficiários, foram formados quatro grupos segundo a renda familiar per capita dentre as citadas faixas que apresentaram salto na probabilidade: (i) indivíduos com renda superior a R\$ 50,00 e inferior a R\$ 55,00; (ii) indivíduos com renda superior a R\$ 55,00 e inferior a R\$ 60,00; (iii) indivíduos com renda superior a R\$ 90,00 e inferior a R\$ 95,00; (iv) indivíduos com renda superior a R\$ 95,00 e inferior a R\$ 100,00. A probabilidade de participação no PBF para os quatro grupos é diferente, para o primeiro grupo, a probabilidade do indivíduo pertencer à uma família beneficiária é aproximadamente 0,43; para o segundo, é 0,35; para o terceiro, é 0,32; para o quarto, é 0,17.

Em posse desses quatro grupos, a comparação desejada será feita entre os grupos “i” e “ii” e entre os grupos “iii” e “iv”. Pois, para avaliar o impacto do PBF no tempo de emprego (desemprego) dos beneficiários, é necessário formar dois grupos em que a única variável que difira entre eles seja a probabilidade de participação no programa. Essa condição é necessária para que se possa atribuir uma possível diferença entre os grupos à participação no programa e não a uma variável correlacionada. Para a formação destes grupos é interessante então que eles possuam renda *per capita* próximas e a partir daí verificar se as covariáveis também apresentam as mesmas proporções para cada categoria. Estas variáveis são de caracterização do indivíduo ou a família. Para isso, foi realizada uma análise descritiva e alguns testes de hipóteses.

3.2 Variáveis de caracterização do domicílio e indivíduo de famílias com renda *per capita* entre R\$ 50,00 e R\$ 55,00 e entre R\$ 55,00 e R\$ 60,00.

- Variáveis: tipo de localidade, abastecimento de água, escoamento sanitário, tipo de domicílio, situação domicílio, número de cômodos.

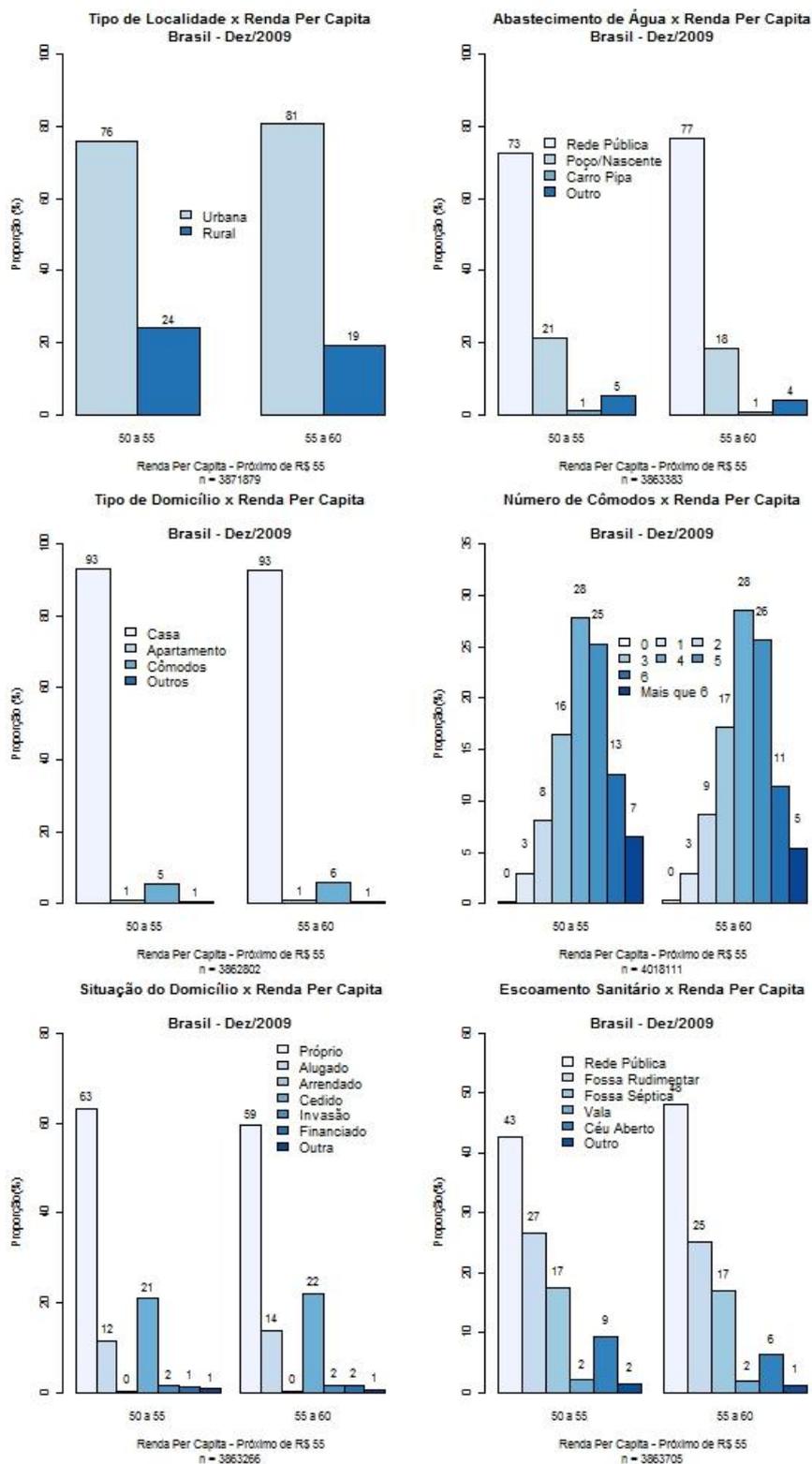


Figura 3

○ Variáveis: sexo, raça, estado civil, escola frequentada, quantidade de pessoas no domicílio, idade.

