



Universidade Federal de Pernambuco  
Centro de Ciências Exatas e da Natureza  
Programa de Pós-graduação em Estatística

Walmir dos Reis Miranda Filho

**Avaliação do Impacto de Políticas de  
Transferência de Renda a partir de Dados  
Amostrais Complexos**

Recife

2017

Walmir dos Reis Miranda Filho

**Avaliação do Impacto de Políticas de Transferência de Renda a partir de Dados Amostrais Complexos**

Dissertação apresentada ao Programa de Pós-graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Orientador: Cristiano Ferraz

Coorientador: Marcel de Toledo Vieira

**Recife**

**2017**

Catálogo na fonte  
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

M672a Miranda Filho, Waldir dos Reis  
Avaliação do impacto de políticas de transferência de renda a partir de dados amostrais complexos / Waldir dos Reis Miranda Filho. – 2017.  
204 f.: il., fig., tab.

Orientador: Cristiano Ferraz.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.  
Inclui referências, apêndices e anexos.

1. Estatística aplicada. 2. Amostragem. I. Ferraz, Cristiano (orientador). II. Título.

310

CDD (23. ed.)

UFPE- MEI 2017-48

**WALMIR DOS REIS MIRANDA FILHO**

**AVALIAÇÃO DO IMPACTO DE POLÍTICAS DE TRANSFERÊNCIA DE RENDA A  
PARTIR DE DADOS AMOSTRAIS COMPLEXOS**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 13 de fevereiro de 2017.

**BANCA EXAMINADORA**

---

Prof. Marcel de Toledo Vieira  
UFJF

---

Prof. Pedro Luis do Nascimento Silva  
ENCE-IBGE

---

Prof. Raydonal Ospina Martínez  
UFPE

*Aos meus pais, Maria e Walmir*

# Agradecimentos

Em primeiro lugar, a Deus. Com sua bondade infinita, me permitiu conhecer pessoas maravilhosas em minha vida que me ajudaram a enfrentar muitas dificuldades até aqui. Sem isso, eu não teria concluído mais esta etapa de minha carreira acadêmica.

Aos meus pais, Maria Aparecida de Campos e Walmir Miranda, os quais me deram todo o suporte necessário. Ainda que não entendessem em totalidade os problemas que passei durante o Mestrado por não estarem comigo em Recife, nunca deixaram de me apoiar ou aconselhar e sempre foram compreensivos e afetivos.

Aos meus orientadores, Cristiano Ferraz e Marcel Vieira, pela ajuda necessária para concluir a dissertação. Sempre foram exigentes quando necessário e solícitos quanto às dúvidas conceituais e na elaboração do texto.

Aos meus colegas do Mestrado e Doutorado em Estatística da UFPE pelos momentos de companheirismo, seja nas tarefas acadêmicas ou nas questões externas à universidade. Eu não teria crescido pessoalmente sem eles. Fiz amizades que levarei para a vida toda, de várias partes do Brasil (Piauí; Ceará; Rio Grande do Norte; Pernambuco; Paraíba; Goiás; Minas Gerais; Santa Catarina e Rio Grande do Sul) e mesmo de fora do país (Colômbia).

Aos professores do Programa de Pós-Graduação em Estatística (PPGE) da UFPE, pela formação adquirida no decorrer do Mestrado. Com eles aprendi bastante. Sempre eram didáticos e disponíveis quando necessitava tirar dúvidas. Devido ao contato direto, sou grato em especial aos mestres (na acepção mais ampla da palavra) Gauss; Francisco Cribari; Alex e Audrey.

À secretaria do PPGE, representada por Valéria e Elaine, impecáveis no atendimento e na solução de dúvidas quanto à documentação necessária desde o início. Ainda, os conselhos fornecidos também contribuíram para a conclusão do Mestrado.

A todos os amigos que fiz fora da UFPE em Recife e que me encorajaram a terminar o Mestrado. Aos professores Luanda e Henrique, os quais me ajudaram na aquisição do *netbook* no qual desenvolvi os trabalhos das disciplinas do Mestrado e boa parte do texto da dissertação enquanto estava em Recife. Aos colegas de quarto Francisco Carlos, Emerson, João Vicente, João Henrique, Robson, Gideão, Pablo, Rodrigo e Luiz Carlos pelo companheirismo e colaboração nos momentos de estudo. Aos evangelistas Lúcio, Maria (Zui) e Edilene, pelas palavras de esperança quando tudo estava difícil. Agradeço também ao professor Luiz, atualmente proprietário de um restaurante próximo à UFPE, que me ajudou muito nos finais de semana do início de Mestrado antes do 1º pagamento da bolsa. A todos eles sou eternamente grato.

Aos amigos e familiares de Juiz de Fora – MG, cidade onde resido, pelo apoio durante o Mestrado, em especial colegas e professores que conheci no ensino médio (Colégio dos Jesuítas) e na graduação (UFJF).

À Pedro Marinho, pelas dicas úteis de comandos do  $\text{\LaTeX}$  e de estilos textuais para o aprimoramento do corpo textual da dissertação.

À Tatiane Abrantes pela tradução da versão em língua espanhola do resumo.

Ao Laboratório de Amostragem e Experimentação do Departamento de Estatística da UFJF, pela permissão do uso de um computador para execução dos comandos necessários para as análises realizadas.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro durante o Mestrado.

*"Wishes can come true.  
But not if you just wait for miracles.  
Miracles are things we make for ourselves.  
Here, and now."  
(Oerba Dia Vanille, Final Fantasy XIII)*

# Resumo

Estudar os efeitos de uma política pública de transferência direta de renda ao longo do tempo é de fundamental interesse para dizer se a mesma teve ou não êxito em mudar a vida de seus beneficiários. Neste sentido, a avaliação de impacto da política no tempo de alocação para o trabalho dos beneficiários, tanto crianças quanto adultos, tem sido usada como critério para decidir sobre seu sucesso. A separação nestes dois grupos etários busca responder a duas questões importantes: se a política conseguiu reduzir o trabalho infantil e se, por outro lado, foi responsável por acomodar os adultos beneficiários. Esta avaliação pode ser feita por modelagem de regressão linear normal com uso do método de Diferença em Diferenças para a resposta de interesse (o tempo alocado para o trabalho). Logo, uma amostra de beneficiários (tratados) e não beneficiários (controles) é extraída para uma pesquisa longitudinal feita em pelo menos dois pontos distintos no tempo: um antes de aplicar a política e outro após. Porém, nem sempre os tratados e controles selecionados são comparáveis na amostra original. Para isso, é necessário parear os tratados com os controles mais próximos, dado um conjunto de características observáveis, por escores de propensão preditos por um modelo de regressão para a atribuição do tratamento (aqui, uma variável binária). Logo, os escores preditos e as observações nas covariáveis especificadas no modelo correspondente, que representam um conjunto de características observáveis, devem estar balanceados pelas duas classes na nova amostra pareada, o que é confirmado por testes de hipóteses. Ainda, ignorar características do plano amostral complexo empregado na pesquisa, como o uso de pesos amostrais desiguais, e tratar os dados como independentes e identicamente distribuídos pode enviesar a estimação do impacto pelo método de Diferença em Diferenças. Assim, a partir de dados amostrais complexos da *Yemen National Social Protection Monitoring Survey*, uma pesquisa longitudinal domiciliar, neste trabalho é avaliado o impacto do *Social Welfare Fund* para o tempo semanal de alocação para o trabalho pelo método de Diferença em Diferenças sob duas abordagens. Na primeira, o plano amostral complexo empregado é completamente ignorado. Na segunda, ele é considerado desde a etapa de pareamento até a avaliação de impacto. Ambas são aplicadas para os dois grupos etários sob estudo. Os escores são modelados por regressão logística e, após encontrar o banco pareado por eles, testes  $\chi^2$  de Pearson para a hipótese nula de independência nas covariáveis e de Mann-Whitney para homogeneidade dos escores são realizados para confirmar o balanceamento.

**Palavras-chave:** Plano amostral complexo. Pareamento por escore de propensão. Balanceamento. Diferença em diferenças.

# Abstract

Studying the effects of a governmental policy of direct cash transfer over time is of fundamental concern to say whether or not it has been successful in changing the lives of its beneficiaries. In this sense, the impact evaluation of the policy on the time allocation to work of its beneficiaries, both children and adults, has been used as criteria to decide on its success. The division in these two age groups seeks to answer two important questions: whether the policy has succeeded in reducing the child labor and, on the other hand, whether it has been responsible for making the adult beneficiaries lazier. This evaluation can be done by normal linear regression modelling with use of the Difference in Differences method for the response of interest (the allocated time to work). Therefore, a sample of beneficiaries (treated) and non-beneficiaries (controls) is extracted for a longitudinal survey carried out in at least two distinct points in time: one before applying the policy and the other after. However, the treated and controls that were selected are not always comparable in the original sample. For this, it is mandatory to match treated with the closest controls, given a set of observable characteristics, by propensity scores predicted by a regression model for the treatment assignment (here, a binary variable). Consequently, the predicted scores and observations in the covariates specified in the corresponding model, which represent a set of observable characteristics, should be balanced by the two classes in the new paired sample, which is confirmed by hypothesis tests. Moreover, ignoring the characteristics of the complex survey design used in the research, such as the presence of unequal sample weights, and treating the data as independent and identically distributed may bias the impact estimation by the Difference in Differences method. Thus, from a complex survey data from the Yemen National Social Protection Monitoring Survey, a longitudinal household survey, this essay assesses the impact of the Social Welfare Fund for the time allocation to work on a week by the Difference in Differences method under two approaches. In the first, the complex survey design employed is completely ignored. In the second, it is considered from the pairing step until the impact assessment. Both are applied for the two age groups under study. The scores are modelled by logistic regression, and after finding the database matched by them, Pearson's  $\chi^2$  tests for the null hypothesis of independence in the covariates and Mann-Whitney for homogeneity of the scores are performed to confirm the balance.

**Keywords:** Complex survey design. Propensity score matching. Balancing. Difference in differences.

# Resumen

Estudiar los efectos de una política pública de transferencia directa de renta a lo largo del tiempo es de fundamental interés para decir si ella tuvo o no éxito en cambiar la vida de sus beneficiarios. En este sentido, la evaluación del impacto de la política en el tiempo de asignación para el trabajo de los beneficiarios, tanto niños como adultos, ha sido usada como criterio para decidir sobre su éxito. La separación en estos dos grupos de edad busca contestar a dos cuestiones importantes: si la política logró reducir el trabajo infantil y si, por otro lado, fue responsable por acomodar los adultos beneficiarios. Esta evaluación puede ser hecha por modelaje de regresión lineal normal con uso del método de Diferencia en Diferencias para la respuesta de interés (el tiempo asignado para el trabajo). Luego, una muestra de beneficiarios (tratados) y no beneficiarios (controles) es extraída para una investigación longitudinal hecha en por lo menos dos puntos distintos en el tiempo: uno antes de aplicar la política y otro después. Sin embargo, ni siempre los tratados y controles seleccionados son comparables en la muestra original. Para ello, es necesario emparejar los tratados con los controles más próximos, dado un conjunto de características observables, por puntajes de propensión predichos por un modelo de regresión para la atribución del tratamiento (aquí, una variable binaria). Luego, los puntajes predichos y las observaciones en las covariables especificadas en el modelo correspondiente, que representan un conjunto de características observables, deben estar balanceados para las dos clases en la nueva muestra emparejada, lo que está confirmado por pruebas de hipótesis. Aún, ignorar características del plano de muestra complejo empleado en la investigación, como el uso de pesos muestrales desiguales, y tratar los datos como independientes y igualmente distribuidos puede sesgar la estimación del impacto por el método de Diferencia en Diferencias. De esta manera, a partir de datos de la muestra compleja de la Yemen National Social Protection Monitoring Survey, una investigación longitudinal domiciliar, en este trabajo se evalúa el impacto del Social Welfare Fund en una franja semanal de tiempo de asignación para el trabajo por el método de Diferencia en Diferencias sobre dos abordajes. En el primero, el plano de muestra complejo empleado es completamente ignorado. En el segundo, el plano amostral está considerado desde la etapa de emparejamiento hasta la evaluación de impacto. Ambos son aplicados para los dos grupos de edad bajo estudio. Los puntajes son modelados por regresión logística y, tras encontrar el banco pareado por ellos, pruebas  $\chi^2$  de Pearson para la hipótesis nula de independencia entre las covariables y de Mann-Whitney para homogeneidad de los puntajes son realizados para confirmar el balanceo.

**Palabras clave:** Plano de muestra complejo. Emparejamiento por puntaje de propensión. Balanceo. Diferencia en diferencias.

# Lista de ilustrações

Figura 1 – Mapa Político do Iêmen à época da YNSPMS (Ferraz & Vieira, 2013) . . . . .	38
Figura 2 – Densidade da variável TSAT nos adultos, caso IID, 1ª rodada. . . . .	197
Figura 3 – Densidade da variável TSAT nos adultos, caso IID, 4ª rodada. . . . .	197
Figura 4 – Densidade da variável TSAT nas crianças, caso IID, 1ª rodada. . . . .	198
Figura 5 – Densidade da variável TSAT nas crianças, caso IID, 4ª rodada. . . . .	198
Figura 6 – Densidade da variável TSAT nos adultos, considerando o PAC da YNSPMS, 1ª rodada. . . . .	199
Figura 7 – Densidade da variável TSAT nos adultos, considerando o PAC da YNSPMS, 4ª rodada. . . . .	199
Figura 8 – Densidade da variável TSAT nas crianças, considerando o PAC da YNSPMS, 1ª rodada. . . . .	200
Figura 9 – Densidade da variável TSAT nas crianças, considerando o PAC da YNSPMS, 4ª rodada. . . . .	200
Figura 10 – Frequência dos escores preditos pelo modelo inicial para o banco original das crianças, caso IID. . . . .	201
Figura 11 – Frequência dos escores preditos pelo modelo inicial para o banco original dos adultos, caso IID. . . . .	202
Figura 12 – Frequência dos escores preditos pelo modelo inicial para o banco original das crianças, considerando o PAC da YNSPMS. . . . .	202
Figura 13 – Frequência dos escores preditos pelo modelo reduzido para o banco original das crianças, considerando o PAC da YNSPMS. . . . .	203
Figura 14 – Frequência dos escores preditos pelo modelo inicial para o banco original dos adultos, considerando o PAC da YNSPMS. . . . .	203
Figura 15 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado das crianças, caso IID. . . . .	204
Figura 16 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado dos adultos, caso IID. . . . .	204
Figura 17 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado das crianças, considerando o PAC da YNSPMS. . . . .	205
Figura 18 – Frequência por classe dos escores preditos pelo modelo reduzido para o banco pareado das crianças, considerando o PAC da YNSPMS. . . . .	205
Figura 19 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado dos adultos, considerando o PAC da YNSPMS. . . . .	205

# Lista de tabelas

Tabela 1	– Resultados do teste da RV no modelo completo para o escore de propensão no grupo das crianças, caso IID . . . . .	101
Tabela 2	– Resultados do teste da RV no modelo apenas com DRC (além do intercepto) para o escore de propensão no grupo das crianças, caso IID	102
Tabela 3	– Estimativas para os coeficientes no modelo apenas com DRC (além do intercepto) para o escore de propensão no grupo das crianças, caso IID	102
Tabela 4	– Resultados do teste da RV no modelo completo para o escore de propensão no grupo dos adultos, caso IID . . . . .	104
Tabela 5	– Resultados do teste da RV no modelo com API; DRFP; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, caso IID . . . . .	104
Tabela 6	– Estimativas para os coeficientes no modelo com API; DRFP; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, caso IID . . . . .	105
Tabela 7	– Resultados do teste da RP no modelo completo para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS . . . . .	106
Tabela 8	– Resultados do teste da RP no modelo apenas com DRC e CRDAI (além do intercepto) para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS . . . . .	107
Tabela 9	– Estimativas para os coeficientes no modelo com DRC e CRDAI (além do intercepto) para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS . . . . .	108
Tabela 10	– Resultados do teste da RP no modelo completo para o escore de propensão no grupo dos adultos, considerando o PAC da YNSPMS . . . . .	109
Tabela 11	– Resultados do teste da RP no modelo com API; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, considerando o PAC da YNSPMS . . . . .	109
Tabela 12	– Estimativas para os coeficientes no modelo com API; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, considerando o PAC da YNSPMS . . . . .	110
Tabela 13	– Resultados dos testes $\chi^2$ de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo das crianças, caso IID . . . . .	112
Tabela 14	– Resultados dos testes $\chi^2$ de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo dos adultos, caso IID . . . . .	113

Tabela 15 – Resultados dos testes $\chi^2$ de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo das crianças, considerando o PAC da YNSPMS . . . . .	114
Tabela 16 – Estimativas para os coeficientes no modelo apenas com DRC (além do intercepto) para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS . . . . .	115
Tabela 17 – Resultados dos testes $\chi^2$ de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo reduzido no grupo das crianças, considerando o PAC da YNSPMS . . . . .	115
Tabela 18 – Resultados dos testes $\chi^2$ de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo dos adultos, considerando o PAC da YNSPMS . . . . .	116
Tabela 19 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo das crianças, caso IID . . . . .	120
Tabela 20 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo dos adultos, caso IID . . . . .	122
Tabela 21 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo das crianças, considerando o PAC da YNSPMS . . . . .	124
Tabela 22 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo dos adultos, considerando o PAC da YNSPMS . . . . .	128

# Lista de abreviaturas e siglas

AAS	Amostragem Aleatória Simples sem Reposição
AAScR	Amostragem Aleatória Simples com Reposição
ABNT	Associação Brasileira de Normas Técnicas
AC	Amostragem por Conglomerados
AE	Área de Enumeração
AES	Amostragem Estratificada Simples
AI	Adulto é Idoso
AMCD	Adulto é Mulher e Chefe do Domicílio
API	Adulto é Portador de Invalidez
APS	Amostragem Poisson Sequencial
AR	Amostragem Representativa
BDH	<i>Bono de Desarrollo Humano</i>
COM	Criança é Orfã de Mãe
COP	Criança é Orfã de Pai
CPI	Criança é Portadora de Invalidez
CRDAI	Criança Reside em Domicílio com Adulto Idoso
CRDAPI	Criança Reside em Domicílio com Adulto Portador de Invalidez
CRDCM	Criança Reside em Domicílio Chefiado por Mulher
CSS	<i>Comprehensive Social Survey</i>
DD	Diferença em Diferenças
DRC	Domicílio Recebe Caridade
DRFP	Domicílio Recebe Fundos de Pensão
DRR	Domicílio Recebe Remessas

EC	Escolaridade da Criança
EC1	Estrato de Controle I
EC2	Estrato de Controle II
EMC	Escolaridade da Mãe da Criança
EMT	Efeito Médio do Tratamento
ENHT	Estimador(es) de Narain-Horvitz-Thompson
ENCEL	<i>Encuesta de Evaluación de los Hogares</i>
EP	Estrutura Populacional
EPA	Efeito do Plano Amostral
EPAA	Efeito do Plano Amostral Ampliado
EPC	Escolaridade do Pai da Criança
EQM	Erro Quadrático Médio
ET	Estrato de Tratamento
GC	Gênero da Criança
GL	Graus de Liberdade
HBS	<i>Household Budget Survey</i>
IC	Idade da Criança
IID	Independentes e Identicamente Distribuídas
IMF	<i>International Monetary Fund</i>
IPC-IG	<i>International Policy Centre for Inclusive Growth</i>
Ipea	Instituto de Pesquisa Econômica Aplicada
ISI	<i>International Statistical Institute</i>
LD	Localização do Domicílio
LST	Linearização em Série de Taylor
MCPE	Matrícula da Criança em Programa Educaional
MFA	<i>Más Familias en Acción</i>

MLG	Modelo(s) Linear(es) Generalizado(s)
MM	Mudança em Mudanças
MP	Máxima Pseudoverossimilhança
MPL	Modelo de Probabilidade Linear
MQO	Mínimos Quadrados Ordinários
MQP	Mínimos Quadrados Ponderados
MR	<i>Missing at Random</i>
MV	Máxima Verossimilhança
MW	Mann-Whitney
NPD	Nível de Pobreza do Domicílio
OIT	Organização Internacional do Trabalho
PAC	Plano Amostral Complexo
PATFI	Pressuposto da Atribuição do Tratamento Fortemente Ignorável
PBE	Programa Bolsa Escola
PBF	<i>Programa Bolsa Família</i>
PCD	Pessoas por Cômodo no Domicílio
PDE	Pareamento pela Distância Euclidiana
PDM	Pareamento pela Distância de Mahalanobis
PI	Pareamento por Intervalo
PIB	Produto Interno Bruto
PMT	<i>Proxy Means Test</i>
PNAD	Pesquisa Nacional por Amostra de Domicílios
PPT	Probabilidades Proporcionais ao Tamanho
PTC	Pressuposto da Tendência Comum
QDPC	Quantidade Descritiva Populacional Correspondente
RAM	<i>Random-access Memory</i>

RB	Replicação <i>Bootstrap</i>
RC	Rodada de Coleta
RP	Razão de Pseudoverossimilhanças
RRB	Replicação Repetida Balanceada
RRJ	Replicação Repetida <i>Jackknife</i>
RSWF	Recebimento do <i>Social Welfare Fund</i>
RV	Razão de Verossimilhanças
SPM	<i>Social Protection Monitoring</i>
SWF	<i>Social Welfare Fund</i>
TCL	Teorema Central do Limite
TD	Topografia do Domicílio
TSAT	Tempo Semanal de Alocação para o Trabalho
UNICEF	<i>United Nations Children's Emergency Fund</i>
UNDP	<i>United Nations Development Programme</i>
UPA	Unidade Primária de Amostragem
USA	Unidade Secundária de Amostragem
VP	Viés Padronizado
WSIWG	<i>What you See Is What you Get</i>
YNSPMS	<i>Yemen National Social Protection Monitoring Survey</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>20</b>
1.1	Organização da Dissertação	24
1.2	Revisão da Literatura em Amostragem	26
1.3	Revisão da Literatura em Avaliação de Impacto	30
1.4	O Iêmen e a YNSPMS	38
1.4.1	O Iêmen: Aspectos Demográficos, Socioeconômicos e Políticos	38
1.4.2	A YNSPMS: Principais Características	40
1.4.3	O Plano Amostral da YNSPMS	42
1.5	Suporte Computacional	44
1.5.1	L <sup>A</sup> T <sub>E</sub> X	45
1.5.2	R	46
<b>2</b>	<b>ESTIMAÇÃO DOS ESCORES DE PROPENSÃO</b>	<b>47</b>
2.1	Definição de Controles e Tratados	47
2.2	O Escore de Propensão	48
2.3	Escolha da Distribuição para o Escore de Propensão	52
2.3.1	Modelos Lineares Generalizados	53
2.3.2	Estimação por Máxima Verossimilhança em MLG	54
2.3.3	Estimação por Máxima Pseudoverossimilhança em MLG	57
2.4	Crítérios para Seleção de Covariáveis	59
2.5	Estimação de Variâncias sob o PAC	64
2.5.1	Linearização em Série de Taylor	65
2.5.2	Métodos de Reamostragem Não-paramétricos	67
2.6	Avaliação da Influência do Plano Amostral	68
2.6.1	O EPA de Kish (1965)	68
2.6.2	O EPAA de Skinner <i>et al.</i> (1989)	69
<b>3</b>	<b>METODOLOGIA PARA AVALIAÇÃO DE IMPACTO</b>	<b>72</b>
3.1	O Pareamento pela Distância Euclidiana	73
3.2	Balanceamento dos Indivíduos Pareados	75
3.2.1	O Teste Qui-quadrado de Pearson	77
3.2.2	O Teste da Soma dos Postos de Mann-Whitney	82
3.3	O Método de Diferença em Diferenças	87
3.3.1	História do Método DD	88
3.3.2	Suposições de Identificação no Método DD	89
3.3.3	Modelagem no Método DD	93

3.3.4	Combinando o Método DD ao Pareamento do Escore . . . . .	95
<b>4</b>	<b>RESULTADOS EMPÍRICOS . . . . .</b>	<b>97</b>
4.1	Modelos para os Escores sem e com PAC na YNSPMS . . . . .	97
4.2	Pareamento e Balanceamento Pós-modelagem . . . . .	111
4.3	Avaliação do Impacto do SWF em TSAT sem e com PAC . . . . .	117
<b>5</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>131</b>
	<b>Referências . . . . .</b>	<b>135</b>
	 <b>APÊNDICES</b>	 <b>144</b>
	<b>APÊNDICE A – CÁLCULO DOS PESOS AMOSTRAIS . . . . .</b>	<b>145</b>
A.1	Cálculo dos Pesos Amostrais Básicos . . . . .	145
A.2	Cálculo dos Pesos Transversais . . . . .	147
A.3	Cálculo dos Pesos Longitudinais . . . . .	148
	 <b>APÊNDICE B – CÓDIGOS COMPUTACIONAIS PARA O R . . . . .</b>	 <b>150</b>
B.1	<b>Código para o Cálculo de Estatísticas Descritivas . . . . .</b>	<b>152</b>
B.1.1	Frequência de Controles e Tratados pela Variável RSWF . . . . .	152
B.1.2	Frequência de Respostas para a Variável TSAT . . . . .	154
B.2	<b>Código para o Cálculo de Estimativas . . . . .</b>	<b>158</b>
B.2.1	Ajuste de Modelos para o Escore de Propensão sem PAC . . . . .	158
B.2.2	Ajuste de Modelos para o Escore de Propensão com PAC . . . . .	168
B.2.3	Pareamento e Balanceamento dos Modelos Ajustados sem PAC . . . . .	172
B.2.4	Pareamento e Balanceamento dos Modelos Ajustados com PAC . . . . .	175
B.2.5	Avaliação de Impacto do SWF em TSAT sem PAC . . . . .	181
B.2.6	Avaliação de Impacto do SWF em TSAT com PAC . . . . .	188
	 <b>ANEXOS</b>	 <b>196</b>
	<b>ANEXO A – HISTOGRAMAS DA DENSIDADE DA VARIÁVEL TSAT . . . . .</b>	<b>197</b>
	<b>ANEXO B – HISTOGRAMAS DA FREQUÊNCIA DOS ESCORES PREDITOS PARA O BANCO ORIGINAL . . . . .</b>	<b>201</b>
	<b>ANEXO C – HISTOGRAMAS DA FREQUÊNCIA DOS ESCORES PREDITOS PARA O BANCO PAREADO . . . . .</b>	<b>204</b>

# 1 Introdução

A Amostragem é uma área importante dentro da Estatística, muito utilizada em pesquisas transversais ou longitudinais quando o objetivo é analisar uma ou mais características a partir de variáveis avaliadas para uma dada população finita (não necessariamente composta por pessoas). Cada observação ou unidade amostral desta população possuirá um valor específico para cada variável pesquisada, mas não é possível realizar uma pesquisa censitária por alguma razão (como restrição orçamentária). Pesquisas amostrais são amplamente utilizadas por órgãos estatísticos oficiais para definir a adoção de políticas públicas e por instituições acadêmicas e privadas, em geral para analisar a opinião pública sobre determinado assunto. Não apenas estatísticos, mas também economistas, sociólogos e psicólogos utilizam bastante as ferramentas da Amostragem. Como principais vantagens da pesquisa amostral, Cochran (1977) elenca:

- Custo reduzido: se os dados são coletados apenas para uma pequena fração da população, claramente os gastos para a realização de uma pesquisa amostral serão menores do que se a mesma fosse censitária. A Amostragem permite que populações com milhões de observações tenham suas informações resumidas por alguns milhares;
- Tempo reduzido: pelas mesmas razões expostas no ponto anterior, a pesquisa amostral terá execução bem mais rápida do que a censitária e, com a velocidade dos computadores modernos, a aplicação de técnicas amostrais é praticamente tão veloz quanto a obtenção de estatísticas descritivas;
- Escopo ampliado: como consequência dos pontos acima, a Amostragem permite contratar e treinar melhor tanto os responsáveis por ir a campo quanto quem fará a análise dos dados obtidos, aumentando a qualidade da informação obtida e dos resultados publicados.

A característica sob análise em uma pesquisa amostral é quantificada por uma estatística, em geral a média; o total ou a proporção de uma quantidade extraída de cada observação. Logo, é função apenas dos dados coletados na amostra (ou seja, de valores para uma ou mais variáveis) e possivelmente de outras quantidades conhecidas (como o tamanho amostral). Assim, a estatística encontrada para uma amostra extraída de sua respectiva população, na grande maioria dos casos, não coincidirá com o seu análogo (denominado parâmetro) obtido para uma pesquisa censitária, sendo a sua *estimativa pontual*.

Por sua vez, a precisão da estimativa pontual é dada pelo erro-padrão associado, uma estimativa do desvio-padrão (raiz quadrada da variância) da estimativa pontual. Com o erro-padrão, é possível construir um intervalo de confiança para o parâmetro

estimado. Tal intervalo contém a estimativa pontual  $a$  e é denotado por  $IC(a, \gamma\%)$ , onde  $\gamma$  é um número positivo entre zero e cem (exclusive<sup>1</sup>) que fornecerá o nível de confiança do intervalo. Ou seja, se pudéssemos extrair  $k$  amostras diferentes da mesma população, todas com o mesmo tamanho, e construir intervalos de forma similar em cada amostra,  $\gamma\%$  destes intervalos conterão o valor verdadeiro do parâmetro. Em geral, são usados os níveis 90%; 95% e 99% de confiança. Usando um estimador consistente, o intervalo resultante pode ser calculado a partir dos quantis da distribuição  $t$  de Student, cujos graus de liberdade dependerão do tamanho amostral e da metodologia usada para estimar as variâncias das estimativas pontuais (Heeringa *et al.*, 2010).

Definido o objetivo da pesquisa (transversal ou longitudinal) e quais estatísticas iremos obter da amostra a ser extraída da população, é necessário antes de ir a campo delinear um plano amostral. Toda pesquisa amostral precisa de um protocolo de instruções para definir como selecionaremos os elementos da população para a amostra. Este protocolo é o plano amostral. Um plano amostral permitirá realizar inferência para as características de interesse apenas se o mesmo atribuir probabilidade de seleção não-nula a todos os elementos da população, suposta finita e de tamanho  $N$  (diferença crucial em relação à inferência clássica). Neste caso, a amostragem é dita probabilística. Definido o conjunto  $\mathcal{S} = \{s_1, \dots, s_L\}$  de todas as  $L = 2^N$  amostras possíveis que podem ser extraídas a uma margem de erro fixada (esta dependerá do tamanho e plano amostrais escolhidos), cada uma das amostras  $s_l$ ,  $l = 1, \dots, L$  terá a sua probabilidade associada  $p(s_l)$ . O plano amostral é, portanto, o mecanismo aleatório que determina as probabilidades  $p(s_l)$ ,  $\forall$  (Särndal *et al.*, 1992).

Como nada é suposto acerca da distribuição dos dados na população, qualquer inferência realizada sob amostragem probabilística será válida apenas para a população finita da qual a amostra foi extraída. Se o objetivo for uma análise mais ampla, que permita estabelecer associações entre duas ou mais variáveis e mesmo relações de causalidade a partir dos dados amostrais, é necessário adotar a abordagem de modelo para a superpopulação. Nela, supomos que o processo estocástico gerador dos dados é um modelo, o qual pode gerar uma sequência de populações finitas, dentre as quais aquela da qual nossa amostra  $s_l$  foi extraída. Tomando uma função  $h(\cdot)$  dos valores desconhecidos  $y_1, \dots, y_N$  na população (um exemplo é a média), é possível inferir sobre os parâmetros do modelo utilizando métodos já existentes para a estimação sob amostragem probabilística de  $h(y_1, \dots, y_N)$ , a quantidade descritiva populacional que passa a ser o novo parâmetro-alvo. Assim, adicionamos a informação sobre o plano amostral contida em  $p(s_l)$  ao modelo que gerou a população finita (Pessoa & Silva, 1998).

<sup>1</sup> Nosso intervalo não pode ser apenas a estimativa pontual (0%) nem cobrir todos os valores possíveis para a mesma (100%).

O mais básico dentre todos os planos amostrais é a Amostragem Aleatória Simples com Reposição (AAScR). Nela, todos os elementos da população têm a mesma probabilidade de seleção e, uma vez amostrados, podem entrar novamente no processo de seleção. Logo, é possível que um mesmo elemento seja selecionado mais de uma vez. Quando o tamanho populacional  $N$  é muito superior comparado ao amostral  $n$ , a AAScR é bem semelhante à Amostragem Aleatória Simples sem Reposição (AASsR, a qual será denotada daqui em diante pela forma reduzida, AAS). Ao contrário da AAScR, na AAS o elemento selecionado não pode entrar novamente no processo de seleção da amostra. Além disso, a AAS exige um conhecimento prévio de todos os elementos da população através de um cadastro. Finalmente, a variância estimada sob um plano AAS aplica um fator de correção para populações finitas, dado por  $1 - n/N$ . Se  $N \gg n$ , então este fator é aproximadamente igual a 1 e por isso se diz, nesta situação, que a AAS é similar à AAScR e à suposição de que as observações são Independentes e Identicamente Distribuídas (IID) na amostra (Lohr, 2010).

Apesar de sua simplicidade, não existem na prática levantamentos amostrais que façam uso exclusivo da AAS em um único estágio. Uma pesquisa amostral retrata uma parte da população em um instante fixado, ou seja, a variável analisada pode mudar em questão de meses para a mesma unidade amostral (como o gasto percentual da renda total de um domicílio com alimentação). Além disso, o orçamento disponível pode não permitir o deslocamento de uma unidade amostral até outra em pouco tempo (razão pela qual a conglomeração é muito utilizada). Mesmo quando a variável é estática por um longo período de tempo ou o suficiente para que o orçamento permita viagens entre todas as unidades amostrais em um período menor, é preferível utilizar estratificação simples para melhorar a precisão das estimativas pontuais (os intervalos de confiança serão mais estreitos). Pode ser necessário também corrigir os pesos amostrais (inicialmente dados pelo inverso das probabilidades de seleção) pela não-resposta unitária (quando uma unidade amostral não responde por algum motivo à pesquisa) e/ou por métodos que incorporem a informação obtida com a amostra, como pós-estratificação e calibração (novamente, para melhorar a precisão) (Heeringa *et al.*, 2010).

Denominamos Plano Amostral Complexo (PAC) qualquer metodologia de Amostragem que envolva um ou mais recursos adicionais além da AAS para selecionar os elementos da população para a amostra, incluindo estratificação; conglomeração; probabilidades desiguais de seleção e amostragem em múltiplos estágios e/ou fases. Atualmente, ainda é bastante comum muitos analistas de dados ignorarem parcial ou completamente o PAC na obtenção das estimativas pontuais e dos intervalos de confiança, supondo que a amostra foi obtida usando apenas AAS. A inferência traçada desta forma pode conduzir a conclusões incorretas, por obter estimativas pontuais em geral enviesadas e com precisão sub ou

superestimada (os intervalos são mais largos ou estreitos do que deveriam)<sup>2</sup>.

Nesta dissertação, o objetivo geral é saber o que acontece quando indevidamente ignoramos o PAC de fato utilizado na avaliação do impacto para um dado tratamento binário, em uma variável resposta de interesse, através do método de Diferença em Diferenças (DD) com pareamento por escore de propensão tomando a distância euclidiana. Conseqüentemente, também é de nosso interesse verificar o quanto a decisão sobre ignorar ou considerar o PAC pode influenciar no poder preditivo dos modelos ajustados para o escore de propensão. Para isso, consideramos dados amostrais complexos da *Yemen National Social Protection Monitoring Survey* (YNSPMS, em português: Pesquisa Nacional de Monitoramento da Proteção Social no Iêmen), uma pesquisa longitudinal domiciliar realizada em quatro rodadas entre 2012 e 2013 pelo IPC-IG (*International Policy Centre for Inclusive Growth*, em português: Centro Internacional de Políticas para o Crescimento Inclusivo) em parceria com o UNDP (*United Nations Development Programme*, em português: Programa das Nações Unidas para o Desenvolvimento) e o governo iemenita (YNSPMS Final Report, 2014a). Nossa variável resposta de interesse é o tempo de alocação para o trabalho nos últimos 7 dias (em horas), tomados como base para uma semana completa. No dicionário de variáveis da YNSPMS, esta variável é denominada `p6a_q_10` (Banco de Dados da YNSPMS, 2014c), mas será denotada de agora em diante por TSAT (Tempo Semanal de Alocação para o Trabalho).

O tratamento para o qual desejamos avaliar o impacto em TSAT é o *Social Welfare Fund* (SWF, em português: Fundo de Bem-estar Social), uma política pública (governamental) de transferência direta de renda. Assim, a avaliação de impacto será dividida em duas partes: em um primeiro momento, supomos que as observações nas variáveis pesquisadas na YNSPMS foram amostradas por AAS, e depois, incorporamos o PAC utilizado através dos pesos amostrais longitudinais para os indivíduos pesquisados nas quatro rodadas. O impacto do SWF será avaliado da 1<sup>a</sup> à 4<sup>a</sup> rodada, das quais tomamos as observações na resposta. Mostraremos que as inferências realizadas nestas duas situações divergem bastante entre si, o que é uma evidência da necessidade de se considerar o PAC também em análises de avaliação do impacto de tratamentos binários, como políticas públicas de transferência direta de renda.

Ressaltamos ainda que inicialmente parearemos indivíduos que receberam o tratamento (tratados) com aqueles que não o receberam (controles) pelo escore de propensão, modelado por um dado conjunto de covariáveis. Verificaremos se as covariáveis utilizadas e os escores preditos estão balanceados no banco pareado, ou seja, se os pares de tratados e controles são similares nas observações das covariáveis e do escore predito, através de testes de hipóteses. Após confirmar o balanceamento, é feita no banco pareado a avaliação de

<sup>2</sup> Em geral, a estratificação produz ganhos de precisão, especialmente se os elementos dentro de cada estrato definido forem homogêneos entre si. Por outro lado, o uso da conglomeração e de pesos desiguais leva à produção de maiores erros-padrão das estimativas obtidas.

impacto pelo método DD, utilizando potenciais covariáveis de confundimento. Para todas estas etapas, dividimos as análises para os casos sem e com a incorporação do PAC. Os cálculos necessários serão feitos através do *software* estatístico R (R Core Team, 2016b).

Além deste objetivo metodológico, também há uma motivação substantiva a ser desenvolvida. Avaliaremos o impacto do SWF para duas subpopulações ou grupos construídos com base na idade dos indivíduos pesquisados. O primeiro grupo é constituído por todos aqueles com idade entre 6 e 14 anos de idade (inclusive) na 1<sup>a</sup> rodada da YNSPMS, o qual denominaremos como grupo das crianças. O segundo, pelos maiores de 14 anos na mesma rodada, sendo o grupo dos adultos. Esta divisão busca respeitar a definição de idade mínima para admissão em emprego dada pela Convenção 138 da Organização Internacional do Trabalho (OIT) (UNICEF Brasil, 2016, artigo 2<sup>o</sup>, parágrafo 3). Ainda, a variável TSAT possui observações para os indivíduos acima de 4 anos. No primeiro grupo etário, queremos verificar se o SWF contribuiu ou não de forma significativa (dado um nível  $\gamma$  de confiança) para a redução do trabalho infantil. No segundo, se contribuiu ou não de forma significativa para a redução da jornada de trabalho ou mesmo para um possível “efeito preguiça”.

## 1.1 Organização da Dissertação

A presente dissertação é composta de cinco capítulos. Neste capítulo introdutório, já apresentamos ao leitor um breve resumo do que será investigado. É abordada aqui uma revisão da literatura na área de Amostragem até os avanços mais recentes. Também serão revistos estudos de avaliação do impacto de políticas de transferência direta de renda (como o SWF) no tempo de alocação para o trabalho, tanto para crianças quanto adultos. Ainda, são detalhados a YNSPMS, o contexto no qual foi realizada e seu plano amostral; bem como os programas computacionais (*software*) utilizados nesta dissertação.

No segundo capítulo abordaremos, sem deixar de inserir no contexto da YNSPMS, os aspectos teóricos da estimação do vetor dos escores de propensão para o caso de observações IID e também sob a ótica da Amostragem. Logo, apresentaremos a maneira como os pesos amostrais são incluídos na modelagem dos escores, já que é necessário considerar o PAC em todas as etapas da avaliação de impacto do SWF se não quisermos ignorá-lo. Como definiremos uma classe de indivíduos tratados e outra de controles, em que cada indivíduo é um residente entrevistado, para ambos os grupos etários ajustaremos modelos de regressão com variável dependente categórica binária para produzir os escores de propensão. Também serão definidos os critérios para inclusão de covariáveis na modelagem. As variâncias dos coeficientes (estimativas pontuais dos parâmetros) associados a cada covariável serão aproximadas pela Linearização em Série de Taylor (LST), embora outros possíveis métodos sejam discutidos de forma mais breve. Por fim, será abordada a influência do PAC através

dos conceitos de Efeito do Plano Amostral (EPA) e Efeito do Plano Amostral Ampliado (EPAA). Como não dispomos de dados populacionais para a YNSPMS, decidiremos pelo uso descritivo do EPAA, o qual veremos que pode utilizar as informações da amostra.