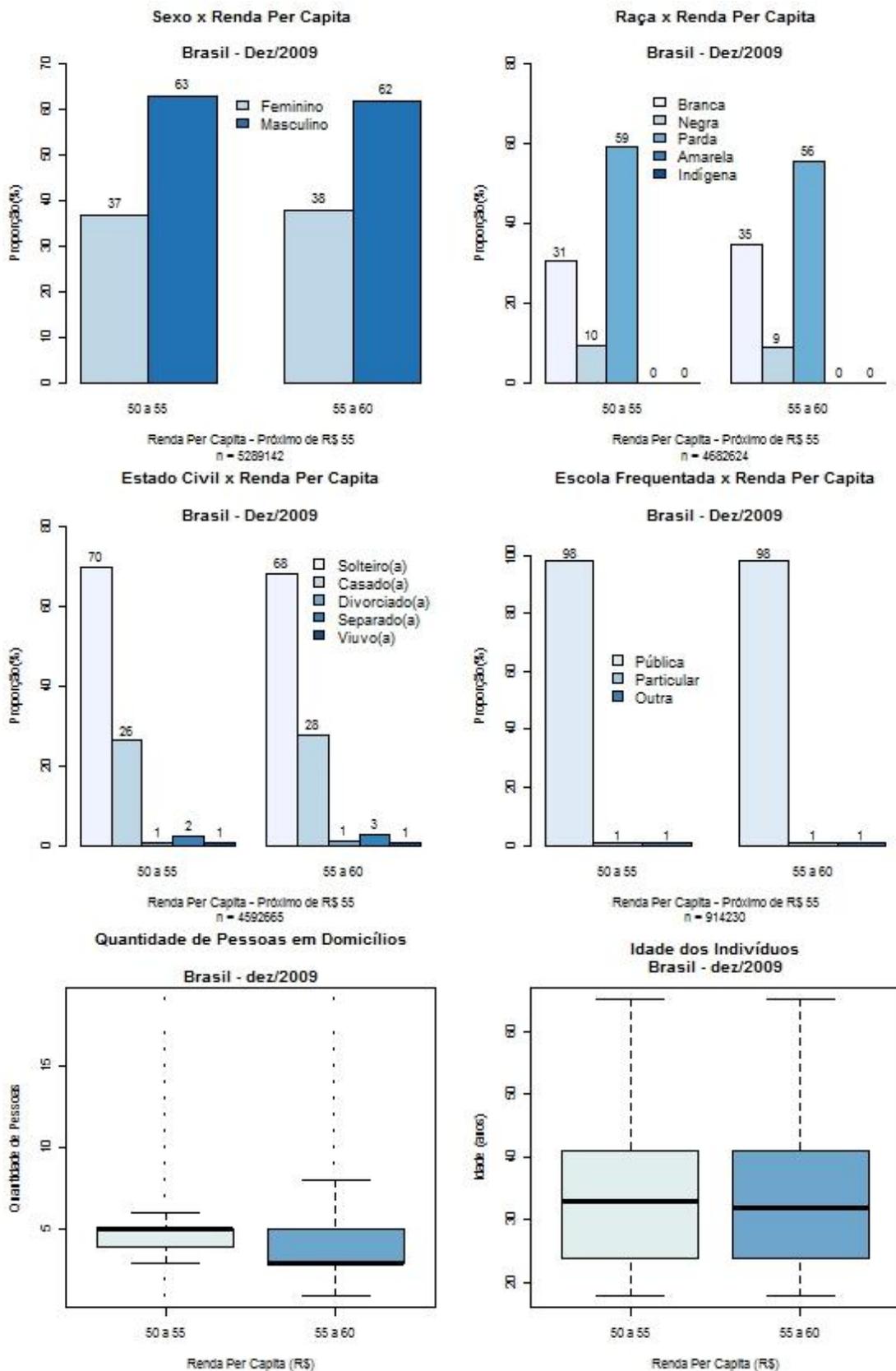


Figura 4

Medidas resumo para variáveis quantitativas: idade e renda *per capita*

Quantidade de Pessoas por Domicílio de famílias com renda *p.c.* entre R\$ 50,00 e R\$ 60,00 - CadÚnico - Brasil - Dez/2009

Quantidade de Pessoas	Renda <i>Per Capita</i> (R\$)	
	50 - 55	55 - 60
Mínimo	1	1
1º Quartil	4	3
Mediana	5	3
Média	4,467	3,73
3º Quartil	5	5
Máximo	19	19

Tabela 4

Idade dos indivíduos com renda *p.c.* entre R\$ 50,00 e R\$ 60,00 - CadÚnico - Brasil - Dez/2009

Idade	Renda <i>Per Capita</i> (R\$)	
	50 - 55	55 - 60
Mínimo	18	18
1º Quartil	24	24
Mediana	33	32
Média	33,63	33,39
3º Quartil	41	41
Máximo	65	65

Tabela 5

Testes-T:

- Hipótese nula: a média da variável testada para o grupo com renda entre R\$ 50,00 e R\$ 55,00 é igual à média para o grupo com renda entre R\$ 55,00 e R\$ 60,00.
- Hipótese alternativa: a média da variável testada para o grupo com renda entre R\$ 50,00 e R\$ 55,00 é diferente da média para o grupo com renda entre R\$ 55,00 e R\$ 60,00.

Teste T para variáveis de caracterização domiciliar e individual de famílias com renda *p.c.* entre R\$50,00 e R\$60,00 - CadÚnico - Brasil - Dez/2009

Variáveis quantitativas	Teste t	Graus de liberdade	p-valor
Quantidade de pessoas	145,02	312214,2	2,20E-16
Idade	7,10	406113,9	1,29E-09

Tabela 6

Testes de Qui-Quadrado

- Hipótese nula: as proporções de cada categoria de cada variável são iguais para indivíduos pertencentes às famílias com renda *per capita* entre R\$ 50,00 e R\$ 55,00 e entre R\$ 55,00 e R\$ 60,00.
- Hipótese alternativa: pelo menos uma das categorias tem proporção diferente para indivíduos pertencentes às famílias com renda *per capita* entre R\$ 50,00 e R\$ 55,00 e entre R\$ 55,00 e R\$ 60,00.

Teste de Qui-quadrado para variáveis de
 caracterização domiciliar e individual para famílias
 com renda *p.c.* entre R\$ 50,00 e R\$ 60,00 - CadÚnico -
 Brasil - Dez/2009

Variáveis qualitativas	Qui-Quadrado	Graus de liberdade	p-valor
Tipo de localidade	499,51	1	2,20E-16
Abastecimento de água	391,5098	3	2,20E-16
Escoamento Sanitário	767,4989	5	2,20E-16
Tipo de domicílio	15,1977	3	9,51E-04
Situação domicílio	297,2555	5	2,20E-16
Número de cômodos	74,5326	5	2,20E-16
Estado civil	66,8607	4	6,34E-12
Raça	379,0816	4	2,20E-16
Sexo	33,0018	1	9,21E-06
Escola frequentada	0,0022	1	9,62E-01

Tabela 7

Como é possível observar nas tabelas dos testes de hipótese, apenas para a variável escola frequentada a hipótese nula dos testes de qui-quadrado e t não foi rejeitada. Para os demais casos, rejeita-se a hipóteses de que as proporções ou médias são iguais para ambos os grupos. Isso ocorre por causa da influência que o tamanho da amostra incide sobre esses testes. Como a amostra do estudo em questão é grande, 164.416 indivíduos com renda entre R\$ 50,00 e R\$ 55,00 e 259.112 indivíduos com renda entre R\$ 55,00 e R\$ 60,00, se encontrou esse resultado.