No terceiro capítulo discutiremos o pareamento pelo escore de propensão, o qual pode tomar um ou mais controles comparáveis para cada indivíduo tratado. Assim, enunciaremos qual algoritmo dentre os disponíveis na literatura será utilizado para realizá-lo e como verificar se o pareamento feito é balanceado usando testes de hipóteses não-paramétricos para independência (em covariáveis categóricas) e homogeneidade (em covariáveis contínuas e no escore predito). Em seguida, abordaremos o método DD restrito aos indivíduos pareados pelo escore de propensão predito. Veremos, através de histogramas da densidade, que a variável na qual avaliaremos o impacto do SWF (TSAT) pode assumir dezenas de valores distintos para os indivíduos de ambos os grupos etários, sem possuir assimetria ou curtose extremas. Isso ocorre tanto no caso IID quanto no caso em que consideramos o PAC. Assim, TSAT pode ser tratada como aproximadamente contínua, embora seja de fato discreta. Desta forma, o método DD se valerá dos pressupostos do modelo de regressão linear normal com função de ligação identidade (em razão de limitações de *software* na opção de distribuições contínuas para dados amostrais complexos), com coeficientes para variáveis indicadoras de atribuição do tratamento (tratado ou controle), rodada (1<sup>a</sup> ou 4<sup>a</sup>), interação entre ambos e possivelmente também para covariáveis de confundimento. Assim como a modelagem do escore de propensão e qualquer outra análise inferencial, a avaliação de impacto pelo método DD pode ser feita ignorando ou incorporando o PAC efetivamente utilizado. Por isso, a teoria descrita contemplará ambas as perspectivas, como feito no capítulo 2.

No quarto capítulo são apresentados os resultados da avaliação de impacto do SWF na variável TSAT para ambos os grupos etários (crianças e adultos), com base na teoria vista nos dois capítulos anteriores. Isto será feito tanto para a situação em que ignoramos o PAC quanto para aquela em que este é incorporado ao processo de estimação. A partir disto, é feita também uma análise da influência do PAC na estimação do impacto avaliado para cada grupo etário considerado. Com isto saberemos o que muda nas conclusões para a variável resposta (em ambos os grupos etários) tomando o EPAA calculado ainda na etapa de modelagem dos escores de propensão e também no final da estimação do impacto pelo método DD. Logo, o EPAA será nosso referencial para descrever o quanto a inferência realizada muda ao se ignorar o PAC. Por fim, também é feita uma comparação com os estudos anteriores revistos no capítulo introdutório.

Finalmente, no quinto e último capítulo são feitas as considerações finais, reunindo as conclusões obtidas no capítulo anterior. Ainda, são traçadas perspectivas e desafios para o futuro da pesquisa em avaliação de impacto com pareamento por escore de propensão quando dados amostrais complexos estão presentes.

## 1.2 Revisão da Literatura em Amostragem

A Amostragem como conhecemos atualmente teve início no final do século XIX com o artigo pioneiro de Kiaer (1895), o qual introduziu o método de Amostragem Representativa (AR). Em sua época foi fortemente criticado por muitos produtores de estatísticas governamentais, especialmente aqueles que compunham o ISI (*International Statistical Institute*, em português: Instituto Internacional de Estatística) por ser o primeiro método a sugerir o cálculo de medidas, para uma dada população, utilizando apenas uma fração de seus elementos. Apesar da luta de Kiaer em convencê-los da validade do método, ele veio a ser reconhecido como científico apenas em 1925 no encontro do ISI em Roma, após o seu falecimento (Kruskal & Mosteller, 1980).

No início do século XX, Bowley (1906) e Fisher (1925) desenvolveram o papel da aleatorização na seleção amostral e métodos para estimação e inferência estatística em planos amostrais com AAS ou estratificação simples. Reunindo tais contribuições, Neyman (1934) definiu uma estrutura para estimar intervalos de confiança (sendo o primeiro a usar esta denominação<sup>3</sup>) para quantidades populacionais. Tais intervalos provinham da distribuição de probabilidade (a posteriori) destas mesmas quantidades em amostras repetidas da população dada (cuja distribuição *a priori* era desconhecida) sem a necessidade da aplicação do teorema de Bayes, ou seja, sem supor nada sobre a distribuição *a priori*. Segundo Neyman, o trabalho de Fisher forneceu uma base matemática para a AR. No mesmo artigo, Neyman também introduziu a estratificação com alocação ótima e a seleção com probabilidades desiguais.

Após o artigo de Neyman, em especial no período após a 2ª Guerra Mundial, surgiram vários novos trabalhos em Amostragem, focados na estimação pontual para medidas populacionais e para a respectiva variância (erro-padrão), necessária para a construção dos intervalos de confiança. Eles também abordavam os seguintes problemas:

- Quais as possíveis fontes de erro em pesquisas amostrais;
- Como lidar com a complexidade da amostra de forma que as estimações feitas não fossem enviesadas;
- Desenvolvimento de novos estimadores consistentes, dentre os quais se destacam os estimadores de Narain-Horvitz-Thompson (ENHT)<sup>4</sup>, de razão e de regressão;
- Como medir a influência de um PAC (EPA e EPAA) nas estimativas obtidas;

As principais contribuições foram reunidas pelos livros de Deming (1950); Hansen

<sup>3</sup> Embora tais intervalos já fossem definidos antes por Fisher, que os chamou limites fiduciais do parâmetro estudado, Neyman preferiu denominá-los intervalos de confiança uma vez que a palavra *fiducial* já era utilizada em outro contexto dentro da Estatística (probabilidade fiducial).

<sup>4</sup> Na literatura de Amostragem é comum citar tal estimador como de Horvitz & Thompson (1952), mas por questão de justiça optou-se por incluir Narain (1951) devido a este ter sido seu primeiro idealizador, enquanto Horvitz & Thompson mostraram que o estimador criado por Narain é o mais eficiente para algumas classes de estimadores.

*et al.* (1953); Sukhatme (1954); Yates (1960); Kish (1965); Cochran (1977); Skinner *et al.* (1989) e Särndal *et al.* (1992). Ainda hoje, as técnicas desenvolvidas por estes autores permanecem como base para quase todas as análises descritivas de dados amostrais.

A partir dos anos 1950, o tratamento analítico dos dados amostrais expandiu consideravelmente com novos desenvolvimentos na teoria estatística e a introdução dos métodos de modelagem, como modelos log-lineares; modelos lineares generalizados (como a regressão logística); modelos de análise de sobrevivência; modelos lineares generalizados mistos; modelos de equações estruturais e modelos de variável latente. Muitos dos modelos citados aplicam o método de Máxima Verossimilhança (MV) para estimar seus respectivos parâmetros (tanto a estimativa pontual quanto a intervalar). Ou seja, partem do pressuposto de que as observações amostrais da medida a ser analisada são independentes e seguem uma mesma distribuição (por exemplo, binomial; multinomial; normal ou gama), ou seja, um único processo gerador de dados para toda a população da qual a amostra foi extraída. Logo, não consideram em princípio nenhuma informação sobre o PAC (quando for o caso) do qual a amostra foi extraída para inferência sobre os parâmetros da população (Heeringa *et al.*, 2010).

Esta abordagem inferencial é dita *baseada no modelo*. Se as observações na população de fato seguem um modelo pré-especificado, então as probabilidades de inclusão (ou seu inverso, os pesos amostrais) determinadas pelo PAC não terão efeito na estimação dos parâmetros. Assim, o principal desafio para a inferência baseada no modelo é a especificação deste. Deve-se então verificar todos os pressupostos do modelo cautelosamente bem como a qualidade do ajuste para os dados da amostra através de gráficos dos resíduos *versus* valores preditos para a variável dependente (por exemplo) e testes de diagnóstico para identificar possíveis pontos aberrantes (em inglês, *outliers*). Como a inferência para as observações não selecionadas na amostra são baseadas unicamente no modelo especificado, é assumido que este pode descrevê-las, o que é um pressuposto muito forte (Lohr, 2010).

Usar inferência baseada no modelo, e somente neste, é de grande interesse, pois permite conectar com maior facilidade os resultados com teorias desenvolvidas pelo investigador (sejam nas áreas de Sociologia, Economia ou Psicologia); é consistente com outras áreas dentro da Estatística e fornece um mecanismo para lidar com o problema da não-resposta. Finalmente, resultados significativos para os testes de hipóteses podem ser obtidos para tamanhos amostrais relativamente pequenos, com erro-padrão baixo se as covariáveis possuírem pouca ou nenhuma correlação entre si. Contudo, muitos dados coletados sob um PAC nem sempre estão em conformidade com as suposições de um modelo pré-especificado. Isto se explica pelo fato de que os pesos amostrais podem depender de covariáveis não pesquisadas na amostra e conseqüentemente ausentes da especificação do modelo. Logo, os pesos possuem uma informação útil (que foi ignorada) para estimar os coeficientes do modelo.

Como bem destacaram Skinner *et al.* (1989), ao trabalhar com dados amostrais complexos, os estatísticos devem estar cientes de que aplicar diretamente a modelagem, ignorando as características do PAC, pode resultar em estimativas enviesadas do erro-padrão. Este viés é devido aos efeitos de estratificação; conglomeração e pesos desiguais. Logo, os intervalos de confiança obtidos serão mais largos ou estreitos do que deveriam e os testes de hipóteses realizados serão, respectivamente, mais conservadores ou liberais<sup>5</sup>, incluindo menos ou mais variáveis no modelo a ser ajustado do que o devido.

Com exceção de planos amostrais relativamente simples e bem restritos, a determinação exata do tamanho e da natureza deste viés (ou de uma possível correção) é difícil de ser expressa analiticamente. Inicialmente, a investigação da influência de um PAC nas estimativas era feita apenas através de estudos empíricos. Comparavam-se as variâncias das estimativas quando as características do plano amostral eram consideradas com as mesmas variâncias obtidas ao ignorar tais características, supondo que foi feita AAS com probabilidades iguais de seleção. Ao longo da 2<sup>a</sup> metade do século XX, foram desenvolvidas na literatura aproximações específicas, dado um modelo, para a estimação de seus parâmetros considerando o PAC. Tais aproximações se dividem em duas novas abordagens para a inferência realizada, ditas *baseada no plano amostral* e *assistida por modelo*.

A inferência baseada no plano amostral, dentro da qual desenvolveremos a avaliação de impacto do SWF para a variável TSAT, toma como parâmetros de interesse as características de uma população suposta *finita*, um subconjunto de uma superpopulação (com infinitas observações). A construção dos testes de hipóteses e intervalos de confiança consiste na amostragem repetida desta população finita e a estrutura probabilística usada para inferência é definida por uma variável aleatória que indica se a observação na população pertence ou não à amostra  $s_l$  de fato extraída. Note que não se pressupõe aqui a existência de um modelo como verdadeiro processo gerador dos dados para a população finita. Logo, os pesos amostrais devem ser usados na estimação de médias, totais, parâmetros do modelo de regressão (dada a variável de interesse) ou de qualquer outra quantidade populacional. Por exemplo, na subseção 2.3.3, discutimos como a estimação por Máxima Verossimilhança é modificada de forma a incorporar os pesos amostrais. Neste sentido, são relevantes as contribuições de Koch & Lemeshow (1972), para técnicas de análise multivariada; Kish & Frankel (1974), para modelos de regressão linear; Rao & Scott (1981), para testes qui-quadrado em tabelas de contingência; e Binder (1983), para modelos lineares generalizados.

<sup>5</sup> Na modelagem estatística, um teste de hipóteses é dito *liberal* se, tomadas  $k$  amostras de tamanho igual da mesma população, a taxa de rejeição da hipótese nula (o coeficiente associado à variável sob teste é nulo) pelo teste, realizado em cada uma das  $k$  amostras, é maior do que o nível de significância (complementar do nível de confiança) do teste. Um teste é *conservador* quando ocorre o fenômeno inverso.

Se algumas variáveis importantes forem excluídas da modelagem, tanto as estimativas pontuais baseadas unicamente em um modelo pré-especificado quanto no plano amostral serão enviesadas, porém as obtidas pela segunda abordagem possuem viés decrescente com o aumento do tamanho amostral. Para a primeira, isto só acontecerá se os pesos amostrais não forem correlacionados com as variáveis excluídas (Lohr, 2010). Ao usar a inferência baseada no modelo, se espera que as estimativas pontuais dos parâmetros populacionais calculadas unicamente pelo modelo sejam semelhantes às obtidas incorporando os pesos amostrais. Assim, o ajuste das observações deve ser feito por ambas as abordagens inferenciais para validar o modelo especificado. Uma diferença entre as estimativas ponderadas e não-ponderadas é um indício de que o modelo proposto não é um bom gerador dos dados populacionais.

Tanto na inferência baseada no modelo quanto na baseada no plano amostral, não podemos estabelecer de imediato relações causais entre uma ou mais covariáveis e a variável resposta de interesse, pois qualquer associação encontrada é válida somente para a população finita da qual a amostra foi retirada e no instante em que a amostragem foi feita. Isto se explica pelo fato de que dados amostrais são, em geral, observacionais. Para isso, devemos realizar um experimento. Sem perda de generalidade, tome duas classes de indivíduos, que serão investigados em pelo menos duas ocasiões, de modo que os indivíduos de ambas as classes sejam idênticos na primeira entrevista (rodada) e, a partir da segunda, uma das classes receba um tratamento e a outra não. A decisão sobre quem recebe o tratamento pode ou não ser aleatória. Na seção 2.2 é detalhado como se procede em cada caso, em especial quando esta decisão não é aleatória por razões éticas (como quem recebe o benefício de uma política de transferência de renda).

Mais recente, a inferência assistida por modelo combina as duas abordagens anteriores. Um modelo é suposto como único e verdadeiro processo gerador dos dados na população para estimar os parâmetros de interesse, porém o modelo incorporará a estrutura populacional, da qual podem fazer parte características do plano amostral como estratificação ou conglomeração na sua especificação (Pessoa & Silva, 1998). As estimativas pontuais e intervalares podem ou não ser obtidas considerando os pesos amostrais. Embora possua algumas vantagens sobre a abordagem agregada, como a diferenciação do efeito de uma covariável sobre a variável dependente em níveis distintos (domicílio e residentes, por exemplo) em modelos hierárquicos, a abordagem com desagregação possui um tratamento matemático bem mais intensivo, especialmente quando há muitas covariáveis no modelo proposto, exigindo a informação em nível individual das observações, indo de encontro à proteção da confidencialidade dos entrevistados na pesquisa amostral. O leitor interessado nesta abordagem pode consultar Skinner *et al.* (1989); Särndal *et al.* (1992); Pessoa & Silva (1998) e Pfeiffermann *et al.* (1998).

### 1.3 Revisão da Literatura em Avaliação de Impacto

A literatura sobre avaliação do impacto de uma política pública de transferência monetária direta em uma dada resposta de interesse é bastante vasta (IFPRI, 2010; Ipea, 2013). Nesta seção, a revisão é restringida para a avaliação do impacto de políticas governamentais de transferência direta de renda no tempo de alocação para o trabalho, em geral (mas não unicamente) tomando a semana como base, de crianças e adultos. Quando o tempo não estiver disponível, a avaliação é feita para a participação da força de trabalho. Muitos estudos dividem a avaliação para as mais diversas subpopulações, não apenas para grupos etários. Mesmo quando estes são estudados, não necessariamente se adota o critério utilizado aqui para divisão dos indivíduos pesquisados entre crianças e adultos no contexto da YNSPMS.

A finalidade desta divisão é, em geral, verificar se a política contribui para a redução do trabalho infantil e conseqüentemente para o maior comparecimento à escola no primeiro grupo e para um “efeito preguiça”, ou mesmo aumento da informalidade, no segundo grupo. Frequentemente, conclusões distintas podem ser obtidas quando os estudos são feitos para as mesmas subpopulações de interesse. Isto ocorre em razão das características da política em questão e de sua população-alvo; da forma como os dados foram coletados e das ferramentas teóricas usadas para avaliação do impacto. É importante ressaltar que os resultados de tais estudos não se restringem à teoria a ser apresentada nos dois próximos capítulos, uma vez que a avaliação pode ser feita com distintas ferramentas estatísticas e econométricas.

Pela natureza deste tipo de pesquisa, o ambiente de estudo é em geral composto por países em desenvolvimento, onde a adoção de políticas de transferência direta de renda como forma de aliviar a extrema pobreza e movimentar o mercado interno são comuns. Cada política possui seu próprio critério de elegibilidade e podem ter critérios adicionais para permanência do domicílio beneficiário, também denominadas condicionalidades. Exemplos disso são a exigência da carteira de vacinação em dia ou de alta frequência escolar para as crianças residentes no domicílio beneficiário, por exemplo. Para o SWF, não é este o caso: uma vez atendidos os critérios de elegibilidade e enquanto se enquadrarem nos mesmos, o domicílio passa a ser beneficiário do programa. Por fim, as covariáveis utilizadas na modelagem (do score e/ou do impacto) podem alterar de acordo com o contexto do estudo.

As políticas governamentais de transferência direta de renda, bem como os estudos acerca dela, têm em grande parte origem na América Latina a partir da década de 1990, dentre elas se destacando o brasileiro *Programa Bolsa Família* (PBF). Pelo fato de ser aplicado em um país de dimensões continentais e com elevado nível de desigualdade social, o PBF tem sido alvo de muitos pesquisadores, tanto dos institutos governamentais quanto das universidades. Outros exemplos de políticas similares em países latinos são o

mexicano *Oportunidades* (escrito em português de maneira similar), o colombiano *Más Familias en Acción* (MFA, em português: Mais Famílias em Ação) e o equatoriano *Bono de Desarrollo Humano* (BDH, em português: Bônus de Desenvolvimento Humano), todos eles contemporâneos do PBF (Ipea, 2013). A partir dos anos 2000, tais políticas começaram a ser adotadas em maior escala por países africanos e asiáticos (Fiszbein & Schady, 2009).

Em especial no caso do PBF, políticas de transferência monetária direta têm sido alvo de debates internos entre os pesquisadores sobre a sua eficiência. Ao mesmo passo em que aliviam a situação de vulnerabilidade das famílias em extrema pobreza no curto prazo, nem sempre estabelecem prazos para saída de quem recebe, o que poderia torná-los dependentes do benefício. Entretanto, uma política social não será emancipatória a curto prazo. Os domicílios beneficiários irão competir por recursos entre si e com aqueles que não a recebem, mas estão imediatamente acima da renda definida como critério para elegibilidade, o que pode deslocar (ou devolver) novos domicílios para condições de extrema pobreza. A política terá sucesso se também for acompanhada de mudanças na estrutura ocupacional da população, garantindo que todas as crianças tenham acesso à educação e melhorando a qualificação profissional dos adultos (Medeiros *et al.*, 2007).

Para as crianças, ao contrário do que se espera, a maioria dos estudos não aponta para uma redução no tempo de alocação para o trabalho (ou seja, uma diminuição do trabalho infantil) ou na probabilidade de participação da força de trabalho. Skoufias & Parker (2001) estudaram o programa mexicano *Progresá*, uma das primeiras políticas públicas de transferência monetária direta na América Latina (implantado em 1997 e precursor do *Oportunidades*, que só viria a surgir alguns anos depois), para avaliar seu impacto na participação da força de trabalho nas crianças dos domicílios beneficiários. Como na maioria das políticas de seu gênero, o objetivo principal do *Progresá* era aumentar o investimento em capital humano (ou seja, em educação; saúde e nutrição) pelas famílias em situação de pobreza. Para atingi-lo, o programa condicionava a transferência à matrícula e frequência regular das crianças residentes nos domicílios beneficiários na escola, bem como o comparecimento da família aos serviços públicos de saúde quando necessário (por exemplo, campanhas de vacinação e atendimento pré-natal).

A avaliação de impacto foi feita com uso isolado do método DD estendido para quatro pontos distintos no tempo, sem combiná-lo com qualquer modelo para o escore de propensão, para a participação da força de trabalho no grupo das crianças. Skoufias & Parker consideraram como crianças todos os indivíduos entre 8 e 17 anos (inclusive). Ainda, os resultados foram separados por gênero (masculino e feminino) e por faixas etárias dentro da amostra de crianças analisada. Os indivíduos tratados são compostos por residentes em comunidades que eram beneficiárias do *Progresá* e os controles, por outras comunidades em situação econômica semelhante que ainda não recebiam o benefício no momento em que a amostra foi coletada (para as quatro rodadas consideradas). A amostra

utilizada foi selecionada através da *Encuesta de Evaluación de los Hogares* (ENCEL, em português: Pesquisa de Avaliação dos Domicílios). A ENCEL é uma pesquisa amostral longitudinal específica para o *Progresá*, realizada em rodadas semestrais. Feita em dois estágios, primeiramente foram selecionados municípios em sete estados mexicanos e em seguida os domicílios dentro de cada município.

Apesar do uso de um PAC na seleção dos indivíduos para a amostra da ENCEL, Skoufias & Parker supõem que as observações são IID (como se o plano amostral utilizado fosse AAS)<sup>6</sup>. A primeira rodada da ENCEL (ocorrida em novembro de 1997) foi realizada antes que qualquer domicílio recebesse o benefício do *Progresá*, o que permitiu avaliar o impacto do programa. Quando agregados por município, indivíduos tratados e controles são semelhantes com base em testes para a hipótese nula de igualdade na média de algumas das variáveis pesquisadas (não rejeitada), mas o mesmo não pode ser dito nos níveis domiciliar e individual. Tendo em vista que a amostra coletada a partir da 1ª rodada da ENCEL é bem grande, facilitando a rejeição dos testes de hipóteses (cerca de 24 mil domicílios, quase quatro vezes se comparada à YNSPMS), não é inválido usar o método DD para avaliação de impacto do *Progresá* (se os dados da variável de interesse forem longitudinais). Como covariáveis de confundimento do modelo *probit* para o método DD, foram considerados o nível educacional; a idade e indicadores para fluência em espanhol ou idioma indígena da mãe e do pai da criança. Skoufias & Parker concluíram que o *Progresá* contribuiu para a redução, embora não estatisticamente significativa (exceto para algumas subpopulações bem específicas), na probabilidade de participação da força de trabalho. Evidentemente, a maior redução ocorre entre a primeira e segunda rodadas da ENCEL.

Ferro & Kasouf (2005) investigaram o brasileiro Programa Bolsa Escola (PBE), um dos precursores do PBF<sup>7</sup>, e outros programas de transferência monetária direta condicionada especificamente à obrigatoriedade da frequência escolar para as crianças dos domicílios beneficiários. Para isso, foram usados dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), uma pesquisa amostral transversal, realizada em 2001. O afastamento pela criança de atividades laborais não era uma condição necessária para concessão deste tipo de benefício, mas em tese este poderia indiretamente reduzir a participação da força de trabalho e o tempo alocado para este ao forçar uma alta frequência escolar da criança. Neste caso, a redução seria um efeito colateral. Assim, as autoras avaliaram o impacto de tais programas (de forma transversal) para estas duas respostas de interesse.

Para a participação da força de trabalho, uma variável binária, foi usado um modelo *probit* ponderado pelo fator de expansão da amostra, ou seja, pelo inverso da probabilidade

<sup>6</sup> De fato, não são incorporadas na estimação do impacto as informações sobre estratificação, conglomeração ou pesos desiguais. É sabido que os municípios foram selecionados com probabilidades proporcionais ao tamanho, mas nada é dito a respeito da variável auxiliar utilizada para isso.

<sup>7</sup> O Programa Bolsa Família foi criado em 2003 como um benefício mais amplo e substituto dos programas federais de transferência direta de renda até então existentes, dentre eles o Programa Bolsa Escola (Ipea, 2013).

de inclusão dos elementos da população na amostra (os pesos amostrais). Já o tempo (semanal) alocado para o trabalho foi modelado por regressão linear com estimação por MQP, usando os pesos amostrais da PNAD na matriz de covariância estimada. Assim, mesmo a avaliação sendo transversal, Ferro & Kasouf consideraram características do PAC utilizado para a PNAD na obtenção das estimativas de cada modelo. Foram consideradas todas as crianças entre 6 e 15 anos (inclusive) de domicílios cuja renda *per capita* mensal não ultrapassava meio salário mínimo. Para ambos os modelos, as variáveis explicativas (covariáveis) são as mesmas: uma variável binária indicadora para o recebimento do benefício pelo domicílio em que a criança reside (a estimativa do coeficiente associado fornecerá o impacto da política); variáveis para características individuais das crianças (gênero, raça, idade) e dos pais ou responsáveis (idade, educação, logaritmo dos salários) e para características geográficas (a unidade da federação, divisão territorial equivalente à província, a qual o domicílio pertence).

Os resultados da modelagem são divididos para duas subpopulações (urbana e rural), uma vez que o teste da Razão de Verossimilhanças (RV) para o modelo *probit* e o teste  $F$  para a regressão linear com estimação por Mínimos Quadrados Ponderados (MQP) foram altamente significativos a 5% para o modelo com a variável indicadora correspondente<sup>8</sup>. Além de conclusões esperadas e recorrentes na literatura para as demais covariáveis, como probabilidade menor de participação da força de trabalho para a criança do gênero feminino, mais jovem e cujo pai ganha menos; Ferro & Kasouf afirmam que o PBE e programas de natureza similar têm impacto negativo e significativo (a 5%) no tempo (semanal) de alocação para o trabalho da criança, tanto no meio urbano quanto rural. Aquelas que residem nos domicílios beneficiários trabalham cerca de 3 horas a menos comparadas àquelas que residem em domicílios não beneficiários. Entretanto, a magnitude dos coeficientes estimados é reduzida e condizente com a hipótese de que o incentivo à participação no programa é maior para as crianças que trabalham em meio período, pois neste caso o rendimento mensal da família é em geral incrementado.

Por outro lado, para aquelas que trabalham em período integral o valor do benefício não compensa a renda que é obtida através do trabalho pela criança e que deixaria de existir se o domicílio em que reside for beneficiário da política. Logo, o PBE (por exemplo) não estaria atingindo de forma eficaz o seu público-alvo. Ainda para a participação da força de trabalho, a probabilidade é maior e significativa (a 5%) para as crianças residentes de domicílios inscritos no PBE. Para este resultado um tanto surpreendente, as autoras argumentam que o programa analisado possui mecanismo de autosseleção, logo é bastante provável que haja heterogeneidade individual que não foi observada. Isto quer dizer que possíveis variáveis omitidas com impacto positivo na participação da força de trabalho (como indicadores da ambição familiar) podem ter anulado o impacto negativo esperado

<sup>8</sup> Apesar do uso de dados amostrais complexos e da incorporação dos pesos associados na estimação de cada modelo, não é dito se tais testes foram corrigidos para também considerarem o PAC da PNAD.

pelo programa. Assim, elas concluem que o PBE contribui para a redução da jornada semanal das crianças entre 6 e 15 anos (inclusive) residentes em domicílios cuja renda *per capita* mensal é inferior a meio salário mínimo e que já trabalham, mas nada pode ser dito para a participação da força de trabalho neste mesmo grupo.

Posteriormente, Cacciamali *et al.* (2010) avaliaram de forma transversal o impacto do Programa Bolsa Família sobre a incidência do trabalho infantil, uma variável binária, a partir de dados da PNAD de 2004 e tomando duas outras variáveis binárias como respostas de interesse para a estimação: uma indicadora para a matrícula escolar da criança e outra para a participação da força de trabalho na semana da pesquisa. Foram consideradas as crianças na faixa etária entre 7 e 15 anos (inclusive), pois no PBF a frequência escolar é obrigatória somente para o ensino fundamental. Ao contrário da maioria dos estudos em avaliação de impacto, os autores modelaram ambas as respostas de interesse simultaneamente, através de um modelo *probit* bivariado<sup>9</sup>.

Como vantagem, este modelo permite supor que as duas respostas (matrícula escolar e participação da força de trabalho) estão relacionadas entre si de alguma forma. Logo, a probabilidade de incidência do trabalho infantil é determinada conjuntamente pela probabilidade da criança estar matriculada na escola e de participar da força de trabalho. Como covariáveis, foram consideradas características individuais das crianças e do responsável principal (aquele que contribui com a maior parte da renda do domicílio). Apesar do uso de dados da PNAD, o PAC usado para selecionar a amostra não foi incorporado de nenhuma forma na obtenção das estimativas, sendo todas as observações tratadas como IID. Os resultados encontrados pelo modelo *probit* bivariado vão ao encontro dos mencionados anteriormente para o grupo das crianças e do que normalmente é relatado na literatura. A probabilidade de incidência do trabalho infantil é maior para a criança do gênero masculino; de maior idade; cuja família é numerosa; residente em domicílio rural e os responsáveis trabalham na informalidade. Por outro lado, esta mesma probabilidade é menor quando o responsável principal é homem; os residentes no domicílio são mais escolarizados (tanto crianças quanto adultos) e a renda familiar *per capita* mensal é maior.

Curiosamente, o coeficiente associado ao PBF é positivo e significativo, indicando que ser beneficiário do programa aumenta a probabilidade de incidência do trabalho infantil. Ao contrário do PBE, o PBF não tem como objetivo primário o combate ao trabalho de menores de idade, embora em tese a interrupção do ciclo geracional pretendida pelo programa passe por uma maior escolarização das crianças dos domicílios beneficiários e na consequente redução do tempo alocado por elas para o trabalho. Cacciamali *et al.* ressaltam ainda que muitas famílias pobres, para garantir sua sobrevivência, ainda preferem manter as crianças com vínculo empregatício como forma de aumentar a renda domiciliar. Este

<sup>9</sup> Além das suposições usuais para o modelo *probit* se cada resposta fosse modelada separadamente, em geral também é exigido que a covariância entre os erros de cada resposta seja constante, não necessariamente nula, para todo indivíduo pesquisado (Cacciamali *et al.*, 2010).

componente, a ambição familiar, é o mesmo descrito por Ferro & Kasouf (2005) para resultado semelhante na participação da força de trabalho nas crianças dos domicílios beneficiários pelo PBE.

Já para os adultos, as conclusões sobre o impacto de uma política de transferência direta de renda na participação da força de trabalho e no tempo de alocação para este nem sempre vão no mesmo sentido. Skoufias & di Maro (2006) estudaram dados do programa mexicano *Progres*a com o objetivo de investigar se ser elegível ou não para o programa afetava a participação dos adultos na força de trabalho. Para isso, consideraram como resposta uma variável binária indicadora desta participação. Se o adulto pesquisado trabalhasse mesmo que no setor informal, a resposta atribuída era positiva. Para avaliar o impacto do Progesa na participação dos adultos na força de trabalho, também foi utilizado o método DD estendido para quatro pontos distintos no tempo, correspondentes a algumas rodadas da *Encuesta de Evaluación de los Hogares* (ENCEL, em português: Pesquisa de Avaliação dos Domicílios) entre 1997 e 1999 (neste caso, não igualmente espaçadas). Novamente, foi utilizado o modelo *probit* e as covariáveis de confundimento são dadas por indicadores demográficos, dentre os quais: idade; o quadrado desta; estado civil, escolaridade e a fluência em idioma indígena. Também foi usado como indicador geográfico a distância do domicílio até à sede do governo municipal correspondente.

Os autores concluíram que o programa não altera a probabilidade de participar da força de trabalho para os adultos em geral, com exceção dos homens entre 35 e 54 anos da 3ª para a 4ª rodada (para os quais há aumento nesta probabilidade) e das mulheres entre 45 e 54 anos da 1ª para a 2ª rodada (para as quais há uma redução). A amostra coletada pela ENCEL é bem maior que a pela YNSPMS (mais de 100 mil indivíduos pesquisados a partir da 1ª rodada) e foi baseada em um censo prévio dos domicílios elegíveis para o Progesa. Em tese, isto permite fazer correções das estimativas para uma população finita não muito superior em magnitude com relação ao tamanho amostral. Porém, o artigo não fornece informações sobre qual PAC foi utilizado, supondo apenas que as observações são IID. Além disso, é usado um modelo não-linear para estimar o impacto pelo método DD, violando o Pressuposto da Tendência Comum (PTC), uma das suposições do método DD (a serem discutidas no terceiro capítulo, na seção 3.3).

No caso do PBF, Tavares (2010) analisou, a partir de dados da PNAD realizada em 2004, a ocupação de trabalho para mães de domicílios cadastrados para receber o PBF e que já recebiam o benefício (não necessariamente chefes de família), tomadas como a classe de indivíduos tratados. Elas foram comparadas com outras três classes de controle: mães de domicílios cadastrados, mas que ainda não recebiam o benefício (semelhantes nas características observáveis em relação às pertencentes a classe de tratamento); mães de domicílios que não estão cadastrados, mas são elegíveis para receber o benefício; e mães de domicílios que não são elegíveis, mas cuja renda mensal *per capita* não ultrapassava a

quantia de 260 reais (R\$ 260,00).

Tavares também adotou a suposição de observações IID, mas inicialmente usou um modelo para a probabilidade de inclusão no PBF (o escore de propensão), tendo como covariáveis indicadores demográficos, muitos dos quais idênticos aos utilizados por Skoufias & di Maro (2006) (as poucas diferenças existentes são explicadas pelo contexto de cada pesquisa), e geográficos, dentre os quais: se o domicílio pertence à zona urbana; se está localizado em uma região metropolitana e se tem acesso a serviços básicos (como água encanada, coleta de lixo e rede de esgoto). Foi tomado o Pareamento pela Distância Euclidiana (PDE, a ser discutido no terceiro capítulo, na seção 3.1) sem reposição ou sobreamostragem de controles como método de pareamento, para encontrar mães não beneficiárias (para cada uma das três classes de controle) comparáveis com as mães da classe de indivíduos tratados.

O balanceamento das três amostras pareadas obtidas foi verificado através do teste  $F$  para existência da regressão (a hipótese nula assume que todos os coeficientes são nulos), ou seja, para todas as covariáveis do modelo conjuntamente o teste não foi significativo ao nível nominal de 5%. Logo, as amostras pareadas pelo modelo especificado para o escore de propensão poderiam ser utilizadas para avaliar o impacto do PBF dada uma resposta de interesse. Além da participação na força de trabalho, também foi considerado por Tavares (2010) o tempo semanal de alocação para o trabalho. No caso da primeira resposta de interesse, foi utilizado o modelo *probit*. Já para a segunda, foi usado um modelo de regressão linear com estimação por Mínimos Quadrados Ordinários (MQO) e correção para censura das mães que não trabalham (cuja jornada potencial é desconhecida). Para ambas as respostas, as covariáveis demográficas usadas como confundimento eram as mesmas. Os respectivos modelos diferiam apenas nas covariáveis geográficas (para a participação na força de trabalho, foram especificadas duas covariáveis a mais).

Como resultado, mães beneficiárias do PBF têm, significativamente, maior probabilidade de participarem da força de trabalho e, quando participam, alocam semanalmente mais tempo para o trabalho (comparadas às pertencentes a qualquer uma das três classes de controles). Portanto, Tavares conclui que o PBF não gera uma acomodação nas mães beneficiárias, mesmo que em maioria sejam responsáveis pelas tarefas domésticas e supervisão das crianças. Argumenta ainda que receber o benefício pode tornar o(a) beneficiário(a) alvo de discriminação, não raro autoimposta, conseqüentemente sendo um estímulo extra para a procura de trabalho. Deve-se notar, porém, que apenas uma pesquisa transversal foi considerada, de modo que nada pode ser dito sobre o impacto do PBF na subpopulação considerada ao longo do tempo, somente no instante da coleta dos dados.

Por sua vez, Foguel & Barros (2010) avaliaram o impacto do PBF na participação da força de trabalho e no tempo (semanal) de alocação para trabalho ao longo do tempo, incluindo dados de cinco rodadas da PNAD (de 2001 a 2005). Na sua definição, a PNAD

garante que os mesmos municípios são pesquisados a cada rodada durante uma mesma década, mas o mesmo não pode ser dito sobre os domicílios ou seus residentes. Logo, a avaliação do impacto é feita para os municípios, reduzindo o número de observações em relação à uma avaliação individual ou domiciliar. Claramente, as rodadas da PNAD realizadas antes da criação do PBF não nos permitem dizer quais domicílios são beneficiários, mas possuem perguntas que em tese ajudam a identificá-los. Para cada rodada (até 2003), foram classificados como beneficiários todos os domicílios cujos residentes não possuíam ativos financeiros (contas bancárias e ações de empresas, por exemplo). Tal estratégia, em conjunto com o questionário da PNAD de 2004, identificou corretamente mais de 90% dos domicílios pesquisados.

Para avaliar o impacto, foi usado apenas o método DD com dois termos adicionais na regressão linear, um para a proporção de beneficiários do PBF no município e outro para o efeito aleatório associado a este nível em cada resposta de interesse (mesmo a participação da força de trabalho sendo uma variável binária). Supondo observações IID, os autores concluíram que o PBF não tem impacto estatisticamente significativo na participação da força de trabalho e tem um impacto negativo e significativo (embora pequeno em magnitude) no tempo de alocação para o trabalho na subpopulação de mulheres. Para os homens, o impacto do PBF na participação da força de trabalho é significativo e positivo (também pequeno em magnitude), e no tempo de alocação para o trabalho, não há impacto significativo.

As conclusões também foram estendidas tomando somente os indivíduos cuja renda *per capita* familiar é menor do que a mesma medida para o município em que residiam, com distinção apenas no impacto do PBF para o tempo de alocação para o trabalho nas mulheres (não significativo). Novamente, características demográficas e geográficas (ao nível municipal) figuram entre as covariáveis utilizadas. Ainda, foi usado um Modelo de Probabilidade Linear (MPL, a ser discutido na seção 2.3) para a participação da força de trabalho. Embora não haja violação de nenhuma das suposições do método DD, usar um MPL requer extremo cuidado com a interpretação das estimativas. Além disso, as conclusões obtidas podem não valer para todos os indivíduos.

Em resumo, os trabalhos apresentados acima sugerem que as políticas públicas condicionadas de transferência monetária direta (como o *Progresas/Oportunidades* e o PBF), não têm impacto negativo na participação da força de trabalho ou no tempo de alocação para o trabalho nos adultos, exceto para algumas subpopulações bem específicas. Portanto, este tipo de intervenção não desencadeia um “efeito preguiça” nos adultos dos domicílios beneficiários. Isto pode ser explicado pelo fato de que a renda transferida por tais políticas é planejada de modo a não gerar dependência exclusiva desta ajuda, e sim servir de complemento à renda familiar quando esta não é suficiente para sustento dos indivíduos do domicílio. Porém, este impacto pode ser positivo para a participação da

força de trabalho no grupo das crianças devido ao componente de ambição familiar, ainda que o impacto seja negativo no tempo alocado para o trabalho.

Logo, neste trabalho procuraremos responder qual é o impacto de um programa, quando não há condicionalidades para a permanência do domicílio na política (como no SWF) e adotando o PAC utilizado na inferência realizada, no tempo (semanal) de alocação para o trabalho em crianças e adultos (TSAT), uma variável que pode ser tratada como aproximadamente contínua. A incorporação do PAC será feita em todas as etapas da avaliação de impacto, inclusive nos testes de balanceamento dos bancos pareados pelos escores de propensão estimados (preditos). Ao contrário da resposta para a participação da força de trabalho (não considerada nesta dissertação), não violaremos o PTC por não ser necessário o uso de um modelo não-linear. Além disso, o risco de inferir incorretamente para alguns indivíduos na população (dadas as observações nas covariáveis de confundimento) será bem menor comparado à situação em que o MPL é usado.

## 1.4 O Iêmen e a YNSPMS

### 1.4.1 O Iêmen: Aspectos Demográficos, Socioeconômicos e Políticos

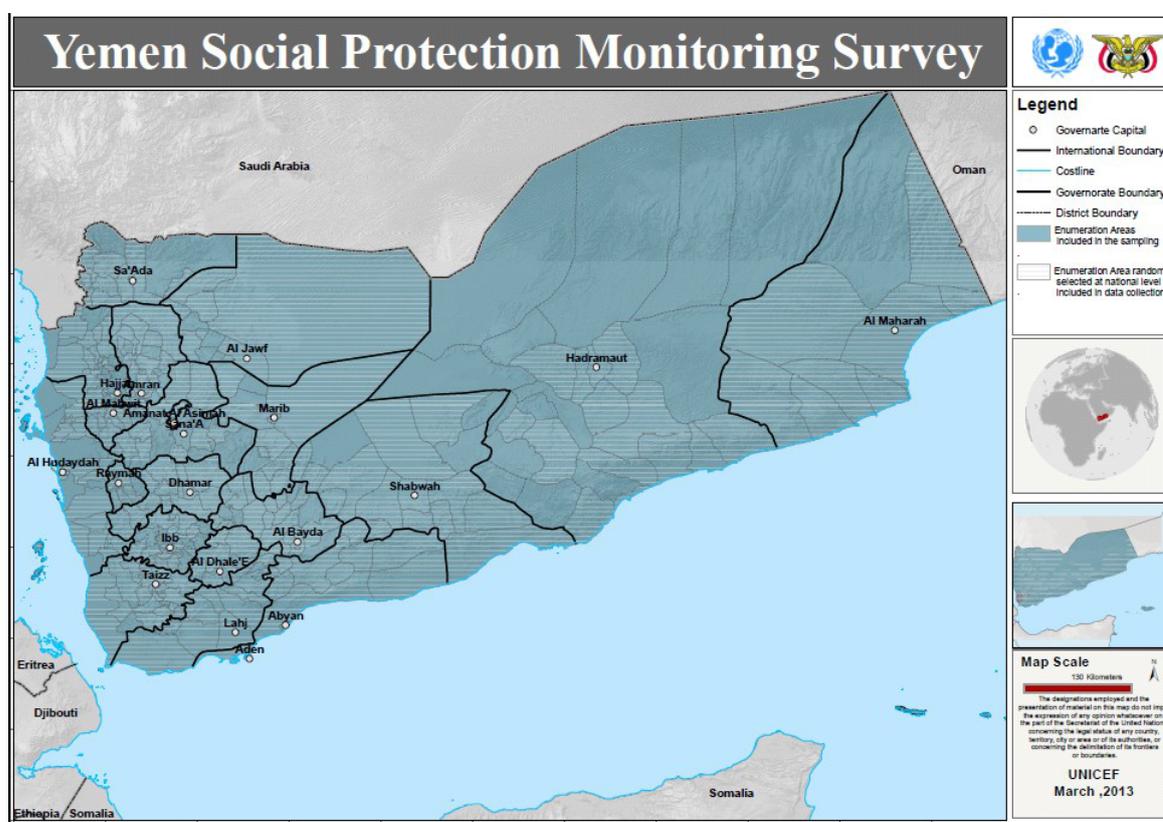


Figura 1 – Mapa Político do Iêmen à época da YNSPMS (Ferraz & Vieira, 2013)

O Iêmen é um país árabe localizado no Oriente Médio, mais precisamente no sul da

Península Arábica, fronteiro à Arábia Saudita e ao Omã. Oficialmente conhecido como República do Iêmen, possui área de 527.968 km<sup>2</sup> incluindo, além da parte continental, um arquipélago formado por aproximadamente 200 ilhas sendo a maior delas a ilha de Socotra. O clima iemenita é em sua maior parte desértico, especialmente na porção leste do país, com tempestades de areia frequentes no verão, sendo mais ameno na costa ocidental (CIA World Factbook, 2016). O país possui sua configuração atual desde 1990, quando os (agora extintos) governos do Iêmen do Norte e do Iêmen do Sul aprovaram a unificação. No ano seguinte, a constituição iemenita foi promulgada após um referendo. O Iêmen é dividido em 22 unidades federativas, sendo 21 províncias (também denominadas governadorados) e um distrito federal (a capital nacional, Sana'a). Sua forma de governo é uma república parlamentarista e pluripartidária, com mandato de sete anos para o presidente (chefe de estado) e direito a uma reeleição. O sufrágio é universal a partir dos 18 anos.