Como é possível observar nas tabelas dos testes de hipótese, apenas para a variável escola frequentada a hipótese nula dos testes de qui-quadrado e t não foi rejeitada. Para os demais casos, rejeita-se a hipóteses de que as proporções ou médias são iguais para ambos os grupos. Isso ocorre por causa da influência que o tamanho da amostra incide sobre esses testes. Como a amostra do estudo em questão é grande, 164.416 indivíduos com renda entre R\$ 50,00 e R\$ 55,00 e 259.112 indivíduos com renda entre R\$ 55,00 e R\$ 60,00, se encontrou esse resultado.

No entanto, dos gráficos acima se percebe que, em termos práticos, as características de domicílio das famílias inseridas no CadÚnico em 2009, com renda familiar *per capita* entre R\$ 50,00 e R\$ 55,00 e entre R\$ 55,00 R\$ 60,00 tem uma diferença mínima. O perfil domiciliar dessas famílias nesta primeira faixa de renda, de R\$ 50,00 a R\$ 55,00 é caracterizado com 76% dos domicílios localizados em área urbana, 73% com abastecimento público de água, 93% dos indivíduos residem em casas, sendo que 63% dos domicílios são habitações próprias, aproximadamente 53%

das residências tem 4 ou 5 cômodos, 43% possuem escoamento sanitário público e em média 4,4 pessoas residindo em cada domicílio. Enquanto que o perfil domiciliar das famílias com renda entre R\$ 55,00 e R\$ 60,00 é caracterizado com 81% dos domicílios localizados em área urbana, 77% com abastecimento público de água, 93% dos indivíduos residem em casas, sendo que 59% dos domicílios são habitações próprias, aproximadamente 58% das residências tem 4 ou 5 cômodos, 48% possuem escoamento sanitário público e em média 3,73 pessoas residindo em cada domicílio.

Assim como para as variáveis de domicílio, as características dos indivíduos para os dois grupos de renda são muito próximas. A composição do perfil dos indivíduos inseridos no CadÚnico com renda *per capita* entre R\$ 50,00 e R\$ 55,00 é de 63% indivíduos do sexo masculino, 59% são pardos, 70% são solteiros, 98% frequentou escola pública e a idade média é 33 anos. Enquanto que o perfil das famílias na segunda faixa de renda *per capita*, entre R\$ 55,00 e R\$ 60,00, é composto por 62% indivíduos do sexo masculino, 56% são pardos, 68% são solteiros, 98% frequentou escola pública e a idade média é 33 anos.

A seguir, tem-se análise descritiva semelhante para os grupos de comparação “iii” e “iv”.

3.3 Variáveis de caracterização do domicílio e indivíduo de famílias com renda *per capita* entre R\$ 90,00 e R\$ 95,00 e entre R\$ 95,00 e R\$ 100,00

- Variáveis: tipo de localidade, abastecimento de água, escoamento sanitário, tipo de domicílio, situação domicílio, número de cômodos.

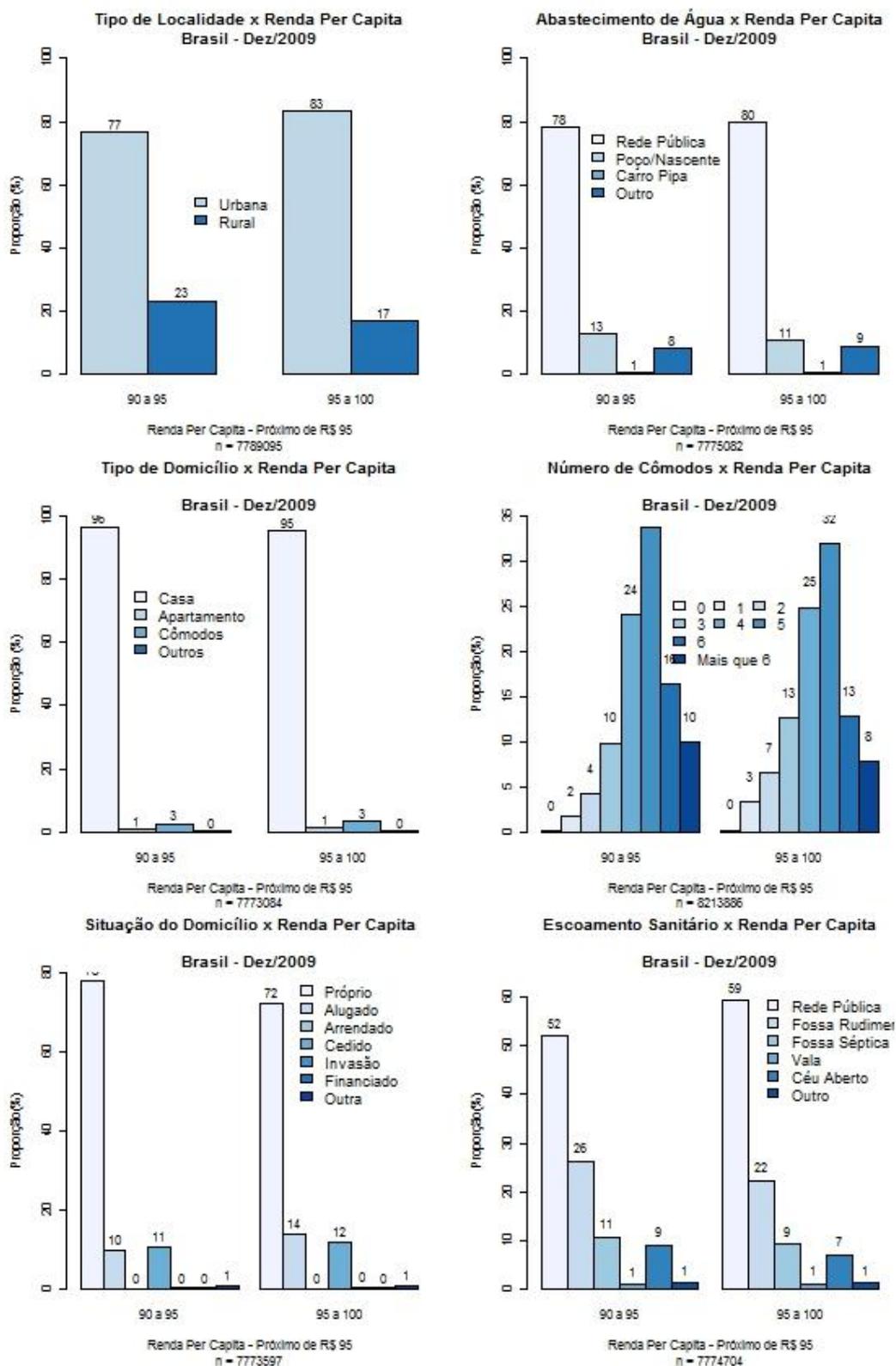


Figura 5

○ Variáveis: sexo, raça, estado civil, escola frequentada, quantidade de pessoas no domicílio, idade.