Atualmente, a população do Iêmen é estimada em 26,7 milhões (2015), sendo a grande maioria islâmica (99,1%, destes 65,0% são sunitas e 35,0% são xiitas)<sup>10</sup> e de etnia árabe. Dada a alta taxa de natalidade (30 nascidos por 1000 habitantes) e a baixa expectativa de vida (cerca de 65 anos), mais da metade da população (62,2%) está na faixa etária entre 0 e 24 anos. Apenas 34,6% da população é urbana, porém esta taxa vem crescendo rapidamente (em média, 4,0% ao ano). Ainda, o analfabetismo atinge 29,9% da população acima dos 15 anos.

Embora o petróleo se constitua na principal fonte de divisas externas para o país, também há jazidas de mármore, sal-gema e, em quantidade mais reduzida, de níquel; cobre; ouro; chumbo e carvão. Outros recursos naturais incluem peixes e um solo fértil na parte oeste. A produção agrícola consiste basicamente de algodão; café; plantas leguminosas e *khat*, considerada uma droga estimulante, porém legalizada e consumida em larga escala no país. Na pecuária, são comuns as criações de ovelhas, cabras e camelos. Destacam-se como atividades industriais a produção de óleo cru (com o consequente refinamento do petróleo); gás natural; têxteis (em grande parte, a partir do algodão); bens de couro e de alumínio; alimentos processados; artesanato e cimento. O país também obtém renda através do reparo de navios comerciais, uma vez que o litoral do sul iemenita é bastante movimentado devido à sua proximidade do Canal de Suez. A maior parte do Produto Interno Bruto (PIB), porém, é gerada através do setor de serviços (70,4%).

Apesar disso, a economia do Iêmen é extremamente frágil, especialmente devido à instabilidade política, e dependente da ajuda de órgãos internacionais, em especial o IMF (*International Monetary Fund*, em português: Fundo Monetário Internacional). Para piorar, os conflitos internos que se iniciaram em 2015 quase paralisaram as exportações; desvalorizaram drasticamente a moeda nacional; aumentaram a taxa de inflação; trouxeram

<sup>10</sup> Outros grupos religiosos incluem judeus, cristãos e hindus, muitos dos quais são refugiados ou estrangeiros com visto temporário.

danos generalizados para a infraestrutura e limitaram severamente a importação de alimentos e combustível. Com isso, cerca de 80% dos iemenitas precisavam de ajuda humanitária ao final deste ano. Outros desafios do país para o longo prazo incluem a redução das altas taxas de crescimento populacional (2,5%) e de desemprego (33,7%) e da escassez hídrica e de alimentos (CIA World Factbook, 2016).

Na época da realização da YNSPMS, o Iêmen foi um dos países participantes da série de eventos conhecida como Primavera Árabe. No país, os protestos em massa começaram no final de janeiro de 2011, com milhares de iemenitas tomando as ruas para exigir mudanças no governo (Bakri & Goodman, 2011). Embora fossem inicialmente pacíficos, os protestos rapidamente evoluíram para conflitos armados, com consequências econômicas e sociais negativas em uma das nações mais pobres do Oriente Médio. Dentre elas, as mais fortes foram o aumento nos preços dos alimentos e combustíveis e a interrupção do fornecimento de água e energia elétrica (ainda que em gravidade menor em relação a 2015). Como resultado, houve uma piora das condições de vida nos domicílios iemenitas, em especial naqueles mais pobres e vulneráveis (YNSPMS Final Report, 2014a).

Não havia nenhum repositório no país que pudesse prover dados sobre as consequências da crise de 2011. Por esta razão, os representantes da UNICEF (*United Nations Children's Emergency Fund*, em português: Fundo das Nações Unidas para a Infância) no Iêmen lançaram uma pesquisa amostral piloto, denominada SPM (*Social Protection Monitoring*, em português: Monitoramento de Proteção Social) em 29 de junho de 2011 com o objetivo de verificar como a população mais vulnerável estava enfrentando a crise e a partir disto, extrair tendências ao longo do tempo. Os dados para esta pesquisa foram coletados de 120 domicílios em 3 províncias (Sana'a; Hodeida e Amran) entre junho de 2011 e fevereiro de 2012 e suas conclusões foram compartilhadas com vários órgãos nacionais e internacionais. A SPM deixou claro que a crise de 2011 trouxe consequências negativas para os domicílios mais pobres, especialmente com relação à segurança alimentar. Assim, uma pesquisa de acompanhamento da proteção social em todo o Iêmen era necessária.

#### 1.4.2 A YNSPMS: Principais Características

A *Yemen National Social Protection Monitoring Survey* é uma pesquisa longitudinal por amostragem, com quatro rodadas trimestrais entre outubro de 2012 e setembro de 2013, realizada em domicílios de todo o território do Iêmen, com exceção de duas províncias<sup>11</sup>. No total, 6.397 domicílios (de uma amostra inicial de 7.560) foram entrevistados. A ideia original era ter um monitoramento em tempo real com um questionário curto semelhante ao que foi feito na pesquisa SPM. Porém após uma consulta com especialistas do governo iemenita, da UNICEF e de outras agências das Nações Unidas, a metodologia de coleta foi

<sup>11</sup> Por razões de segurança, nenhuma entrevista pôde ser feita para as províncias de Sa'ada (durante toda a pesquisa) e Al-Jawf (na última rodada), sendo este o principal fator para a redução da amostra inicial.

modificada em uma pesquisa amostral completamente focada na proteção social. Por ser uma pesquisa domiciliar, a YNSPMS não incluiu dados sobre comunidades de refugiados; nômades e pessoas vivendo em locais fora de suas residências, como hotéis, dormitórios, prisões e hospitais (YNSPMS Final Report, 2014a).

A YNSPMS tinha por base dois objetivos primordiais. O primeiro era fornecer dados atualizados sobre a situação socioeconômica da população mais pobre e vulnerável após a Primavera Árabe. O segundo, medir a influência, para alguns indicadores de desenvolvimento, do programa de transferência de renda *Social Welfare Fund* a partir da evidência produzida pelos domicílios com beneficiários do programa. Para a amostra pesquisada, o YNSPMS proverá dados sobre o SWF e as características do domicílio (por exemplo, o tipo de material usado na construção e o uso de eletricidade e água potável) e de seus ocupantes (como educação; saúde materna e infantil; fontes de renda e segurança alimentar). O SWF é um programa de transferência de renda sem condicionalidades financiado pelo governo iemenita para os cidadãos que não podem, de forma temporária ou permanente, sustentar a si mesmos e cujos familiares não são capazes de oferecer suporte. É o mais importante programa desta natureza no Iêmen e um dos maiores na região geográfica que compreende o Oriente Médio e o norte da África. Foi criado em 1996 como um mecanismo compensatório para a população mais humilde, amenizando as consequências negativas da retirada de subsídios para compra de alimentos.

O programa é constituído por duas categorias de beneficiários. A categoria social é composta por idosos (homens acima de 60 anos e mulheres acima de 55 anos); inválidos (inclui não apenas deficientes físicos e mentais, mas também pessoas com doenças crônicas tais como diabetes, asma e reumatismo) e órfãos (abaixo dos 18 anos ou entre 18 e 25 anos se estiver matriculado(a) em uma universidade ou escola técnica). A categoria econômica é composta por desempregados (entre 18 e 60 anos) e por mulheres solteiras, viúvas ou divorciadas (acima de 18 anos ou também abaixo desta idade se for viúva ou divorciada com pelo menos um filho) que não contam com outra pessoa dentro do domicílio para sustentar a família e seu nível de renda está abaixo do benefício oferecido pelo SWF. A YNSPMS mostrou que o SWF é favorável aos mais pobres, sendo em muitos casos a única fonte de renda do domicílio, sem o qual não seria possível comprar mesmo os itens alimentícios mais básicos.

A população residente em domicílios com pelo menos um beneficiário do SWF aumentou de 30% na primeira rodada do YNSPMS para 35% na quarta rodada. Tal expansão foi devida à entrada de novos beneficiários no programa. Estes foram identificados pela CSS (*Comprehensive Social Survey*, em português: Pesquisa Social Abrangente), realizada em 2008, e selecionados através do PMT (*Proxy Means Test*, traduzido para o português como Teste de Elegibilidade Multidimensional<sup>12</sup>), mas sua incorporação ao

<sup>12</sup> O PMT permite identificar a população em situação de pobreza a partir de características individuais

SWF ocorreu apenas a partir da primeira rodada da YNSPMS. Com o uso do PMT e dos dados do YNSPMS, o percentual de pobreza na população iemenita foi estimado em 45%. Em seu relatório final, os beneficiários existentes antes de 2008 (portanto, antes da 1ª rodada da YNSPMS) são referidos como *antigos beneficiários*, e os selecionados pelo PMT neste ano (mas beneficiários apenas a partir da realização da YNSPMS), como *novos beneficiários*. Estes últimos representam cerca de 33% do total de domicílios beneficiários do SWF (YNSPMS Final Report, 2014a).

### 1.4.3 O Plano Amostral da YNSPMS

O plano amostral da YNSPMS foi baseado na necessidade de verificar se o SWF teve êxito em amenizar as dificuldades dos mais vulneráveis. Logo, foi dada maior probabilidade de seleção aos domicílios mais pobres de acordo com o mapa da pobreza de 2007 para o Iêmen e aos beneficiários ou potenciais beneficiários do SWF. Todos os indicadores calculados no relatório final da YNSPMS usaram uma amostra com os 6.397 domicílios entrevistados nas quatro rodadas, correspondente a cerca de 90% da amostra selecionada na 1ª rodada (7.152 domicílios). Evidentemente, como o plano amostral empregado na YNSPMS era um PAC, pesos amostrais transversais e longitudinais são considerados (YNSPMS Final Report, 2014a). Como é recomendado antes de toda pesquisa amostral em larga escala, antes da realização da YNSPMS foi feita uma pesquisa amostral piloto, no caso a SPM (descrita brevemente na subseção 1.4.1). Foram coletados dados de 120 domicílios em 3 das 21 províncias (Sana'a, Hodeida e Amran), dos quais 40 foram identificados como beneficiários do SWF. Em cada província, subdistritos de baixa renda foram selecionados de acordo com o mapa da pobreza construído através da HBS (*Household Budget Survey*, em português: Pesquisa de Orçamento Domiciliar), realizada entre 2005 e 2006.

O questionário da SPM incluiu perguntas sobre características do domicílio e de seus ocupantes, abordando nutrição e segurança alimentar; proteção e saúde infantil; e presença de água encanada e saneamento básico. Adicionalmente, numa fase mais avançada e introspectiva foram feitas questões sobre matrícula escolar; comparecimento a exames médicos; registros de nascimentos e possíveis ocorrências de mutilações genitais em crianças do sexo feminino<sup>13</sup>. Os dados foram coletados durante 12 rodadas quinzenais de junho a dezembro de 2011 e duas rodadas adicionais, mensais, em janeiro e fevereiro de 2012. Foi revelado que os indicadores de segurança alimentar e desnutrição infantil, em todas as rodadas, eram preocupantes, mesmo nos domicílios beneficiários do SWF.

---

ou domiciliares correlacionadas com a renda. Como vantagens, possui baixo custo de implementação; não exige comprovação de renda pelo entrevistado e pode ainda ser replicado: domicílios semelhantes receberão a mesma avaliação ainda que se feita por membros diferentes do grupo de pesquisa em dias diferentes, evitando a corrupção ou politização de um dado programa (Faria et. al., 2007).

<sup>13</sup> Infelizmente, tal prática ainda é comum no Iêmen e em vários países do norte africano, não raro realizada sem anestesia (UNICEF, 2013).

Para se ter uma ideia, 66% dos domicílios relataram não consumir nenhum tipo de carne e 33% disseram que pelo menos um membro já teve que dormir estando faminto. Com relação à proteção infantil, a sensação de segurança caiu drasticamente com os conflitos de 2011: o percentual de crianças que tinham medo ao sair de casa para brincar devido à violência chegou a 25%. A crise de 2011 também afetou o fornecimento de água, com 31% dos domicílios consumindo menos água devido ao aumento dos preços. Desta forma, os resultados da SPM indicaram uma necessidade de uma pesquisa nacional para saber como os domicílios vulneráveis estavam lidando com os efeitos dos conflitos de 2011 e investigar o impacto de mecanismos de proteção social, em especial o SWF.

Alinhada ao objetivo de avaliar a qualidade do SWF, a YNSPMS também teve como meta permitir a avaliação comparativa entre beneficiários e não beneficiários do programa ao longo do tempo. Assim, a amostra da YNSPMS foi extraída seguindo um plano amostral estratificado em dois estágios<sup>14</sup>.

No 1º estágio, Áreas de Enumeração (AE) constituíram as Unidades Primárias de Amostragem (UPA), sendo implementada a Amostragem por Conglomerados (AC) com probabilidades desiguais de seleção, considerando as 21 províncias iemenitas como estratos. Em cada província, 30 AE's foram selecionadas. A seleção foi feita usando o esquema de Probabilidades Proporcionais ao Tamanho (PPT) definidas por Amostragem Poisson Sequencial (APS). A variável cujo tamanho foi utilizado é o número médio de pessoas pobres por AE, considerando o total no distrito (unidade geográfica inferior à província, as 21 províncias possuíam ao todo 333 distritos) ao qual a AE pertence. Assim, AE's dentro de um mesmo distrito tiveram probabilidades de seleção iguais entre si. A alocação, em cada estrato, do tamanho amostral no 1º estágio foi feita de maneira uniforme. A finalidade desta decisão era produzir estimativas do contraste entre subpopulações de domicílios beneficiários e não beneficiários do SWF com níveis similares de precisão para cada estrato (Ferraz & Vieira, 2013).

No 2º estágio, domicílios dentro da AE selecionada no estágio anterior foram escolhidos por Amostragem Estratificada Simples (AES), sendo 12 os domicílios selecionados em cada AE amostrada, constituindo as Unidades Secundárias de Amostragem (USA). A alocação do tamanho amostral por estrato neste último estágio permitiu que maiores probabilidades de seleção fossem dadas aos correspondentes contrafatuais dos domicílios beneficiários do SWF. Ainda no 2º estágio, uma operação de listagem foi conduzida em cada AE selecionada no 1º estágio, entre 1º e 22 de setembro de 2012 (pouco antes da 1ª rodada). O objetivo era identificar e classificar todos os domicílios localizados nas AE's escolhidas em três grupos (YNSPMS Final Report, 2014a):

<sup>14</sup> No YNSPMS Final Report, os estágios são erroneamente denominados de fases, cujo significado em Amostragem é distinto do que foi de fato realizado. Uma explicação detalhada sobre a amostragem em duas (ou mais) fases pode ser encontrada em Lohr (2010, capítulo 12).

- Estrato de Tratamento (ET): domicílios com pelo menos um beneficiário do SWF e ao menos um pagamento já recebido;
- Estrato de Controle I (EC1): domicílios com pelo menos um morador registrado ou selecionado para o SWF, porém sem qualquer beneficiário residindo no domicílio;
- Estrato de Controle II (EC2): domicílios sem nenhum residente registrado no SWF.

Os domicílios listados, dentro de cada AE previamente selecionada, foram então aleatoriamente selecionados de acordo com a seguinte regra: cinco domicílios são alocados para o estrato ET; outros cinco, para o EC1; e os últimos dois, para o EC2. Esta alocação atinge então o objetivo de encontrar um correspondente contra-factual para os beneficiários. Este critério de classificação não é o que será utilizado para definir os indivíduos controles e tratados nesta dissertação, servindo apenas como instrumento para seleção amostral no 2º estágio.

Além das províncias de Saa'da (em toda a pesquisa) e de Al-Jawf (na quarta e última rodada), quatro AE's em outras províncias também não tiveram seus domicílios listados por razões de segurança, reduzindo a amostra inicial de 7.560 para 7.152 domicílios na 1ª rodada. Dos 7.152 domicílios entrevistados na 1ª rodada, 6.968 (97,5%) o foram com sucesso nesta rodada. A taxa de resposta não variou entre as zonas urbana e rural. Dos domicílios entrevistados com êxito, 26 foram excluídos devido a problemas de falsificação na listagem em duas AE's, descobertos apenas na 2ª rodada (entre janeiro e março de 2013). Com esta exclusão, os pesos amostrais para os domicílios na província em que ocorreu esta exclusão foram reajustados. A amostra final, de 6.397 domicílios entrevistados em todas as rodadas, corresponde a 84,6% da amostra inicial planejada de 7.560 domicílios.

Os pesos amostrais, tanto transversais (para indicadores de uma rodada específica) quanto longitudinais (para indicadores que representem todas as rodadas), foram calculados para refletir não apenas o plano amostral da YNSPMS. Eles são ajustados para considerar também a não-resposta unitária (um domicílio em específico dentre os selecionados no 2º estágio não participou da pesquisa) nas quatro rodadas<sup>15</sup>. Novamente, é importante ressaltar que qualquer inferência realizada com os dados da YNSPMS deve usar os pesos amostrais fornecidos e considerar o plano amostral para produzir estimativas não enviesadas.

## 1.5 Suporte Computacional

Este trabalho utilizou ferramentas computacionais para a elaboração do corpo textual, sem deixar de atender ao rigor matemático necessário, e para a produção e análise de resultados dos métodos estatísticos utilizados tendo como base os dados da YNSPMS. Com relação ao *hardware*, os cálculos apresentados foram feitos em dois computadores (de

<sup>15</sup> Para mais detalhes, consulte o apêndice A.

forma não simultânea). O primeiro deles era residencial, com 8.00 GB de memória RAM (*Random-access Memory*, em português: Memória de Acesso Aleatório) e um processador Intel® Core™ i3-4170 de velocidade 3.70 GHz e com base em x64. O segundo, de propriedade do Laboratório de Amostragem e Experimentação do Departamento de Estatística da Universidade Federal de Juiz de Fora, com características similares exceto pelo processador (Intel® Core™ i5-3330 de velocidade 3.00 GHz). Quanto ao *software*, basicamente foram utilizados dois programas. A construção do corpo textual da dissertação foi feita através do processador de texto L<sup>A</sup>T<sub>E</sub>X com auxílio do TeXstudio versão 2.10.8 como ambiente integrado de escrita. Para a aplicação da teoria estatística a ser discutida, dentro da área de Amostragem, foi utilizado o *software* estatístico R versão 3.2.4 (R Core Team, 2016b), também acompanhado de um ambiente integrado (RStudio versão 0.99.491).

### 1.5.1 L<sup>A</sup>T<sub>E</sub>X

O L<sup>A</sup>T<sub>E</sub>X é um sistema tipográfico muito utilizado para a produção de documentos científicos, especialmente aqueles que fazem uso constante de fórmulas matemáticas, a partir de um *layout* pré-definido. É também apropriado para a criação de outros tipos de documentos, desde simples cartas até livros completos. Criado por Leslie Lamport em 1985, o L<sup>A</sup>T<sub>E</sub>X usa o T<sub>E</sub>X, uma linguagem de programação para edição de textos com muitas equações, como sistema de formatação. O T<sub>E</sub>X fora desenvolvido por Donald Knuth em 1977 como uma ferramenta computacional capaz de evitar a deterioração da qualidade tipográfica de livros e artigos (Oetiker *et al.*, 2015).

Ao contrário dos processadores de textos amplamente utilizados, como *Microsoft Word* ou *LibreOffice Writer*, o L<sup>A</sup>T<sub>E</sub>X não usa a abordagem WSIWG (*What you See Is What you Get*, em português: O Que Você Vê É O Que Você Obtém). Assim, não é possível visualizar a versão impressa do documento enquanto se escreve o texto. Apesar desta pequena desvantagem, o L<sup>A</sup>T<sub>E</sub>X é mais prático do que os processadores de textos convencionais baseados na abordagem WSIWG. As principais razões para isto são:

- Prevenção de erros de formatação ao forçar o autor do documento a declarar a estrutura lógica deste antes de começar a escrita do texto;
- Possibilidade de referenciar objetos, ou seja, a criação de ligações internas no documento para figuras, tabelas, seções, equações e mesmo citações;
- Simplicidade para gerar notas de rodapé, sumário e bibliografias;
- Gratuidade e Liberdade para alteração do código-fonte;
- Possibilidade de uso em uma grande variedade de sistemas operacionais, com muitas opções de ambiente integrado.

Nesta dissertação, se optou por manter o texto compatível com as normas da ABNT (Associação Brasileira de Normas Técnicas). Para atingir tal objetivo, foi utilizado o pacote abnTeX2 versão 1.9.6, um conjunto de ambientes e comandos personalizados da

classe *memoir* para a criação de documentos técnicos e científicos em consonância com as normas ABNT(sítio online do projeto abnTeX2, 2016)<sup>16</sup>.

### 1.5.2 R

O R é uma linguagem de programação interpretada para manipulação e armazenamento de dados; cálculos matemáticos; operações matriciais e produção de gráficos, direcionado principalmente à análise estatística. É também um *software* gratuito e de código aberto com uma grande variedade de bibliotecas e em constante desenvolvimento, distribuído através da licença GNU pelo sítio <<https://www.R-project.org/>>. Embora o uso das linhas de comando seja obrigatório, as funções nativas do R são bastante intuitivas e flexíveis, permitindo a criação de estruturas condicionais e recursivas sem muito esforço. (Venables *et al.*, 2016).

O R foi desenvolvido inicialmente por Ross Ihaka e Robert Gentleman no Departamento de Estatística da Universidade de Auckland (Nova Zelândia) no começo da década de 1990. O *software* foi muito influenciado por duas linguagens de programação pré-existentes: a linguagem S de Rick Becker; Allan Wilks e John Chambers, e a linguagem Scheme de Guy Steele e Gerald Sussman, esta última muito utilizada na Engenharia. Em aparência, R e S se assemelham bastante, mas a semântica do R é mais próxima à da linguagem Scheme. Desde 1997, um grupo (o R Core Team) foi criado para manter e melhorar o código-fonte das funções nativas do R e de bibliotecas amplamente utilizadas. Atualmente, o R pode ser executado nos sistemas operacionais derivados do Unix; Windows e Mac (Hornik, 2016).

---

<sup>16</sup> No momento em que a referência foi consultada, havia um total de 11 normas específicas, com as mais antigas estabelecidas em 2002.

## 2 Estimação dos Escores de Propensão

Neste capítulo abordaremos de forma mais aprofundada o cálculo dos escores de propensão para inferência sob o pressuposto de observações IID e também quando um PAC é considerado, no contexto da pesquisa YNSPMS, uma vez que avaliaremos o impacto do programa SWF sob ambas as perspectivas. Tais escores serão dados pela probabilidade predita de receber o tratamento (participar do SWF), para cada indivíduo (residente do domicílio) na amostra, dados os valores nas covariáveis selecionadas. Desta forma, o modelo utilizado terá uma variável dependente binária, igual a zero se o indivíduo for controle e a um se for tratado. O critério adotado para definição de controles e tratados bem como a escolha de covariáveis e da função de ligação para o modelo serão melhor detalhados neste capítulo. Finalmente, são feitas considerações sobre a aproximação da variância estimada para estimativas pontuais considerando a influência do PAC adotado, quando queremos incorporá-lo na análise.

### 2.1 Definição de Controles e Tratados

Como vamos avaliar o impacto do SWF, independentemente da variável resposta escolhida e da perspectiva adotada (AAS *versus* PAC), é necessário criar uma variável categórica nominal binária, que assuma apenas os valores 0 ou 1. Se um indivíduo for classificado como tratado, a nova variável assumirá o valor 1. Senão, ela assumirá o valor 0, e o indivíduo será classificado como controle. Os controles e tratados devem ser comparáveis, ou seja, as respostas nas covariáveis para eles precisam ser o mais similares possível para que possamos avaliar o impacto do SWF. Na próxima seção, veremos que isso é análogo a ter distribuição semelhante no escore de propensão predito para as subpopulações de controles e tratados quando o experimento não for aleatorizado (como na pesquisa YNSPMS). Um indivíduo será classificado como controle se, em todas as rodadas da pesquisa amostral, o domicílio em que reside não possuir nenhum beneficiário residente. Um indivíduo será classificado como tratado ou de tratamento se, na primeira rodada da pesquisa, o domicílio em que reside não possui nenhum beneficiário residente e, em uma ou mais das rodadas posteriores (a partir da segunda), havia pelo menos um beneficiário residente.

Note que um indivíduo também será tratado se este não for o beneficiário direto (basta residir no mesmo domicílio que quem recebe de fato o benefício) e se havia um beneficiário residente a partir da segunda rodada, mas não em alguma das rodadas posteriores a esta (basta o domicílio ter um beneficiário em pelo menos uma das rodadas a partir da segunda). Note ainda que indivíduos cujo domicílio já possuía um beneficiário residente na 1ª rodada não serão considerados na avaliação do impacto, uma vez que

o ponto de partida é distinto e, para eles, não há informação sobre qualquer ocasião em que o respectivo domicílio não possuía nenhum beneficiário (YNSPMS Final Report, 2014a). Na pesquisa YNSPMS, iremos utilizar `p6b_q_15`, uma variável binária indicadora do recebimento ou não do SWF pelo domicílio (Banco de Dados da YNSPMS, 2014c) para realizar a classificação. Por razões práticas, ela será denotada de agora em diante por RSWF (Recebimento do *Social Welfare Fund*). Portanto, os controles serão todos os indivíduos que responderam negativamente a esta variável em todas as quatro rodadas e os tratados, aqueles que responderam negativamente na primeira rodada e afirmativamente em pelo menos uma das rodadas posteriores a partir da segunda. Desta forma, temos para cada tratado o seu respectivo controle contrafactual.

No total, a YNSPMS possui 47 006 indivíduos entrevistados em todas as quatro rodadas. Para que as análises ignorando e considerando o PAC sejam comparáveis, são retirados aqueles cujos pesos amostrais longitudinais da 1<sup>a</sup> à 4<sup>a</sup> rodada são faltantes, o que nos fornece 46 992 indivíduos. Destes, 19 118 (40.68%) serão classificados como controles e 6 169 (13.13%), como tratados. Note que apenas pouco mais da metade dos indivíduos puderam ser classificados. Dividindo pelos dois grupos etários (entre 6 e 14 anos, inclusive, e acima de 14 anos), temos 5 166 controles e 1 713 tratados no primeiro grupo e 10 620 controles e 3 598 tratados no segundo. Restringindo para quem respondeu à variável TSAT na 1<sup>a</sup> e na 4<sup>a</sup> rodadas, temos 493 controles e 118 tratados no grupo das crianças e 4 053 controles e 1 116 tratados no grupo dos adultos. Além disso, dado o domicílio, todos os seus residentes foram classificados corretamente de maneira uniforme (como controles ou tratados), não havendo nenhum domicílio com residentes classificados de maneira distinta<sup>1</sup>. Também não há quem respondeu negativamente em todas as rodadas para `p6b_q_15` e o domicílio já tenha recebido o SWF antes da 1<sup>a</sup> rodada. Isto pode ser verificado através da variável `bitreat` (se alguém dentro do domicílio já recebeu o SWF). Assim, o critério adotado para classificação em controles e tratados pode ser utilizado para a avaliação de impacto do SWF.

## 2.2 O Escore de Propensão

O conceito *escore de propensão* foi introduzido pela primeira vez no artigo de Rosenbaum & Rubin (1983) como ferramenta na área de estudos experimentais e para a situação em que há apenas duas possíveis classificações com relação ao tratamento recebido. Note que é possível ter dois tratamentos aplicados a todos os indivíduos, com cada indivíduo recebendo um único tratamento. Aqueles que recebem um dos tratamentos realizam o mesmo papel dos controles definidos na seção anterior no contexto da YNSPMS quando o outro tratamento é que está sob estudo pelo pesquisador. Dada uma variável resposta de interesse  $R$ , qualquer inferência causal para o efeito do tratamento nos indivíduos de

<sup>1</sup> O código utilizado para a produção dos resultados enunciados se encontra na seção B.1 do apêndice.

fato tratados (aqueles que residem em domicílios que passaram a receber o SWF) envolve especular também sobre qual seria este mesmo efeito nos controles.

Em princípio, cada indivíduo  $i$  da amostra terá uma única resposta  $r_{1i}$  se receber o tratamento de interesse (denotado por 1) ou outra resposta  $r_{0i}$  se não recebê-lo (a ausência do tratamento de interesse é denotada por 0). Partindo desta definição, o efeito do tratamento é dado por uma comparação envolvendo as duas respostas, como uma diferença ( $r_{1i} - r_{0i}$ ) ou razão ( $r_{1i}/r_{0i}$ ). Assumimos também que o tratamento (ou a ausência deste) que um indivíduo recebe não influenciará o que qualquer um dos outros indivíduos receberem<sup>2</sup>. Uma vez que um mesmo indivíduo de fato receberá ou não o tratamento, teremos apenas uma das duas respostas. Logo, as comparações realizadas terão um grau de incerteza. Assim, estimar o efeito do tratamento (o impacto do SWF) é um problema de não-resposta, já que  $r_{1i}$  ou  $r_{0i}$  não é observado.

Seja  $T$  a variável (binária) indicadora do tratamento de interesse, logo  $T_i = 1$  se o indivíduo  $i$  recebeu o tratamento (reside em domicílio com beneficiário do SWF) e  $T_i = 0$  se não o recebeu. Sejam  $\mathbf{x}_i$  um vetor de  $p$  covariáveis fixas ao longo do tempo ou observadas antes da decisão de receber ou não o tratamento (para que as duas classes sejam comparáveis) e  $X$  a matriz cujas linhas representam estes  $n$  vetores (o  $n^o$  de indivíduos pesquisados). Suponha que a indexação dos  $n$  indivíduos foi feita aleatoriamente, não havendo informação fornecida pelo valor de  $i$ . Logo, a informação observada para o indivíduo  $i$  está contida apenas no vetor  $\mathbf{x}_i$ .

Se o experimento for aleatorizado, os resultados nas duas classes (controle e tratamento) poderão ser diretamente comparáveis, pois supomos que todos indivíduos sob estudo são similares entre si. Porém, a avaliação do impacto de políticas de transferência de renda é, em geral, realizada a partir de um experimento não-aleatorizado, no qual os indivíduos que recebem o tratamento diferem dos que não o recebem. A comparação direta não pode mais ser feita. Por razões éticas, nem sempre se escolhe de forma aleatória na população quais domicílios serão ou não alvo da política, a menos que os domicílios da população considerada sejam bastante homogêneos com relação às condições socioeconômicas de seus residentes.

Para experimentos não-aleatorizados, Rosenbaum & Rubin (1983) definem o *escore de balanceamento*  $b(\mathbf{x}_i)$ , uma função das covariáveis observadas tal que a distribuição condicional de  $\mathbf{x}_i$ , dado  $b(\mathbf{x}_i)$ , é a mesma para controles e tratados (ou seja, é independente de  $T$ ), ou seja,  $\forall i$ . Em notação matemática,  $\mathbf{x}_i \perp\!\!\!\perp T_i | b(\mathbf{x}_i)$ . O escore de balanceamento mais trivial é a função identidade  $b(\mathbf{x}_i) = \mathbf{x}_i$ . Em experimentos aleatorizados,  $b(\mathbf{x}_i)$  é conhecido e dado por um único valor, não sendo necessário estimá-lo. Isso se explica pelo

<sup>2</sup> Segundo Rosenbaum & Rubin (1983), tal pressuposto não é crucial para obter a expressão do escore de propensão mas, como veremos adiante (equação (2.1)), facilita seu cálculo. Podemos modificá-la assumindo dependência de  $T$  com relação a  $i$ .

fato de que a distribuição de  $T$  é dada por um mecanismo aleatório. Ou seja, todos os  $n$  indivíduos têm probabilidade positiva de serem controles ou tratados. Contudo, em experimentos não-aleatorizados  $b(\mathbf{x}_i)$  é desconhecido e deve ser estimado por um modelo estatístico, visto que a atribuição do tratamento é determinística.

Além disso, em experimentos aleatorizados as respostas  $r_1$  e  $r_0$  são condicionalmente independentes de  $T$  dado  $X$ , ou seja,  $R_i \perp\!\!\!\perp T_i | \mathbf{x}_i$  (hipótese de ignorabilidade ou independência condicional nas covariáveis), com  $0 < \Pr(T_i = 1 | \mathbf{x}_i) < 1, \forall i$  (hipótese de sobreposição). Isto permite incluir na matriz  $X$  variáveis que influenciem simultaneamente a atribuição do tratamento (1 ou 0) a cada  $i$  e a variável resposta  $R$ . Esta independência condicional nem sempre é assegurada, em experimentos não-aleatorizados, para todas as variáveis observadas na pesquisa, mas apenas para um subconjunto de  $q \leq p$  variáveis. Nesta situação, a atribuição do tratamento é dita ser *fortemente ignorável* com relação às  $q$  covariáveis (excluímos de  $X$  as colunas correspondentes às outras  $p - q$  variáveis), ou abreviadamente, PATFI (Pressuposto da Atribuição do Tratamento Fortemente Ignorável). Por esta razão, as  $q$  variáveis escolhidas para cálculo do escore de balanceamento  $b(\mathbf{x}_i)$  em experimentos não-aleatorizados devem ser mensuradas antes da atribuição do tratamento para que os indivíduos sejam comparáveis. Na YNSPMS, estes são representados pelos que não residem em domicílios com beneficiários do SWF na 1ª rodada. Uma parte destes indivíduos - os tratados - mudarão esta característica a partir da 2ª rodada.

Considere agora a distribuição amostral de  $T$  dadas as  $q$  variáveis escolhidas para  $X$  e denote por  $\Pr(T_i = 1 | \mathbf{x}_i) = c(\mathbf{x}_i)$  a probabilidade do indivíduo  $i$  receber o tratamento. Como assumimos que a atribuição do tratamento é estocasticamente independente nos  $n$  indivíduos, temos que

$$\Pr(T_1, \dots, T_n | \mathbf{x}_1, \dots, \mathbf{x}_n) = \prod_{i=1}^n c(\mathbf{x}_i)^{t_i} [1 - c(\mathbf{x}_i)]^{1-t_i} \quad (2.1)$$

em que  $c(\mathbf{x}_i)$  é o *escore de propensão* para o indivíduo  $i$ .

A partir das definições acima para os escores de balanceamento e propensão, temos dois resultados importantes:

*Teorema 1.* (Rosenbaum & Rubin, 1983, p. 44) Seja  $b(\mathbf{x}_i)$  uma função qualquer de  $\mathbf{x}_i$ . Então,  $b(\mathbf{x}_i)$  é um escore de balanceamento, ou seja,  $\mathbf{x}_i \perp\!\!\!\perp T_i | b(\mathbf{x}_i), \forall i$ , se, e somente se, o escore de propensão  $c(\mathbf{x}_i)$  é tal que  $c(\mathbf{x}_i) = f[b(\mathbf{x}_i)]$  para alguma função  $f$ . Nesta situação,  $b(\mathbf{x}_i)$  é dito ser mais *fino* que  $c(\mathbf{x}_i)$ .

*Prova.* ( $\Rightarrow$ ) Vamos primeiro supor que  $c(\mathbf{x}_i) = f[b(\mathbf{x}_i)]$ . Como  $c(\mathbf{x}_i) = \Pr(T_i = 1 | \mathbf{x}_i)$ , para mostrar que  $b(\mathbf{x}_i)$  é um escore de balanceamento, é suficiente mostrar que

$$\Pr[T_i = 1 | b(\mathbf{x}_i)] = c(\mathbf{x}_i) \quad (2.2)$$

Pela definição de  $c(\mathbf{x}_i)$ , temos ao tomar sua esperança condicional a  $b(\mathbf{x}_i)$  que

$$\mathbb{E}[c(\mathbf{x}_i) | b(\mathbf{x}_i)] = \Pr(T_i = 1 | b(\mathbf{x}_i))$$

Porém,  $c(\mathbf{x}_i) = f[b(\mathbf{x}_i)]$ , logo

$$\mathbb{E}[c(\mathbf{x}_i)|b(\mathbf{x}_i)] = c(\mathbf{x}_i)$$

Portanto,  $b(\mathbf{x}_i)$  é um escore de balanceamento.

( $\Leftarrow$ ) Suponha por absurdo que  $b(\mathbf{x}_i)$  é um escore de balanceamento, mas que  $c(\mathbf{x}_i)$  não pode ser escrito como uma função  $f$  de  $b(\mathbf{x}_i)$ . Então, existem  $\mathbf{x}_k$  e  $\mathbf{x}_l$ , com  $k \neq l$ , tais que  $c(\mathbf{x}_k) \neq c(\mathbf{x}_l)$  mas  $b(\mathbf{x}_k) = b(\mathbf{x}_l)$ .

Porém, pela definição do escore de propensão,  $\Pr(T_k = 1|\mathbf{x}_k) \neq \Pr(T_l = 1|\mathbf{x}_l)$ , logo  $T$  e  $X$  não são independentes dado  $b(\mathbf{x}_i)$ . Portanto,  $b(\mathbf{x}_i)$  não é um escore de balanceamento. Para sê-lo, é necessário que  $c(\mathbf{x}_i) = f[b(\mathbf{x}_i)]$ .  $\square$

Como corolário do teorema acima, temos que o escore de propensão também é um escore de balanceamento, bastando tomar a função identidade:  $c(\mathbf{x}_i) = b(\mathbf{x}_i)$ . Isto implica que se um par controle-tratado é homogêneo com relação ao escore de propensão  $c(\cdot)$ , então os indivíduos deste par terão a mesma distribuição nas covariáveis consideradas.

*Teorema 2.* (Rosenbaum & Rubin, 1983, p. 45) Se o PATFI é válido para as  $q$  covariáveis escolhidas para  $X$ , ou seja,  $(r_1, r_0) \perp\!\!\!\perp T_i|\mathbf{x}_i$  com  $0 < \Pr(T_i = 1|\mathbf{x}_i) < 1 \forall i$ , então o PATFI também é válido para qualquer escore de balanceamento  $b(\mathbf{x}_i)$ :  $(r_1, r_0) \perp\!\!\!\perp T_i|b(\mathbf{x}_i)$  com  $0 < \Pr[T_i = 1|b(\mathbf{x}_i)] < 1 \forall i$  e  $\forall b(\cdot)$ .

*Prova.* A desigualdade dado  $b(\mathbf{x}_i)$  segue diretamente da desigualdade dado  $\mathbf{x}_i$ . Logo, basta provar que

$$\Pr[T_i = 1|r_1, r_0, b(\mathbf{x}_i)] = \Pr[T_i = 1|b(\mathbf{x}_i)] \quad (2.3)$$

O que, pela equação (2.2) do teorema anterior, é equivalente a mostrar que

$$\Pr[T_i = 1|r_1, r_0, b(\mathbf{x}_i)] = c(\mathbf{x}_i)$$

Analogamente ao que foi feito na prova do teorema anterior, pela definição de  $c(\mathbf{x}_i)$  e sabendo que  $c(\mathbf{x}_i) = f[b(\mathbf{x}_i)]$ , temos ao tomar sua esperança condicional às respostas  $r_1$  e  $r_0$  e ao escore de balanceamento  $b(\mathbf{x}_i)$  que:

$$\begin{aligned} c(\mathbf{x}_i) &= \mathbb{E}[c(\mathbf{x}_i)|r_1, r_0, b(\mathbf{x}_i)] \\ &= \mathbb{E}[\Pr(T_i = 1|\mathbf{x}_i)|r_1, r_0, b(\mathbf{x}_i)] \\ &= \mathbb{E}[\Pr(T_i = 1|r_1, r_0, \mathbf{x}_i)|r_1, r_0, b(\mathbf{x}_i)] \\ &= \Pr[T_i = 1|r_1, r_0, b(\mathbf{x}_i)] \end{aligned}$$

Logo, o PATFI é válido para qualquer escore de balanceamento  $b(\mathbf{x}_i)$ .  $\square$

O teorema anterior garante que, se as covariáveis escolhidas influenciam ao mesmo tempo na atribuição do tratamento para cada indivíduo  $i$  e na correspondente resposta

( $r_1$  ou  $r_0$ ), então qualquer escore de balanceamento (incluindo o de propensão) aplicado a estas mesmas covariáveis terá o mesmo comportamento.

Definido o escopo teórico para o escore de propensão, discutiremos como modelá-lo em duas etapas: escolha do modelo estatístico apropriado com relação à distribuição suposta para a variável dependente (note que  $T$  é binária), e seleção de covariáveis dentre as variáveis pesquisadas na amostra para o modelo. Como toda análise estatística a ser descrita nesta dissertação, dividiremos cada etapa da modelagem do escore para o caso IID e para a situação em que o PAC é considerado, com maior enfoque nesta última perspectiva.

## 2.3 Escolha da Distribuição para o Escore de Propensão

Como  $T$  (RSWF), a variável da atribuição de tratamento, é categórica binária (igual a 1 se o indivíduo é tratado e a 0 se é controle), não é recomendável ajustá-la por um modelo de regressão linear clássico. É verdade que na literatura há uma sugestão neste sentido: o Modelo de Probabilidade Linear (MPL). O MPL assume que para cada indivíduo  $i$ ,  $i = 1, \dots, n$ , a variável dependente  $Y_i$  segue uma distribuição de Bernoulli com parâmetro  $p_i$ , em que  $p_i$  é a probabilidade de que  $Y_i$  seja igual a 1 e seu complementar a probabilidade de que seja igual a zero. Logo, seu valor esperado, igual a  $p_i$ , deve pertencer ao intervalo  $I = [0, 1]$ . Também é pressuposta linearidade de  $Y_i$  com relação aos parâmetros da regressão ( $Y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$ ),  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)'$ . Entretanto, esta abordagem tem sérios problemas teóricos (Gujarati & Porter, 2009):

- O termo de erro  $\varepsilon_i$  também seguirá obrigatoriamente uma distribuição de Bernoulli. O modelo linear de regressão clássica não exige uma distribuição específica dos erros para garantir que as estimativas pontuais dos coeficientes (parâmetros da regressão) sejam não enviesadas. Porém, não podemos inferir de forma análoga à situação em que  $Y_i$  é contínua em  $\mathbb{R}$  se a amostra for pequena. Para grandes amostras, podemos usar resultados assintóticos garantidos pelo Teorema Central do Limite (TCL);
- A interpretação do parâmetro  $\beta_j$ ,  $j = 1, \dots, q$  com relação à variação na média de  $Y_i$  quando variamos a estimativa correspondente em uma unidade (com os demais parâmetros fixados) é incorreta, pois supõe que as variações na média são constantes independentemente do ponto inicial em que ocorreu a variação unitária da estimativa;
- Finalmente, este modelo permite que valores preditos para  $Y_i$  estejam fora do intervalo  $I$ , o que também é incorreto e invalida qualquer análise feita a partir do MPL.

Para resolver o problema da modelagem de variáveis binárias, foi proposta na literatura uma nova abordagem: o uso de funções de distribuição aplicadas ao preditor linear  $X\boldsymbol{\beta}$ . Como estas funções têm sua imagem restrita ao intervalo  $I = [0, 1]$  para qualquer valor em  $\mathbb{R}$  e estabelecem uma relação não-linear entre as covariáveis e a variável

dependente (binária), elas se tornaram em uma alternativa correta do ponto de vista teórico e amplamente utilizadas em análises estatísticas.

Em princípio, qualquer função de distribuição pode ser usada para ligar o preditor linear  $\boldsymbol{\eta} = X\boldsymbol{\beta}$  ao valor esperado  $\mathbb{E}(Y_i) = p_i$ , mas algumas têm sido mais utilizadas, como por exemplo as funções de distribuição logística e gaussiana ou normal (os modelos correspondentes são denominados *logit* e *probit*). Nesta dissertação, optamos pela escolha da primeira para modelar o escore de propensão, uma vez que a regressão logística permite a interpretação dos coeficientes em termos de probabilidades e também da razão de chances, sendo por esta razão a mais utilizada na literatura sobre modelagem do escore de propensão. Além disso, o modelo *logit* possui resultados similares quando comparado a modelos com outras funções de distribuição, como o modelo *probit*, na modelagem do escore de propensão (Caliendo & Kopeinig, 2008) quando há apenas dois possíveis valores para  $T$  (tratamento binário). Consequentemente, o nível de tolerância para pareamento dos escores preditos de indivíduos tratados com os de controles (a ser discutido no próximo capítulo) é em geral dado por uma função do desvio padrão dos escores preditos pelo modelo *logit* para todos os indivíduos da amostra na ampla maioria dos *software* estatísticos<sup>3</sup>.

Para estimar os parâmetros  $\beta_j$  da regressão, é necessário inverter a função de distribuição, obtendo a correspondente função quantílica. Logo, a função quantílica é uma *função de ligação*. Os modelos envolvendo funções de ligação entre  $\mathbb{E}(Y_i)$  e  $\boldsymbol{\eta}$ , não apenas para variáveis dependentes binárias mas também categóricas de maneira geral; discretas e mesmo contínuas (independentemente do suporte), são denominados Modelos Lineares Generalizados (MLG).

### 2.3.1 Modelos Lineares Generalizados

Apresentados inicialmente por Nelder & Wedderburn (1972), os MLG foram a primeira proposta de uma modelagem mais ampla no sentido de permitir a regressão de uma variável com observações IID não importando a sua natureza (categórica, discreta ou contínua). Como condição, basta que a distribuição da variável dependente pertença à família exponencial. Logo, o modelo de regressão linear normal clássico é um caso particular da classe dos MLG, em que a função de ligação é a identidade. Nelder & Wedderburn também apresentam novas funções de ligação para a distribuição normal ( $Y_i$  é real) bem como funções de ligação específicas para as distribuições gama ( $Y_i$  é real positiva); Poisson ( $Y_i$  é discreta não-negativa) e binomial ( $Y_i$  é categórica binária). Para modelar o escore de propensão, usaremos esta última. Se as observações  $Y_i$  são supostas IID, o vetor de parâmetros  $\boldsymbol{\beta}$  é estimado por Máxima Verossimilhança (MV). Após a introdução dos MLG's, várias extensões foram propostas na literatura. Dentre elas, se destacam:

<sup>3</sup> O pacote `Matching` (Sekhon, 2011), usado para o pareamento dos escores neste trabalho, trabalha com esta definição para o nível de tolerância

- Os modelos de quase-verossimilhança (Wedderburn, 1974), que especificam uma relação envolvendo apenas a média e a variância das observações na variável dependente, sem precisar de sua forma distribucional, permitindo que as observações pudessem ser correlacionadas;
- Os modelos de dispersão (Jørgensen, 1983), que alteram a forma escrita na família exponencial da distribuição normal, introduzindo outras possíveis distribuições para o erro da variável dependente;
- Os modelos não-lineares da família exponencial (Cordeiro & Paula, 1989), que permitem o uso de funções não-lineares  $h(\cdot)$  para o preditor ( $\boldsymbol{\eta} = h(X\boldsymbol{\beta})$ ); Os modelos lineares generalizados mistos (Breslow & Clayton, 1993), que incluem efeitos aleatórios gaussianos (normais) no preditor linear.

O desenvolvimento teórico abaixo para a estimação por MV se restringiu aos MLG originais, por esta classe ser suficiente para modelar o escore de propensão e estimar o impacto do SWF pelo método DD.

### 2.3.2 Estimação por Máxima Verossimilhança em MLG

A teoria apresentada para a estimação por MV em MLG segue como fonte básica o livro de Paula (2013) e busca resumir os principais conceitos-chave. Sejam  $(y_1, \dots, y_n)'$  o vetor de observações da variável dependente e  $(Y_1, \dots, Y_n)'$  o vetor das respectivas variáveis aleatórias. Suponha que cada variável  $Y_i$ ,  $i = 1, \dots, n$  tenha função densidade (se contínua) ou de probabilidade (se discreta ou categórica) pertencente à família exponencial, podendo ser expressa por

$$f(y_i|\theta_i, \phi) = \exp \{ \phi [y_i\theta_i - b(\theta_i)] + c(y_i, \phi) \} \quad (2.4)$$

onde  $\phi > 0$  ( $\phi^{-1} > 0$ ) é o parâmetro de precisão (dispersão). Embora tanto  $\theta_i$  quanto  $\phi$  possam ser estimados por MV, nos concentraremos na estimação do primeiro, e por consequência dos parâmetros da regressão  $\beta_j$ ,  $j = 1, \dots, q$ , supondo que  $\phi$  é conhecido.

Seja  $\mathbb{E}(Y_i) = \mu_i$ ,  $i = 1, \dots, n$  e  $\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i|\theta_i, \phi)$  a função de verossimilhança para  $\boldsymbol{\theta}$ . Suponha também que as seguintes condições de regularidade sejam atendidas (Cox & Hinkley, 1974, p. 280-281):

- O espaço paramétrico  $\Theta = (\theta_1, \dots, \theta_n)$  tem dimensão finita; é fechado; compacto e contém o vetor de valores verdadeiros dos parâmetros;
- Para todas as variáveis aleatórias  $Y_1, \dots, Y_n$ , suas distribuições de probabilidade são identificáveis: dados dois pontos distintos de  $\Theta$ , elas nunca serão iguais;
- As derivadas até a 3ª ordem (inclusive) de  $\ell(\theta_i|y_i, \phi) = \log [f(y_i|\theta_i, \phi)]$  com respeito a  $\theta_i$  existem na vizinhança do valor verdadeiro de  $\theta_i$  quase certamente  $\forall i$ ;
- Nesta mesma vizinhança, o valor absoluto da derivada de 3ª ordem vezes  $n^{-1}$  (inverso do tamanho amostral) é limitado por uma função de  $Y_i$  cuja esperança é finita;

- A esperança da derivada de 1ª ordem de  $\ell(\theta_i|y_i, \phi)$  com respeito a  $\theta_i$ , também conhecida como função escore (não confundir com o escore de propensão), é nula  $\forall i$ ;
- A esperança da derivada de 2ª ordem de  $\ell(\theta_i|y_i, \phi)$  com respeito a  $\theta_i$  (duas vezes,  $\forall i = 1, \dots, n$ ), também conhecida como informação de Fisher, é finita e positiva.

Das condições de regularidade acima, decorre que (Jørgensen, 1987, p. 130-131)

$$\mathbb{E}(Y_i) = \mu_i = \frac{db(\theta_i)}{d\theta_i} = b'(\theta_i) \quad (2.5)$$

$$Var(Y_i) = \phi^{-1} \frac{d^2b(\theta_i)}{d\theta_i^2} = \phi^{-1} b''(\theta_i) = \phi^{-1} Var(\mu_i) = \phi^{-1} d\mu_i/d\theta_i \quad (2.6)$$

$$\lim_{\phi \rightarrow \infty} \left[ \sqrt{\phi} \cdot (Y - \mu_i) \right] \xrightarrow{d} N(0, Var(\mu_i)) \quad (2.7)$$

onde  $Var(\mu_i)$  é a função de variância e  $b'(\theta_i)$  e  $b''(\theta_i)$  são, respectivamente, as derivadas de 1ª e 2ª ordem de  $b(\theta_i)$ .

Cada distribuição de probabilidade possuirá a sua respectiva função de variância. Note que a abordagem assintótica em (2.7) difere da usual, em que  $n$  (e não  $\phi$ ) tende ao infinito. Assim, os MLG são definidos por dois componentes: um aleatório, expresso em (2.4), e outro determinístico,  $g(\mu_i) = \eta_i = \mathbf{x}_i\boldsymbol{\beta}$ . A função de ligação  $g(\cdot)$  deve ser monótona; diferenciável e tal que  $g : I \mapsto \mathbb{R}$ , em que  $I$  é o intervalo de variação da variável dependente. Sendo  $\phi$  conhecido, o logaritmo da função de verossimilhança de um MLG com observações independentes entre si é dado por (Paula, 2013, p. 8)

$$\ell(\boldsymbol{\theta}|y, \phi) = \ell(\boldsymbol{\beta}) = \sum_{i=1}^n \ell(\theta_i|y_i, \phi) = \sum_{i=1}^n \phi [y_i\theta_i - b(\theta_i)] + \sum_{i=1}^n c(y_i, \phi) \quad (2.8)$$

Para obter os estimadores por MV de  $\boldsymbol{\beta}$ , é necessário antes calcular a função escore  $\mathbf{U}_\beta(\boldsymbol{\theta})$ , um vetor  $q \times 1$ , e a matriz de informação (esperada) de Fisher  $\mathbf{K}_{\beta\beta}(\boldsymbol{\theta})$ , de ordem  $q$ , para este vetor de parâmetros. Consideremos então  $\boldsymbol{\theta} = \boldsymbol{\beta}$ . A função escore para  $\boldsymbol{\beta}$  pode ser obtida derivando  $\ell(\boldsymbol{\theta})$  com relação a  $\boldsymbol{\beta}$  e é dada por (Paula, 2013, p. 21)

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \phi \cdot X' \Omega^{1/2} V^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \quad (2.9)$$

em que  $X$  é a matriz  $n \times q$  de covariáveis da regressão com posto coluna completo;  $\Omega = \text{diag}\{\omega_1, \dots, \omega_n\}$ ;  $\omega_i = (d\mu_i/d\eta_i)^2 / Var(\mu_i)$ ;  $V = \text{diag}\{Var(\mu_1), \dots, Var(\mu_n)\}$ ;  $\mathbf{y} = (y_1, \dots, y_n)'$  e  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ .

Tomando as derivadas de 2ª ordem da função de log-verossimilhança com relação a  $\boldsymbol{\beta}$  e ao seu transposto, temos que a matriz  $q \times q$  de informação (esperada) de Fisher, denotada por  $\mathbf{K}_{\beta\beta}(\boldsymbol{\theta})$ , é expressa por (Paula, 2013, p. 22)

$$\mathbf{K}_{\beta\beta}(\boldsymbol{\theta}) = \mathbb{E} \left[ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right] = \phi X' \Omega X \quad (2.10)$$

Se  $Y_i$  segue uma distribuição normal  $N(\mu_i, \sigma^2)$  com  $\phi = \sigma^{-2}$ ; funções de ligação  $\theta_i = \mu_i$  (identidade) e de variância  $Var(\mu_i) = 1 \forall i$ , temos  $\mathbf{U}_\beta = \sigma^{-2}X'(\mathbf{y} - \boldsymbol{\mu})$  e  $\mathbf{K}_{\beta\beta} = \sigma^{-2}X'X$ . Por outro lado, se  $Y_i$  segue uma distribuição de Bernoulli com parâmetro  $\mu_i = p_i$ ;  $\phi = 1$ ; funções de ligação  $\theta_i = \log[\mu_i/(1 - \mu_i)]$  e de variância  $Var(\mu_i) = \mu_i(1 - \mu_i) \forall i$ , temos  $\mathbf{U}_\beta = X'(\mathbf{y} - \boldsymbol{\mu})$  e  $\mathbf{K}_{\beta\beta} = X'VX$ . Em ambos os casos, é possível expressar a função escore e a matriz de informação de Fisher em termos de  $\boldsymbol{\beta}$ , bastando tomar  $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$  e  $\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})/[1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})]$ , respectivamente. Consideramos estas duas distribuições em específico pelo fato de que serão usadas na modelagem do escore de propensão (regressão logística ou modelo *logit*) e na estimação do impacto do SWF (modelo de regressão linear normal) quando supusermos que as observações amostrais são IID.

Por fim, o estimador de MV de  $\boldsymbol{\beta}$  é obtido, para o caso geral, através do processo iterativo de Newton-Raphson (Nelder & Wedderburn, 1972). Nem sempre tal estimador terá forma fechada, basta tomar como exemplo a regressão logística ( $Y_i$  segue uma distribuição de Bernoulli, com uso da função de ligação canônica). Assim, a função escore  $\mathbf{U}_\beta$  é expandida em torno de uma estimativa inicial,  $\boldsymbol{\beta}^{(0)}$ . Os detalhes do processo iterativo podem ser encontrados em Paula (2013, p. 25–26). Entretanto, dependendo do modelo a ser ajustado, a maximização da função de log-verossimilhança pode ser feita de forma analítica. Para o modelo normal com preditor linear  $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$ , o estimador por MV  $\hat{\boldsymbol{\beta}}$  tem forma fechada, pois

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) &= \left(\frac{2\pi}{\phi}\right)^{-n/2} \exp -\frac{\phi}{2} \sum_{i=1}^n (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 \\ \Rightarrow \ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = \ell(\boldsymbol{\beta}) &= -\frac{n}{2} [\log(2\pi) - \log(\phi)] - \frac{\phi}{2} (\mathbf{y} - X\boldsymbol{\beta})'(\mathbf{y} - X\boldsymbol{\beta}) \\ \Rightarrow \mathbf{U}_\beta = \mathbf{0} \Leftrightarrow -\frac{\phi}{2}(-2X'\mathbf{y} + 2X'X\hat{\boldsymbol{\beta}}) &\Rightarrow \hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y} \end{aligned}$$

Assintoticamente, temos que  $\hat{\boldsymbol{\beta}} \xrightarrow{d} N_q(\boldsymbol{\beta}, K_{\beta\beta})$  (Sen & Singer, 1993, p. 309). Com este resultado importante, testes de hipóteses para verificar a significância de um ou mais coeficientes de um MLG ajustado, quando comparado ao ajuste sem a(s) variável(eis) sob teste, podem ser construídos. Na literatura, há quatro propostas amplamente conhecidas: o teste da Razão de Verossimilhanças; o teste de Wald; o teste escore e o teste gradiente. O leitor interessado pode consultar o desenvolvimento matemático no caso geral em Paula (2013, seção 1.7) para os três primeiros testes e em Lemonte (2016, capítulos 1 e 2) para o teste gradiente. Todas as estatísticas associadas a esses testes convergem em distribuição para uma  $\chi_q^2$ . Na seção 2.4, discutiremos a seleção de covariáveis através dos testes da RV; de Wald e escore quando consideramos o plano amostral. Antes de desenvolvermos as respectivas expressões, é necessário introduzir as mudanças na estimação dos coeficientes em MLG quando as observações são extraídas através de um PAC.

### 2.3.3 Estimação por Máxima Pseudoverossimilhança em MLG

Suponha agora que as observações  $y_1, \dots, y_n$  compõem uma amostra  $s$  extraída sob um PAC de uma população finita de tamanho  $N$ . Para cada  $y_i$ ,  $i = 1, \dots, n$ , sejam  $x_{i1}, \dots, x_{iq}$  o respectivo vetor de valores nas  $q$  covariáveis e  $w_i$  o peso amostral (possivelmente corrigido por não-resposta) associado ao indivíduo  $i$  selecionado. Suponha também que os pesos amostrais estão corretamente escalonados, ou seja,  $\sum_{i \in s} w_i = N$ . Assuma que as observações  $y_i, x_{i1}, \dots, x_{iq}$  são geradas independentemente  $\forall i$  de uma distribuição com densidade desconhecida  $k(y, \mathbf{x})$  (modelo para a superpopulação ou verdadeiro). Esta última suposição é menos restritiva do que pode parecer num primeiro momento: é possível gerar populações com estruturas de correlação para as observações (não mais independentes entre si) através de variáveis geográficas (Lumley & Scott, 2014).

Nosso desejo é ajustar um modelo paramétrico  $f(y|\mathbf{x}, \boldsymbol{\theta})$  para a densidade (ou função de probabilidade) condicional de  $Y$  dado  $X$ , mas não iremos assumir que este contenha o modelo verdadeiro  $k(y|\mathbf{x})$  (ao contrário do que foi feito em 2.3.2). O melhor ajuste será dado pelo modelo que minimiza a divergência de Kullback-Leibler entre  $f(y|\mathbf{x}, \boldsymbol{\theta})$  e  $k(y|\mathbf{x})$ , o que é equivalente a impor  $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ , onde  $\boldsymbol{\theta}^*$  é solução para a função escore da superpopulação (Lumley & Scott, 2014, p. 2)

$$\mathbf{U}(\boldsymbol{\theta}) = \mathbb{E}_k \left\{ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} = \mathbf{0} \quad (2.11)$$

e, conseqüentemente, o estimador por MV de  $\boldsymbol{\theta}$ .

Para explicar isto, note que  $\forall i$  a divergência de Kullback-Leibler para as duas densidades  $f(y|\mathbf{x}, \boldsymbol{\theta})$  e  $k(y|\mathbf{x})$ , denotada por  $KL(k(y|\mathbf{x}), f(y|\mathbf{x}, \boldsymbol{\theta}))$ , é definida como (Claeskens & Hjort, 2008, p. 24)

$$KL(k(y|\mathbf{x}), f(y|\mathbf{x}, \boldsymbol{\theta})) = \int k(y|\mathbf{x}) \log \left[ \frac{k(y|\mathbf{x})}{f(y|\mathbf{x}, \boldsymbol{\theta})} \right] dy \quad (2.12)$$

Se aplicarmos a Lei Forte dos Grandes Números tomando  $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(y_i|\mathbf{x}_i, \boldsymbol{\theta})$ , a função de log-verossimilhança, temos que (Claeskens & Hjort, 2008, p. 25)

$$\frac{\ell(\boldsymbol{\theta})}{n} \xrightarrow{q.c.} \mathbb{E}_k [\log f(Y|X, \boldsymbol{\theta})] = \int k(y|\mathbf{x}) \log [f(y|\mathbf{x}, \boldsymbol{\theta})] dy \quad (2.13)$$

desde que a integral seja finita, ou seja, a média do logaritmo das densidades individuais  $f(y_i|\mathbf{x}_i, \theta_i)$  converge quase certamente (com probabilidade 1) para a esperança do logaritmo de  $f(Y|X, \boldsymbol{\theta})$ . Assim,  $\boldsymbol{\theta}^*$  converge quase certamente para  $\boldsymbol{\theta}_0$ , em que  $\boldsymbol{\theta}_0$  é o vetor no espaço paramétrico  $\Theta$  que minimiza a distância  $KL(k(y|\mathbf{x}), f(y|\mathbf{x}, \boldsymbol{\theta}))$ . Se o modelo paramétrico ajustado estiver totalmente correto, então  $k(y|\mathbf{x}) = f(y|\mathbf{x}, \boldsymbol{\theta}_0)$  e a distância de Kullback-Leibler será nula.

Entretanto, como nossa amostra foi obtida a partir de um PAC (as observações não são mais IID) de uma população finita, devemos estimar a função escore em (2.11)

considerando os pesos amostrais. Logo, o respectivo ENHT é dado por (Lumley & Scott, 2014, p. 2)

$$\widehat{\mathbf{U}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i \in s} w_i \frac{\partial \log f(y_i | \mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{N} \sum_{i \in s} w_i \mathbf{U}_i(\boldsymbol{\theta}) \quad (2.14)$$

A solução  $\hat{\boldsymbol{\theta}}$  para (2.14) é o *estimador por Máxima Pseudoverossimilhança* (MPV) de  $\boldsymbol{\theta}^*$ . Sob as mesmas condições de regularidade enunciadas em 2.3.2, temos que  $\hat{\boldsymbol{\theta}}$  também é consistente para  $\boldsymbol{\theta}^*$  à medida que  $n$  e  $N$  tendem ao infinito. Seja

$$\hat{\mathbf{J}}(\boldsymbol{\theta}) = -\frac{\partial \widehat{\mathbf{U}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = -\frac{1}{N} \sum_{i \in s} w_i \frac{\partial^2 \log f(y_i | \mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \quad (2.15)$$

o análogo da matriz de informação observada de Fisher e  $I(\boldsymbol{\theta}) = -\partial \mathbf{U}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}' = \mathbb{E} [\hat{\mathbf{J}}(\boldsymbol{\theta})]$  a respectiva matriz de informação esperada. Sob condições adicionais (Fuller, 2009, teorema 1.3.9), temos que o estimador por MPV  $\hat{\boldsymbol{\theta}}$  é assintoticamente normal, ou seja

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N(\mathbf{0}, \text{Var}(\boldsymbol{\theta}^*)) \quad (2.16)$$

à medida que  $n$  e  $N$  tendem ao infinito. Além disso,  $\hat{\mathbf{J}}(\boldsymbol{\theta})$  converge em probabilidade para  $I(\boldsymbol{\theta})$ . Podemos estimar  $\text{Var}(\boldsymbol{\theta}^*)$  por  $\widehat{\mathbf{V}}(\hat{\boldsymbol{\theta}}) = \text{var}(\hat{\boldsymbol{\theta}})$ , a qual nem sempre será obtida analiticamente, mas pode ser aproximada através da Linearização em Série de Taylor (LST, a ser discutida na seção 2.5) por (Pessoa & Silva, 1998, p. 80)

$$\widehat{\mathbf{V}}(\hat{\boldsymbol{\theta}}) \approx [\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}})]^{-1} \text{var} \left[ \sum_{i \in s} w_i \mathbf{U}_i(\hat{\boldsymbol{\theta}}) \right] [\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}})]^{-1} \quad (2.17)$$

onde

$$\text{var} \left[ \sum_{i \in s} w_i \mathbf{U}_i(\hat{\boldsymbol{\theta}}) \right] = \sum_{i \in s} \sum_{j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} [\mathbf{U}_i(\hat{\boldsymbol{\theta}})] [\mathbf{U}'_j(\hat{\boldsymbol{\theta}})]$$

é a matriz de variância (estimada) do estimador para o total populacional dos vetores de funções escores considerando o PAC e

$$I(\boldsymbol{\theta}^*) = \left. \frac{\partial T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \sum_{i=1}^N \left. \frac{\partial \mathbf{U}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

$$\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}) = \left. \frac{\partial \widehat{T}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \sum_{i \in s} w_i \left. \frac{\partial \mathbf{U}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

são a matriz de informação avaliada em  $\boldsymbol{\theta}^*$ , se os dados para a população estivessem disponíveis, e a respectiva estimativa, mas avaliada em  $\hat{\boldsymbol{\theta}}$ .

Para a ampla classe dos MLG com função de ligação canônica, temos que as equações (2.14) e (2.15) podem ser reescritas como (Lumley & Scott, 2014, p. 6)

$$\widehat{\mathbf{U}}_{\beta}(\boldsymbol{\theta}) = \sum_{i \in s} w_i [y_i - \mu_i] \mathbf{x}_i$$

$$\hat{\mathbf{J}}_{\beta\beta}(\boldsymbol{\theta}) = \sum_{i \in s} w_i \frac{d\mu_i}{d\theta_i} \mathbf{x}_i \mathbf{x}'_i = X' D_{\mu} X$$

onde  $\theta_i = \eta_i = \mathbf{x}_i\boldsymbol{\beta}$  é o preditor linear;  $\mu_i = g^{-1}(\mathbf{x}_i\boldsymbol{\beta})$  é o inverso da função de ligação aplicado no preditor linear e  $D_{\boldsymbol{\mu}}$  é a matriz com elementos diagonais  $w_i [d\mu_i/d\theta_i]$ .

No caso particular da regressão logística, que será utilizada para modelar o escore de propensão, fórmulas específicas de  $\widehat{\mathbf{U}}_{\boldsymbol{\beta}}$  e  $\hat{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta})$  considerando pesos amostrais complexos são desenvolvidas por Roberts *et al.* (1987, p. 2). Como nosso objetivo é avaliar o impacto do SWF da 1ª à 4ª rodada da pesquisa YNSPMS, devemos tomar como  $w_i$  os pesos amostrais longitudinais para todas as rodadas, mesmo que na modelagem do escore de propensão usemos apenas dados da 1ª rodada para as covariáveis (respeitando o PATFI). Isto se explica pelo fato de que definimos nossos controles e tratados com base no que foi respondido, para as quatro rodadas, na variável `p6b_q_15`.

Como principais vantagens em relação à estimação tradicional por MV (amostra com observações IID), as estimativas obtidas por MP são baseadas nas características do plano amostral para a variância assintótica dos estimadores de  $\boldsymbol{\theta}$ . Além disso, as estimativas obtidas (tanto pontual quanto do erro-padrão) serão robustas no sentido de que, em geral, a quantidade  $\boldsymbol{\theta}^*$  da população finita permanece um alvo válido para inferência mesmo quando o modelo especificado por  $f(y_i|\mathbf{x}_i, \boldsymbol{\theta})$  não descreve adequadamente a distribuição de  $Y_i, i = 1, \dots, n$ . Entretanto, como todo procedimento que incorpora a informação do PAC, a estimação por MP requer o conhecimento dos pesos amostrais e de pertinência a estratos e conglomerados (em todos os estágios) das observações selecionadas. Este conhecimento nem sempre está disponível por razões operacionais ou de confidencialidade (Pessoa & Silva, 1998).

## 2.4 Critérios para Seleção de Covariáveis

Para atender ao PATFI nas covariáveis selecionadas para modelar o escore de propensão ( $R_i \perp T_i|\mathbf{x}_i$ , com  $0 < \Pr(T_i = 1|\mathbf{x}_i) < 1$ ), alguns cuidados devem ser tomados. Somente indicadores socioeconômicos usados para definir quais domicílios podem receber o SWF ou variáveis que ao menos estejam fortemente correlacionadas com estes podem ser especificados no modelo do escore de propensão, desde que também influenciem a variável resposta  $R$  (TSAT) para a qual o impacto do SWF será avaliado. Como exemplos de indicadores, temos o nível de pobreza do domicílio (domicílios mais pobres em geral possuem adultos desempregados ou com renda baixa) e a presença de idosos (mulheres acima de 55 anos ou homens acima de 60 anos) no domicílio. Ainda, tais indicadores devem ser mensurados antes da atribuição do tratamento (aqui, na 1ª rodada da YNSPMS) ou fixos ao longo do tempo, para que seja possível avaliar o impacto do SWF em ambas as classes, controle e tratamento (Caliendo & Kopeinig, 2008).

Finalmente, adicionar muitas covariáveis ao modelo não é recomendado. Se  $\hat{c}(\mathbf{x}_i)$  é muito próximo ou igual a 0 ou 1 para alguns valores de  $\mathbf{x}_i$ , não será possível realizar

pareamento para os respectivos indivíduos, visto que eles sempre farão parte de uma das duas classes (controle ou tratamento). Nesta situação, estamos violando a restrição imposta para  $\mathbf{x}_i$  quando dissemos que o PATFI é válido para  $X$ . Logo, um componente aleatório é necessário no modelo para garantir que indivíduos idênticos possam ser observados em ambas as classes: dado um conjunto mensurável  $C$ ,  $\Pr(\mathbf{x}_i \in C | T_i = 1) > 0 \Leftrightarrow \Pr(\mathbf{x}_i \in C | T_i = 0) > 0$ . Assim, a restrição  $0 < \Pr(T_i = 1 | \mathbf{x}_i) < 1$  implica que qualquer medida condicional de  $\mathbf{x}_i$  dado  $T_i = 0$  ou  $T_i = 1$  será absolutamente contínua  $\forall i$  (Heckman *et al.*, 1998). Ainda, a presença de muitas covariáveis não significativas a um dado nível nominal aumenta o erro-padrão dos coeficientes associados a cada covariável no modelo, mesmo se esta for significativa (Caliendo & Kopeinig, 2008).

Variáveis significativas que obedecem às restrições feitas acima devem estar presentes no modelo do escore de propensão. Logo, é necessário encontrar o equilíbrio entre um modelo parcimonioso, mas que ao mesmo tempo seja informativo sobre a probabilidade de se receber o tratamento sob estudo, sem a omissão de covariáveis importantes. Como nossas observações foram extraídas a partir de um PAC, não podemos utilizar testes usuais para comparação de MLG's encaixados diferindo apenas nas variáveis testadas (RV, Wald, escore ou gradiente) se quisermos considerar o plano. Nesta situação, os respectivos testes corrigidos para dados amostrais complexos irão substituí-los. Na literatura, já existem propostas de correção neste sentido para os testes da RV, de Wald e escore. Com relação ao suporte computacional, os dois primeiros estão disponíveis na linguagem R.

Escrevendo o vetor  $\boldsymbol{\theta} = X\boldsymbol{\beta}$  como uma partição na forma  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$ , com  $\boldsymbol{\theta}_1$  sendo um vetor de dimensão  $q$  (correspondente às  $q$  variáveis para as quais testaremos a qualidade do ajuste), tome como hipótese nula  $H_0 : \boldsymbol{\theta}_1^* = \boldsymbol{\theta}_{1;0}$ . O teste bilateral<sup>4</sup> de Wald pode ser escrito, de forma a incorporar o PAC empregado, como (Lumley & Scott, 2014, p. 4)

$$W_n = n(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1;0})\hat{V}_{11}^{-1}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1;0}) \quad (2.18)$$

onde  $\hat{V}_{11}^{-1}$  é matriz bloco diagonal  $q \times q$  de  $\hat{V}(\hat{\boldsymbol{\theta}})$ , correspondente aos  $q$  parâmetros sob teste. Sob as condições enunciadas em 2.3.3,  $W_n$  converge em distribuição para  $\chi_q^2$  quando  $H_0$  é verdadeira.

Apesar da correção implementada para planos amostrais,  $W_n$  continua desvantajosa neste contexto quando comparada aos demais testes, com mais problemas além dos já conhecidos para a estatística Wald usual (Thomas & Rao, 1987), dentre os quais:

- Ausência de invariância sob transformações não-lineares no(s) parâmetro(s) testados;

<sup>4</sup> Um teste de hipóteses é dito *bilateral* se a sua hipótese nula especificar uma igualdade entre duas características populacionais e a hipótese alternativa, a diferença entre elas (nesta situação, é o seu complementar). Logo, um teste de hipóteses significativo a um nível  $\alpha$  confirma a rejeição da hipótese nula e aceitação da alternativa. Nesta dissertação, todos os testes de hipóteses apresentados estão na forma bilateral.

- É necessário conhecer a matriz de covariância estimada para o(s) parâmetro(s) testado(s);
- Comportamento inadequado em pequenas amostras, com taxas de rejeição muito acima do nível verdadeiro de significância, o complementar do nível de confiança (quando realizado várias vezes em amostras simuladas);
- Em pesquisas amostrais com mais de um estágio, os graus de liberdade da matriz de covariância estimada serão bem menores do que sob a suposição de amostragem IID, pois dependem do nº de UPA's, e não de observações, aumentando o erro-padrão das estimativas pontuais;
- Se houver muitas covariáveis, é possível que haja multicolinearidade quase exata e a matriz de covariância estimada seja difícil (se não impossível) de ser invertida.

Por outro lado, o teste da RV tem propriedades melhores do que o teste de Wald, o tornando preferível em relação a este. É invariante sob transformações não-lineares; não exige o cálculo da matriz de covariância estimada e possui taxas de rejeição mais próximas do nível verdadeiro de significância. Para construir seu análogo para dados amostrais complexos a partir da mesma hipótese nula enunciada no teste de Wald, basta tomar a função de máxima pseudoverossimilhança (Lumley & Scott, 2014, p. 5)

$$\hat{\ell}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i \in s} w_i \log [f(y_i | \mathbf{x}_i, \boldsymbol{\theta})] = \frac{1}{N} \sum_{i \in s} w_i \ell_i(\boldsymbol{\theta}) \quad (2.19)$$

A estatística do teste corrigido, denominado Razão de Pseudoverossimilhanças (RP), é dada por (Lumley & Scott, 2014, p. 5)

$$A_n = -2n \left[ \hat{\ell}(\hat{\boldsymbol{\theta}}_0) - \hat{\ell}(\hat{\boldsymbol{\theta}}) \right] \quad (2.20)$$

em que  $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}'_{1;0}, \boldsymbol{\theta}'_2)'$  e  $\hat{\boldsymbol{\theta}}_0$  é a solução de  $\widehat{\boldsymbol{U}}_2(\boldsymbol{\theta}_0) = \mathbf{0}$ , onde

$$\widehat{\boldsymbol{U}}_2(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i \in s} w_i \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_2}$$

é a função escore estimada (pelo ENHT).

A estatística  $A_n$  converge em distribuição, sob  $H_0$ , para  $\sum_{j=1}^q \delta_j Z_j^2$  à medida que  $n$  e  $N$  tendem ao infinito. Aqui,  $Z_1, \dots, Z_q$  são variáveis aleatórias independentes  $N(0, 1)$  e  $\delta_1, \dots, \delta_q$  são autovalores da matriz  $\Delta = \left( I_{11}^* - I_{12}^* I_{22}^{*-1} I_{21}^* \right) V_{11}$ , em que  $V_{11} = V_{11}(\boldsymbol{\theta}^*)$  é a matriz de covariância assintótica de  $\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1;0})$  e

$$I^* = I^*(\boldsymbol{\theta}) = \begin{bmatrix} I_{11}^* & I_{12}^* \\ I_{21}^* & I_{22}^* \end{bmatrix}$$

é a matriz de informação esperada de Fisher. No apêndice de seu artigo, Lumley & Scott (2014) mostram as condições para a validade deste resultado. Note que a matriz de informação (observada) ponderada  $\hat{J}(\boldsymbol{\theta})$  é uma média estimada baseada em todas as  $n$

observações, de modo que não há problemas de estabilidade (o erro-padrão das estimativas pontuais é baixo) com  $\hat{J}_{22}^{-1}$  ao estimar o termo  $I_{11}^* - I_{12}^* I_{22}^{*-1} I_{21}^*$ . Devido a esta e as demais vantagens já citadas com relação ao teste (corrigido para planos amostrais) de Wald, o teste da RP foi escolhido como critério para seleção de covariáveis no modelo do escore de propensão para RSWF, tomando como nível de significância  $\alpha = 1 - \gamma = 5\%$ .

Encontrar uma distribuição de referência para a qual  $A_n$  (pelo menos) convirja sob  $H_0$  não é tão trivial quanto para  $W_n$ , uma vez que  $A_n$  converge para uma combinação linear de variáveis aleatórias gaussianas (normais) padrão, independentes e elevadas ao quadrado (também denominadas formas quadráticas em variáveis aleatórias normais). Se os valores dos coeficientes  $\delta_1, \dots, \delta_q$  são conhecidos, é possível usar rotinas computacionais para calcular numericamente os quantis da distribuição assintótica de  $A_n$  e conseqüentemente obter os valores-p e intervalos de confiança desejados. Em geral, é utilizada uma aproximação de Satterthwaite (1946) que corresponda à média e variância da distribuição empírica de  $A_n$  para uma  $\chi^2$  escalonada. Se a pesquisa amostral é de larga escala, aproximações adequadas para a distribuição de  $A_n$  são obtidas exceto possivelmente nos quantis extremos (Lumley & Scott, 2014). Aproximações mais precisas podem ser obtidas ao integrar numericamente a função característica (Davies, 1980) ou somar uma série infinita de percentis  $F$  (Farebrother, 1984). Com menos esforço computacional, uma aproximação ponto de sela (Kuonen, 1999) também produz resultados melhores que a aproximação de Satterthwaite.

Evidentemente, em grande parte das aplicações os coeficientes  $\delta_1, \dots, \delta_q$  não são conhecidos, mas estimados através de  $V_{11}$ , a matriz de covariância assintótica de  $\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{1,0})$ . Porém, devido às características do PAC, as estimativas para  $V_{11}$  e de quaisquer quantidades derivadas dela podem ter alta variabilidade (Lumley & Scott, 2014). Para lidar com isto, podemos seguir uma abordagem similar à de Thomas & Rao (1987) e tratar  $\hat{V}(\hat{\boldsymbol{\theta}})$  como uma variável aleatória escalonada e independente de  $\boldsymbol{\theta}$ . Logo,  $F_n = A_n / \sum_{j=1}^q \hat{\delta}_j$  é a razão de duas formas quadráticas em variáveis aleatórias normais. Baseando-se na aproximação de Satterthwaite (1946), dividimos numerador e denominador pela soma verdadeira dos coeficientes,  $\sum_{j=1}^q \delta_j$ . Logo, a distribuição de  $A_n / \sum_{j=1}^q \delta_j$  é aproximada por uma  $\chi_{\nu_1}^2 / \nu_1$ , com  $\nu_1 = q / (1 + c^2)$  e  $c^2 = \sum_{j=1}^q (\delta_j - \bar{\delta}_j)^2 / (q \bar{\delta}_j^2)$ . Substituindo  $\delta_j$  por  $\hat{\delta}_j$ , obtemos a estimativa para  $c^2$ . Por sua vez, a distribuição de  $\sum_{j=1}^q \hat{\delta}_j / \sum_{j=1}^q \delta_j$  é aproximada por uma  $\chi_{\nu_2}^2 / \nu_2$ , com  $\nu_2 = kq / (1 + c^2)$ . Podemos novamente estimar  $c^2$  ou fixar  $\nu_2 = k$ , desta forma a aproximação para  $\sum_{j=1}^q \hat{\delta}_j / \sum_{j=1}^q \delta_j$  seria a mais conservadora possível. A escolha de  $k$  é, em geral, dada pelo número de UPA's menos o de estratos no 1º estágio<sup>5</sup>. Como resultado destas duas aproximações, a distribuição de  $F_n$  converge para uma  $F(\nu_1, \nu_2)$ ,

<sup>5</sup> A derivação da aproximação dos graus de liberdade pelo nº de UPA's menos o de estratos para estimar variâncias sob dados amostrais complexos pode ser encontrada em Korn & Graubard (1999, seção 5.2). Esta aproximação também é usada para os graus de liberdade da distribuição da estatística  $t$  em um MLG com estimação por MP, a ser visto no capítulo 4.

por ser a razão de duas distribuições qui-quadrado divididas pelos seus respectivos graus de liberdade<sup>6</sup>. Alternativamente, também se pode usar uma aproximação ponto de sela para a distribuição de  $A_n/\chi_{\nu_2}^2$ .

Note que ambas as aproximações ignoram a variabilidade adicional ao usar  $\sum_{j=1}^q \hat{\delta}_j$  no lugar de  $\sum_{j=1}^q \delta_j$ . Lumley & Scott (2014) argumentam que usar  $\nu_2 = k$  compensa melhor a ausência desta variabilidade adicional, especialmente quando  $k$  é relativamente baixo<sup>7</sup>. Nos casos extremos em que  $c^2 = 0$  (todos os coeficientes  $\delta_j$  são iguais) ou em que  $c^2 = q - 1$  (somente um dos coeficientes é diferente de zero),  $F_n$  segue exatamente uma distribuição  $F(\nu_1, \nu_2)$ , com  $\nu_1 = q$  e  $\nu_2 = qk$  no primeiro caso extremo e  $\nu_1 = 1$  e  $\nu_2 = k$  no segundo. Assim, se garante que a distribuição de  $F_n$  estará sempre entre uma  $F(1, k)$  e  $F(q, qk)$ . Por fim, é possível estimar os quantis da distribuição nula (sob  $H_0$ ) de  $A_n$  via replicação *bootstrap* (Beaumont & Bocci, 2009). Para garantir que a distribuição nula está de fato sendo simulada, os autores substituem a hipótese nula por  $H_0 : \theta_1^* = \hat{\theta}_1$ , obtendo resultados similares aos da aproximação de Satterthwaite aplicada para gerar a distribuição  $F(\nu_1, \nu_2)$  desenvolvida anteriormente.