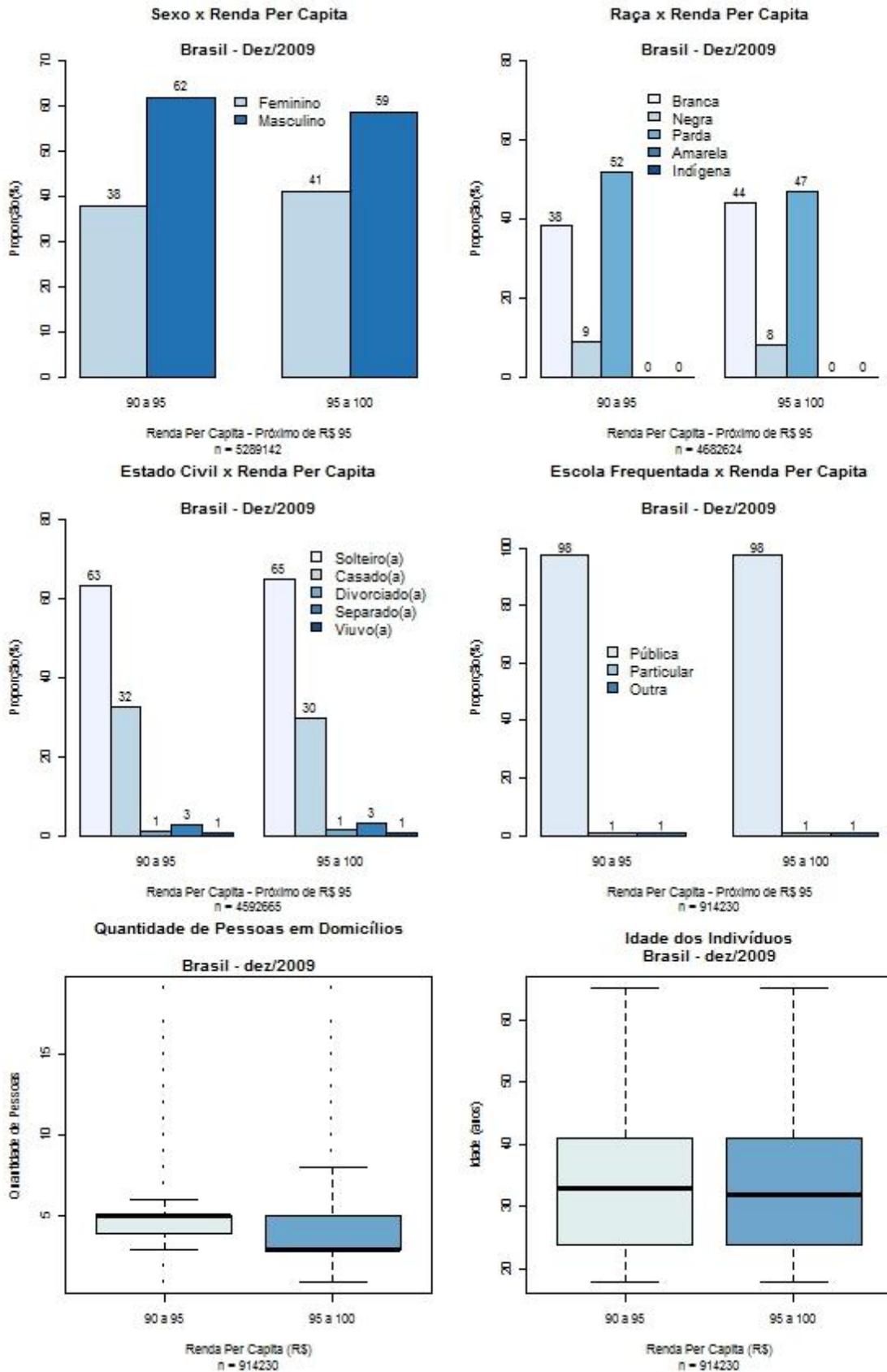


Figura 6

Medidas resumo para variáveis quantitativas: idade e renda *per capita*

Quantidade de Pessoas por Domicílio de famílias com renda p.c. entre R\$ 90,00 e R\$ 100,00 - CadÚnico - Brasil - Dez/2009

Quantidade de Pessoas	Renda Per Capita (R\$)	
	90 - 95	95 - 100
Mínimo	1	1
1º Quartil	4	3
Mediana	5	3
Média	4,467	3,73
3º Quartil	5	5
Máximo	19	19

Tabela 8

Idade dos indivíduos com renda p.c.entre R\$ 90,00 e R\$ 100,00 - CadÚnico - Brasil - Dez/2009

Idade	Renda Per Capita (R\$)	
	90 - 95	95 - 100
Mínimo	18	18
1º Quartil	24	24
Mediana	33	32
Média	33,63	33,39
3º Quartil	41	41
Máximo	65	65

Tabela 9

Teste-T:

- Hipótese nula: a média da variável para o grupo com renda entre R\$ 90,00 e R\$ 95,00 é igual à média para o grupo com renda entre R\$ 95,00 e R\$ 100,00
- Hipótese alternativa: a média da variável testada para o grupo com renda entre R\$ 90,00 e R\$ 95,00 é diferente da média par ao grupo com renda entre R\$ 95,00 e R\$ 100,00

Teste T para variáveis de caracterização domiciliar e individual de famílias com renda p.c. entre R\$ 90,00 e R\$ 100,00 - CadÚnico - Brasil - Dez/2009

Variáveis quantitativas	Teste t	Graus de liberdade	p-valor
Quantidade de pessoas	145,016	312214,2	2,20E-16
Idade	7,0953	406113,9	1,29E-09

Tabela 10

Testes de Qui-Quadrado

- Hipótese nula: as proporções de cada categoria de cada variável são iguais para indivíduos pertencentes às famílias com renda *per capita* entre R\$ 90,00 e R\$ 95,00 e entre R\$ 95,00 e R\$ 100,00.
- Hipótese alternativa: pelo menos uma das categorias tem proporção diferente para indivíduos pertencentes às famílias com renda *per capita* entre R\$ 90,00 e R\$ 95,00 e entre R\$ 95,00 e R\$ 100,00.