Embora ainda não esteja implementado computacionalmente no R e nos *software* estatísticos de grande circulação, a correção para planos amostrais do teste escore, originalmente proposta por Rao *et al.* (1998) e denominada de teste pseudoescore<sup>8</sup>, é baseada em  $\widehat{U}(\hat{\theta})$ . Tomando para este vetor a mesma partição usada para  $\theta$  e sob a escolha do método de replicação *jackknife* para estimação da matriz de covariâncias (assim como a LST, também a ser discutido na seção 2.4), o teste pseudoescore compartilha da mesma propriedade de invariância sob transformações não-lineares com o teste da RP. Entretanto, o teste pseudoescore sofre dos mesmos problemas de estabilidade que o teste de Wald quando os graus de liberdade para estimação de variâncias são baixos. Apesar disso, não deixa de ser atraente, pois é necessário ajustar apenas o modelo sob  $H_0$ , reduzindo em tese o esforço computacional.

Escolhido um conjunto de covariáveis para a matriz  $X$  com base no teste da RP, também é necessário verificar, se houver modelos concorrentes que atendem ao PATFI, qual especificação minimiza o erro de classificação, prevendo com o maior (menor) percentual de acerto (erro) possível o tratamento dado ao indivíduo  $i$ . Modelos concorrentes podem existir se a especificação completa não respeitar a hipótese de sobreposição para algum

<sup>6</sup> A distribuição  $F$  é bem definida para qualquer valor real positivo dos graus de liberdade associados, não necessariamente inteiros como nas distribuições  $\chi^2$  e  $t$ . Em verdade, valores fracionários para os graus de liberdade são muito comuns quando a distribuição  $F$  é usada para convergência (Johnson *et al.*, 1995).

<sup>7</sup> O pacote `survey` do R (Lumley, 2016) realiza testes da RP, para MLG ajustados sob a existência de um PAC, usando uma aproximação ponto de sela com  $\nu_2 = k$ .

<sup>8</sup> Optou-se por utilizar a nomenclatura *pseudoescore* para se referir ao teste escore corrigido para dados amostrais complexos ao invés da original, *quase-escore* (Rao *et al.*, 1998) para manter a coerência com métodos estatísticos já adaptados para dados amostrais complexos e para evitar qualquer associação com os modelos de quase-verossimilhança.

$\hat{c}(\mathbf{x}_i)$ . Seja  $\mathbb{E}(T) = (\sum_{i=1}^n w_i t_i) / (\sum_{i=1}^n w_i) = \bar{t}$  o ENHT para a média de  $T$  nos  $n$  indivíduos pesquisados dado o plano amostral (Horvitz & Thompson, 1952, p. 668). Se o escore de propensão estimado para o indivíduo  $i$ ,  $\hat{c}(\mathbf{x}_i)$ , for maior do que  $\bar{x}_T$ , o modelo prevê sua classificação como tratado. Se  $\hat{c}(\mathbf{x}_i)$  for menor do que ou igual a  $\bar{x}_T$ , o modelo prevê sua classificação como controle. Após verificar o valor verdadeiro de  $T_i$ , calculamos a proporção de acertos na previsão de controles e tratados, dado o modelo especificado. Ao final, escolhemos aquele que maximiza (minimiza) a proporção de acertos (erros). Para ignorar o plano amostral, basta impor  $w_i = 1$  para todo indivíduo  $i$ . No quarto capítulo, detalhamos para os dados da YNSPMS todo o processo de seleção das covariáveis para a especificação do modelo para o escore de propensão.

Observe que critérios de informação tradicionalmente utilizados em modelos estatísticos, como o de Akaike (1973), não são apropriados para a seleção de modelos para o escore de propensão, pois tais critérios tendem a escolher modelos com parâmetros em excesso (mesmo que algumas das respectivas estimativas não sejam significativas) e que não são parcimoniosos (Claeskens & Hjort, 2008). Como dito anteriormente, quanto mais covariáveis forem especificadas no modelo para o escore de propensão, maior é a chance de que o escore predito  $\hat{c}(\mathbf{x}_i)$  esteja muito próximo ou igual a 0 ou 1 para alguns valores de  $\mathbf{x}_i$ , impossibilitando o pareamento para os respectivos indivíduos.

## 2.5 Estimação de Variâncias sob o PAC

Não apenas o ajuste de um MLG, mas toda análise estatística que envolve dados amostrais complexos e busca realizar inferência para parâmetros sob estudo requer a estimação da variabilidade amostral para os estimadores pontuais (estatísticas) de interesse. Para planos amostrais com AAS ou AES (em um ou mais estágios), expressões exatas para estimadores simples (como a média amostral) já estão definidas na literatura (Cochran, 1977). Entretanto, em geral as observações são selecionadas através de um PAC com tamanhos diferentes para os conglomerados ou com pesos desiguais para os elementos da população. Neste caso, muitas estatísticas de interesse não serão funções lineares simples dos dados observados, o que impede a estimação de forma analítica das respectivas variâncias.

Assim, a adoção de métodos alternativos que aproximem a variância estimada se faz necessária. Neste sentido, são comumente utilizados a Linearização em Série de Taylor (LST, exemplificada na explicação sobre a estimação por MP) do estimador, obtendo a aproximação correspondente de sua variância, e os métodos de reamostragem (criação de réplicas de subamostras a partir da amostra original) não-paramétrica. Dentre estes, os mais conhecidos são: a Replicação em Semiamostras ou Replicação Repetida Balanceada (RRB); a Replicação Repetida *Jackknife* (RRJ, citada brevemente quando falamos do

teste pseudoescore) e a Replicação *Bootstrap* (RB). Para tais métodos, pesos amostrais replicados (referentes aos indivíduos reamostrados em cada réplica gerada) devem ser criados como substitutos dos pesos originais, os quais não são alterados quando estimamos variâncias por LST. Nas duas subseções a seguir, detalhamos o processo de estimação para cada método citado.

### 2.5.1 Linearização em Série de Taylor

A Linearização em Série de Taylor (LST) é um método no qual, para encontrar a variância estimada de um estimador não-linear, é feita uma aproximação deste por uma função linear das observações. Desta forma, se estima a variância sob tal aproximação linear, não produzindo por si só um estimador pronto para a variância. O tratamento matemático apresentado a seguir será restringido ao caso em que a população é finita e há apenas um parâmetro a ser estimado, o qual pode ser função de um ou mais parâmetros populacionais (Seção 6.3 de Wolter, 2007).

Sejam  $N$  o tamanho de uma dada população finita;  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)'$  um vetor  $p$ -dimensional de parâmetros populacionais e  $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_p)'$  o respectivo vetor de estimadores, baseados numa amostra  $s$  de tamanho  $n(s)$ . A forma dos estimadores  $\hat{\lambda}_i, i = 1, \dots, p$  dependerá do PAC utilizado para gerar  $s$ . Na grande maioria das aplicações,  $\boldsymbol{\lambda}$  será um vetor de totais, médias ou proporções para  $p$  diferentes variáveis amostrais. Todos os elementos de  $\hat{\boldsymbol{\lambda}}$  devem ser pelo menos assintoticamente não enviesados e consistentes.

Suponha que nosso parâmetro populacional de interesse seja  $\theta = g(\boldsymbol{\lambda})$  e adotemos como estimador  $\hat{\theta} = g(\hat{\boldsymbol{\lambda}})$ . As duas principais etapas a serem resolvidas são: (i) encontrar uma aproximação para a variância de  $\hat{\theta}$  sob o PAC; e (ii) construir um estimador consistente para a variância de  $\hat{\theta}$ . Se a função  $g(\boldsymbol{l})$ , onde  $\boldsymbol{l}$  é qualquer vetor de valores que  $\boldsymbol{\lambda}$  e  $\hat{\boldsymbol{\lambda}}$  possam assumir, tiver derivadas contínuas de 2ª ordem em uma região do espaço paramétrico  $\Lambda$  contendo  $\boldsymbol{\lambda}$  e  $\hat{\boldsymbol{\lambda}}$ , temos pela expansão em série de Taylor que (Wolter, 2007, p. 230)

$$\hat{\theta} - \theta = \sum_{j=1}^p \frac{\partial g(\boldsymbol{\lambda})}{\partial l_j} (\hat{\lambda}_j - \lambda_j) + R_{n(s)}(\hat{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) \quad (2.21)$$

onde

$$R_{n(s)}(\hat{\boldsymbol{\lambda}}, \boldsymbol{\lambda}) = \frac{1}{2!} \sum_{j=1}^p \sum_{i=1}^p \frac{\partial^2 g(\ddot{\boldsymbol{\lambda}})}{\partial l_j \partial l_i} (\hat{\lambda}_j - \lambda_j) (\hat{\lambda}_i - \lambda_i)$$

e  $\ddot{\boldsymbol{\lambda}}$  é entre  $\hat{\boldsymbol{\lambda}}$  e  $\boldsymbol{\lambda}$ .

A aplicação da expansão em série de Taylor representada pela equação (2.21) é útil para aproximar variâncias quando pressupomos que a população amostrada é finita. Em geral, o termo  $R_{n(s)}(\hat{\boldsymbol{\lambda}}, \boldsymbol{\lambda})$  é desprezado enquanto coeficiente da diferença  $\hat{\theta} - \theta = g(\hat{\boldsymbol{\lambda}}) - g(\boldsymbol{\lambda})$ .

Na prática, aproximar a variância de  $\hat{\theta}$ ,  $Var(\hat{\theta})$ , é equivalente a aproximar o Erro Quadrático Médio (EQM) de  $\hat{\theta}$ ,  $EQM(\hat{\theta})$ . Sabemos que  $EQM(\hat{\theta}) = Var(\hat{\theta}) + \{\mathbb{E}(\hat{\theta} - \theta)\}^2$ , em que o segundo termo do lado direito representa o quadrado do viés de  $\hat{\theta}$ . Porém, a aproximação para este termo é de ordem inferior à de  $Var(\hat{\theta})$ , de modo que  $EQM(\hat{\theta})$  e  $Var(\hat{\theta})$  serão iguais na aproximação de 1ª ordem.

Assim, o EQM de  $\hat{\theta}$  é aproximado em 1ª ordem por (Wolter, 2007, p. 230)

$$\begin{aligned} EQM(\hat{\theta}) &= \mathbb{E}\left\{[g(\hat{\boldsymbol{\lambda}}) - g(\boldsymbol{\lambda})]^2\right\} \\ &= Var\left\{\sum_{j=1}^p \frac{\partial g(\boldsymbol{\lambda})}{\partial l_j} (\hat{\lambda}_j - \lambda_j)\right\} \\ &= \sum_{j=1}^p \sum_{i=1}^p \frac{\partial g(\boldsymbol{\lambda})}{\partial l_j} \frac{\partial g(\boldsymbol{\lambda})}{\partial l_i} Cov\{\hat{\lambda}_j, \hat{\lambda}_i\} \\ &= \mathbf{d}\Sigma_{n(s)}\mathbf{d}' \end{aligned} \quad (2.22)$$

onde  $\Sigma_{n(s)}$  é a matriz de covariância de  $\hat{\boldsymbol{\lambda}}$  e  $\mathbf{d}$  é um vetor  $1 \times p$  com elementos  $d_j = \partial g(\boldsymbol{\lambda})/\partial l_j$ ,  $j = 1, \dots, p$ . Aproximações de ordens maiores são possíveis se estendermos a expansão em série de Taylor e mantivermos os termos adicionais na aproximação. Entretanto, muitas aplicações envolvendo grandes pesquisas amostrais com dados complexos (como a YNSPMS) têm mostrado que a aproximação de 1ª ordem produz resultados satisfatórios, exceto quando a população é altamente assimétrica.

Com isto, finalmente será possível estimar a variância de  $\hat{\theta}$ . Devemos então substituir  $\mathbf{d}$  e  $\Sigma_{n(s)}$  por suas respectivas estimativas amostrais. Suponha que exista um estimador para  $\Sigma_{n(s)}$ , denotado por  $\hat{\Sigma}_{n(s)}$ , especificado de acordo com o PAC. Logo, o estimador da LST para  $EQM(\hat{\theta})$ , e conseqüentemente para  $Var(\hat{\theta})$ , é dado por (Wolter, 2007, p. 231)

$$var_{LST}(\hat{\theta}) = \hat{\mathbf{d}}\hat{\Sigma}_{n(s)}\hat{\mathbf{d}}' \quad (2.23)$$

onde, analogamente,  $\hat{\mathbf{d}}$  é um vetor  $1 \times p$  com elementos  $\hat{d}_j = \partial g(\hat{\boldsymbol{\lambda}})/\partial l_j$ .

Em geral,  $var_{LST}(\hat{\theta})$  é um estimador não enviesado tanto de  $EQM(\hat{\theta})$  quanto de sua aproximação  $\hat{\mathbf{d}}\hat{\Sigma}_{n(s)}\hat{\mathbf{d}}'$ . Porém, ele será consistente desde que  $\hat{\boldsymbol{\lambda}}$  e  $\hat{\Sigma}_{n(s)}$  sejam consistentes para  $\boldsymbol{\lambda}$  e  $\Sigma_{n(s)}$ , respectivamente.

Para amostras de populações muito reduzidas ( $N < 20$ ), a aproximação de variâncias via LST de 1ª ordem não deve ser usada, pois a convergência pode ocorrer de forma bastante lenta ou sequer ser atingida dependendo do tamanho amostral  $n(s)$  escolhido. Isto ocorre pelo fato de que não é possível estabelecer a ordem dos demais termos (ignorados) da expansão sem assumir um modelo de superpopulação ou uma seqüência de populações finitas crescentes em tamanho.

## 2.5.2 Métodos de Reamostragem Não-paramétricos

O primeiro método envolvendo reamostragem não-paramétrica, o qual serviu de inspiração para os métodos RRB; RRJ e RB, foi desenvolvido por Mahalanobis (1958). O método de Mahalanobis requer a seleção de uma amostra completamente probabilística de tamanho  $n$  como um conjunto de  $c = 1, \dots, C$  réplicas amostrais independentes a partir de um plano amostral comum a todas elas. As estimativas, para cada réplica, de  $\theta$  (parâmetro de interesse) e de sua variância com base no plano amostral são dadas respectivamente por (Heeringa *et al.*, 2010, p. 74)

$$\hat{\theta}_{\text{rep}} = \sum_{c=1}^C \hat{\theta}_c / C \quad (2.24)$$

$$\text{var}(\hat{\theta}_{\text{rep}}) = \sum_{c=1}^C (\hat{\theta}_c - \hat{\theta}_{\text{rep}})^2 / [C(C - 1)] \quad (2.25)$$

Apesar de sua simplicidade, o método de Mahalanobis é raramente usado na prática, pois se um grande número de réplicas é usado para obter os graus de liberdade adequados para estimação da variância, a exigência de que cada réplica seja uma “miniatura” da amostra completa restringe a eficiência do plano amostral ao limitar o  $n^\circ$  de estratos que podem ser empregados em cada estágio do processo de seleção. Assim, se um plano amostral estratificado altamente eficiente é usado para selecionar a amostra, o  $n^\circ$  de réplicas independentes bem como os graus de liberdade para estimação da variância podem ser bastante reduzidos (Heeringa *et al.*, 2010).

Durante o final da década de 1950 e a década de 1960, a ideia de Mahalanobis foi estendida para métodos baseados na replicação simples, porém com uso de procedimentos mais eficientes para criar as réplicas, aumentando os graus de liberdade e melhorando a precisão das estimativas, o que contribuiu para a redução de custos na realização de uma pesquisa amostral (a amostra a ser selecionada não precisava ser muito grande). Os métodos de replicação RRB; RRJ e RB surgiram neste contexto. Uma explicação detalhada para a estimação de variâncias por estes três métodos pode ser encontrada em Wolter (2007) e Heeringa *et al.* (2010). Para todos eles, as variâncias estimadas serão assintoticamente não-enviesadas e consistentes. Resumidamente, cada um deles utiliza a seguinte sequência de passos:

- Obtenha  $C$  réplicas (subamostras) a partir da amostra original com base nas regras para o método escolhido (RRB; RRJ ou RB);
- Em cada réplica, corrija os pesos amostrais originais dos indivíduos selecionados, criando  $C$  conjuntos de novos pesos amostrais (observações não pertencentes a uma dada réplica terão peso amostral nulo na mesma);
- Calcule estimativas pontuais ponderadas do parâmetro de interesse para a amostra original e em seguida para cada réplica separadamente, usando os respectivos pesos amostrais;

- Obtenha a variância estimada tomando as estimativas pontuais (na amostra original e nas réplicas) calculadas no passo anterior, de acordo com o método escolhido;
- Construa intervalos de confiança e testes de hipóteses para o parâmetro de interesse com graus de liberdade corrigidos pelo método escolhido.

Nesta dissertação, foi escolhido o uso da LST para aproximar a variância estimada, uma vez que temos uma amostra de AE's (UPA's); domicílios (USA's) e indivíduos bem grande, mesmo dentro de cada província (nível de estratificação no 1º estágio), na YNSPMS. Outra razão para isso é o fato de que o método LST apresenta resultados pelo menos semelhantes (se não melhores), em termos de viés e EQM, quando comparado aos métodos de reamostragem não-paramétrica, mas com esforço computacional bem menor (Krewski & Rao, 1981; Rao & Wu, 1985). Assim, as estimativas pontuais, tanto dos parâmetros para o modelo do escore de propensão quanto para a modelagem por DD, serão acompanhadas do respectivo erro-padrão calculado através da LST quando considerarmos o PAC. A partir das variâncias estimadas, resta apenas quantificar a influência do plano amostral, o que será discutido na seção a seguir.

## 2.6 Avaliação da Influência do Plano Amostral

Na literatura de Amostragem são duas as quantidades usadas para avaliar a influência do plano amostral sobre uma dada estimativa pontual de um parâmetro de interesse: o Efeito do Plano Amostral (EPA) e o Efeito do Plano Amostral Ampliado (EPAA). O objetivo desta seção é esclarecer as principais diferenças entre ambas a partir dos trabalhos que as introduzem, uma vez que mesmo programas computacionais nem sempre fazem uma interpretação totalmente correta destas quantidades<sup>9</sup>.

### 2.6.1 O EPA de Kish (1965)

A primeira definição de uma quantidade para representar a influência de um PAC na precisão das estimativas obtidas com os dados da pesquisa amostral correspondente foi feita por Leslie Kish em seu livro *Survey Sampling* (1965, p. 258). Denominada Efeito do Plano Amostral (EPA, em inglês: *design effect* ou, de forma abreviada: *deff*), e embora tenha sido definida inicialmente apenas para a média amostral  $\bar{y}$ , ela pode ser avaliada para qualquer estimador pontual  $\hat{\theta}$  de um parâmetro populacional  $\theta$  (Pessoa & Silva, 1998). Escolhido  $\hat{\theta}$ , o Efeito do Plano Amostral de Kish  $EPA(\hat{\theta})$  será dado pela variância deste estimador sob o Plano Amostral Complexo,  $Var_{PAC}(\hat{\theta})$ , dividida pela variância do mesmo estimador calculada sob a hipótese de AAS  $Var_{AAS}(\hat{\theta})$ , ambas calculadas para o mesmo

<sup>9</sup> No pacote `survey` do R (Lumley, 2016), o argumento que permite avaliar a influência do plano amostral (`deff`) diz calcular o EPA, porém na verdade calcula o EPAA, já que o primeiro é calculado apenas com as informações do plano amostral (antes de extrair a amostra), como veremos adiante.

tamanho amostral  $n$ , logo

$$EPA(\hat{\theta}) = \frac{Var_{PAC}(\hat{\theta})}{Var_{AAS}(\hat{\theta})} \quad (2.26)$$

O valor do EPA é único para cada estimador e resultado das influências combinadas de estratificação, conglomeração e pesos desiguais. Quanto maior elas forem, mais longe o valor de  $EPA(\hat{\theta})$  estará de 1. Porém, note que o EPA considera em seu cálculo apenas as variâncias verdadeiras de  $\theta$  sob o PAC e sob AAS, sem considerar qualquer informação da amostra, supondo que a mesma ainda não foi selecionada (Skinner *et al.*, 1989).

### 2.6.2 O EPAA de Skinner *et al.* (1989)

Apesar de ser a primeira medida criada para quantificar a influência de um PAC na precisão das estimativas pontuais, o EPA de Kish considera apenas as variâncias verdadeiras de um dado estimador sob o plano amostral usado para selecionar a amostra e sob AAS (e não os estimadores das respectivas variâncias). Além disso, considera somente o plano amostral em si mesmo como causa das distorções na precisão das estimativas pontuais. Desta forma, o EPA despreza os dados obtidos na amostra e a estrutura da população a qual ela pertence. Como consequência, ignora a influência de tal estrutura combinada ao plano amostral utilizado: para qualquer variável analisada, as observações a serem amostradas são tratadas como completamente independentes entre si, uma vez que se baseia apenas nas distribuições induzidas por aleatorização dos planos amostrais comparados: o PAC efetivamente utilizado e o plano sob AAS (Pessoa & Silva, 1998).

Isto não será verdade se, por exemplo, a população for dividida naturalmente em conglomerados e todos os elementos dentro de cada conglomerado selecionado são amostrados, como gêmeos; casais ou, no caso da pesquisa YNSPMS, pessoas vivendo no mesmo domicílio. Nesta situação, a correlação para uma dada variável entre os elementos de um conglomerado específico (também denominada correlação intraconglomerado<sup>10</sup>) será alta: adultos que residem no mesmo domicílio influenciam uns aos outros nas suas respostas. Pensando nisso, Skinner *et al.* propuseram em seu livro *Analysis of Complex Surveys* (1989, p. 24) uma medida mais abrangente para avaliar a influência de um PAC, também única para cada estimativa e fixada para o tamanho amostral  $n$ , denominada Efeito do Plano Amostral Ampliado ou da Má Especificação (EPAA, em inglês: *misspecification effect* ou, de forma abreviada: *meff*).

Inicialmente, considere que haja uma Estrutura Populacional (EP) implícita para as observações usadas no cálculo do estimador pontual  $\hat{\theta}$  (por exemplo, pessoas que

<sup>10</sup> Definida por Cochran (1977, p. 209), é dada pela variância das observações entre conglomerados distintos dividida pela soma desta mesma variância com a das observações dentro do conglomerado especificado. Sempre será um valor entre 0 e 1, maior quanto menor for a variância dentro do conglomerado.

moram no mesmo domicílio na população-alvo da YNSPMS) e seja  $v_0 = var_{\text{IID}}(\hat{\theta})$  um estimador consistente da variância de  $\hat{\theta}$  calculado sob o pressuposto (incorreto) de que as observações são IID. Devido ao uso de um PAC e também à existência da EP, a estimativa de  $v_0$  se distanciará da variância verdadeira (calculada sob ambas as características) de  $\hat{\theta}$ ,  $Var_{\text{PAC-EP}}(\hat{\theta})$ .

Para avaliar se tal afastamento será grande ou pequeno, é necessário considerar o comportamento de  $v_0$  com relação à distribuição verdadeira da variância de  $\hat{\theta}$  e localizar  $Var_{\text{PAC-EP}}(\hat{\theta})$  também com relação a esta última. Como em geral é difícil obter analiticamente a distribuição verdadeira, é tomada uma medida central de  $v_0$ , a qual será comparada a  $Var_{\text{PAC-EP}}(\hat{\theta})$ . Assim, o Efeito do Plano Amostral Ampliado de Skinner *et al.*  $EPAA(\hat{\theta}, v_0)$  será uma medida não apenas da influência do PAC, mas também da especificação incorreta da EP, dada por

$$EPAA(\hat{\theta}, v_0) = \frac{Var_{\text{PAC-EP}}(\hat{\theta})}{\mathbb{E}_{\text{PAC-EP}}(v_0)} \quad (2.27)$$

onde  $\mathbb{E}_{\text{PAC-EP}}(v_0)$  denota a esperança (medida central escolhida) verdadeira de  $v_0$ . Assim,  $EPAA(\hat{\theta}, v_0)$  dirá em que intensidade  $v_0$  subestima ou superestima  $Var_{\text{PAC-EP}}(\hat{\theta})$ . Quanto mais longe de 1 for o valor de  $EPAA(\hat{\theta}, v_0)$ , maior é a consequência de se ignorar o PAC e a especificação correta da EP.

Para estimadores de variáveis que dizem respeito apenas à unidade selecionada no último estágio de uma pesquisa amostral (na YNSPMS, o domicílio amostrado no 2º estágio), não faz sentido calcular a correlação intraconglomerado, pois não haverá uma EP dissociada do PAC (o conglomerado no último estágio terá tamanho unitário). Desta forma, o EPAA coincidirá com o EPA se não usar nenhuma informação da amostra. Caso contrário, o EPA crescerá de forma bem mais modesta com o aumento da correlação intraconglomerado do que o EPAA. Assim, aceitar apenas o cálculo do EPA em detrimento do EPAA (ignorando a estrutura populacional) resultará na subestimação da variância do estimador  $\hat{\theta}$ , que cresce com o aumento da correlação intraconglomerado. Na YNSPMS e nas pesquisas amostrais domiciliares em geral, o domicílio surge naturalmente como conglomerado devido ao planejamento amostral, mas sem a realização de uma nova etapa de seleção para amostrar os seus residentes (Pessoa & Silva, 1998).

O EPAA é uma função não linear das influências da estratificação, conglomeração, pesos desiguais e das interações entre estes, por isso sua expressão nem sempre é obtida de forma analítica. Neste caso, aproximações devem ser utilizadas a partir dos dados amostrais disponíveis, logo devemos estimar  $Var_{\text{PAC-EP}}(\hat{\theta})$  e  $EPAA(\hat{\theta}, v_0)$ , denotando suas estimativas respectivamente por  $var_{\text{PAC-EP}}(\hat{\theta})$  e  $epaa(\hat{\theta}, v_0)$ . O cálculo do EPAA estimado nos diz até que ponto o PAC utilizado produziu perda de eficiência (se a inflação das variâncias estimadas devido à conglomeração e aos pesos desiguais superam o ganho obtido em precisão com a estratificação, o que é mais comum) e permite identificar

características da EP que também podem afetar a precisão das estimativas pontuais dos parâmetros de interesse. Assim, quando dividimos o tamanho  $n$  da amostra selecionada através do PAC por  $epaa(\hat{\theta})$ , temos o tamanho amostral *efetivo*  $n_{ef}$  de um plano AAS que tenha o mesmo nível de precisão. Desta forma, no quarto capítulo são apresentadas estimativas para o EPAA no modelo do escore de propensão e no modelo do método DD para avaliar o impacto do SWF na variável TSAT, tomando a LST como método para aproximação da variância estimada. Ainda, o EPAA pode ser usado para adaptar testes de hipóteses (paramétricos e não-paramétricos) que envolvam dados amostrais complexos desde que a amostra selecionada seja grande (Pessoa & Silva, 1998; Heeringa *et al.*, 2010).

Suponha agora que há não apenas um, mas  $p$  parâmetros sob interesse dispostos em um vetor  $\theta$ . Sejam  $\hat{\theta}$  o estimador de  $\theta$  e  $\hat{V}_0 = var_{IID}(\hat{\theta})$  a matriz de covariância estimada  $p \times p$  para  $\hat{\theta}$  sob o pressuposto de que as observações são IID. Podemos estender a equação (2.27) definindo o Efeito Multivariado do Plano Amostral Ampliado de  $\hat{\theta}$  e  $\hat{V}_0$ , dado por (Pessoa & Silva, 1998, p. 65)

$$EMPAA(\hat{\theta}, \hat{V}_0) = E = \left[ \mathbb{E}_{PAC-EP}(\hat{V}_0) \right]^{-1} Var_{PAC-EP}(\hat{\theta}) \quad (2.28)$$

em que  $\mathbb{E}_{PAC-EP}(\hat{V}_0)$  e  $Var_{PAC-EP}(\hat{\theta})$  são calculados com respeito à distribuição de aleatorização induzida pelo PAC utilizado.

### 3 Metodologia para Avaliação de Impacto

Suponha que modelamos o escore de propensão (com regressão logística) para todos os indivíduos em cada grupo etário (ou qualquer subpopulação utilizada), também na situação em que o PAC é considerado. A avaliação do impacto de uma política requer que os indivíduos sejam comparáveis ao mesmo tempo em que pertencem a classes diferentes (conforme definição de controles e tratados feita na 2.1). Ou seja, para cada indivíduo tratado, é necessário existir um controle que possua perfil o mais semelhante possível nas covariáveis do modelo especificado para o escore de propensão (consequentemente, com o escore predito mais próximo). Dizemos que a amostra de controles e tratados deve ser *balanceada*.

Assim, a amostra coletada para o grupo das crianças (de 6 a 14 anos) e dos adultos (15 anos ou mais), será reduzida de maneira a preservar somente os indivíduos pareados. Cada par é formado por um indivíduo tratado e por outro controle com o escore de propensão mais próximo. A forma mais comum de verificar se a nova amostra pareada é balanceada para as duas classes é através de testes de hipóteses, os quais também serão corrigidos para a situação em que consideramos o plano amostral, nos valores de cada covariável e do escore predito. Tais testes tomam como hipótese nula a equivalência, entre controles e tratados, para uma medida de locação (como média ou mediana) ou característica da distribuição (como independência ou homogeneidade). Somente após conferir se houve balanceamento em todas as covariáveis isoladamente e nos escores preditos, o que garante a existência dos contrafatuais para cada tratado selecionado para o pareamento, é que podemos utilizar um método para avaliação do impacto da política. Como comparação, os testes também podem ser feitos na amostra original com a finalidade de justificar o pareamento, mostrando que esta era desbalanceada.

Na literatura estatística e econométrica, várias técnicas foram propostas para parear tratados e controles através do escore de propensão e/ou dos valores nas covariáveis, de maneira a determinar qual é o controle mais parecido dado um indivíduo tratado. Neste trabalho, listamos três das técnicas mais conhecidas e aplicadas, cuja metodologia não depende diretamente da variável resposta escolhida (TSAT). Basta que as covariáveis do modelo especificado para o escore de propensão influenciem simultaneamente a atribuição do tratamento e a resposta. São elas: o Pareamento pelo Vizinho mais Próximo ou pela Distância Euclidiana (PDE, como a denominaremos daqui em diante); o Pareamento pela Distância de Mahalanobis (PDM) e o Pareamento por Estratificação, Subclassificação ou Intervalo (PI)<sup>1</sup>.

<sup>1</sup> Novamente para evitar confusão de conceitos e esclarecer a ideia por trás de cada técnica, se optou pela mudança de parte da terminologia clássica.

Mais simples dentre os listados, o PDE basicamente toma o escore predito do indivíduo tratado e procura pelo controle com o escore predito mais próximo pela distância euclidiana. Assim, um par é formado com um indivíduo tratado e outro controle com a menor diferença entre todos os controles que poderiam ser pareados. Esta busca é análoga se procurarmos pelo controle com as respostas mais similares nas covariáveis. O PDM se diferencia do PDE apenas por aplicar a distância de Mahalanobis no lugar da euclidiana para formar os pares, tomando tanto os valores no escore de propensão quanto nas covariáveis e dada por (Rosenbaum & Rubin, 1985, p. 36)

$$d(\mathbf{v}_{i|t=1}, \mathbf{v}_{i|t=0}) = (\mathbf{v}_{i|t=1} - \mathbf{v}_{i|t=0})' C^{-1} (\mathbf{v}_{i|t=1} - \mathbf{v}_{i|t=0}) \quad (3.1)$$

em que  $\mathbf{v}_{i|t=1}$  e  $\mathbf{v}_{i|t=0}$  são vetores de valores, dado que o indivíduo  $i$  é tratado ou controle respectivamente, tais que  $\mathbf{v}_i = (\mathbf{x}'_i, \hat{c}(\mathbf{x}_i))'$  e  $C$  é a matriz de covariância amostral de  $\mathbf{v}_i$  apenas para os controles da amostra. Assim, o PDM é um método de pareamento computacionalmente mais intensivo do que o PDE.

No PI, a ideia é particionar os valores dos escores preditos em um número pré-definido de intervalos, a fim de que tratados e controles tenham, na média, escores de propensão similares em cada intervalo. Logo, o principal desafio deste método é a escolha do nº de intervalos. Sob normalidade, cinco intervalos são suficientes para remover grande parte (cerca de 95%) do viés associado à uma única covariável na amostra completa (Cochran & Chambers, 1965). Entretanto, nem sempre é possível pressupor normalidade ou mesmo simetria na distribuição dos escores preditos, como veremos mais adiante para os modelos do escore de propensão em ambos os grupos etários da YNSPMS.

Pelo fato de exigir pouco esforço computacional quando há muitas observações e também por não ter seu desempenho afetado pela forma da distribuição dos escores preditos, se optou nesta dissertação por parear tratados e controles definidos para a YNSPMS pelo PDE. De fato, este método de pareamento é um dos mais utilizados na literatura<sup>2</sup>. Na próxima seção, detalharemos um pouco mais esta técnica, com possíveis modificações para redução de viés e da variabilidade no conjunto de dados pareados (as quais também podem ser aplicadas nas demais técnicas de pareamento).

### 3.1 O Pareamento pela Distância Euclidiana

Embora seja um método bastante direto, o PDE pode ser realizado de várias maneiras distintas. A primeira modificação diz respeito à seleção com ou sem reposição do controle a ser pareado com o indivíduo tratado. Especialmente em pequenas amostras, se trata de uma escolha entre a redução no viés (com reposição) ou na variabilidade (sem reposição) dos escores entre as duas classes. Quando a distribuição dos escores preditos

<sup>2</sup> Luo *et al.* (2010, tabela 2) fazem uma revisão dos artigos que utilizam o PDE como método de pareamento por escore de propensão para tratados e controles.

difere bastante entre tratados e controles, é preferível selecionar com reposição (Caleido & Kopeinig, 2008). Por exemplo, se tivermos muitos tratados com altos escores de propensão mas poucos controles com esta característica, não fazer a reposição resultará na formação de pares com diferenças altas, aumentando o viés e piorando a qualidade do pareamento. Com a reposição, um mesmo indivíduo controle pode ser usado mais de uma vez. Como a amostra de tratados e controles na YNSPMS é grande (em ambos os grupos etários), a variabilidade na seleção de controles com reposição será praticamente a mesma se feita sem reposição. Além disso, selecionar com reposição não dependerá da ordem em que os controles foram escolhidos quando todos estes possuírem escores preditos distintos entre si (a seleção sem reposição depende da disposição dos tratados no banco original, exigindo aleatoriedade). Finalmente, a seleção sem reposição é inviável na situação em que os tratados são maioria no banco de dados. Por estas razões, se optou nesta dissertação por fazer seleção de controles com reposição.

Uma segunda modificação é a sobreamostragem de controles para um único indivíduo tratado, tomando uma medida ponderada dos escores preditos de um  $n^o$  pré-definido  $M$  de controles mais próximos como representativa do contrafactual a ser pareado. O uso ou não de sobreamostragem também envolve uma troca entre viés e variabilidade. Tomar dois ou mais controles para construir um contrafactual do respectivo indivíduo tratado com escore ponderado claramente reduzirá a variância por usar mais informação, porém aumentará o viés, ou seja, a distância entre as médias dos escores preditos para tratados e controle. De maneira geral, tanto a eficiência relativa (razão do erro-padrão do escore de propensão com apenas um controle por tratado sobre o erro-padrão com mais de um controle por tratado) quanto o viés aumentam a cada vez que mais controles são considerados. Porém, o crescimento da eficiência relativa, dada por  $\sqrt{2M/(M+1)}$ , é limitado no infinito pelo fator  $\sqrt{2}$  (Smith, 1997). Assim, usar muitos controles não é recomendado, mesmo em pequenas amostras. Além de quantificar  $M$ , a sobreamostragem também exige um critério para ponderar os escores dos controles escolhidos. Por exemplo os pesos podem ser uniformes, triangulares (Caliendo & Kopeinig, 2008) ou podemos usar os próprios pesos amostrais longitudinais. Como a amostra de controles é grande em ambos os grupos etários da YNSPMS e considerando que a variabilidade é um “mal menor” comparada ao viés, escolhemos apenas um único controle como contrafactual para cada tratado.

Finalmente, é possível reduzir ainda mais o viés do PDE com reposição e um único controle como contrafactual por indivíduo tratado impondo um nível máximo de tolerância<sup>3</sup> como função dos escores preditos para todos os indivíduos. Entretanto, o tamanho deste nível deve ser escolhido com cautela, uma vez que níveis muito baixos de tolerância podem ocasionar a perda de tratados para os quais não há controle cuja diferença entre os escores

<sup>3</sup> Na literatura do pareamento por escore de propensão, tal nível é conhecido como *caliper* (em português: paquímetro) (Caliendo & Kopeinig, 2008).

atenda ao nível imposto, aumentando a variabilidade entre as classes. Através de estudos de simulação com observações IID (sem o uso de um PAC), Austin (2011) e Wang *et al.* (2013) mostram, respectivamente para as situações em que há dois ou três tratamentos, que o nível ótimo de tolerância (com o menor erro quadrático médio) é dado por 0.2 vezes o desvio padrão dos escores preditos pelo modelo *logit*<sup>4</sup>. Assim, fixamos este valor como o nível de tolerância inicial para o pareamento em ambos os grupos etários nesta dissertação. Uma vantagem adicional do PDE sem sobreamostragem e com nível de tolerância é o atendimento da condição de suporte comum para os escores preditos entre as duas classes (Caliendo & Kopeinig, 2008), que pode ser verificada através de histogramas dos escores preditos para cada classe.

## 3.2 Balanceamento dos Indivíduos Pareados

A ideia básica ao verificar o balanceamento é checar se, após o pareamento, há diferenças entre as duas classes (de controles e tratados) na distribuição das covariáveis ou dos escores preditos. Se houver, dizemos que o pareamento não teve sucesso em construir os contrafatuais de cada indivíduo tratado. Assim, devemos voltar à etapa de modelagem do escore de propensão, reespecificando o modelo com interações entre covariáveis ou potências inteiras positivas de uma mesma covariável (se numérica), por exemplo (Caliendo & Kopeinig, 2008).

Não há um consenso na literatura sobre qual metodologia é a melhor para confirmar se uma amostra pareada é balanceada. Além de métodos inferenciais, nos quais baseamos nesta dissertação para julgar a existência do balanceamento, métodos puramente descritivos também são utilizados. Dentre estes, o mais conhecido é o cálculo da Diferença ou Viés Padronizado (VP). Em cada covariável numérica  $k$ ,  $k = 1, \dots, q$ , do modelo especificado o VP é definido para amostras com observações IID como (Caliendo & Kopeinig, 2008, p. 48)

$$VP = 100 \frac{\bar{\mathbf{x}}_{k1} - \bar{\mathbf{x}}_{k0}}{\sqrt{0.5 [\text{var}(\mathbf{x}_{k1}) + \text{var}(\mathbf{x}_{k0})]}} \quad (3.2)$$

onde  $\mathbf{x}_{k0}$  e  $\mathbf{x}_{k1}$  são o vetor de observações para os controles e tratados, respectivamente, na covariável  $k$ .

O VP é calculado para as amostras original e pareada. Tomando a razão entre estas duas medidas, temos a magnitude da redução do viés na covariável  $k$  ao realizar o pareamento. Como não é um método estatístico, a desvantagem de calcular o VP é justamente a ausência de uma evidência clara de seu sucesso, ainda que em geral valores do VP abaixo de 3% ou 5% para todas as covariáveis especificadas sejam considerados suficientes para dizer que houve balanceamento (Caliendo & Kopeinig, 2008). Ainda, ele

<sup>4</sup> Em termos absolutos, Austin (2011) mostra que fixar o nível de tolerância em 0.02 ou 0.03 também produz os menores índices de erro quadrático médio.

só pode ser usado se todas as covariáveis especificadas no modelo do escore de propensão forem numéricas.

Dentre os métodos inferenciais utilizados para checagem do balanceamento, o mais intuitivo consiste na realização de testes  $t$  para duas amostras, cuja hipótese nula  $H_0$  é a igualdade entre as médias, em uma dada covariável categórica (com dois níveis) ou numérica, ou ainda no escore predito, para as amostras das duas classes. Note que o escore pode ser usado como uma quantidade representativa de todas as covariáveis. Suponha que ignoramos a existência de um PAC e assumimos normalidade na distribuição das proporções da covariável (se categórica) ou de seus valores (se numérica, assim como o escore predito). Temos a hipótese nula  $H_0 : \bar{\mathbf{X}}_{k0} = \bar{\mathbf{X}}_{k1}$  e a estatística de teste (Welch, 1947, p. 32)

$$t_{k;o} = \frac{\bar{\mathbf{x}}_{k1} - \bar{\mathbf{x}}_{k0}}{\sqrt{\frac{\text{var}(\mathbf{x}_{k1})}{n_1} + \frac{\text{var}(\mathbf{x}_{k0})}{n_0}}} \quad (3.3)$$

onde  $n_0$  e  $n_1$  são o nº de observações amostradas classificadas como controles ou tratados. A estatística  $t_{k;o}$  segue aproximadamente uma distribuição  $t$  com graus de liberdade  $\nu_o$  dados por (Welch, 1947, p. 32)

$$\nu_{k;o} = \frac{\left[ \frac{\text{var}(\mathbf{x}_{k1})}{n_1} + \frac{\text{var}(\mathbf{x}_{k0})}{n_0} \right]^2}{\frac{[\text{var}(\mathbf{x}_{k1})]^2}{n_1^2 \nu_1} + \frac{[\text{var}(\mathbf{x}_{k0})]^2}{n_0^2 \nu_0}} \quad (3.4)$$

com  $\nu_0 = n_0 - 1$  e  $\nu_1 = n_1 - 1$  representando os graus de liberdade associados às estimativas  $\text{var}(\mathbf{x}_{k0})$  e  $\text{var}(\mathbf{x}_{k1})$  das variâncias nas classes de controles e tratados, respectivamente. Note que não supomos nada sobre as variâncias (das covariáveis ou do escore) na população de controles ou tratados. Assim, o teste acima também é chamado *teste t de Welch*, pois se diferencia do teste  $t$  clássico para duas amostras por não impor homoscedasticidade<sup>5</sup>.

Por sua vez, na amostra pareada temos tamanhos semelhantes para as duas classes ( $n_0 = n_1$ ). Substituindo esta igualdade nas equações 3.3 e 3.4, temos que

$$t_{k;p} = \frac{\bar{\mathbf{x}}_{k1} - \bar{\mathbf{x}}_{k0}}{\sqrt{\frac{\text{var}(\mathbf{x}_{k1}) + \text{var}(\mathbf{x}_{k0})}{n_0}}}$$

$$\nu_{k;p} = \frac{[\text{var}(\mathbf{x}_{k1}) + \text{var}(\mathbf{x}_{k0})]^2}{[\text{var}(\mathbf{x}_{k1})]^2 + [\text{var}(\mathbf{x}_{k0})]^2} \nu_0$$

Observe que usamos um teste  $t$  para duas amostras (e não um teste  $t$  pareado) mesmo que a amostra aqui seja dita *pareada*, pois estamos comparando diferenças (na

<sup>5</sup> O teste  $t$  de Welch foi a primeira solução, ainda que aproximada, para o problema de Behrens-Fisher. É frequentemente aplicado quando ambas as amostras são grandes. Para pequenas amostras, há três soluções exatas na literatura, todas desenvolvidas posteriormente (Dudewicz *et al.*, 2007).

covariável ou no escore predito) entre controles e tratados para uma medida de locação (Rosenbaum & Rubin, 1985). Entretanto, por supor normalidade nas covariáveis ou nos escores sob teste de ambas as classes e sua hipótese nula contemplar apenas um aspecto distribucional (no caso, a média), o teste  $t$  não será usado como critério para decidir se houve ou não balanceamento após o pareamento.

Assim, nesta dissertação foram adotados dois testes estatísticos não-paramétricos (ou seja, que não fazem qualquer suposição sobre a distribuição dos dados na população) com hipóteses nulas mais restritas do que a igualdade apenas em uma medida de locação. Para covariáveis categóricas, foi usado como critério o teste  $\chi^2$  de Pearson para a hipótese nula de independência entre as observações de ambas as classes (Baser, 2006; Judkins *et al.*, 2007). Para covariáveis numéricas e o escore predito consideramos o teste da soma dos postos de Mann-Whitney<sup>6</sup>, denominado de agora em diante de teste MW (Austin, 2011), para a hipótese nula de homogeneidade entre as populações de controles e tratados. O teste MW é, em geral, mais poderoso que o teste  $t$  para várias classes de distribuições não-normais (Blair & Higgins, 1980).

Tanto o teste  $\chi^2$  de Pearson quanto o teste MW possuem correções para a situação em que o PAC é considerado, assim como o teste  $t$ . Como nos testes para seleção de covariáveis descritos no capítulo anterior e em qualquer teste de hipóteses estatístico, é necessário escolher previamente o nível  $\alpha$  de significância, aqui fixado em 5%. Porém, o objetivo do teste de balanceamento é não rejeitar a hipótese nula, permitindo a avaliação do impacto pelo método DD com a amostra pareada de controles e tratados.

### 3.2.1 O Teste Qui-quadrado de Pearson

Sendo um dos primeiros testes não-paramétricos desenvolvidos na literatura, o teste qui-quadrado ( $\chi^2$ ) de Pearson pode ser feito tomando três hipóteses nulas distintas, todas envolvendo variáveis categóricas (nominais ou ordinais). São elas: *qualidade do ajuste*; *homogeneidade* e *independência*.

O teste  $\chi^2$  de Pearson para a hipótese nula de (boa) qualidade do ajuste equivale a testar se as proporções amostrais para cada categoria de uma única variável sob teste são iguais às respectivas proporções na população. Por outro lado, os testes para as hipóteses nulas de homogeneidade e independência são utilizados quando há duas ou mais variáveis categóricas distintas, exigindo tabelas de múltiplas entradas para sua realização. Supondo que haja duas variáveis categóricas dispostas em uma tabela de contingência, o teste  $\chi^2$  de Pearson para a hipótese nula de homogeneidade possui ideia parecida ao do teste para qualidade do ajuste. Porém, agora se compara a distribuição das proporções para as

<sup>6</sup> Também denominado teste da soma dos postos de Wilcoxon ou de Wilcoxon-Mann-Whitney, sua construção é baseada no teste de postos sinalizados de Wilcoxon para igualdade na média, porém com uma hipótese nula mais restritiva (Mann & Whitney, 1947). Ambos os testes não exigem como pressuposto normalidade das observações nas duas amostras.

categorias de uma das variáveis (coluna da tabela) nos diferentes níveis (representados pelas categorias dispostas nas linhas da tabela) da outra variável. Por sua vez, o teste  $\chi^2$  para a hipótese nula de independência investiga as distribuições conjuntas das duas variáveis categóricas sob análise. No caso clássico de observações IID, os testes para homogeneidade e independência possuem a mesma estatística de teste e convergem assintoticamente para a mesma distribuição. Para esta situação, uma explicação geral do teste  $\chi^2$  para cada hipótese, com ideia para construção da respectiva estatística, pode ser encontrada em Pessoa & Silva (1998, capítulos 7 e 8).

Porém, se a amostra foi obtida através de um PAC, a distribuição assintótica utilizada no caso IID (para as três hipóteses consideradas) não é mais válida. Testes análogos conduzidos em tabelas de contingência para amostras com conglomeração, por exemplo, tendem a ser mais liberais (rejeitam mais do que deveriam). Isto é explicado pela correlação intraconglomerado positiva de cada variável categórica usada para definir a tabela de contingência. Observe também que o teste  $\chi^2$  de Pearson, para as hipóteses de homogeneidade e independência, é definido sobre vetores de proporções para distribuições multinomiais. Sejam  $L$  e  $C$  o nº de categorias nas linhas e colunas da tabela de contingência, respectivamente. Adotando a hipótese nula de independência, temos uma única distribuição multinomial com vetor de proporções com comprimento  $LC$ . Para a hipótese nula de homogeneidade, temos  $L$  multinomiais com vetores de proporções com comprimento  $C$  cada. O processo de contagem gerador das distribuições multinomiais pressupõe que as observações individuais são IID, uma suposição válida apenas sob AAScR ou AAS, se o tamanho amostral  $n$  é suficientemente grande (Pessoa & Silva, 1998).

Assim, as estatísticas de teste têm formulações distintas quando o PAC é considerado, logo a escolha da hipótese nula a ser testada passa a ser crucial. Para dados amostrais complexos, o desenvolvimento matemático do teste  $\chi^2$  de Pearson para as três hipóteses pode ser consultado em Holt *et al.* (1980). Nesta dissertação, nos atemos apenas à estatística com correção de 1ª e 2ª ordens de Rao-Scott  $F$ -ajustada (Rao & Scott, 1979; Thomas & Rao, 1987) do teste  $\chi^2$  de Pearson para a hipótese nula de independência. Para justificar o uso desta configuração, note que não temos dados populacionais para as variáveis pesquisadas na YNSPMS, inviabilizando o uso de testes para qualidade do ajuste. Além disso, o tamanho empírico do teste (a taxa de rejeição da hipótese nula) é menos vulnerável aos efeitos de estratificação; conglomeração e probabilidades desiguais de seleção, portanto mais próximo do nível nominal  $\alpha$  de significância, para a hipótese nula de independência (Holt *et al.*, 1980).

Finalmente, a aplicação das correções de Rao-Scott e do ajuste para a distribuição  $F$  diminuem ainda mais a diferença entre o tamanho do teste e o nível  $\alpha$  (Thomas *et al.*, 1996). Em comparação com outras correções, como a do ajuste pelo EPA (ou EPAA, se as variâncias verdadeiras não forem conhecidas) médio das proporções estimadas em

cada célula (Fellegi, 1980), o desempenho da correção de Rao-Scott de 1ª ordem já é melhor, independentemente da hipótese nula testada (Holt *et al.*, 1980). O uso de correções para a influência do PAC é necessário, uma vez que suas características, em particular a conglomeração (Cohen, 1976), afetam a distribuição da estatística do teste  $\chi^2$  de Pearson. Esta converge para uma soma ponderada de distribuições  $\chi^2$  com cada uma possuindo um grau de liberdade (a ponderação dependerá da influência do plano amostral), qualquer que seja a hipótese nula de interesse (Rao & Scott, 1984). Quando testarmos o balanceamento no caso IID, também o faremos sob a hipótese nula de independência, para que os resultados (com e sem o PAC incorporado) sejam comparáveis.

Considerando o PAC, a hipótese nula de independência para o teste  $\chi^2$  de Pearson em uma tabela de contingência  $L \times C$  ainda é a mesma na situação em que as observações são IID. Sejam  $Y_L$  e  $Y_C$  as duas variáveis categóricas sob teste, com  $L$  e  $C$  categorias, respectivamente. Sejam também  $p_{l;c}$  as proporções de cada célula  $(l, c)$  da tabela;  $p_{l+} = \sum_{c=1}^C p_{l;c}$  a proporção marginal da linha  $l$  e  $p_{+c} = \sum_{l=1}^L p_{l;c}$  a proporção marginal da coluna  $c$ . Formalmente, estamos testando a hipótese nula  $H_0 : p_{l;c} = p_{l+}p_{+c}$  para  $l = 1, \dots, L-1; c = 1, \dots, C-1$ . Reescrevendo  $H_0$  usando contrastes de proporções, temos que  $H_0 : f_{l;c} = p_{l;c} - p_{l+}p_{+c} = 0$ . Agora, considere o vetor  $\mathbf{f}$  formado pelos  $(L-1)(C-1)$  contrastes  $f_{l;c}$  ordenados pelas linhas da tabela de contingência, isto é,  $\mathbf{f} = (f_{1;1}, \dots, f_{1;C-1}, f_{2;1}, \dots, f_{L-1;C-1})'$ . Note que  $\mathbf{f} = h(\mathbf{p}) = h(p_{1;1}, \dots, p_{1;C-1}, p_{2;1}, \dots, p_{L-1;C-1})$  é uma função não linear do vetor  $\mathbf{p}$  de proporções da tabela de contingência (sem as últimas linha e coluna). O teste para a hipótese de independência pode também ser definido em termos da distância entre uma estimativa consistente de  $\mathbf{f}$  e o vetor nulo de mesmo comprimento. Tal estimativa é denotada por  $\hat{\mathbf{f}} = (\hat{f}_{1;1}, \dots, \hat{f}_{1;C-1}, \hat{f}_{2;1}, \dots, \hat{f}_{L-1;C-1})'$ , com  $\hat{f}_{l;c} = \hat{p}_{l;c} - \hat{p}_{l+}\hat{p}_{+c}$  (Pessoa & Silva, 1998, p. 128).

Para estimar  $\hat{p}_{l;c}$ , uma vez que a amostra de tamanho  $n$  extraída não é mais IID, considere o ENHT do tamanho na população de cada célula  $(l; c)$ , denotado por  $N_{l;c\text{-NHT}} = \sum_{s_{l;c}} w_i$ , em que  $w_i$  é o peso da observação  $i$  na amostra  $s$ . Consideremos também o ENHT do total populacional, denotado por  $N_{\text{NHT}} = \sum_s w_i$ , tomando agora todas as observações em  $s$ . Temos que  $\hat{p}_{l;c}$  é dado pela razão entre os dois totais enunciados, isto é,  $\hat{p}_{l;c} = N_{l;c\text{-NHT}}/N_{\text{NHT}}$ . Logo, a frequência na célula  $(l; c)$  é estimada por  $\hat{n}_{l;c} = n\hat{p}_{l;c}$ . Devido à consistência para o ENHT, temos que  $n^{1/2}(\hat{\mathbf{p}} - \mathbf{p})$  converge assintoticamente para uma distribuição normal multivariada  $N(\mathbf{0}; V(\hat{\mathbf{p}}))$  e, conseqüentemente,  $n^{1/2}(\hat{\mathbf{f}} - \mathbf{f})$  converge assintoticamente para uma  $N(\mathbf{0}; V(\hat{\mathbf{f}}))$ , com  $V(\hat{\mathbf{f}}) = \Delta h(\hat{\mathbf{p}})V(\hat{\mathbf{p}})\Delta h(\hat{\mathbf{p}})'$ , onde  $\Delta h(\mathbf{p}) = \partial h(\mathbf{p})/\partial \mathbf{p}$  é uma matriz jacobiana  $(L-1) \times (C-1)$  de derivadas parciais (Holt *et al.*, 1980).

Se o tamanho populacional verdadeiro for conhecido (informação da qual não dispomos para a YNSPMS), é possível ainda aplicar uma correção para populações finitas em  $N_{l;c\text{-NHT}}$  (Quatember, 2015). Sem aplicar nenhuma correção adicional além da

incorporação dos pesos amostrais, temos que a estatística do teste  $\chi^2$  de Pearson para a hipótese nula de independência é dada por (Pessoa & Silva, 1998, p. 129<sup>7</sup>)

$$X_I^2 = n \sum_{l=1}^L \sum_{c=1}^C \frac{(\hat{p}_{l;c} - \hat{p}_{l+}\hat{p}_{+c})^2}{\hat{p}_{l+}\hat{p}_{+c}} \quad (3.5)$$

$$= n \hat{\mathbf{f}}' [\widehat{\mathbf{P}}_{0;f}]^{-1} \hat{\mathbf{f}} \quad (3.6)$$

em que a expressão em (3.6) é nada mais do que uma representação matricial equivalente para (3.5), com

$$\begin{aligned} \widehat{\mathbf{P}}_{0;f} &= \Delta h(\hat{\mathbf{p}}) \widehat{\mathbf{P}}_0 \Delta h(\hat{\mathbf{p}})' \\ \widehat{\mathbf{P}}_0 &= \text{diag}(\hat{\mathbf{p}}_0) - \hat{\mathbf{p}}_0 \hat{\mathbf{p}}_0' \end{aligned}$$

onde  $\hat{\mathbf{p}}_0$  é o vetor de componentes  $\hat{p}_{l+}\hat{p}_{+c}$  e  $\text{diag}(\hat{\mathbf{p}}_0)$  representa a matriz cuja diagonal principal consiste destes mesmos componentes e os demais elementos são nulos. Sob a hipótese nula de independência, a matriz estimada  $\widehat{V}_0(\hat{\mathbf{p}})$  da covariância multinomial de  $\hat{\mathbf{p}}$ , o vetor de proporções estimadas, é dada por  $\widehat{\mathbf{P}}_0/n$ . Por fim,  $\Delta h(\hat{\mathbf{p}})$  é uma matriz jacobiana de derivadas parciais avaliadas em  $\mathbf{p} = \hat{\mathbf{p}}$ .

Seja  $\widehat{V}(\hat{\mathbf{f}}) = \Delta h(\hat{\mathbf{p}}) \widehat{V}(\hat{\mathbf{p}}) \Delta h(\hat{\mathbf{p}})'$  a matriz estimada da covariância do vetor  $\hat{\mathbf{f}}$  de contrastes estimados<sup>8</sup> (observe que  $\widehat{V}(\hat{\mathbf{p}})$  é a matriz estimada da covariância para as proporções estimadas, aproximada por LST). Considere a matriz  $\widehat{E}$  de estimativas dos efeitos multivariados do plano amostral (ampliado), definida pela equação (2.28), aqui dada por (Pessoa & Silva, 1998, p. 130)

$$\begin{aligned} \widehat{E} &= n [\widehat{\mathbf{P}}_{0;f}]^{-1} \widehat{V}(\hat{\mathbf{f}}) \\ &= [n^{-1} \Delta h(\hat{\mathbf{p}}) \widehat{\mathbf{P}}_0 \Delta h(\hat{\mathbf{p}})']^{-1} \Delta h(\hat{\mathbf{p}}) \widehat{V}(\hat{\mathbf{p}}) \Delta h(\hat{\mathbf{p}})' \\ &= [\Delta h(\hat{\mathbf{p}}) \widehat{V}_0(\hat{\mathbf{p}}) \Delta h(\hat{\mathbf{p}})']^{-1} \Delta h(\hat{\mathbf{p}}) \widehat{V}(\hat{\mathbf{p}}) \Delta h(\hat{\mathbf{p}})' \\ &= [\Delta h(\hat{\mathbf{p}})']^{-1} [\widehat{V}_0(\hat{\mathbf{p}})]^{-1} [\Delta h(\hat{\mathbf{p}})]^{-1} \Delta h(\hat{\mathbf{p}}) \widehat{V}(\hat{\mathbf{p}}) \Delta h(\hat{\mathbf{p}})' \\ &= [\Delta h(\hat{\mathbf{p}})']^{-1} [\widehat{V}_0(\hat{\mathbf{p}})]^{-1} \widehat{V}(\hat{\mathbf{p}}) \Delta h(\hat{\mathbf{p}})' \end{aligned} \quad (3.7)$$

Pela equação (3.7) podemos perceber que, se as observações fossem de fato IID, a matriz  $\widehat{E}$  seria igual à identidade, pois nesta situação  $\widehat{V}_0(\hat{\mathbf{p}}) = \widehat{V}(\hat{\mathbf{p}})$ . A correção de 1<sup>a</sup> ordem de Rao-Scott para a estatística (3.5) é dada por (Pessoa & Silva, 1998, p. 130)

$$X_{I;RS1}^2 = \frac{X_I^2}{\bar{e}^*} \quad (3.8)$$

em que  $\bar{e}^*$  é um estimador da média  $\bar{e}$  dos autovalores da matriz  $\widehat{E}$ . Sem precisar calculá-la, podemos estimar  $\bar{e}$  usando somente os valores do EPAA estimado nas células ( $epaa_{l;c}$ ) e

<sup>7</sup> Há uma pequena correção na fórmula da estatística fornecida pelos autores, que pode ser conferida em Holt *et al.* (1980, p. 314) e Rao & Scott (1981, p. 226).

<sup>8</sup> Novamente, uma correção se faz necessária em relação ao texto de Pessoa & Silva (1998), a qual pode ser conferida em Holt *et al.* (1980, p. 313).

marginais da tabela de contingência ( $epaa_{l+}$  e  $epaa_{+c}$ ) por (Pessoa & Silva, 1998, p. 130)

$$\begin{aligned} \bar{e}^* &= \frac{1}{(L-1)(C-1)} \sum_{l=1}^L \sum_{c=1}^C \frac{\hat{p}_{l;c}(1-\hat{p}_{l;c})}{\hat{p}_{l+}\hat{p}_{+c}} epaa_{l;c} \\ &\quad - \sum_{l=1}^L (1-\hat{p}_{l+}) epaa_{l+} - \sum_{c=1}^C (1-\hat{p}_{+c}) epaa_{+c} \end{aligned}$$

Sob  $H_0$  (independência), a distribuição assintótica de  $X_{I;RS1}^2$  é uma  $\chi^2$  com  $(L-1)(C-1)$  graus de liberdade. Se a matriz  $\hat{E}$  for conhecida, é possível aplicar também a correção de 2ª ordem de Rao-Scott tomando os elementos de sua diagonal principal, visto que a soma dos autovalores de uma matriz quadrada (como  $\hat{E}$ ) é sempre igual ao seu traço (Harville, 2001, p. 545-546). Por definição, este operador é dado pela soma dos elementos da diagonal principal de uma matriz quadrada. Ainda, a soma dos quadrados de cada autovalor é igual ao traço do quadrado da matriz correspondente. Com estas propriedades e inspirada pela aproximação de Satterwhaite (1946), a correção de 2ª ordem de Rao-Scott para  $X_I^2$  é dada por (Pessoa & Silva, 1998, p. 130)

$$X_{I;RS2}^2 = \frac{X_I^2}{\bar{e}^*(1+a^2)} \tag{3.9}$$

em que

$$\begin{aligned} \bar{e}^* &= \frac{tr(\hat{E})}{[(L-1)(C-1)]} \\ a^2 &= \frac{\sum_{k=1}^{(L-1)(C-1)} e_k^2}{(\bar{e}^*)^2 (L-1)(C-1)} - 1 = \frac{tr(\hat{E}^2)}{(\bar{e}^*)^2 (L-1)(C-1)} - 1 \end{aligned}$$

O termo  $a^2$  é uma estimativa para o quadrado do coeficiente de variação dos autovalores  $e_k$  de  $\hat{E}$ . Assintoticamente, a estatística  $X_{I;RS2}^2$  converge para uma distribuição  $\chi^2$  com  $(L-1)(C-1)/(1+a^2)$  graus de liberdade. Ainda, ela pode ser reescrita de forma a facilitar seu cálculo como

$$X_{I;RS2}^2 = X_I^2 \frac{tr(\hat{E})}{tr(\hat{E}^2)}$$

Finalmente, o ajuste para a distribuição  $F$ , que pode ser feito em qualquer estatística que convirja assintoticamente para uma distribuição  $\chi^2$ , tem como objetivo reduzir ainda mais a influência do PAC na estatística do teste  $\chi^2$  de Pearson (não importando a hipótese nula adotada) quando o EPAA estimado varia drasticamente entre as proporções estimadas. Thomas *et al.* (1996) mostram que, para valores altos de  $a$ , a estatística do teste  $\chi^2$  de Pearson para a hipótese nula de independência com correção de 1ª e 2ª ordens de Rao-Scott  $F$ -ajustada, aqui denotada por  $X_{I;RS2-F}^2$ , possui tamanho empírico mais próximo do nível nominal  $\alpha$  de significância com relação a  $X_{I;RS2}^2$ , exceto para valores de  $a$  perto de zero

(situação em que  $X_{I;RS2-F}^2$  é mais conservadora). A estatística  $F$ -ajustada é dada por (Thomas & Rao, 1987, p. 636<sup>9</sup>)

$$X_{I;RS2-F}^2 = \frac{X_{I;RS2}^2}{(L-1)(C-1)} \quad (3.10)$$

cuja distribuição converge assintoticamente, sob  $H_0$ , para  $(1+a^2)^{-1}$  vezes uma distribuição  $F$  com  $[(L-1)(C-1)/(1+a^2); (m-H)(L-1)(C-1)/(1+a^2)]$  graus de liberdade, sendo  $m$  o n.º de UPA's e  $H$  o n.º de estratos no 1.º estágio.

### 3.2.2 O Teste da Soma dos Postos de Mann-Whitney

Para covariáveis numéricas ou o escore predito, não podemos aplicar diretamente o teste  $\chi^2$  de Pearson a menos que seja feita uma divisão dos valores em um n.º fixo de intervalos, com cada intervalo representando uma categoria. Entretanto, esta divisão implicará em perda de informação das ordens dos valores na covariável numérica ou no escore predito para as observações que pertencerem a um mesmo intervalo. Como veremos adiante para os modelos do escore de propensão na YNSPMS, o próprio escore predito será nossa única variável numérica (as covariáveis escolhidas serão todas categóricas). Se a hipótese nula de independência não é rejeitada em nenhuma das covariáveis (se todas forem categóricas), podemos utilizar testes para a hipótese nula de homogeneidade, nas classes de controles e tratados, dos escores preditos. Em geral, testes de homogeneidade em variáveis numéricas exigem o pressuposto de independência entre as observações da amostra.

Como os valores do escore predito são contínuos no intervalo  $(0, 1)$ , na literatura há três testes não-paramétricos amplamente conhecidos e apropriados para testar a hipótese nula de homogeneidade (as amostras de ambas as classes possuem a mesma distribuição no escore predito). Eles correspondem às versões para duas amostras dos testes de Kolmogorov-Smirnov (Seção 5.4 de Hollander *et al.*, 2014); de Cramér-von-Mises (Anderson, 1962) e de Anderson-Darling (Scholz & Stephens, 1987). Como principal vantagem em relação ao teste da soma dos postos de Mann-Whitney (MW), estes três testes consideram a diferença de magnitude entre os escores preditos através da sua função de distribuição. É possível construir para cada um a respectiva correção para dados amostrais complexos (Wang, 2012). Entretanto, os testes propostos por Wang são limitados a amostras estratificadas simples, sem conglomeração ou nova seleção de unidades amostrais em 2 ou mais estágios (como na YNSPMS). Por outro lado, o teste MW (Mann & Whitney, 1947) já foi estendido para planos amostrais com múltiplos estágios (Lumley & Scott, 2013)<sup>10</sup>.

<sup>9</sup> A fórmula original foi desenvolvida para o teste de qualidade do ajuste sem estratificação da amostra, aqui adaptada para a hipótese nula de independência com estratificação.

<sup>10</sup> Como vantagem adicional, o teste MW está implementado no pacote `survey` do R (Lumley, 2016), referencial neste *software* para análises de dados amostrais complexos.

Escolhido o teste MW para a hipótese nula de homogeneidade nos escores preditos para ambas as classes de controles e tratados, também descreveremos para este teste a sua construção quando as observações são supostas IID. Formalmente, a hipótese nula para o teste MW não é definida da mesma forma que no teste  $\chi^2$  de Pearson para homogeneidade em tabelas de contingência, por exemplo. Para começar, a natureza das variáveis que cada teste lida nem sempre é a mesma (somente variáveis categóricas ordinais podem ser usadas para ambos os testes). Além disso, o teste MW compara na mesma variável duas amostras distintas, mas não necessariamente definidas por uma segunda variável categórica. Por fim, mesmo no caso em que as observações analisadas pertençam a uma variável categórica ordinal, o teste MW utiliza os postos (a ordem) das observações e não proporções com relação ao total da amostra. Os postos são baseados na combinação das duas amostras originais, e não em cada amostra separadamente, sempre começando do valor 1 e assumindo valores intermediários entre dois inteiros se houver empates (as observações empatadas terão o mesmo posto). Como pressupostos, o teste MW exige apenas que as observações possam ser ordenadas, por isso devem ser pelo menos categóricas ordinais, e que sejam independentes (Corder & Foreman, 2009).

O segundo pressuposto é garantido se não rejeitarmos a hipótese nula de independência no teste  $\chi^2$  de Pearson para nenhuma covariável (se todas forem categóricas) do modelo para o escore de propensão<sup>11</sup>, razão pela qual o teste MW no escore predito é realizado por último para verificar se houve ou não balanceamento. O nível nominal  $\alpha$  de significância também é fixado aqui em 5%<sup>12</sup>. Após ordenar as observações na amostra combinada e obter o respectivo posto, queremos testar a hipótese nula  $H_0$ : *os postos relativos às observações de uma das duas amostras originais não são maiores (ou menores) do que os postos relativos à outra amostra.*

Suponha inicialmente que nossa amostra seja IID. Seja  $T$  tal que  $t \in \{0; 1\}$  a variável de classificação utilizada para todos os indivíduos na população, dividindo também a amostra selecionada ( $T$  pode ser baseada, por exemplo, na atribuição do tratamento para os indivíduos da YNSPMS). Sejam também  $n_0$  e  $n_1$  os tamanhos amostrais, respectivamente, na 1ª e 2ª amostras  $s_0$  e  $s_1$  (partições correspondentes às classes de controles e tratados na YNSPMS), logo a amostra combinada tem tamanho  $n = n_0 + n_1$ . Considere  $M_0 = \sum_{i=1}^n d_i \mathbb{I}(i \in s_0)$  e  $M_1 = \sum_{i=1}^n d_i \mathbb{I}(i \in s_1)$  as respectivas somas dos postos, onde  $d_i$  é o posto da observação do indivíduo  $i$  com relação a uma variável  $Y$  (por exemplo, o escore predito) na amostra combinada, e suponha que não há empates<sup>13</sup> ( $d_i$  é inteiro e positivo  $\forall i$ ). Matematicamente, podemos reescrever  $H_0 : M_0/n_0 = M_1/n_1$  ou ainda  $H_0 : F_{0;Y}(y) = F_{1;Y}(y) \forall y$ , em que  $F_{T=t;Y}(\cdot)$  denota a função de distribuição ou

<sup>11</sup> Se alguma covariável do modelo para o escore de propensão for contínua ou discreta, é possível categorizá-la da mesma forma em ambas as classes e em seguida usar o teste  $\chi^2$  de Pearson.

<sup>12</sup> É possível que o teste  $\chi^2$  de Pearson não rejeite a hipótese nula em nenhuma das covariáveis do modelo para o escore de propensão, mas o teste MW a rejeite nos escores preditos.

<sup>13</sup> Para o caso em que há empates, o leitor pode consultar Putter (1955).

probabilidade (se a variável analisada não fosse contínua) condicional (dada a classe). A razão para isso é que se a média dos postos é igual, então nenhuma das observações (no escore predito) é estocasticamente maior (ou menor) para uma das amostras originais. Conseqüentemente,  $s_0$  e  $s_1$  pertencem à mesma população e são homogêneas entre si. Definida  $H_0$ , calcule as estatísticas  $U_0$  e  $U_1$  (também correspondentes a  $s_0$  e  $s_1$ ) tais que (Corder & Foreman, 2009, p. 58)

$$U_0 = n_0 n_1 + \frac{n_0(n_0 + 1)}{2} - M_0 \quad (3.11)$$

$$U_1 = n_0 n_1 + \frac{n_1(n_1 + 1)}{2} - M_1 \quad (3.12)$$

e escolha como estatística de teste  $U$  a menor delas. Para valores de  $n_0$  e  $n_1$  menores ou iguais a 20, Corder & Foreman (2009) tabelam os valores críticos de  $U$ .

Para obtê-los, Mann & Whitney (1947) inicialmente substituem, na amostra combinada, o valor da observação por 0 se ela veio da 1ª amostra original (por exemplo) e por 1 se ela veio da 2ª amostra, supondo que  $U_1 = U$  foi a estatística escolhida (caso contrário, inverta a atribuição dos valores). Os autores mostram que  $U$  pode ser dado pelo nº de vezes que o valor 1 imediatamente antecede o valor 0 na nova seqüência. Sob  $H_0$ , cada uma das  $(n_0 + n_1)!/(n_0!n_1!)$  seqüências de  $n_0$  0's e  $n_1$  1's possíveis de serem construídas tem a mesma probabilidade de ocorrência. Definindo por  $p_{n_0;n_1}(U)$  a soma das probabilidades de seqüências em que o valor 1 imediatamente antecede o valor 0 na nova seqüência exatamente  $U$  vezes, é possível obter a seguinte relação de recorrência (Mann & Whitney, 1947, p. 51<sup>14</sup>)

$$p_{n_0;n_1}(U) = \frac{n_0}{n_0 + n_1} p_{n_0-1;n_1}(U - n_1) + \frac{n_1}{n_0 + n_1} p_{n_0;n_1-1}(U) \quad (3.13)$$

Usando a relação em (3.13), Mann & Whitney (1947) calculam  $p_{n_0;n_1}(U)$  até  $n_1 \leq n_0 \leq 8$ . Eles também concluem que  $p_{n_0;n_1}(U) = p_{n_1;n_0}(U)$ , por isso não fará diferença escolher entre  $U_0$  ou  $U_1$  como estatística de teste  $U$ ; derivam expressões até o 4º momento para  $U$  e mostram que esta estatística é consistente e converge para uma distribuição normal quando  $n_0$  e  $n_1$  tendem ao infinito. Portanto, em grandes amostras como as das classes de controles e tratados na YNSPMS em ambos os grupos etários (ignorando o plano amostral em um primeiro momento), podemos utilizar a seguinte aproximação para um teste  $z$  (Corder & Foreman, 2009, p. 59)

$$z = \frac{U - \mathbb{E}_{n_0;n_1}(U)}{\sqrt{\text{Var}_{n_0;n_1}(U)}} \quad (3.14)$$

onde

$$\mathbb{E}_{n_0;n_1}(U) = \frac{n_0 n_1}{2}$$

$$\text{Var}_{n_0;n_1}(U) = \frac{n_0 n_1 (n_0 + n_1 + 1)}{12}$$

<sup>14</sup> Há uma pequena errata na fórmula original, devidamente corrigida em Berry *et al.*, 2012, p. 11.

e por fim comparando  $z$  com o valor crítico de acordo com o nível nominal de significância  $\alpha$  fixado antes de realizar o teste em uma tabela de quantis da distribuição normal padrão. Como  $Z \sim N(0, 1)$ , também podemos tomar o quadrado de  $z$  e compará-lo com o valor crítico em uma tabela de quantis da distribuição qui-quadrado com 1 grau de liberdade.

O teste da soma dos postos de Mann-Whitney é um caso particular do teste  $H$  de Kruskal-Wallis (Kruskal & Wallis, 1952) com  $k = 2$  amostras. Para a situação em que incorporamos o PAC, o teste MW também passará por adaptações. O uso de amostragem complexa altera boa parte da otimalidade assintótica (consistência e convergência em distribuição) de estatísticas para testes envolvendo soma de funções (lineares ou não) de postos das observações. Isto não ocorre apenas com o teste MW (Lumley & Scott, 2013), que utiliza a função identidade. Contudo, esta limitação não afetará o uso de tais testes em dados amostrais complexos, desde que o tamanho amostral seja grande o suficiente para permitir inferências baseadas no TCL. Além disso, em geral nada se sabe sobre a distribuição dos dados nas duas classes avaliadas.

Suponha agora que nossa amostra  $s$  de tamanho  $n$  foi extraída de uma população finita de tamanho  $N$ , sendo esta um subconjunto de uma superpopulação, através de um PAC. Novamente, seja  $T$  tal que  $t \in \{0; 1\}$  a variável de classificação utilizada para todos os indivíduos na população, dividindo também a amostra selecionada em  $s_0$  e  $s_1$ , e  $Y$  a variável (pelo menos categórica ordinal) na qual queremos testar a homogeneidade das amostras separadas por  $T$ . Assim como feito para o caso IID e sem perda de generalidade, tomemos a situação em que não há empates nos postos das observações de  $Y$  e que  $Y$  é contínua. Logo, cada indivíduo  $i$  na população está associado a um par de valores  $(y_i, t_i)$  gerado independentemente de qualquer outro com função de distribuição marginal  $F_Y$  na superpopulação. Nosso objetivo é testar a hipótese nula  $H_0 : F_{0;Y}(y) = F_{1;Y}(y) \forall y$ , o que equivale a testar a homogeneidade entre as amostras divididas nas duas classes.

Sejam  $N_0$  e  $N_1$  o n.º de unidades na população finita tais que  $t_i = 0$  e  $t_i = 1$  respectivamente e  $d_i$  o posto de cada observação  $y_i$  nesta mesma população. Agora, denote a função de distribuição empírica na população finita de  $Y$  por  $F_N(y) = N^{-1} \sum_{i=1}^N \mathbb{I}(y_i \leq y)$ . Podemos reescalonar os postos  $d_i$  de modo que  $d_i = F_N(y_i)$ . Temos que a estatística para o teste MW na população finita pode ser definida pela diferença na média dos postos em cada classe (Lumley & Scott, 2013, p. 833)

$$U_N = \frac{1}{N_0} \sum_{i:t_i=0} d_i - \frac{1}{N_1} \sum_{i:t_i=1} d_i \quad (3.15)$$

Observe que  $U_N$  pode ser positiva ou negativa. Na segunda situação, invertemos a ordem dos termos da diferença de modo que  $U_N$  volte a ser positiva. Se  $Y$  fosse discreta, substituiríamos  $d_i$  por  $d_i^* = [F_N(y) + F_N(y^-)]/2$ , onde  $F_N(y^-) = \sum_{i=1}^N \mathbb{I}(y_i < y)/N$ .

Sejam  $\pi_i$  a probabilidade de seleção, através do PAC, de cada indivíduo  $i$  da população finita e  $w_i$  o seu peso amostral, igual ao inverso de  $\pi_i$  se não houver ajustes

por não-resposta e pós-estratificação (por exemplo). Para cada indivíduo amostrado em  $s$ , observamos os valores correspondentes de  $Y$  e  $T$ . Como reescalamos os postos  $d_i$  na população finita, podemos estimá-los com os dados amostrais por  $\hat{d}_i = \hat{F}_n(y_i)$ , em que  $\hat{F}_n(y) = \hat{N}^{-1} \sum_{i \in s} w_i \mathbb{I}(y_i \leq y)$  é um ENHT de  $\hat{F}_N(y)$ . Analogamente ao que foi feito para as estimativas das proporções  $\hat{p}_{l;c}$  na tabela de contingência, podemos tomar  $\hat{N} = \sum_{i \in s} w_i$ , o ENHT do total populacional se este não estiver disponível. Logo,  $\hat{F}_n(y)$  é um estimador consistente para  $F_N(y)$  e  $F_Y(y)$ , as funções de distribuição na população finita e na superpopulação, respectivamente. Com isso, podemos estimar a estatística  $U_N$  por (Lumley & Scott, 2013, p. 833)

$$\hat{U}_n = \frac{1}{\hat{N}_0} \sum_{i \in s_0} w_i \hat{d}_i - \frac{1}{\hat{N}_1} \sum_{i \in s_1} w_i \hat{d}_i \quad (3.16)$$

em que  $\hat{N}_0 = \sum_{i \in s_0} w_i$  e  $\hat{N}_1 = \sum_{i \in s_1} w_i$  são os ENHT correspondentes para os tamanhos populacionais  $N_0$  e  $N_1$ .

Se a estimativa do posto  $\hat{d}_i$  fosse uma quantidade fixa associada ao indivíduo  $i$  na população finita, então  $\hat{U}_n$  seria dada pela diferença entre duas médias estimadas e a inferência baseada nessa estatística seria direta. Contudo,  $\hat{d}_i$  depende não apenas de um PAC, mas também dos valores  $y_i$  para outros indivíduos selecionados na amostra  $s$ . Se substituirmos  $\hat{F}_n(y_i)$  por  $a_i = \hat{F}_Y(y_i)$  na igualdade  $\hat{d}_i = \hat{F}_n(y_i)$ , então a estimativa dos postos não dependerá do plano amostral. Aplicando esta substituição, não há qualquer mudança na distribuição assintótica, sob  $H_0$ , para a estatística  $U_N$  referente à população finita. Para explicar isto, tomemos a convergência fraca (em distribuição) de  $N^{1/2} [F_N(y) - F_Y(y)]$  para uma ponte browniana (um processo estocástico com valor esperado nulo e incrementos não-independentes entre si) através do método delta descrito em van der Vaart & Wellner (1996, seção 3.9). Adaptemos esta ideia para a situação em que há um PAC. Logo,  $\hat{U}_n$  terá (sob  $H_0$ ) a mesma distribuição assintótica nula que  $\tilde{U}_n$ , dada por (Lumley & Scott, 2013, p. 834)

$$\tilde{U}_n = \frac{1}{\hat{N}_0} \sum_{i \in s_0} w_i a_i - \frac{1}{\hat{N}_1} \sum_{i \in s_1} w_i a_i \quad (3.17)$$

Tomando  $\hat{\Psi}_n(y) = \hat{F}_{0;n}(y) - \hat{F}_{1;n}(y)$ , em que  $\hat{F}_{0;n}(y)$  e  $\hat{F}_{1;n}(y)$  são ENHT avaliados apenas nas amostras correspondentes  $s_0$  e  $s_1$  para  $F_{0;Y}(y)$  e  $F_{1;Y}(y)$ , e seguindo o exemplo 20.12 de van der Vaart (1998), temos que

$$n^{1/2} (\hat{U}_n - \tilde{U}_n) = n^{1/2} \int [\hat{F}_n(y) - F_Y(y)] d\hat{\Psi}_n(y) \quad (3.18)$$

Considere agora os seguintes pressupostos (Lumley & Scott, 2013, p. 839–840):

- P1** Os pares de valores  $(y_i, t_i)$  na população finita são gerados de forma independente a partir de alguma distribuição conjunta fixada, porém desconhecida. Ainda, a variável  $Y$  é absolutamente contínua, com função de distribuição marginal  $F_Y$ ;

**P2** Se  $Y$  é limitada e **P1** não é violado, então  $n^{1/2} (\hat{F}_n(y) - F_Y(y))$  converge fracamente para  $W_\pi(y)$ , um processo gaussiano (normal) com amarras (em inglês, *tied-down*), valor esperado nulo e caminho contínuo com  $v \rightarrow \infty$ , em que  $v$  indexa uma sequência de populações finitas.

Temos o seguinte teorema:

*Teorema 3.* (Lumley & Scott, 2013, p. 834) Seja  $\varphi_Y = \int y d\Psi_Y(y)$ . Sob os pressupostos **P1** e **P2**,  $n^{1/2} (\hat{U}_n - \varphi_Y)$  e  $n^{1/2} (\tilde{U}_n - \varphi_Y)$  são assintoticamente normais com média (valor esperado) nula. Mais ainda: sob  $H_0 : F_{0;Y}(y) = F_{1;Y}(y) \forall y$ , então  $\varphi_Y = 0$  e  $n^{1/2} (\hat{U}_n - \tilde{U}_n)$  converge para 0 em probabilidade.

A prova do teorema anterior, estendida para funções  $h(\cdot)$  deriváveis e cuja derivada é limitada e contínua em  $(0; 1)$  dos postos  $d_i$  (não apenas a identidade, associada ao teste MW), pode ser encontrada no apêndice de Lumley & Scott (2013, p. 840–841). Agora,  $\tilde{U}_n$  está resumida a uma diferença entre duas médias de postos estimadas em cada classe. A equivalência assintótica de  $\hat{U}_n$  e  $\tilde{U}_n$  garantida pelo teorema 3 permite que troquemos  $a_i$  por  $\hat{d}_i$  sem alterar distribuição assintótica sob  $H_0$ . Por sua vez, a variância de  $\tilde{U}_n$ ,  $\tilde{V}_n(\mathbf{a})$  (note que  $\tilde{U}_n$  é função de  $a_i$ ), pode ser estimada por  $\hat{V}_n = \tilde{V}_n(\hat{\mathbf{D}})$ . Assim, o teste MW adaptado para dados amostrais complexos é baseado na estatística  $z_n = \hat{U}_n / \sqrt{\hat{V}_n}$ , que converge para uma distribuição normal padrão sob  $H_0$ <sup>15</sup>.