Teste Qui-Quadrado para variáveis de caracterização domiciliar e individual de famílias com renda p.c. entre R\$ 90,00 e R\$ 100,00 - CadÚnico - Brasil -

Variáveis qualitativas	Qui-Quadrado	Graus de liberdade	p-valor
Tipo de localidade	3298,834	1	2,20E-16
Abastecimento de água	677,008	3	2,20E-16
Escoamento Sanitário	2894,034	5	2,20E-16
Tipo de domicílio	434,4071	3	2,20E-16
Situação domicílio	2367,855	5	2,20E-16
Número de cômodos	4397,026	5	2,20E-16
Estado civil	375,084	4	2,20E-16
Raça	1206,731	4	2,20E-16
Sexo	391,3193	1	2,20E-16
Escola frequentada	0,0505	1	8,22E+03

Tabela 11

Assim como ocorreu para os grupos “i” e “ii”, os testes de hipótese para grupos de renda *per capita* de R\$ 90,00 – R\$ 95,00 e R\$ 95,00 – R\$ 100,00 tiveram as hipóteses nulas rejeitadas, exceto para a variável Escola Frequentada. Novamente esse fato se deve ao tamanho da amostra, são 255.544 indivíduos com renda *per capita* de R\$ 90,00 a R\$ 95,00 e 1.478.241 indivíduos com renda *per capita* de R\$ 95,00 a R\$ 100,00.

Como visto nos grupos comparados anteriormente, dos gráficos acima se percebe que, em termos práticos, as características de domicílio das famílias inseridas no CadÚnico em 2009, com renda familiar *per capita* entre R\$ 90,00 e R\$ 95,00 e entre R\$ 95,00 R\$ 100,00 não diferem. O perfil domiciliar das famílias que pertencem ao grupo de renda *per capita*, mais baixa, de R\$ 90,00 a R\$ 95,00 é caracterizado por 77% dos domicílios localizados em área urbana, 78% com abastecimento público de água, 52% possuem escoamento sanitário público, 96% dos indivíduos residem em casas, sendo que 78% dos domicílios são habitações próprias, aproximadamente 53% das residências tem 4 ou 5 cômodos e em média 4,467 pessoas residindo em cada domicílio. Enquanto que o perfil domiciliar das famílias que pertencem ao grupo de renda *per capita*, mais alta, de R\$ 95,00 a R\$ 100,00 é caracterizado por 83% dos domicílios localizados em área urbana, 80% com abastecimento público de água, 59% possuem escoamento sanitário público, 95% dos indivíduos residem em casas, sendo que 72% dos domicílios são habitações próprias, aproximadamente 57% das residências tem 4 ou 5 cômodos e em média 3,73 pessoas residindo em cada domicílio.

As características dos indivíduos para os dois grupos de renda também são semelhantes. A composição do perfil dos indivíduos inseridos no CadÚnico com renda

per capita entre R\$ 90,00 e R\$ 95,00 é de 62% indivíduos do sexo masculino, 52% são pardos, 63% são solteiros, 98% frequentou escola pública e a idade média é 33,63 anos. Enquanto que a composição do perfil dos indivíduos que pertencem ao grupo de renda *per capita*, mais alta, de R\$ 95,00 a R\$ 100,00 é de 59% indivíduos do sexo masculino, 47% são pardos, 65% são solteiros, 98% frequentou escola pública e a idade média é 33,39 anos.

Apesar de apresentarem comportamento similares na prática, para os quatro grupos, quando comparados dois a dois com as rendas próximas, tem-se que em geral os grupos com maior participação no PBF são mais precários, o que geraria um viés de ficar empregado menos tempo e desempregado mais tempo quando comparados ao grupo com menor probabilidade de participação no PBF para uma renda próxima.

3.4 Avaliação do tempo de emprego (desemprego) formal

A ideia inicial do estudo foi desenvolver um modelo de regressão descontínua, pois a atribuição do tratamento, participação no PBF, é uma função descontínua da renda *per capita* e, além disso, foi possível formar dois conjuntos de grupos onde a única diferença relevante entre eles é a probabilidade de participação no PBF.

Como visto no capítulo 3, o objetivo da regressão descontínua é estimar o efeito médio do tratamento, usando indivíduos pouco abaixo do limite de atribuição como um controle para os que estão um pouco acima. Porém, devido à forte presença de dados censurados nas variáveis de emprego, não foi possível calcular o tempo médio de desemprego, pois seria necessária a estimação das médias por um modelo paramétrico, cuja validade é questionável. Com isso, optou-se por analisar o comportamento do tempo de desemprego nos quatro grupos, respeitando os pressupostos da regressão descontínua, mas utilizando gráficos de sobrevivência.

Sejam os conjuntos de grupos:

- i. $W_i = 0$, se o indivíduo i tem renda familiar *per capita* entre R\$ 50,00 e R\$ 55,00
- ii. $W_i = 1$, se o indivíduo i tem renda familiar *per capita* entre R\$ 55,00 e R\$ 60,00

- iii. $Y_i = 0$, se o indivíduo i tem renda familiar *per capita* entre R\$ 90,00 e R\$ 95,00
- iv. $Y_i = 1$, se o indivíduo i tem renda familiar *per capita* entre R\$ 95,00 e R\$ 100,00

Na regressão descontínua, se o impacto de alguma variável não observável correlacionada com a variável usada na atribuição do tratamento é suave, a seguinte suposição é razoável para um ε pequeno:

$$E[Y_1^C | T, X < \bar{X} + \varepsilon, X > \bar{X} - \varepsilon] = E[Y_1^C | C, X < \bar{X} + \varepsilon, X > \bar{X} - \varepsilon],$$

onde X é a variável subjacente e \bar{X} é o limite de atribuição. Essa suposição implica que sob um intervalo ε de \bar{X} , o viés de seleção é zero e é a base do “delineamento da estimativa da regressão descontínua”.

Apesar de não ter sido possível calcular a média do tempo de desemprego, com os gráficos apresentados no capítulo 3, observa-se que sob um intervalo pequeno de renda *per capita*, $\varepsilon = 5$, o comportamento dos dois conjuntos de grupos formados são semelhantes, mas em algumas características o grupo com maior participação no PBF apresenta resultados mais precários, o que geraria um viés de ficar empregado menos tempo e desempregado mais tempo em relação ao grupo com menor participação no PBF.

Com as suposições da Regressão Descontínua respeitadas, é possível fazer a análise com os gráficos de sobrevivência.

Gráficos de Sobrevivência: proporção cumulativa permanecendo empregado/desemprego

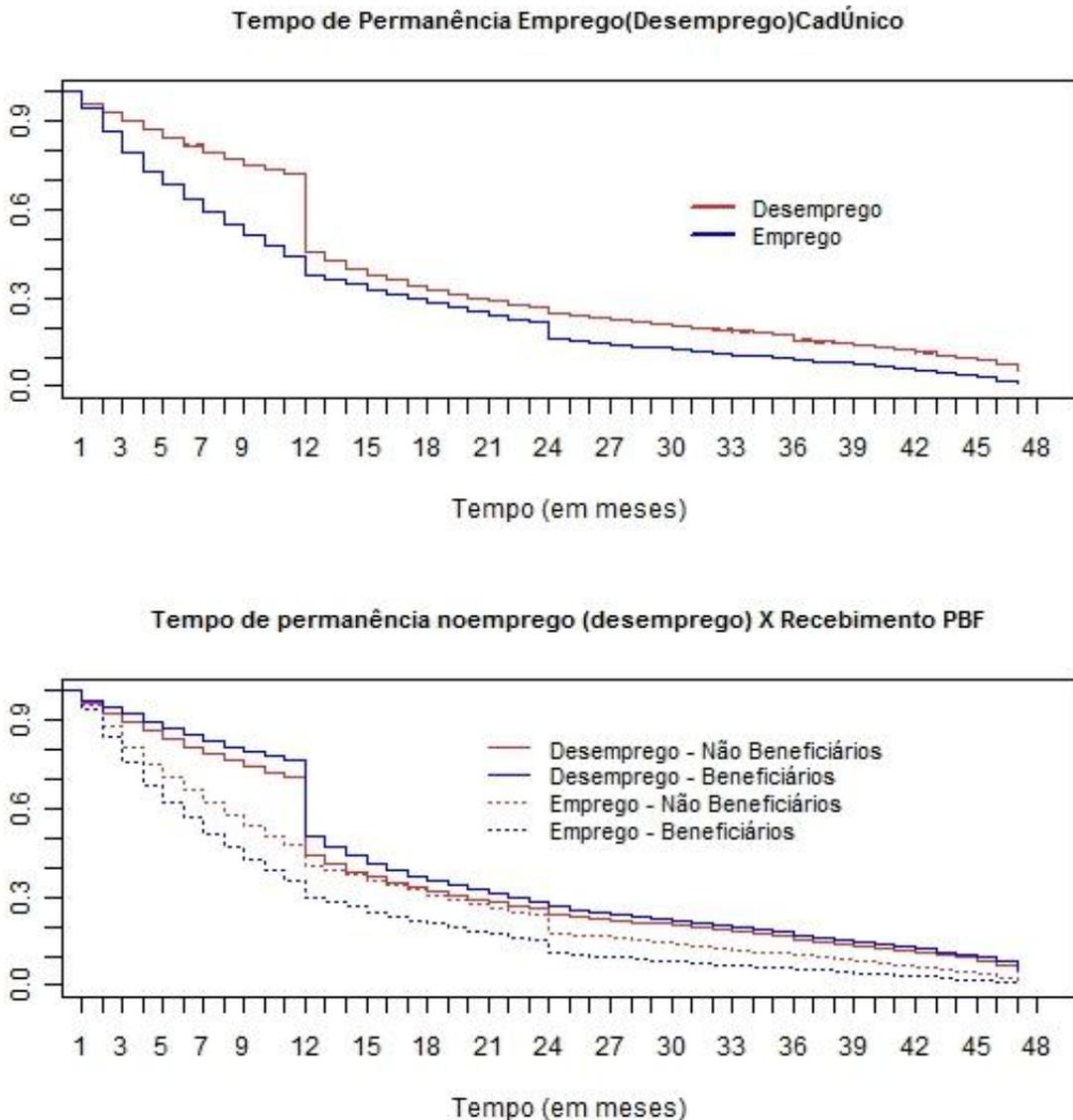


Figura 7

Pelo primeiro gráfico apresentado na Figura 7, que representa o tempo de permanência no emprego (desemprego) formal dos indivíduos inseridos no CadÚnico, pode-se afirmar que ao longo dos 48 meses de estudo, a permanência no estado de desemprego é maior do que no estado de emprego, em relação ao mercado formal. Deste gráfico tem-se que 50% dos inseridos no CadÚnico ficam pelo menos 12 meses empregado e também tem-se que 50% ficam pelo menos 12 meses desempregados.

O segundo gráfico mostra a curva de sobrevivência para o tempo de emprego e desemprego, para os inseridos no CadÚnico, separados em beneficiários e não beneficiários. Desta curva tem-se que, em geral, não beneficiários permanecem menos

tempo na situação de desemprego que os beneficiários. E o oposto ocorre para o estado de emprego, ou seja, os não beneficiários permanecem na situação de emprego mais tempo que os beneficiários. Todavia, essa diferença não pode ser atribuída à participação no bolsa família, visto que existem diversas outras covariáveis que possuem comportamentos divergentes entre esses grupos podendo estar relacionado à alguma(s) delas essa diferença no tempo de emprego e desemprego. Do gráfico temos que 40% dos não beneficiários ficam no mínimo 13 meses no estado de desemprego enquanto que para beneficiários esse tempo é em torno de 16 meses. Para o tempo de emprego, tem-se que 40% dos não beneficiários estiveram pelos menos 13 meses neste estado enquanto que para os beneficiários este tempo foi de 10 meses.