Fixemos um nível nominal  $\alpha$  e tomemos uma amostra fornecida por algum método de pareamento (aqui, o PDE com reposição de controles selecionados; um controle por indivíduo tratado e nível máximo de tolerância). Suponha que o teste  $\chi^2$  de Pearson para a hipótese nula de independência em cada covariável do modelo especificado para o escore de propensão e o teste MW para a hipótese nula de homogeneidade no escore predito (dadas as classes de controles e tratados) não foram significativos. Logo, a amostra pareada pode ser usada para a avaliação de impacto da política (o SWF) na variável resposta de interesse (TSAT). Nesta dissertação, tal avaliação será feita pelo método de Diferença em Diferenças (DD). Assim, os testes de balanceamento servem como refinamento para o modelo do escore de propensão, o qual deve ser reespecificado se algum resultado para tais testes for significativo.

### 3.3 O Método de Diferença em Diferenças

O método de Diferença em Diferenças (DD) é uma das ferramentas mais usadas na literatura econométrica para estimar efeitos ou impactos causais de uma política, quando nem todos os indivíduos da amostra são afetados simultaneamente ou da mesma maneira

<sup>15</sup> No pacote `survey`, a função que realiza o teste MW considerando o PAC utilizado toma o valor crítico a partir de uma distribuição  $t$  com  $m - H$  graus de liberdade (o  $n^\circ$  de UPA's menos o de estratos no 1º estágio).

pela política considerada (Lechner, 2010). A escolha do método DD é adequada quando há informação sobre os indivíduos selecionados antes da atribuição do tratamento. A variável resposta modelada pelo método DD (aqui, TSAT) será utilizada para avaliar o impacto do tratamento, ou seja, avaliar o impacto da política (no caso, o SWF). Em sua forma mais simples (na qual baseamos o desenvolvimento teórico desta seção), para apenas duas classes derivadas da atribuição do tratamento e acompanhadas em dois pontos distintos do tempo, o método DD compara quatro grupos diferentes de indivíduos. Destes, três não são afetados pelo tratamento: os controles e tratados pré-tratamento, e os controles pós-tratamento. Para avaliar o impacto do SWF em TSAT, tomamos como pontos distintos a primeira e a quarta (última) rodadas.

Assim, no método DD supomos que as duas partições de cada classe (de controles e tratados), correspondentes aos dois pontos distintos de tempo considerados, estão sujeitas à mesma tendência ao longo do tempo para a variável resposta, já que o tratamento não interfere no período antes de sua atribuição. Com isso, uma estimativa para o impacto do tratamento em um período no qual é sabido que este não ocorreu pode ser utilizada para remover o efeito de covariáveis ou fatores de confundimento aos quais os controles e tratados podem estar sujeitos após a atribuição do tratamento. Isto quer dizer que usamos a mudança na média da variável resposta para os controles entre os dois instantes em que a amostra foi pesquisada e adicionamos esta mudança à média da resposta para os tratados *antes* de receberem o tratamento para obter a média nestes mesmos indivíduos se eles *não* o recebessem.

### 3.3.1 História do Método DD

O primeiro estudo científico que fez uso explícito do método DD foi realizado por Snow (1855), que estava interessado em saber se a cólera era uma enfermidade transmissível pelo ar ou pela água. O tratamento considerado por Snow foi uma mudança no fornecimento de água para um dos distritos de Londres, a qual vinha anteriormente do rio Tâmsa no centro londrino (muito poluído na época) e passou a vir de um ponto mais próximo à nascente e conseqüentemente bem mais limpo e saudável do rio. Snow demonstrou que os distritos que ainda recebiam a água contaminada apresentaram taxas de mortalidade devido à cólera bem maiores quando comparadas ao distrito no qual houve a mudança no fornecimento.

As conclusões de Snow tornam-se ainda mais confiáveis uma vez que ele tomou domicílios de distritos distintos, porém bem próximos entre si, cujos moradores compartilhavam a mesma qualidade do ar e não sabiam da mudança no fornecimento para um dos distritos. Através de um simples teste de sal, ele conseguiu identificar de onde vinha a água de cada domicílio, tanto para os que passaram a receber diretamente da nascente quanto os que ainda recebiam a água poluída. Devido a isso, Snow praticamente eliminou os erros de

classificação quanto ao fornecimento da água ao domicílio e pôde evidenciar uma relação entre a impureza da água e a prevalência da cólera. Por fim, ele também possuía dados da taxa de mortalidade para os distritos pesquisados antes da mudança no fornecimento para um distrito específico. Isto permitiu corrigir as estimativas tomando como fatores de confundimento características dos domicílios que também poderiam levar a diferenças nas taxas de mortalidade por cólera. Neste sentido, o primeiro uso do método DD foi um importante marco nos estudos em saúde coletiva tanto do ponto de vista científico quanto prático.

Na Economia, o método DD começou a ser usado a partir do início do século XX. Obenauer & von der Nienburg (1915) analisaram o impacto da introdução de um salário mínimo no setor de varejo do estado de Oregon (EUA). A classe de indivíduos tratados consistia de um grupo de empregados da cidade de Portland (a maior do estado). Estes foram comparados com os empregados que trabalhavam no setor de varejo da cidade de Salem (a capital de Oregon), que possuía características econômicas similares à de Portland e portanto comparável a esta. Logo, os trabalhadores de Salem compunham a classe de controles. Os autores documentaram as mudanças na média de muitas variáveis respostas para ambas as classes antes e depois da introdução do salário mínimo em Portland.

Estas primeiras aplicações do método DD se destacam pelo fato de não exigirem um grande esforço computacional para obter as estimativas, uma vez que as classes de controles e tratados definidas são muito semelhantes entre si. Assim, covariáveis de confundimento não são necessárias e a inferência realizada é bem mais simples, se limitando a variações na resposta média para alterações no instante do tempo e/ou na atribuição do tratamento. Ao longo da 2ª metade do século XX, surgiram vários trabalhos na literatura econômica que usavam o método DD para avaliar o impacto de mudanças na legislação ou no ambiente de trabalho. Eles sempre definiam períodos pré-tratamento (antes da mudança) e grupos de comparação não afetados pelo tratamento para formar a classe de controles. Como exemplos de mudanças avaliadas dado um grupo alvo, além da imposição de um salário mínimo, temos a adoção de programas de treinamento (Ashenfelter, 1978; Heckman & Hotz, 1989); a imigração (Card, 1990); a implantação de medidas para trabalhadores temporariamente incapacitados de trabalhar (Meyer *et al.*, 1995) e claro, de políticas de transferência de renda. Para este último caso, uma revisão da literatura já foi feita no capítulo introdutório, para a situação em que é avaliado o impacto da política no tempo de alocação para o trabalho.

### 3.3.2 Suposições de Identificação no Método DD

Sabemos que a variável de atribuição do tratamento  $T$  é binária e tal que  $t \in \{0; 1\}$  e que nossa amostra foi coletada em dois pontos distintos do tempo. Sejam  $R$  a variável resposta de interesse e  $M$  tal que  $m \in \{0; 1\}$  o momento da coleta, onde  $m_i = 0$  indica

que a resposta  $r_i$  do indivíduo  $i$  foi coletada no período pré-tratamento e  $m_i = 1$ , que  $r_i$  foi coletada no período pós-tratamento. Assuma que a atribuição do tratamento ocorre entre estes dois períodos e que as observações amostrais são IID. Estamos interessados em avaliar o impacto médio em  $R$  quando  $T$  varia de 0 para 1. Para simplificar a notação, somente  $r_i$  será indexada pelo indivíduo correspondente. Logo, defina  $r_i^{m;t}$  a resposta que seria observada dados  $m$  e  $t$  para o indivíduo  $i$ , dita *resposta potencial*. A resposta de fato observada é denotada somente por  $r_i^m$  (não é indexada por  $t$ ). Seja  $X$  a matriz das possíveis variáveis de confundimento consideradas dentre as demais que foram pesquisadas. Assuma que  $X$  não depende de  $m$ , ou seja, seus valores não variam ao longo do tempo. Embora o método DD seja amplamente utilizado dentro da especificação de um modelo de regressão linear, iniciaremos o estudo de suas suposições de identificação através de uma modelagem não-paramétrica. Esta permite que o impacto do tratamento seja naturalmente heterogêneo para todos os indivíduos da população (Lechner, 2010).

A ideia principal da estratégia de identificação do método DD é calcular a diferença na média das respostas para os controles e tratados no instante  $m = 1$  (pós-tratamento) e subtrair deste primeiro resultado a mesma diferença no instante  $m = 0$ , condicionada aos valores de  $X$ . Para que tal estratégia tenha sucesso em identificar o impacto causal médio a ser avaliado, as seguintes suposições devem ser respeitadas (Lechner, 2010, p. 176–181):

**S1** Para cada indivíduo  $i$  na população, há uma única resposta potencial  $r_i^{m;t}$ , dados  $m$  e  $t$ , que pode ser observada para todos os indivíduos na população. Estabelecida por Rubin (1977), implica que as duas classes (de controles e tratados) estão completamente representadas e que não há interação entre os indivíduos na população. Matematicamente, podemos representar **S1** por

$$R_i^m = t_i R_i^{m;1} + (1 - t_i) R_i^{m;0} \quad \forall i \quad (3.19)$$

Se **S1** é violada, não é possível observar nenhuma das duas respostas potenciais, dada a classe na qual o indivíduo amostrado está contido. Isto pode ocorrer na avaliação do impacto de programas de treinamento para desempregados, por exemplo. Quando há muitos tratados na amostra, os controles (que não são treinados) terão mais dificuldades em encontrar um emprego no período pós-tratamento com relação a uma situação hipotética em que o programa não existe (Lechner, 2010);

**S2** A matriz  $X$  das possíveis variáveis de confundimento é exógena com respeito à atribuição do tratamento, ou seja,  $X_{T=1} = X_{T=0} = X$ ,  $\forall \mathbf{x}_i \in \chi$ , em que  $\chi$  denota o subespaço de  $X$  de interesse da avaliação do impacto. Logo,  $T$  não pode influenciar nos valores de  $X$ . Esta suposição é garantida se as variáveis em  $X$  assumem valores constantes no tempo para cada indivíduo  $i$ , pois a atribuição do tratamento varia com o tempo;

**S3** O tratamento não tem impacto nas observações da resposta no período pré-tratamento. Esta suposição é violada se os indivíduos que pertencerão à classe dos tratados sou-

berem que receberão o tratamento antes de sua atribuição. A condição exigida aqui é similar à **S2** (exogeneidade nas demais variáveis observadas), mas agora aplicada na variável resposta de interesse;

**S4** As diferenças nas respostas potenciais esperadas para todos os indivíduos no período pré-tratamento, ao longo do tempo e condicionadas a  $X$ , serão independentes da classe a que o indivíduo pertencerá no período pós-tratamento. Também conhecida como Pressuposto da Tendência Comum (PTC), esta suposição é a chave do método DD para estabelecer a inferência causal desejada. Em notação matemática, podemos representar **S4** por

$$\begin{aligned} \mathbb{E}\left(R_i^{1;0} \mid \mathbf{x}_i, T_i=1\right) - \mathbb{E}\left(R_i^{0;0} \mid \mathbf{x}_i, T_i=1\right) &= \mathbb{E}\left(R_i^{1;0} \mid \mathbf{x}_i, T_i=0\right) - \mathbb{E}\left(R_i^{0;0} \mid \mathbf{x}_i, T_i=0\right) \\ &= \mathbb{E}\left(R_i^{1;0} \mid \mathbf{x}_i\right) - \mathbb{E}\left(R_i^{0;0} \mid \mathbf{x}_i\right) \end{aligned} \quad (3.20)$$

Na situação hipotética em que os tratados não recebem o tratamento, as classes definidas por  $T = 0$  e  $T = 1$  experienciarão a mesma tendência cronológica na variável resposta condicionada a  $X$ . Logo, a matriz  $X$  deve incluir qualquer variável pesquisada que resulte em diferenças na tendência cronológica para  $R$  nas duas classes nesta situação, para diferentes valores de  $\mathbf{x}_i$  e supondo que a distribuição de  $\mathbf{x}_i$  difere nas duas classes. Por esta razão dizemos que a variável incluída é de confundimento. Disto decorre que qualquer desvio da tendência cronológica para  $R$  nos tratados em relação à tendência nos controles será consequência direta do efeito ou impacto do tratamento e não de quaisquer outras características dos indivíduos nas duas classes;

**S5** O viés na resposta observada  $r_i$  é constante ao longo do tempo, independentemente da classe a que o indivíduo  $i$  pertencerá no período pós-tratamento. Esta suposição é equivalente à **S4**, a partir de uma leitura alternativa da intuição por trás do método DD. Aqui, desejamos estimar o impacto do tratamento nos dois períodos considerados, mas assumimos (erroneamente) exogeneidade em  $X$  (com respeito a  $T$ ) sem verificá-la. Se **S3** é válida, então haverá um impacto adicional além daquele diretamente associado ao tratamento na resposta observada  $r_i^0$  (no período pré-tratamento, ele seria nulo se a condicionalidade em  $X$  estivesse garantida). Isto implica que a estimativa do impacto é enviesada e inconsistente. Se, e somente se, tal viés for constante ao longo do tempo, é possível corrigir a estimativa do impacto em  $r_i^1$ , a resposta observada no período pós-tratamento, no qual estamos interessados. Formalmente, temos que

$$\begin{aligned} B_0(\mathbf{x}_i) &= \mathbb{E}\left(R_i^{0;0} \mid \mathbf{x}_i, T_i=1\right) - \mathbb{E}\left(R_i^{0;0} \mid \mathbf{x}_i, T_i=0\right) \\ &= \mathbb{E}\left(R_i^{1;0} \mid \mathbf{x}_i, T_i=1\right) - \mathbb{E}\left(R_i^{1;0} \mid \mathbf{x}_i, T_i=0\right) = B_1(\mathbf{x}_i) \end{aligned} \quad (3.21)$$

em que  $B_m(\mathbf{x}_i)$  denota o viés na resposta no momento  $m$  dado  $\mathbf{x}_i$ .

A equivalência entre **S4** e **S5** pode ser provada matematicamente ao tomar a diferença entre os vieses, pois

$$\begin{aligned} B_1(\mathbf{x}_i) - B_0(\mathbf{x}_i) &= \\ &= \left[ \mathbb{E} \left( R_i^{1;0} \mid \mathbf{x}_i, T_i = 1 \right) - \mathbb{E} \left( R_i^{1;0} \mid \mathbf{x}_i, T_i = 0 \right) \right] - \left[ \mathbb{E} \left( R_i^{0;0} \mid \mathbf{x}_i, T_i = 1 \right) - \mathbb{E} \left( R_i^{0;0} \mid \mathbf{x}_i, T_i = 0 \right) \right] \\ &= \left[ \mathbb{E} \left( R_i^{1;0} \mid \mathbf{x}_i, T_i = 1 \right) - \mathbb{E} \left( R_i^{0;0} \mid \mathbf{x}_i, T_i = 1 \right) \right] - \left[ \mathbb{E} \left( R_i^{1;0} \mid \mathbf{x}_i, T_i = 0 \right) - \mathbb{E} \left( R_i^{0;0} \mid \mathbf{x}_i, T_i = 0 \right) \right] \end{aligned}$$

Destas suposições, fica claro que a identificação ocorre se a diferença contrafactual  $\mathbb{E} \left( R_i^{1;0} \mid \mathbf{x}_i, T_i = 1 \right) - \mathbb{E} \left( R_i^{0;0} \mid \mathbf{x}_i, T_i = 1 \right)$  for igual à diferença observável  $\mathbb{E} \left( R_i^{1;0} \mid \mathbf{x}_i, T_i = 0 \right) - \mathbb{E} \left( R_i^{0;0} \mid \mathbf{x}_i, T_i = 0 \right)$ . Logo, é necessário que cada valor nas variáveis em  $X$  seja observado para todos os indivíduos nos quatro grupos comparados. Isto é garantido pelo pressuposto (adicional) de suporte comum para todos os grupos. No contexto da metodologia DD, podemos formalizá-lo como (Lechner, 2010, p. 181)

$$\begin{aligned} \Pr [M_i \cdot T_i = 1 \mid \mathbf{x}_i, (M_i, T_i) \in \{(m, t), (1, 1)\}] < 1; \quad \forall (m, t) \in \{(0, 1), (0, 0), (1, 0)\}; \\ \forall \mathbf{x}_i \in \chi \end{aligned} \quad (3.22)$$

Assim como a suposição **S2**, o pressuposto do suporte comum não depende da resposta de interesse e pode ser avaliado para cada variável pesquisada. Em virtude da grande quantidade de dados, tanto no grupo das crianças quanto no dos adultos e para ambas as classes, vamos assumir a validade deste pressuposto em todas as variáveis pesquisadas na YNSPMS.

Diferentemente de outras estratégias não-paramétricas de identificação, como o próprio pareamento e o uso de variáveis instrumentais, a validade das suposições de identificação do método DD são influenciadas pela escala das observações na resposta quando esta segue uma distribuição que depende de  $M$  e  $T$ . Isto quer dizer que, se **S1**–**S5** são válidas para a escala original de  $R$ , elas podem não o ser para transformações monotônicas de  $R$ , mesmo sem supor nenhum modelo paramétrico com covariáveis de confundimento para o método DD. Lechner (2010, p. 184–186) apresenta um exemplo de violação em **S4** para respostas medidas na escala logarítmica.

Como veremos mais adiante que os parâmetros do modelo de regressão usado no método DD estarão na mesma escala da resposta, não é necessário adaptar o método de modo que a resposta não seja funcionalmente dependente. Apesar disso, sugestões neste sentido já foram propostas, como o método de Mudança em Mudanças (MM) proposto por Athey & Imbens (2006). Ainda, vamos combinar o pareamento do escore de propensão (usando o PDE) com o método DD. Assumimos que não apenas a média, mas também a distribuição das respostas potenciais condicionadas a  $\mathbf{x}_i$ , não dependem da classe de tratamento a que o indivíduo  $i$  pertence. Logo, é um pressuposto mais forte do que o PTC (suposição **S4**), com duas vantagens adicionais: identificar a distribuição contrafactual

completa (não somente as esperanças contrafatuais) e tornar a identificação invariante para mudanças na escala de  $R$  (Lechner, 2010).

### 3.3.3 Modelagem no Método DD

A estimação pelo método DD se baseia na comparação entre controles e tratados nos períodos pré e pós-tratamento. Isto é feito através do cálculo da diferença entre as médias das respostas observadas para os tratados e controles nestes dois períodos. Assume-se que a heterogeneidade em variáveis que não foram pesquisadas é invariante com o tempo e não correlacionada com a atribuição do tratamento. Este pressuposto é mais flexível do que a exogeneidade condicional exigida em **S2** (Khandker *et al.*, 2010).

A maioria das aplicações envolvendo o uso do método DD usam como especificação um modelo de regressão linear com respostas (pelo menos aproximadamente) contínuas. A equação para estimação é especificada como (Khandker *et al.*, 2010, p. 72)

$$R_i^m = \beta_0 + \lambda_1 t_i + \lambda_2 m_i + \lambda_3 t_i m_i + \epsilon_i^m \quad (3.23)$$

em que o coeficiente  $\lambda_3$ , associado à interação entre as variáveis de atribuição do tratamento  $T$  e do momento  $M$  em que a resposta foi observada, fornece o impacto médio da intervenção (aqui, do programa SWF) para um aumento em uma unidade na resposta de interesse, pois (Khandker *et al.*, 2010, p. 73)

$$\begin{aligned} \mathbb{E}(R_i^{1:1} | \mathbf{x}_i, T_i=1) - \mathbb{E}(R_i^{0:1} | \mathbf{x}_i, T_i=1) &= (\beta_0 + \lambda_1 + \lambda_2 + \lambda_3) - (\beta_0 + \lambda_1) \\ \mathbb{E}(R_i^{1:0} | \mathbf{x}_i, T_i=0) - \mathbb{E}(R_i^{0:0} | \mathbf{x}_i, T_i=0) &= (\beta_0 + \lambda_2) - (\beta_0) \end{aligned}$$

e, tomando a diferença destas duas expressões, isolamos o coeficiente  $\lambda_3$ , resultado de uma diferença em diferenças.

Em adição ao termo de interação, são incluídos em (3.23) coeficientes associados às variáveis  $T$  e  $M$  separadamente para detectar os respectivos efeitos médios. Desde que tenhamos dados disponíveis de quatro grupos para comparação, não é necessário que a pesquisa seja longitudinal para implementar o método DD. Basta que a variável  $M$  possa ser reinterpretada como um indicador geográfico, por exemplo. Além disso, é possível generalizar o termo  $\beta_0$  como  $X\beta$  para incluir variáveis de confundimento. Além das suposições de identificação enunciadas na subseção anterior, ao adotar uma especificação linear, também é necessário garantir que (Khandker *et al.*, 2010, p. 73)

- S6** O modelo descrito em (3.23), com ou sem variáveis de confundimento, está corretamente especificado;
- S7** O termo de erro  $\epsilon$  não tem nenhuma correlação com as demais variáveis do modelo ( $T$ ;  $M$ ;  $M \cdot T$  e as presentes na matriz  $X$ ).

Neste trabalho, assumimos que a resposta  $R$  é aproximadamente normal (ligando-a com o preditor linear através da função identidade), embora qualquer outra distribuição

contínua na reta real pudesse ser utilizada. Logo, na situação em que as observações são IID, um teste  $t$  é suficiente para concluir sobre a significância dos coeficientes em (3.23) dado um nível nominal  $\alpha$ , que podem ser estimados por Mínimos Quadrados Ordinários (MQO) ou por MV. Quando os dados são obtidos através de um PAC, é possível recorrer à estimação por Mínimos Quadrados Ponderados (MQP), usando os pesos amostrais na matriz de covariância estimada, ou por MP apresentada na subseção 2.3.3. Como estamos adotando a abordagem inferencial baseada no plano amostral, estimaremos os coeficientes por MV quando o PAC é ignorado e por MP quando queremos considerá-lo. Para o caso específico do modelo de regressão linear normal com dados amostrais complexos, Pessoa & Silva (1998, p. 87–92) detalham a estimação dos parâmetros e das respectivas estimativas da variância por MP. Neste caso, a construção dos testes  $t$  e dos intervalos de confiança é análoga ao caso de observações IID, com a diferença de que o método utilizado para estimar a variância também determinará os graus de liberdade. Para a LST, eles serão dados pelo  $n^\circ$  de UPA's menos o  $n^\circ$  de estratos no 1º estágio, semelhante ao que foi feito para o teste  $\chi^2$  de Pearson com correção de 2ª ordem de Rao-Scott  $F$ -ajustado (Lohr, 2010).

Como vantagens motivadoras para seu uso neste trabalho, o método DD permite obter facilmente as estimativas pontuais e do erro-padrão dos coeficientes do modelo. Ainda, pode ser estendido para situações mais gerais, que cobrem vários períodos distintos de tempo; mais de dois tratamentos ou muitas covariáveis usadas como confundimento. Qualquer que seja a extensão, o custo computacional para realizar a inferência necessária é baixo comparado aos resultados proporcionados e conclusões extraídas (Lechner, 2010). Por outro lado, especificar um modelo linear paramétrico para estimação pelo método DD também traz alguns riscos. Primeiramente, o PTC (suposição **S4**) será mais restrito, pois não estará condicionado à matriz  $X$ , e sim ao termo  $X\beta$ . Logo, qualquer desvio da reta estimada será captado pelo termo de erro e invalidará as estimativas (pontual e intervalar) obtidas se for grande. A variável resposta a ser de interesse no método DD não pode ter suporte limitado, pois não há garantia de que a média predita das respostas potenciais estará dentro deste suporte. O MPL, descrito na seção 2.3, é um exemplo deste problema.

Além disso, se for especificado um modelo não-linear projetado para um suporte limitado em particular e com sua própria função de ligação, o PTC será violado se o objetivo do estudo for inferir também sobre  $\lambda_1$ , o impacto isolado do tratamento. Por esta razão, a resposta de interesse deve ser uma variável (pelo menos aproximadamente) contínua. Tomemos apenas os indivíduos que foram classificados como controles ou tratados com base na variável RSWF e que efetivamente participam da força de trabalho, cuja informação na YNSPMS é extraída da variável `p6a_q_2`, indicadora desta participação nos últimos 30 dias (Banco de Dados da YNSPMS, 2014c), com pesos amostrais longitudinais da 1ª à 4ª rodada<sup>16</sup>. Para estes indivíduos, a variável escolhida como resposta, TSAT,

<sup>16</sup> O levantamento dos indivíduos que se enquadram nesta situação está descrito na subseção B.1.2 do

possui distribuição aproximadamente contínua em ambos os grupos etários, ainda que originalmente discreta. Isto é comprovado pelos histogramas das densidades para o caso em que ignoramos o PAC (figuras 2 a 5 do anexo A), e para o caso em que o PAC é considerado (figuras 6 a 9 do anexo A).

Ainda, o percentual de não-resposta é muito baixo, tanto para crianças (0,33% na 1ª e 0% na 4ª rodada) quanto para adultos (1,68% na 1ª e 1,26% na 4ª rodada). Para os histogramas das figuras 6 a 9, as densidades em cada intervalo foram calculadas a partir do total populacional estimado, considerando os pesos amostrais longitudinais da 1ª à 4ª rodada, no mesmo intervalo para a variável TSAT (Lumley, 2016)<sup>17</sup>. Até o presente momento, o R, assim como a maioria dos *software* estatísticos de grande circulação, utiliza apenas modelos da família gaussiana (normal) quando assumimos que a variável resposta é contínua para modelagens que incorporam o PAC<sup>18</sup>. Por esta razão, escolhemos um modelo de regressão linear normal, embora os histogramas indiquem que TSAT é assimétrica, em especial no grupo das crianças.

Embora variáveis com suporte limitado não possam ser usadas no método DD por não respeitarem todas as suposições de identificação definidas na seção 3.3.2, é possível modificar estas mesmas suposições de modo a viabilizar uma análise semelhante. Isto é feito supondo que a resposta é uma função não-linear monótona; estritamente crescente e invertível de uma variável dependente (resposta) latente (não-mensurável diretamente) e assumindo tendência comum na função não-linear dada das respostas esperadas condicionais. Claramente, as estimativas obtidas serão distintas de uma análise (errônea) com as suposições originais do método DD, pois este é funcionalmente dependente (Lechner, 2010).

### 3.3.4 Combinando o Método DD ao Pareamento do Escore

Tanto o método DD quanto o pareamento do escore de propensão, independentemente da maneira escolhida para parear, estão sujeitos a obterem estimativas enviesadas quando aplicados isoladamente aos dados. Assim, uma forma de aumentar ainda mais a robustez das estimativas é combinar ambas as metodologias. Nesta situação, primeiramente os indivíduos são pareados com base em  $q$  covariáveis dentre as observadas na pesquisa. Em seguida, realizamos testes de balanceamento para saber se cada uma das covariáveis do modelo para o escore de propensão é ou não balanceada e se conjuntamente (usando os valores dos escores preditos) as mesmas covariáveis são balanceadas. Como comentado anteriormente, isto é feito de modo a encontrar os controles contrafatuais dos tratados.

---

apêndice.

<sup>17</sup> Por causa disso, foi necessário retirar um indivíduo do grupo das crianças nesta análise, pois este era o único de seu estrato no 1º estágio, inviabilizando a estimação não só dos totais mas também de qualquer outro parâmetro quando consideramos o PAC.

<sup>18</sup> A função `svyglm`, responsável por esta análise no pacote `survey`, é diretamente dependente da função nativa `glm` (Lumley, 2016).

Finalmente, o método DD é aplicado somente nos indivíduos pareados e cujo balanceamento foi comprovado para um conjunto de covariáveis. Como estas estão balanceadas em princípio apenas para as observações no período pré-tratamento, é necessário reinserir as covariáveis balanceadas como de confundimento no modelo de regressão linear do método DD. A influência de cada covariável na resposta de interesse pode mudar com o tempo ou com a atribuição do tratamento sob estudo. Ainda, o simples pareamento pelo score de propensão não assume que características não pesquisadas na amostra, mesmo que supostas constantes no intervalo de tempo da pesquisa, possam explicar o impacto do tratamento nas respostas, o que é considerado pelo método DD. Por outro lado, este não permite verificar suas suposições de identificação, algumas das quais similares às necessárias para a construção do score de propensão, e garantir que todos os (ou uma parte dos) controles existentes no banco são os melhores contrafatuais possíveis para os tratados. Assim, uma metodologia complementa a outra (Gertler *et al.*, 2011).

Para dados longitudinais como os da pesquisa YMSPMS, o uso da combinação é direto, uma vez que apenas um pareamento (entre controles e tratados) é necessário. Para dados transversais, entretanto, o processo de pareamento é mais elaborado. Primeiramente, se deve encontrar para os tratados no período pós-tratamento os seus similares (que pertencem à mesma classe) no período pré-tratamento. Em seguida, o mesmo é feito para a classe de controles. Por fim, são pareados os dois conjuntos obtidos correspondentes a cada classe. Dependendo do método utilizado para o pareamento, a avaliação do impacto pelo método DD produzirá estimativas distintas quando comparadas à situação em que os dados são longitudinais (Blundell & Dias, 2000). Assim, o uso de uma pesquisa longitudinal facilita ainda mais a análise.

## 4 Resultados Empíricos

Grande parte dos estudos revistos na seção 1.3 adota a suposição de observações IID para os indivíduos pesquisados. Mesmo quando um PAC é empregado antes da seleção amostral, não é comum o uso das informações deste no cálculo de estatísticas descritivas ou de estimativas pontuais e intervalares, as quais serão enviesadas. Logo, é incorreto ignorar a existência de um PAC como mecanismo gerador dos dados amostrais quando o objetivo é extrair conclusões a partir da amostra selecionada para a população sob estudo. Além disso, muitos estudos tomam variáveis categóricas (em geral binárias) como respostas de interesse. Isto limita as conclusões com relação à metodologia a ser utilizada para avaliar o impacto (vimos no capítulo anterior que o método DD exige a suposição de variáveis contínuas). Para a avaliação longitudinal do impacto de uma política pública de transferência direta de renda (o SWF) no tempo (semanal) de alocação para o trabalho, tomaremos diretamente o nº de horas trabalhadas nos últimos sete dias, representado pela variável TSAT. Ainda, compararemos a análise supondo observações IID com a aquela em que se incorpora o PAC da YNSPMS em todas as etapas, desde a modelagem do escore de propensão até a estimação do impacto. Desta forma, teremos um referencial para estudos semelhantes em que a informação do PAC está disponível.

### 4.1 Modelos para os Escores sem e com PAC na YNSPMS

Vimos na seção 2.3 que qualquer função de distribuição pode ser utilizada para modelar uma variável binária, como RSWF, a partir de um preditor linear. Pela facilidade na interpretação, se optou pelo modelo *logit*, ou seja, pela regressão com distribuição logística para RSWF. Além da escolha de um modelo para obter os escores de propensão, ou preditos, para cada indivíduo na amostra, também é necessário selecionar covariáveis de modo que a especificação respeite o PATFI. Logo, as covariáveis devem ser mensuradas antes do momento da atribuição do tratamento e devem influenciar tanto esta variável quanto a resposta de interesse, TSAT. Por esta razão, somente variáveis relacionadas diretamente aos critérios de elegibilidade do YNSPMS ou que em tese fossem fortemente correlacionadas com tais critérios foram consideradas.

Como faremos modelos para dois grupos ou subpopulações distintos baseados na idade dos respondentes, a escolha de covariáveis candidatas aos respectivos modelos será distinta. Tanto para as crianças quanto para os adultos, foram construídas novas codificações para variáveis originalmente pesquisadas na YNSPMS. Como nosso objetivo primário é verificar a diferença na análise quando o PAC é incorporado na inferência realizada, para um mesmo grupo etário se partiu das mesmas candidatas em duas situações:

quando as observações são supostas IID, e quando as características do PAC não são ignoradas. Relembremos que nesta última todas as variâncias estimadas serão aproximadas através do método da LST. Ao todo, teremos quatro modelos para o escore de propensão distintos. Por fim, cada um destes modelos tomará somente os indivíduos classificados como tratados ou controles segundo os critérios enunciados na seção 2.1 e que possuem respostas válidas para a variável TSAT na 1<sup>a</sup> e 4<sup>a</sup> rodadas da YNSPMS, de modo que a avaliação do impacto seja possível. Observe que esta análise se aplica apenas para quem estava participando da força de trabalho especificamente nestas duas rodadas, ou seja, para quem respondeu afirmativamente à variável `p6a_q_2`, indicadora desta participação nos últimos 30 dias (Banco de Dados da YNSPMS, 2014c).

Para o grupo das crianças, foram tomadas como candidatas à especificação do modelo de regressão logística para o escore de propensão as seguintes variáveis:

- COM: Criança é Órfã de Mãe;
- COP: Criança é Órfã de Pai;
- CRDAI: Criança Reside em Domicílio com Adulto Idoso;
- CRDAPI: Criança Reside em Domicílio com Adulto Portador de Invalidez;
- CRDCM: Criança Reside em Domicílio Chefiado por Mulher, desde que esta seja adulta e solteira, divorciada ou viúva;
- DRC: Domicílio Recebe Caridade
- DRFP: Domicílio Recebe Fundos de Pensão;
- DRR: Domicílio Recebe Remessas (de parentes ou amigos);
- NPD: Nível de Pobreza do Domicílio.

Todas as variáveis acima influenciam em tese tanto a atribuição do tratamento (RSWF) quanto a resposta TSAT. Ainda, tomaremos seus valores apenas na 1<sup>a</sup> rodada da YNSPMS, de forma que o PATFI seja respeitado. As duas primeiras variáveis, COM e COP, são o exato complementar de `p2_q_07` e `p2_q_09`, os nomes das variáveis indicadoras da mãe e do pai do indivíduo (criança) estarem vivos, respectivamente, na YNSPMS. A presença de crianças ou adolescentes menores de 18 anos órfãos de mãe ou pai no domicílio é um dos critérios de elegibilidade do SWF, logo estas duas variáveis podem explicar a variabilidade do escore a ser predito. Em um país que passou por uma série crise política no período da pesquisa, há uma tendência de aumento na quantidade de crianças órfãs, especialmente do pai, sendo uma razão prática para a inclusão destas variáveis no modelo para o escore de propensão. Para estas e quaisquer variáveis construídas daqui em diante, os comandos utilizados podem ser encontrados na subseção B.2.1 do apêndice.

As variáveis CRDAI; CRDAPI e CRDCM foram construídas usando como auxiliares as informações das respostas para os adultos nas variáveis `p2_q_03`; `p2_q_04`; `p2_q_05`; `p2_q_11` e `p2_q_16`. Domicílios chefiados por mulheres solteiras; divorciadas (que não se casaram novamente) ou viúvas, desde que maiores de 18 anos ou menores, mas com

pelo menos um filho também se enquadram entre os elegíveis para receber o benefício do SWF. O mesmo também vale para domicílios em que residem pessoas idosas (homens acima dos 60 anos e mulheres acima dos 55 anos) ou adultos com algum tipo de invalidez que limita as atividades do dia-a-dia. Restringimos a primeira variável apenas para as mulheres adultas e sem levar em consideração se aquelas que são divorciadas ou viúvas contraíram novo matrimônio, pois casos de domicílios chefiados por mulheres abaixo de 18 anos, divorciadas ou viúvas, são bem raros ou inexistentes. Para identificar os domicílios cujas crianças residiam conjuntamente aos adultos elegíveis para o SWF, foi tomada a variável `f1`, indicadora do código domiciliar.

Por sua vez, as variáveis `DRC`; `DRFP`; `DRR` e `NPD` correspondem diretamente aos nomes `remit`; `charit`; `pension` e `poor`, variáveis anteriormente presentes no banco de dados da YNSPMS. Embora a renda dos indivíduos residentes nos domicílios amostrados tenha sido pesquisada, as variáveis que a medem diretamente, representadas pelos nomes `p6_a_12` e `p6_a_13` (pagamento do trabalho em dinheiro e em espécie, respectivamente) possuem um alto índice de respostas faltantes, mesmo no grupo dos adultos. De fato, geralmente a pessoa entrevistada não se sente confortável em revelar o quanto ganha para seu sustento. Por esta razão, as substituímos pelas quatro variáveis mencionadas no início deste parágrafo, as quais estão fortemente correlacionadas com a renda domiciliar, também um critério de elegibilidade para o SWF. Todas elas são referentes ao domicílio em que o indivíduo reside, logo podem ser usadas na modelagem do score tanto para o grupo das crianças quanto para o dos adultos, como veremos a seguir.

Para o grupo dos adultos, temos as seguintes candidatas para a especificação do modelo de regressão logística para o score de propensão

- `AI`: Adulto é Idoso;
- `AMCD`: Adulto é Mulher e Chefe do Domicílio, desde que seja solteira, divorciada ou viúva;
- `API`: Adulto é Portador de Invalidez;
- `DRC` (`charit`): Domicílio Recebe Caridade
- `DRFP` (`pension`): Domicílio Recebe Fundos de Pensão;
- `DRR` (`remit`): Domicílio Recebe Remessas (de parentes ou amigos);
- `NPD` (`poor`): Nível de Pobreza do Domicílio.

As variáveis `AI`; `AMCD` e `API` foram construídas a partir das variáveis `p2_q_03`; `p2_q_04`; `p2_q_05`; `p2_q_11` e `p2_q_16`, sendo posteriormente usadas como auxiliares para construirmos as variáveis `CRDCM`; `CRDAI` e `CRDAPI` para a modelagem no grupo das crianças. Também usamos no modelo para o score de propensão no grupo dos adultos as variáveis `DRR`; `DRC`; `DRFP` e `NPD`, por se referirem ao domicílio e não ao indivíduo em si. Assim como no grupo das crianças, todas as variáveis escolhidas influenciam tanto a atribuição do tratamento (`RSWF`) quanto a resposta `TSAT` e consideraremos apenas os

valores na 1ª rodada da YNSPMS, de forma que o PATFI seja respeitado.

Todas as variáveis candidatas elencadas acima, para ambos os grupos etários, são categóricas binárias (com categorias para resposta negativa ou positiva), exceto pela variável NPD, que possui quatro categorias definidas com base na renda *per capita* domiciliar, a saber: “pobreza extrema”; “pobreza moderada”; “vulnerável” e “não-pobre”. Logo, NPD também se distingue das demais por ser uma variável categórica ordinal (YNSPMS Final Report, 2014a).

Dadas as variáveis candidatas à especificação do modelo para o escore de propensão em cada grupo etário, é necessário primeiramente ajustar todas elas (modelo completo) e verificar a significância de cada covariável (ao nível nominal de 5%) no modelo pelo teste da RV. Retiramos então todas aquelas que não forem significativas e conseqüentemente explicam muito pouco ou quase nada da variabilidade dos escores preditos pelo modelo inicial e fazemos um novo ajuste apenas com as variáveis significativas remanescentes como covariáveis. A significância das variáveis no modelo reduzido pode ou não mudar, e se alguma covariável passar a ser não significativa, uma nova triagem e um terceiro modelo é ajustado sem especificá-la. Este processo é repetido até que encontremos um modelo em que todas as covariáveis são significativas. Entretanto, as variáveis retiradas inicialmente podem voltar a serem especificadas no modelo desde que, quando adicionadas isoladamente ao conjunto de variáveis significativas, mudem seu comportamento anterior e passem a ser significativas pelo teste da RV. Este processo de seleção das covariáveis é similar a uma combinação dos métodos *backward* e *stepwise* (Draper & Smith, 1998), com a diferença de que estamos retirando todas as variáveis não significativas de uma única vez, as quais podem ou não retornar à especificação do modelo ao serem adicionadas uma a uma, usando ainda os resultados do teste da RV (RP) e sem aplicar qualquer critério de informação. Desta forma, atendemos à literatura para seleção de covariáveis em modelos para o escore de propensão, exposta na seção 2.4.

Assim, é possível que surjam modelos concorrentes com a mesma quantidade de covariáveis especificadas. Quando for este o caso, adotaremos aquele que prevê melhor a classificação do indivíduo como controle ou tratado, conforme descrito na seção 2.4. Como os resultados do teste da RV dependem do ordenamento das covariáveis especificadas, para os quatro modelos completos iniciais listaremos as covariáveis em ordem crescente de variabilidade interna. Desta forma, serão especificadas por último aquelas que possuem maior variabilidade interna, pois tendem a possuir os maiores desvios e portanto maior significância dado um nível nominal (os valores-p são menores do que tal nível) em testes como o da RV (RP)<sup>1</sup>. Observe que todas as variáveis candidatas são categóricas, logo

<sup>1</sup> A função `anova()`, nativa do R, calcula o desvio primeiramente para a última covariável do modelo especificado. É de nosso interesse manter as variáveis que forneçam mais informação sobre a probabilidade de ser beneficiário do SWF. Aquelas que têm maior variabilidade interna tendem a fazê-lo e por isso são especificadas por último.

aquelas que possuem mais categorias (como NPD) serão especificadas por último. O código em R para o ajuste e seleção de modelos para o escore de propensão considerados nesta dissertação, com base nos testes da RV e RP, pode ser encontrado nas subseções B.2.1 e B.2.2 do apêndice, para os casos em que o PAC é ignorado e em que ele é incorporado, respectivamente. Para os quatro modelos considerados, os resultados dos testes serão tabelados quando todas as variáveis candidatas são especificadas como covariáveis (completo) e quando uma especificação final a ser usada para o pareamento dos escores preditos for escolhida.

Começamos pelo grupo das crianças e para o caso em que ignoramos completamente o PAC, tratando as observações amostrais como IID. Apresentamos na tabela 1 os resultados da análise de desvios pelo teste da RV para o modelo completo, ou seja, com todas as respectivas variáveis candidatas especificadas como covariáveis (denotado por `m1.c` na subseção B.2.1 do apêndice). Para esta e as demais tabelas produzidas, também são apresentados os graus de liberdade para os testes em cada variável (abreviadamente, GL). Sempre que uma covariável for significativa ao nível nominal de 5%, seu valor-p será acompanhado do símbolo \* (asterisco).