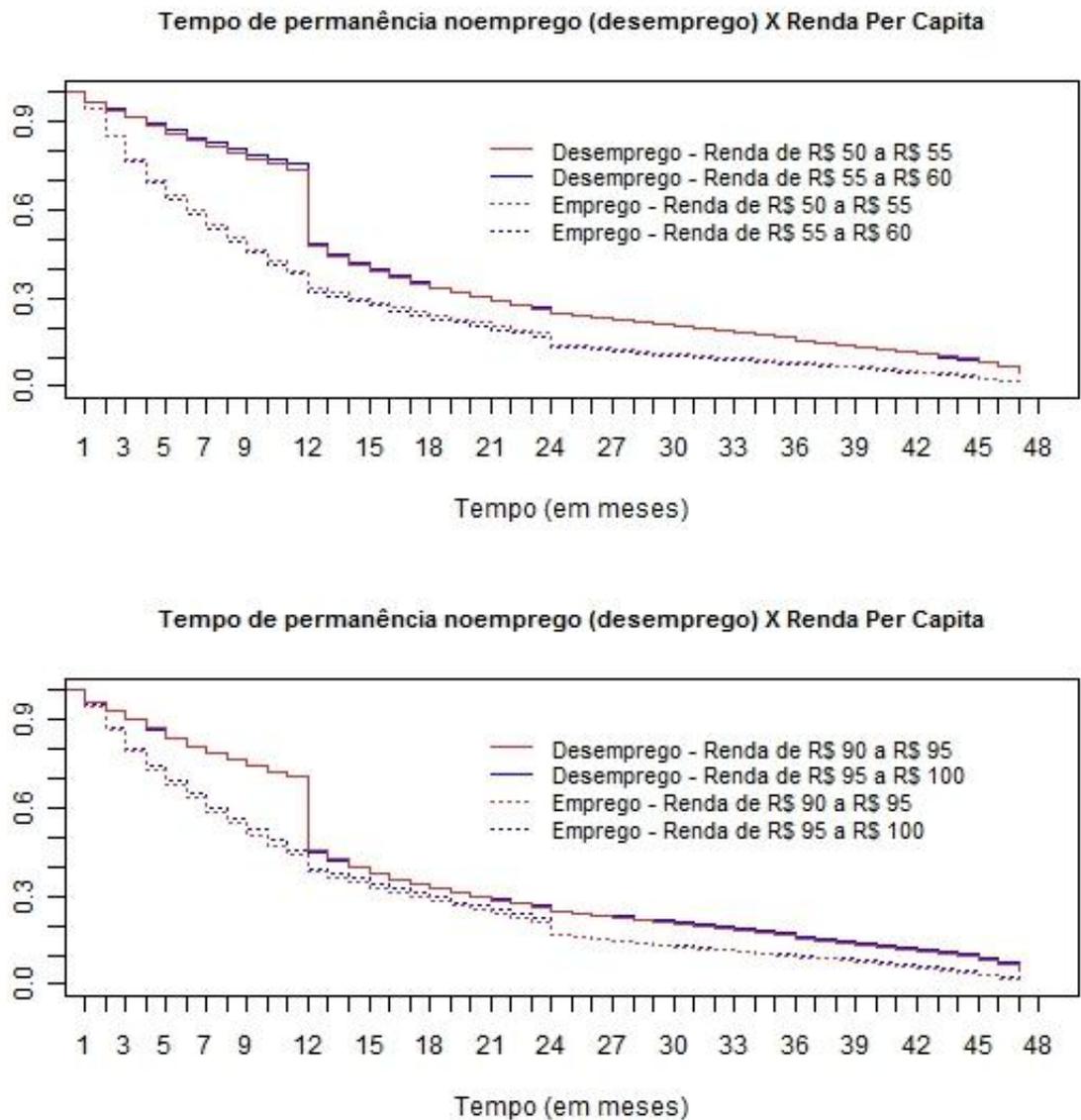


Figura 8

Para verificar a influência da participação no PBF no tempo de emprego (desemprego), foram realizados então, o terceiro e quarto gráficos, para rendas de R\$ 50,00 a R\$ 60,00 e R\$ 90,00 a R\$ 100,00, respectivamente, formando grupos em que a diferença, em termos práticos, entre eles, é a probabilidade de atribuição.

Para as rendas de R\$ 50,00 a R\$ 60,00, é nítido pelo terceiro gráfico que a diferença entre estes grupos é mínima tanto para o tempo de emprego quanto de desemprego, tornando as curvas praticamente sobrepostas. O que sugere que o tempo de emprego (desemprego) não está relacionado com a participação no PBF. Do gráfico temos que 40% tanto para os indivíduos com renda *per capita* entre R\$ 50,00 e R\$ 55,00 e entre R\$ 55,00 e R\$ 60,00 o tempo no estado de emprego é de 11 meses e para o tempo no estado de desemprego tem-se que 40% dos indivíduos nestas faixas permanecem pelo menos 15 meses desempregado.

Para o quarto gráfico segue a mesma interpretação de que o tempo de emprego (desemprego) não está relacionado com a participação no PBF. Comparando esses dois últimos gráficos observa-se que para essa faixa de renda o tempo médio de desemprego é inferior ao da renda mais baixa e o tempo de emprego é mais alto. Assim sendo, do gráfico, temos que 40%, tanto para os indivíduos com renda *per capita* entre R\$ 90,00 e R\$ 95,00 e entre R\$ 95,00 e R\$ 100,00, o tempo no estado de emprego é de 12 meses e para o estado de desemprego tem-se 14 meses.

A partir dessas Tabelas 12 e 13 tem-se que mesmo excluindo da base trabalhada indivíduos completamente censurados, ou seja, que ficaram 48 meses no mesmo estado em relação ao tempo de desemprego ou emprego formal a conclusão do estudo não mudaria, pois a proporção de indivíduos com essas mesmas características é semelhante para as faixas de renda comparadas, como mostrado nas tabelas seguintes, sugerindo mais uma vez que a participação no PBF não influencia no tempo de desemprego.

Tabela resumo dos dados completamente censurados para famílias com renda *p.c.* entre R\$ 50,00 e R\$ 60,00 e entre R\$ 90,00 e R\$ 100,00 - CadÚnico - Brasil - Jan/2006 a Dez/2009

Beneficiários 48 meses empregados	57.171
Não beneficiários 48 meses empregados	375.963
Beneficiários 48 meses desempregados	7.778.773
Não beneficiários 48 meses desempregados	10.375.242
Indivíduos empregados 48 meses com renda <i>p.c.</i> entre R\$50,00 e R\$55,00	4.292
Indivíduos empregados 48 meses com renda <i>p.c.</i> entre R\$55,00 e R\$60,00	10.142
Indivíduos desempregados 48 meses com renda <i>p.c.</i> entre R\$50,00 e R\$55,00	314.682
Indivíduos desempregados 48 meses com renda <i>p.c.</i> entre R\$55,00 e R\$60,00	630.686
Indivíduos empregados 48 meses com renda <i>p.c.</i> entre R\$90,00 e R\$95	14.011
Indivíduos empregados 48 meses com renda <i>p.c.</i> entre R\$95,00 e R\$100,00	20.966
Indivíduos desempregados 48 meses com renda <i>p.c.</i> entre R\$90,00 e R\$95	397.412
Indivíduos desempregados 48 meses com renda <i>p.c.</i> entre R\$95,00 e R\$100	690.460
Total de beneficiários	9.211.368
Total de não beneficiários	14.701.258
Total de indivíduos com renda <i>p.c.</i> entre R\$50,00 e R\$55,00	388.020
Total de indivíduos com renda <i>p.c.</i> entre R\$55,00 e R\$60,00	809.799
Total de indivíduos com renda <i>p.c.</i> entre R\$90,00 e R\$95,00	554.912
Total de indivíduos com renda <i>p.c.</i> entre R\$95,00 e R\$100,00	974.198

Tabela 12

Tabela resumo das proporções dos dados completamente censurados para famílias com renda p.c. entre R\$ 50,00 e R\$ 60,00 e entre R\$ 90,00 e R\$ 100,00 - CadÚnico - Brasil - Jan/2006 a Dez/2009

Total de beneficiários 48 meses empregados/Total de beneficiários	0,0062
Total de não beneficiários 48 meses empregados/Total de não beneficiários	0,0256
Total de beneficiários 48 meses desempregados/Total de beneficiários	0,8445
Total de não beneficiários 48 meses desempregados/Total de não beneficiários	0,7057
Total de indivíduos empregados 48 meses com renda p.c entre R\$50,00 e R\$55,00/Total de indivíduos com renda p.c entre R\$50,00 e R\$55,00	0,0111
Total de indivíduos empregados 48 meses com renda p.c entre R\$55,00 e R\$60,00/Total de indivíduos com renda p.c. entre R\$55,00 e R\$60,00	0,0125
Total de indivíduos desempregados 48 meses com renda p.c entre R\$50,00 e R\$55,00/ Total de indivíduos com renda p.c entre R\$50,00 e R\$55,00	0,8110
Total de indivíduos desempregados 48 meses com renda p.c entre R\$55,00 e R\$60,00 /Total de indivíduos com renda p.c. entre R\$55,00 e R\$60,00	0,7788
Total de indivíduos empregados 48 meses com renda p.c entre R\$90,00 e R\$95,00/Total de indivíduos com renda p.c entre R\$90,00 e R\$95,00	0,0252
Total de indivíduos empregados 48 meses com renda p.c entre R\$95,00 e R\$100,00/Total de indivíduos com renda p.c. entre R\$95,00 e R\$100,00	0,0215
Total de indivíduos desempregados 48 meses com renda p.c entre R\$90,00 e R\$95,00/ Total de indivíduos com renda p.c entre R\$90,00 e R\$95,00	0,7162
Total de indivíduos desempregados 48 meses com renda p.c entre R\$95,00 e R\$100,00 /Total de indivíduos com renda p.c. entre R\$95,00 e R\$100,00	0,7087

Tabela 13

4 Conclusão

As políticas públicas tem se tornado instrumentos importantes em toda América Latina em busca da extinção da fome e pobreza nos países subdesenvolvidos. No Brasil, isso é notado, pelos inúmeros programas de transferência de renda existentes, além daqueles que promovem a distribuição de alimentos ou outros utensílios necessários à sobrevivência com dignidade.

Segundo o 4º Relatório Nacional de Acompanhamento dos Objetivos de Desenvolvimento do Milênio, o programa de maior cobertura e eficiência na redução da miséria no Brasil é o Programa Bolsa Família, que beneficia famílias na linha da pobreza e extrema pobreza, privilegiando aquelas que têm maior número de filhos na infância e adolescência.

O objetivo desse estudo foi promover uma avaliação do Programa Bolsa Família a partir da análise do impacto que a participação no programa provoca no tempo de emprego (desemprego) formal. A motivação surgiu das frequentes discussões a respeito do comodismo e desencorajamento em busca de emprego que o programa pode causar nos beneficiários.

“O fato de os beneficiários do Bolsa-Família ficarem menos tempo no emprego do que os trabalhadores de baixa renda não atendidos pelo programa e de, quando o perdem, demorarem mais para encontrar outra ocupação com carteira assinada é mais um dos graves problemas deixados pelo governo anterior(...)” (Estado de S. Paulo, 2011).

Sendo assim, os indivíduos foram separados em grupos semelhantes, onde a diferença relevante entre eles era apenas a probabilidade de participação no PBF, a fim de minimizar o efeito que covariáveis poderiam causar no tempo de emprego (desemprego).

A análise foi realizada a partir de grupos com renda *per capita* entre R\$ 50,00 e R\$ 60,00 e entre R\$ 90,00 e R\$ 100,00. Com os resultados obtidos, percebe-se que, para as duas comparações, o fato de participar ou não do PBF não influencia no tempo de emprego (desemprego), pois o comportamento de beneficiários e não beneficiários, que possuem características sociais semelhantes, foi o mesmo.

Esse resultado implica que ao contrário da crença de que a participação no PBF provoca a acomodação dos beneficiários no estado de emprego (desemprego), o recebimento do benefício não gerou um impacto mensurável no tempo de emprego (desemprego) formal neste estudo.

5 Trabalhos Futuros

Completar o estudo apresentado, com um trabalho de estatística inferencial, elaborando intervalos de confiança para os gráficos de sobrevivência, a fim de validar o resultado final deste trabalho.

6 Bibliografia

- IMBENS, Guido W.; WOOLDRIDGE, Jeffrey M. **Recent Developments in the Econometrics of Program Evaluation.** Journal of Economic Literature, 2009, 47:1, 5-86.
- Duflo, Esther; Glennerster Rachel e Kremer Michael. **Using Randomization in Development Economics Research: A Toolkit.** BREAD Working Paper No. 136, December 2006
- Doksum, Kjell. 1974. **Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case.** Annals of Statistics, 2(2): 267-77.
- HECKMAN, J.; ICHIMURA H.; SMITH J. & TODD P. **Characterizing Selection Bias Using Experimental Data.** Econometrica 66:1017-99, 1998.
- IMBENS, Guido W.; WOOLDRIDGE, Jeffrey M. **Recent Developments in the Econometrics of Program Evaluation.** Journal of Economic Literature, 2009, 47:1, 5-86.
- Lehmann, Erich L. 1974. **Nonparametrics: Statistical Methods Based on Ranks.** San Francisco: Holden-Day.
- RAVALLION, M. **The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation.** The World Bank Review 15(1):115-140,2000
- Rosenbaum, Paul R., and Donald B. Rubin. 1983b. **The Central Role of the Propensity Score in Observational Studies for Causal Effects.** Biometrika, 70(1): 41–55.
- www.mds.gov.br. Desenvolvido pelo Ministério do Desenvolvimento Social e Combate à Fome. Acesso em: out. 2010.
- Colosismo, Enrico Antônio; Giolo, Suely Ruiz. **Análise de Sobrevida Aplicada.**
- www.estadao.com.br/estadao/hoje/20110119/not_imp668070.0.php. Desenvolvido por Estado de S.P.. Acesso em: jan. 2011
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

- Ipea. Brasília, 2010. **Relatório Nacional do Acompanhamento** – Objetivos de Desenvolvimento do Milênio.
- Ipea, MDS, 2009. Relatório Técnico. RAIS x CadÚnico: Beneficiários do Bolsa Família no Mercado Formal de Trabalho. Alexandre R. Leichsenring