Tabela 1 – Resultados do teste da RV no modelo completo para o escore de propensão no grupo das crianças, caso IID

	GL	Desvio	Desvio Residual	Valor-p ( $\chi^2$ , em %)
Modelo nulo	–	–	599.66	–
CRDCM	1	0.42	599.24	51.84
COM	1	2.36	596.88	12.48
DRFP	1	1.81	595.07	17.87
COP	1	0.03	595.04	86.65
DRC	1	15.66	579.38	<0.01*
CRDAI	1	1.31	578.07	25.27
CRDAPI	1	1.06	577.01	30.36
DRR	1	0.10	576.91	75.60
NPD	3	7.87	569.04	4.88*

Pelos resultados da tabela 1, temos ao especificar como covariáveis todas as candidatas que somente DRC e NPD são significativas, com desvios bem maiores do que os das demais covariáveis. Logo, um novo ajuste é feito sem as sete covariáveis que não foram significativas no modelo completo. Contudo, ao especificar DRC e NPD (nesta ordem), apenas a primeira continua significativa ao nível nominal de 5% (este novo modelo foi denotado por `m2.c`). Ou seja, NPD só é significativa quando incluímos mais covariáveis não significativas no modelo, tornando-o menos parcimonioso.

Para o modelo em que apenas DRC é especificada (denotado por `m3.c`), ela permanece como covariável significativa e portanto estará presente no modelo inicial para o escore de propensão no grupo das crianças. Denominamos “modelo inicial” aquele que será

usado para pareamento dos indivíduos tratados, o qual pode ser alterado dependendo dos resultados nos testes para balanceamento. Ainda, quando as sete variáveis que não foram significativas na especificação do modelo completo voltam a ser especificadas, uma de cada vez, conjuntamente a DRC, elas continuam não sendo significativas (modelos m4.c.1 a m4.c.7). Portanto, ficamos com o modelo em que apenas DRC, junto ao intercepto, é especificada. A tabela 2 apresenta os resultados da análise de desvios pelo teste da RV para este modelo.

Tabela 2 – Resultados do teste da RV no modelo apenas com DRC (além do intercepto) para o escore de propensão no grupo das crianças, caso IID

	GL	Desvio	Desvio Residual	Valor-p ( $\chi^2$ , em %)
Modelo nulo	–	–	599.66	–
DRC	1	14.81	584.85	0.01*

Como este foi o modelo escolhido, também apresentamos as correspondentes estimativas dos coeficientes (pontual e do erro-padrão associado), bem como os resultados do teste  $z^2$ , na tabela 3. Para este e os modelos a serem escolhidos nos demais casos, a interpretação da estimativa pontual de cada coeficiente é feita em termos da probabilidade de que um indivíduo respondente da categoria associada, quando esta não for de referência no modelo. Esta probabilidade é dada pela função de distribuição logística acumulada ( $F(t) = 1/(1 + e^{-t}) = e^t/(e^t + 1)$ ), em que  $t$  é dado pela combinação linear de interesse do preditor  $X\beta$ , de acordo com a(s) categoria(s) sob estudo. Para facilitar a notação e a interpretação, de agora em diante a denominaremos por  $\text{expit}(t)$ , por ser a função inversa da tradicional função  $\text{logit}(t)$ .

Tabela 3 – Estimativas para os coeficientes no modelo apenas com DRC (além do intercepto) para o escore de propensão no grupo das crianças, caso IID

Termo da Regr.	Est. Pontual	Erro-Pad.	Estat. $z$	Valor-p (%)
Intercepto	-1.5497	0.1101	-14.08	<0.01*
DRC = “sim”	1.3490	0.3364	4.01	<0.01*

Como o modelo especifica apenas uma covariável binária (DRC), poderíamos concluir sobre a significância da covariável diretamente pelas estimativas dos coeficientes da regressão. Observe que a categoria de referência na variável DRC é a resposta negativa para o recebimento de renda através de caridade pelo domicílio em que a criança reside. Para qualquer covariável especificada e independentemente do modelo, a categoria de referência adotada sempre é a que menos se enquadra nos critérios de elegibilidade para

<sup>2</sup> O modelo em questão é MLG da família binomial, em que o parâmetro de dispersão  $\phi$  não é estimado. Assim, a função responsável pelo ajuste no R, `glm()` (nativa, do pacote `stats`), fixa este parâmetro como igual a 1, permitindo o uso de um teste  $z$  (R Core Team, 2016b).

o SWF. Segundo a tabela 3, as crianças que residem em domicílios que recebem renda através de caridade têm uma probabilidade igual a  $\text{expit}(-1.5497 + 1.3490) \approx 0.4500$  ou 45,00% de serem beneficiárias do SWF. Esta probabilidade é aproximadamente 2.57 vezes maior do que a relativa às crianças que residem em domicílios que não recebem renda por esta modalidade:  $\text{expit}(-1.5497) \approx 0.1751$  ou 17,51%. Este resultado está de acordo com a literatura, pois domicílios que dependem desta modalidade de renda tendem a ser muito pobres e conseqüentemente elegíveis para políticas públicas de transferência direta de renda.

Como o modelo possui uma única covariável, a interpretação dos coeficientes não depende das demais variáveis observadas na 1ª rodada da YNSPMS. Com relação aos valores dos escores preditos, todos eles estão estritamente entre 0 e 1, respeitando também a hipótese de sobreposição do PATFI (figura 10 do anexo B). Como critério, é exigido que os escores sejam maiores do que 0.01 e menores do que 0.99, o que ocorre (todos são maiores do que 0.1 e menores do que 0.5). Logo, poderemos usá-los para o pareamento dos indivíduos tratados. Embora estejamos no caso IID e com uma amostra relativamente grande (611 indivíduos no grupo das crianças), não foram encontradas mais covariáveis significativas a serem especificadas no modelo inicial para o escore de propensão no grupo das crianças. Assim, tal modelo é pouco informativo, pois há apenas dois valores distintos para os escores preditos. Por outro lado, a inclusão de um número reduzido de covariáveis facilita a não-rejeição nos testes de balanceamento após fazermos o pareamento e permite incluir mais covariáveis de confundimento (incluindo as que não foram significativas na modelagem dos escores) na avaliação de impacto para TSAT, como veremos nas seções 4.2 e 4.3.

Passando agora para o grupo dos adultos e ainda tratando as observações amostrais como IID, temos na tabela 4 os resultados da análise de desvios pelo teste da RV para o modelo completo correspondente (denotado por **m1.a** na subseção B.2.1 do apêndice). Neste modelo, as covariáveis API; DRFP; DRC e NPD são significativas ao nível nominal de 5%. Para NPD, o desvio calculado é bastante superior com relação às demais covariáveis, inclusive as que foram significativas.

Fazendo um novo ajuste especificando apenas API; DRFP; DRC e NPD (nesta ordem) como covariáveis, temos que todas elas permanecem significativas e conseqüentemente estarão presentes no modelo inicial para o escore de propensão no grupo dos adultos (não é necessário retirar mais nenhuma covariável). A tabela 5 apresenta os resultados da análise de desvios pelo teste da RV para este modelo (denotado por **m2.a**). De forma análoga ao que ocorreu no grupo das crianças, ao reespecificar as três covariáveis que não foram significativas no modelo completo, uma de cada vez e conjuntamente às covariáveis que o foram (modelos **m3.a.1** a **m3.a.3**), elas continuam não sendo significativas. Portanto, adotamos o modelo em que, junto ao intercepto, as covariáveis API; DRFP; DRC e NPD

Tabela 4 – Resultados do teste da RV no modelo completo para o escore de propensão no grupo dos adultos, caso IID

	GL	Desvio	Desvio Residual	Valor-p ( $\chi^2$ , em %)
Modelo nulo	–	–	5393.05	–
AMCD	1	0.03	5393.02	87.36
API	1	6.47	5386.55	1.10*
DRFP	1	18.65	5367.90	<0.01*
AI	1	1.51	5366.39	21.84
DRC	1	20.94	5345.45	<0.01*
DRR	1	3.36	5342.09	6.67
NPD	3	112.21	5229.88	<0.01*

são especificadas.

Tabela 5 – Resultados do teste da RV no modelo com API; DRFP; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, caso IID

	GL	Desvio	Desvio Residual	Valor-p ( $\chi^2$ , em %)
Modelo nulo	–	–	5393.05	–
API	1	6.48	5386.57	1.09*
DRFP	1	18.67	5367.90	<0.01*
DRC	1	20.95	5346.95	<0.01*
NPD	3	111.09	5235.86	<0.01*

Assim como feito no modelo escolhido para os escores no grupo das crianças (caso IID), também apresentamos para este modelo as correspondentes estimativas dos coeficientes (pontual e do erro-padrão associado), bem como os resultados do teste  $z$ , na tabela 6. Temos ao todo sete coeficientes para serem estimados, pois a variável NPD possui quatro categorias (as demais possuem apenas duas), logo é dividida na especificação do modelo em três variáveis binárias. A categoria de referência nas variáveis API; DRFP e DRC é sempre a resposta negativa: o indivíduo adulto não é portador de invalidez; reside em um domicílio que não recebe fundos de pensão ou renda através de caridade. Na variável NPD, a categoria de referência original é residir em um domicílio de pobreza extrema, mas a que adotaremos é residir em um domicílio não-pobre.

Pela tabela 6, adultos que residem em domicílios que recebem fundos de pensão ou renda através de caridade, ou que se encontram em situação de pobreza extrema (comparados aos pertencentes a qualquer das outras três categorias de NPD), estão mais propensos a serem beneficiários do SWF. Mantidas as respostas nas categorias de referência adotadas para as demais covariáveis, o domicílio receber fundos de pensão garante ao adulto uma probabilidade igual a  $\text{expit}[-0.6313 + (-1.0194) + 0.5779] \approx 0.2549$  ou 25,49% de ser beneficiário do SWF. Para os adultos residentes em domicílios que recebem renda através de caridade, esta probabilidade é ligeiramente maior:  $\text{expit}[-0.6313 + (-1.0194) + 0.6134] \approx$

Tabela 6 – Estimativas para os coeficientes no modelo com API; DRFP; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, caso IID

Termo da Regr.	Est. Pontual	Erro-Pad.	Estat. $z$	Valor-p (%)
Intercepto	-0.6313	0.0779	-8.10	<0.01*
API = “sim”	-0.4133	0.2030	-2.04	4.17*
DRFP = “sim”	0.5779	0.1295	4.46	<0.01*
DRC = “sim”	0.6134	0.1261	4.86	<0.01*
NPD = “pob. mod.”	-0.7840	0.1027	-7.64	<0.01*
NPD = “vulner.”	-0.7097	0.1120	-6.34	<0.01*
NPD = “não-pob.”	-1.0194	0.0958	-10.64	<0.01*

0.2617 ou 26,17%. Mesmo comparados aos que residem em domicílios de pobreza moderada, os adultos que estão em situação de extrema pobreza têm maior probabilidade de receber o benefício ( $\text{expit}(-0.6313) \approx 0.3472$  ou 34,72% contra  $\text{expit}[-0.6313 + (-0.7840)] \approx 0.1954$  ou 19,54%).

Por outro lado, adultos portadores de alguma forma de invalidez estão menos propensos a serem beneficiários do SWF comparados aos que não possuem nenhuma:  $\text{expit}[-0.6313 + (-1.0194) + (-0.4133)] \approx 0.1126$  ou 11,26% contra  $\text{expit}[-0.6313 + (-1.0194)] \approx 0.1610$  ou 16,10%. Como esta característica é um dos critérios de elegibilidade para o SWF, temos evidência de que o programa não possui uma boa focalização para os portadores de invalidez. Com relação aos valores dos escores preditos para o grupo dos adultos no caso IID, eles também estão estritamente entre 0 e 1, pois todos são maiores do que 0.1 e menores do que 0.5 (figura 11 do anexo B). Logo, poderemos usá-los para o pareamento dos indivíduos tratados. Ainda, tais escores são bem mais informativos do que os para o grupo das crianças, pois o modelo que os gerou contém mais covariáveis. Quanto menos parcimonioso for o modelo inicial, maior é a facilidade com que uma covariável deverá ser excluída posteriormente devido à rejeição nos testes de balanceamento para os indivíduos pareados. Se este for o caso, um novo modelo deve ser escolhido para refazer o pareamento e os testes de balanceamento.

A partir de agora, vamos refazer a análise que produziu o ajuste dos dois modelos considerando o PAC empregado para extrair a amostra da YNSPMS. Qualquer processo de estimação e da inferência subsequente para a população (de crianças ou adultos que participam da força de trabalho) que vise incorporar dados amostrais complexos precisa ser adaptado para características como estratificação, conglomeração e uso de pesos desiguais na amostra. Quando há estratificação, não será possível estimar nenhuma quantidade populacional (parâmetro) para indivíduos que são membros únicos de seu estrato (temos informação apenas do 1º nível de estratificação). Se houverem muitos indivíduos nesta situação, uma solução é uni-los aos pertencentes ao estrato mais similar àquele ao qual pertenciam originalmente e estimar novamente o parâmetro desejado. Assim, não se perde

informação sobre as características destes indivíduos<sup>3</sup>. Não há estratos com um único indivíduo na amostra considerada de adultos (5 169 indivíduos) e, na de crianças, apenas um indivíduo dentre os 611 considerados é único em seu estrato (no caso, a província de Shabwa). Por esta razão, optamos por excluí-lo para permitir a estimação dos escores preditos ao considerar o PAC (veja a subseção B.1.2 para mais detalhes).

Começemos novamente pelo grupo das crianças (com um indivíduo a menos em relação ao caso IID), mantendo as mesmas variáveis como candidatas à especificação. Temos os resultados do teste da RP nas covariáveis do modelo de regressão logística completo (com todas as candidatas como covariáveis, denotado por `m1.cp` na subseção B.2.2 do apêndice) para o escore de propensão, considerando dados amostrais complexos (logo, com estimação dos coeficientes por MP), na tabela 7. Note que a distribuição de referência para a estatística do teste da RP é uma  $F$  corrigida com aproximação por ponto de sela, como descrito na seção 2.4. Como estamos considerando o PAC, também são apresentados os valores do EPAA estimado para cada desvio em uma coluna adicional.

Tabela 7 – Resultados do teste da RP no modelo completo para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS

	GL	Desvio	EPAA	Desvio Residual	Valor-p ( $F$ , em %)
Modelo nulo	–	–	–	445.48	–
CRDCM	(1, 191)	0.12	0.14	445.36	36.06
COM	(1, 190)	1.43	0.50	443.93	9.41
DRFP	(1, 189)	3.29	1.82	440.64	18.32
COP	(1, 188)	0.79	0.95	439.85	36.46
DRC	(1, 187)	72.51	8.64	367.34	0.45*
CRDAI	(1, 186)	46.77	5.72	320.57	0.50*
CRDAPI	(1, 185)	2.47	0.79	318.10	8.16
DRR	(1, 184)	0.01	1.44	318.09	91.99
NPD	(3, 181)	16.15	2.86	301.94	13.36

Assim como no caso IID, duas covariáveis foram significativas (ao nível nominal de 5%) no modelo completo. Porém, temos uma primeira diferença: agora são as covariáveis DRC e CRDAI. Isto é, quando consideramos o PAC, a covariável NPD já não é significativa neste primeiro modelo. O EPAA estimado para cada desvio varia bastante, mas para as covariáveis significativas esta quantidade é bem maior do que o valor unitário. Removendo esta e as demais covariáveis não-significativas e ajustando um novo modelo somente com DRC e CRDAI (denotado por `m2.cp`), temos que ambas as covariáveis continuam significativas. O que é outra distinção do caso IID, quando foi necessário excluir mais uma covariável. Desta forma, DRC e CRDAI estarão presentes no modelo inicial.

<sup>3</sup> Embora preserve bem mais informação quando uma suposta exclusão afeta boa parte da amostra, realocar o indivíduo isolado em um novo estrato requer conhecimento das características geográficas, demográficas e socioeconômicas do estrato ao qual pertencia originalmente e dos demais definidos pelo PAC.

Quando adicionamos, uma por vez, as outras sete covariáveis e ajustamos os respectivos modelos (totalizando sete ajustes com três covariáveis especificadas), mantendo a ordem de especificação do modelo completo, nenhuma delas passa a ser significativa (modelos `m3.cp.1` a `m3.cp.7`). Logo, nosso modelo inicial para o escore de propensão no grupo das crianças, considerando o PAC, especificará somente DRC e CRDAI como covariáveis (além do intercepto). Para este modelo, apresentamos na tabela 8 os resultados da análise de desvios pelo teste da RP.

Tabela 8 – Resultados do teste da RP no modelo apenas com DRC e CRDAI (além do intercepto) para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS

	GL	Desvio	EPAA	Desvio Residual	Valor-p ( $F$ , em %)
Modelo nulo	–	–	–	445.48	–
DRC	(1, 191)	69.05	8.70	376.43	0.57*
CRDAI	(1, 190)	43.07	5.85	333.36	0.76*

Mesmo com a retirada das demais variáveis, os valores do EPAA estimado para os desvios de DRC e CRDAI praticamente não se alteram com relação ao seu comportamento para o modelo completo, sendo ligeiramente maiores. Como o EPAA estimado é superior ao valor unitário, podemos dizer que os efeitos da conglomeração e do uso de pesos amostrais desiguais superaram o efeito da estratificação feita no 1º estágio, ocasionando perda de eficiência na estimação dos desvios. Apesar disto, temos um modelo inicial mais informativo comparado ao obtido para o caso IID, com uma covariável a mais. Ou seja, considerar o PAC nos ajudou a identificar um componente importante para explicar a variabilidade dos escores preditos, o qual seria desprezado se tratássemos as observações amostrais como IID. A seguir, apresentamos na tabela 9 as estimativas dos coeficientes do modelo (pontual e do erro-padrão associado), com os respectivos resultados do EPAA estimado<sup>4</sup> e do teste  $t$  (pois agora estamos trabalhando com dados amostrais complexos), com graus de liberdade dados pelo  $n^\circ$  de UPA's menos o de estratos no 1º estágio (Korn & Graubard 1999)<sup>5</sup>.

Em nossa análise manteremos as categorias de referência originais nas covariáveis do modelo ajustado (respostas negativas para DRC e CRDAI). Pela tabela 9, ao considerar o PAC efetivamente utilizado, as estimativas pontuais dos coeficientes associados ao intercepto e à covariável DRC aumentam em magnitude quando comparadas às mesmas estimativas em um modelo com especificação semelhante, mas sem a suposição da existência de um PAC (denotado por `m2.cp.s`), e também às estimativas no modelo inicial do caso

<sup>4</sup> Até o presente momento, nenhuma das funções do pacote `survey` calcula de forma automática o EPAA estimado para coeficientes estimados por MP, mas este pode ser facilmente obtido pela razão entre a variância estimada do coeficiente no modelo que incorpora o PAC e no modelo que o ignora (ambos com a mesma especificação).

<sup>5</sup> Optou-se por não registrá-los para cada coeficiente na tabela, visto que o cálculo é similar ao feito para o observado no teste da RP

Tabela 9 – Estimativas para os coeficientes no modelo com DRC e CRDAI (além do intercepto) para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS

Termo da Regr.	Est. Pontual	Erro-Pad.	EPAA	Estat. $t$	Valor-p (%)
Intercepto	-2.8317	0.2744	5.95	-10.32	<0.01*
DRC = “sim”	2.2945	0.7000	4.25	3.28	0.12*
CRDAI = “sim”	2.2372	0.7927	3.29	2.82	0.53*

IID para as crianças. O erro-padrão também cresce, resultando em valores superiores a 1 para a estimativa do EPAA, assim como na análise de desvios pelo teste da RP. Mesmo com isso, ao incorporar o PAC todos os coeficientes estimados são significativos ao nível nominal de 5%.

Podemos também concluir pela tabela 9 que as crianças residentes em domicílios que recebem renda através de caridade têm uma probabilidade igual a  $\text{expit}(-2.8317+2.2945) \approx 0.3688$  ou 36,88% de serem beneficiárias do SWF, maior do que para aquelas residentes em domicílios que não recebem renda por esta modalidade ( $\text{expit}(-2.8317) \approx 0.0556$  ou apenas 5,56%), mantidas as demais covariáveis constantes e na categoria de referência (para CRDAI, a resposta negativa). Por sua vez, as crianças que residem junto a um adulto idoso têm uma probabilidade igual a  $\text{expit}(-2.8317+2.2372) \approx 0.3556$  ou 35,56% de serem contempladas pelo programa, também maior em relação à das crianças que não residem, novamente mantendo as demais covariáveis constantes e na categoria de referência (para DRC, a resposta negativa). Esta probabilidade indica que a focalização do SWF precisa melhorar para o subgrupo das crianças que participam da força de trabalho e residem junto a um idoso (esta última característica torna o domicílio elegível para o SWF). Os escores preditos, embora possuam suporte maior comparado ao caso IID, estão estritamente entre 0 e 1: todos são maiores do que 0.05 e menores do que 0.85 (código na subseção B.2.2 e histograma na figura 12 do anexo B).

Finalmente, para o grupo dos adultos (com o mesmo banco do caso IID) e mantendo as mesmas variáveis candidatas, temos na tabela 10 os resultados do teste da RP nas covariáveis do modelo completo para o escore de propensão, considerando o PAC da YNSPMS (denotado por `m1.ap` na subseção B.2.2 do apêndice). Neste modelo, temos que as covariáveis API; DRC e NPD são significativas ao nível nominal de 5%. Ou seja, é perdida uma covariável (DRFP) em comparação com o caso IID.

Ajustando um novo modelo (denotado por `m2.ap`) especificando somente as três covariáveis que foram significativas no modelo completo, temos que todas elas permanecem com este comportamento e por esta razão estarão presentes no modelo inicial para o escore de propensão no grupo dos adultos. Na tabela 11, são apresentados os resultados da análise de desvios pelo teste da RP para o modelo reduzido. Quando adicionamos as covariáveis

Tabela 10 – Resultados do teste da RP no modelo completo para o escore de propensão no grupo dos adultos, considerando o PAC da YNSPMS

	GL	Desvio	EPAA	Desvio Residual	Valor-p ( $F$ , em %)
Modelo nulo	–	–	–	3657.34	–
AMCD	(1, 542)	1.32	0.92	3656.02	23.44
API	(1, 541)	17.96	0.98	3638.06	<0.01*
DRFP	(1, 540)	1.75	5.22	3636.31	55.98
AI	(1, 539)	16.92	4.49	3619.39	5.46
DRC	(1, 538)	103.63	16.33	3515.76	1.26*
DRR	(1, 537)	4.17	6.41	3511.59	42.00
NPD	(3, 534)	332.25	11.68	3179.34	0.04*

que não foram significativas no modelo completo, uma por vez e na ordem em que foram especificadas originalmente (modelos **m3.ap.1** a **m3.ap.4**), elas permanecem não sendo significativas. Logo, escolhemos o modelo que especifica, além do intercepto, API; DRC e NPD como covariáveis.

Tabela 11 – Resultados do teste da RP no modelo com API; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, considerando o PAC da YNSPMS

	GL	Desvio	EPAA	Desvio Residual	Valor-p ( $F$ , em %)
Modelo nulo	–	–	–	3657.34	–
API	(1, 542)	18.07	0.98	3639.27	<0.01*
DRC	(1, 541)	103.81	17.74	3535.46	1.66*
NPD	(3, 538)	326.45	12.04	3209.01	0.05*

Para uma das covariáveis (API), o EPAA estimado para o desvio correspondente foi ligeiramente menor que 1, indicando um pequeno ganho de eficiência quando o PAC e a EP são considerados (lembramos que a YNSPMS entrevistou todos os moradores de um domicílio amostrado). Para as outras duas (DRC e NPD), ocorre o contrário, visto que os valores do EPAA correspondente são bem elevados. Com relação ao modelo completo, novamente os valores do EPAA praticamente são os mesmos. Por fim, apresentamos para o modelo escolhido as estimativas de seus coeficientes, além dos resultados para o EPAA estimado e o teste  $t$  (pois incorporamos o PAC), na tabela 12. Ao todo, seis coeficientes serão estimados. Para as três variáveis associadas, a interpretação dos coeficientes para a subsequente estimação das probabilidades desejadas tomará as mesmas categorias de referência adotadas no caso IID (respostas negativas para API e DRC; residir em um domicílio não-pobre em NPD).

Pela tabela 12, ao contrário do que ocorreu para o modelo escolhido no grupo das crianças, considerar o PAC da YNSPMS não necessariamente causou um aumento na magnitude das estimativas pontuais dos coeficientes. Isso acontece tanto quando elas são

Tabela 12 – Estimativas para os coeficientes no modelo com API; DRC e NPD (além do intercepto) para o escore de propensão no grupo dos adultos, considerando o PAC da YNSPMS

Termo da Regr.	Est. Pontual	Erro-Pad.	EPAA	Estat. $t$	Valor-p (%)
Intercepto	-0.5314	0.5238	6.78	-1.01	31.08
API = “sim”	-1.6726	0.5965	2.94	-2.80	0.52*
DRC = “sim”	1.4630	0.6988	5.55	2.09	3.68*
NPD = “pob. mod.”	-1.9718	0.5821	5.68	-3.39	0.08*
NPD = “vulner.”	-1.4263	0.5825	5.21	-2.45	1.47*
NPD = “não-pob.”	-2.2025	0.5489	5.74	-4.01	<0.01*

comparadas às mesmas estimativas em um modelo com especificação semelhante, mas que ignora o PAC (denotado por `m2.ap.s`), quanto quando são comparadas às estimativas no modelo inicial do caso IID para os adultos (tomando apenas o intercepto e as covariáveis presentes em ambos os modelos). Por outro lado, o erro-padrão cresce para todos os coeficientes, conseqüentemente o EPAA estimado é sempre maior do que o valor unitário. Embora considerar o PAC tenha produzido maiores erros-padrão, somente o coeficiente associado ao intercepto não é significativo ao nível nominal de 5%. Por esta razão, não vamos considerá-lo na interpretação dos coeficientes, pois não foi rejeitada pelo teste  $t$  a hipótese nula para o coeficiente associado ao intercepto (com valor-p bem acima do nível de significância).

Assim, podemos concluir pela tabela 12 que os adultos residentes em domicílios que recebem renda através de caridade ou que se encontram em situação de pobreza extrema têm maior probabilidade de serem beneficiários do SWF. Para estes dois subgrupos, as probabilidades são de  $\text{expit}[1.4630 + (-2.2025)] \approx 0.3231$  ou 32,31% e  $\text{expit}(0) = 0.5$  ou 50%, respectivamente, mantidas constantes as demais categorias de referência. Estes resultados indicam que a focalização do SWF não é boa com relação aos adultos residentes em domicílios extremamente pobres ou que dependem da caridade de terceiros. Da mesma forma, os adultos que possuem alguma forma de invalidez estão menos propensos a receberem o benefício, cuja probabilidade é de apenas  $\text{expit}[-1.6726 + (-2.2025)] \approx 0.0203$  ou 2,03%. Esta probabilidade é ainda menor do que a encontrada para este mesmo subgrupo no caso IID (cujo modelo inicial incluía também a covariável DRFP), indicando que o SWF praticamente não auxilia os adultos que se encaixam apenas neste critério de elegibilidade (ser portador de alguma forma de invalidez). Por sua vez, os escores preditos do modelo escolhido para os adultos, também com suporte maior comparado ao caso IID, estão estritamente entre 0 e 1: todos estão entre 0.01 e 0.72, exclusive (código na subseção B.2.2 e histograma na figura 14 do anexo B).

Tendo os quatro modelos iniciais para o escore de propensão em ambos os grupos etários, já pudemos perceber diferenças claras entre a análise sem considerar o PAC e

aquela em que este é incorporado antes mesmo de prosseguir para o pareamento dos indivíduos tratados com controles comparáveis. Tanto no grupo das crianças quanto no dos adultos, as covariáveis especificadas não são as mesmas. Quando o PAC é considerado, há mais covariáveis na especificação do modelo no grupo das crianças e menos no grupo dos adultos, cujo modelo também não apresentou intercepto significativo ao nível nominal de 5%. Por fim, os resultados do EPAA estimado, tanto para os desvios no teste da RP quanto nos coeficientes e em ambos os grupos, indicam que em geral as variâncias são subestimadas ao se ignorar completamente o PAC. Na próxima seção, encontraremos para cada modelo o banco pareado pelos respectivos escores preditos, confirmando ou não o balanceamento nas covariáveis (através do teste  $\chi^2$  de Pearson) e nos escores preditos (através do teste MW).

## 4.2 Pareamento e Balanceamento Pós-modelagem

O pareamento para os indivíduos tratados em cada um dos quatro modelos iniciais para o escore de propensão especificados na seção anterior seguirá o protocolo definido na seção 3.1. Isto é, utilizaremos o PDE com reposição de indivíduos controles selecionados; um indivíduo controle pareado por tratado (sem sobreamostragem) e com nível máximo de tolerância igual a 0.2 vezes o desvio padrão dos escores preditos pelo modelo de regressão logística, que de fato foi a utilizada para estimar os escores. Sempre que houver empates entre dois ou mais controles a serem selecionados, situação extremamente comum quando há poucas covariáveis especificadas e grande parte ou a totalidade delas é categórica, a definição de qual controle será pareado é feita aleatoriamente. Por esta razão, é necessário impor uma semente no processo de pareamento para que os resultados obtidos através do banco pareado, incluindo os testes de balanceamento, sejam reproduzíveis. No R, todas estas características podem ser configuradas com o uso da função `Match()` do pacote `Matching` (Sekhon, 2011). O código para pareamento dos escores nos quatro modelos pode ser consultado nas subseções B.2.3 e B.2.4 do apêndice.

Construídos os bancos pareados, procedemos para os testes de balanceamento nos valores das covariáveis e dos escores preditos a fim de confirmar se os indivíduos tratados e os controles pareados com estes são de fato comparáveis. Como todas as covariáveis são categóricas, o balanceamento em cada covariável isoladamente será verificado através do teste  $\chi^2$  de Pearson para a hipótese nula de independência. Uma vez confirmado o balanceamento em todas as covariáveis, o teste da soma dos postos de Mann-Whitney para a hipótese nula de homogeneidade nos dirá se o balanceamento ocorreu nos escores preditos. Diferentemente dos testes aplicados na modelagem do escore de propensão, nosso objetivo será atingido se os testes para balanceamento não rejeitarem suas respectivas hipóteses nulas. Caso isso aconteça, o modelo para o escore de propensão deve ser refeito, substituindo o modelo inicial por outro que esteja encaixado dentro deste. Neste contexto,

também é possível que surjam dois ou mais modelos concorrentes, com a mesma quantidade de covariáveis especificadas. Novamente, optaremos pelo que apresentar a maior (menor) taxa de acerto (erro) na classificação dos indivíduos em tratados e controles a partir dos escores preditos correspondentes. A título de comparação, os testes de balanceamento também serão aplicados nos bancos originais, como forma de justificar o pareamento dos indivíduos antes da avaliação de impacto na resposta de interesse.

Começando pelo caso em que tratamos as observações amostrais como IID (código na subseção B.2.3 do apêndice) e tomando os escores preditos do respectivo modelo no grupo das crianças, temos ao realizar o pareamento que todos os indivíduos tratados são pareados com algum controle, resultando em um novo banco com 118 pares. Apesar da seleção dos indivíduos controles ter sido feita com reposição, um total de 103 controles distintos entre si são utilizados<sup>6</sup>. Antes dos testes de balanceamento, é necessário conferir se o suporte dos escores preditos é comum para indivíduos tratados e controles, o que é confirmado pelo histograma da figura 15 (anexo C). Dado o método de pareamento utilizado, esta condição é cumprida como esperado. Fazendo os testes  $\chi^2$  de Pearson para a única covariável do modelo (DRC) e MW para os escores preditos, temos que nenhum dos testes é rejeitado e ambos estão balanceados: os valores em DRC são independentes e os escores preditos são homogêneos nas duas classes, como mostra a tabela 13. De novo, sempre que algum valor-p for significativo (ao nível nominal de 5%), este será acompanhado do símbolo \* (asterisco). Note que os dois testes são rejeitados quando aplicados aos dados do banco original, uma prova de que o pareamento é necessário antes de realizar qualquer avaliação de impacto.

Tabela 13 – Resultados dos testes  $\chi^2$  de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo das crianças, caso IID

Variável (Banco)	Estat. $\chi^2$ (Pearson)	GL	Valor-p (%)	Balanc.
DRC (Original)	18.12	1	<0.01*	Negativo
DRC (Pareado)	<0.01	1	>99.99	Positivo
Variável (Banco)	Estat. $\chi^2$ (MW)	GL	Valor-p (%)	Balanc.
Escore (Original)	18.09	1	<0.01*	Negativo
Escore (Pareado)	<0.01	1	>99.99	Positivo

Ainda supondo que as observações amostrais são IID, mas agora tomando os escores preditos do respectivo modelo no grupo dos adultos, temos ao realizar o pareamento que 5 dos 1116 indivíduos tratados não foram pareados com nenhum controle, devido ao nível de tolerância imposto, e por esta razão não estarão presentes no banco pareado. Logo, um total de 1111 pares foram formados, com 952 controles distintos (85,69% do total). Com

<sup>6</sup> Este elevado índice de controles distintos nos pares formados (87,29% dos 118) é consequência do gerador pseudo-aleatório padrão do R, o algoritmo Mersenne-Twister (Matsumoto & Nishimura, 1998), cujo período é extremamente elevado.

relação ao suporte dos escores preditos no banco pareado, ele é comum para as duas classes (tratados e controles) de indivíduos, como indicado pelo histograma da figura 16 (anexo C). Fazendo o teste  $\chi^2$  de Pearson para cada uma das quatro covariáveis do modelo (API; DRFP; DRC e NPD) e o teste MW para os escores preditos, temos que as observações nas quatro covariáveis e os escores estão balanceados nas duas classes, como mostra a tabela 14. Novamente, os dois testes são rejeitados quando aplicados aos dados do banco original.

Tabela 14 – Resultados dos testes  $\chi^2$  de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo dos adultos, caso IID

Variável (Banco)	Estat. $\chi^2$ (Pearson)	GL	Valor-p (%)	Balanc.
API (Original)	5.95	1	1.48*	Negativo
DRFP (Original)	20.92	1	<0.01*	Negativo
DRC (Original)	21.85	1	<0.01*	Negativo
NPD (Original)	122.53	3	<0.01*	Negativo
API (Pareado)	<0.01	1	>99.99	Positivo
DRFP (Pareado)	<0.01	1	>99.99	Positivo
DRC (Pareado)	<0.01	1	>99.99	Positivo
NPD (Pareado)	<0.01	3	>99.99	Positivo
Variável (Banco)	Estat. $\chi^2$ (MW)	GL	Valor-p (%)	Balanc.
Escore (Original)	127.64	1	<0.01*	Negativo
Escore (Pareado)	<0.01	1	>99.99	Positivo

Tanto o banco pareado das crianças quanto o dos adultos no caso IID poderão ser utilizados para a avaliação de impacto do SWF em TSAT para o grupo etário correspondente. Observe que os valores-p dos testes de balanceamento são praticamente iguais a 1, mesmo nas observações da covariável NPD (que tem quatro categorias distintas). Isto é explicado pelo fato de que ambos os bancos são relativamente grandes e poucas covariáveis são testadas (no banco pareado das crianças, apenas uma), sendo todas categóricas. Com isso, há poucas combinações possíveis para os valores nestas covariáveis e consequentemente nos escores preditos. Além disso, todas as observações nos indivíduos pareados (tratados e controles) são supostas IID, com o mesmo peso amostral. Devido a este segundo fator, quando fizermos os testes de balanceamento (devidamente adaptados para dados amostrais complexos) no banco pareado pelos modelos estimados por MP, será mais difícil de rejeitar a hipótese nula de independência em cada covariável ou de homogeneidade nos escores preditos para as duas classes. Os valores-p serão menores em decorrência dos pesos amostrais distintos para indivíduos do mesmo par (comparáveis), considerados ao fazer os testes adaptados.

Consideremos agora o PAC utilizado na YNSPMS no pareamento pelo escore de propensão (código na subseção B.2.4 do apêndice) e comecemos novamente pelo grupo das crianças. Temos ao tomar os escores preditos do modelo correspondente (estimado por MP) que o pareamento realizado teve sucesso em parear um indivíduo controle para

todos os tratados, formando 118 pares ao total. Ou seja, o indivíduo excluído antes da estimação por MP que pertencia à província de Shabwa era controle. Foram usados 103 controles distintos (87,29% do total), a mesma proporção comparada à do pareamento no mesmo grupo supondo observações IID. Antes de verificar se o suporte dos escores preditos é comum para ambas as classes de tratados e controles, é necessário identificar no banco pareado, se houver, indivíduos que são únicos em seus respectivos estratos, de modo que seja possível fazer os testes de balanceamento adaptados para dados amostrais complexos. De fato, para o estrato correspondente à província de Áden, há um único indivíduo pareado (controle). Desta forma, o removemos do banco pareado conjuntamente ao indivíduo tratado com o qual formava par. Com os 117 pares restantes, temos que os escores preditos possuem suporte comum para as duas classes de controles e tratados, como mostrado pelo histograma da figura 17 (C). Por fim, os resultados dos testes  $\chi^2$  de Pearson com correção de 2ª ordem de Rao-Scott  $F$ -ajustado e MW adaptados para dados amostrais complexos nas observações das duas covariáveis do modelo (DRC e CRDAI) e nos escores preditos, respectivamente, são expostos na tabela 15.

Tabela 15 – Resultados dos testes  $\chi^2$  de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo das crianças, considerando o PAC da YNSPMS

Variável (Banco)	Estat. $F$ (Pearson)	GL	Valor-p (%)	Balanc.
DRC (Original)	12.41	(1, 192)	0.05*	Negativo
CRDAI (Original)	13.09	(1, 192)	0.04*	Negativo
DRC (Pareado)	0.94	(1, 115)	33.54	Positivo
CRDAI (Pareado)	41.10	(1, 115)	<0.01*	Negativo
Variável (Banco)	Estat. $t$ (MW)	GL	Valor-p (%)	Balanc.
Escore (Original)	2.73	191	0.69*	Negativo
Escore (Pareado)	1.87	114	6.36	Positivo

Como os pesos amostrais são distintos, mesmo havendo muitos pares de tratados e controles cujas observações são idênticas em todas as covariáveis (consequentemente, com o mesmo escore predito), temos que a covariável CRDAI não está balanceada. Para ela, o valor-p do teste  $\chi^2$  de Pearson adaptado para dados amostrais complexos é extremamente reduzido. De fato, mesmo sendo significativa no modelo inicial, a variabilidade interna de CRDAI é muito baixa e não consegue sequer amenizar a influência dos pesos amostrais no resultado do teste. Assim, iremos reduzir nosso modelo inicial, especificando agora somente a covariável DRC. Para este novo modelo, de agora em diante chamado “modelo reduzido”, apresentamos as estimativas (pontual e erro-padrão) dos coeficientes associados ao intercepto e à covariável DRC, bem como os valores do EPAA estimado (obtido com auxílio da tabela 3) e os resultados do teste  $t$  na tabela 16.

Note que não é necessária uma análise de desvios pelo teste da RP, pois DRC é a única covariável (binária) do modelo. Os coeficientes, que ainda são todos significativos

Tabela 16 – Estimativas para os coeficientes no modelo apenas com DRC (além do intercepto) para o escore de propensão no grupo das crianças, considerando o PAC da YNSPMS

Termo da Regr.	Est. Pontual	Erro-Pad.	EPAA	Estat. $t$	Valor-p (%)
Intercepto	-2.4540	0.3147	2.86	-7.80	<0.01*
DRC = “sim”	2.7923	0.9766	2.90	2.86	0.47*

ao nível nominal de 5%, têm um ligeiro aumento em relação aos seus valores na tabela 9 (quando CRDAI é especificada no modelo), logo as probabilidades associadas também crescerão. Por exemplo, para a criança residente em domicílio que recebe renda através de caridade, a probabilidade de ser beneficiária do SWF é agora de  $\text{expit}(-2.4540 + 2.7923) \approx 0.5838$  ou 58,38%. Por sua vez, as estimativas de erro-padrão cresceram, mas o EPAA estimado diminuiu (ainda que esteja acima de 1), pois este agora é obtido da mesma especificação no caso IID, na qual as estimativas do erro-padrão também crescem quando CRDAI é retirada. Como os escores preditos pelo modelo reduzido estão estritamente entre 0 e 1 (figura 13 do anexo C), podemos usá-los para construir um banco pareado.

Ao realizar o pareamento, novamente todos os 118 indivíduos tratados são pareados com algum controle (agora com 104 controles distintos, ou 88,14%). No banco pareado para o modelo reduzido, desta vez há somente um indivíduo no estrato correspondente à província da cidade de Sana’a. Como este era originalmente tratado, o retiramos, assim como o controle com o qual estava pareado. Ainda, o suporte dos escores preditos é comum para controles e tratados no banco pareado (figura 18 do anexo C). Refazendo os testes de balanceamento no novo banco pareado para as crianças (considerando o PAC), temos agora que nenhuma das respectivas hipóteses nulas é rejeitada, como mostra a tabela 17.

Tabela 17 – Resultados dos testes  $\chi^2$  de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo reduzido no grupo das crianças, considerando o PAC da YNSPMS

Variável (Banco)	Estat. $F$ (Pearson)	GL	Valor-p (%)	Balanc.
DRC (Original)	12.41	(1, 192)	0.05*	Negativo
DRC (Pareado)	2.86	(1, 112)	9.34	Positivo
Variável (Banco)	Estat. $t$ (MW)	GL	Valor-p (%)	Balanc.
Escore (Original)	1.79	191	0.08	Positivo
Escore (Pareado)	1.42	111	0.16	Positivo

Observe que o valor-p da estatística  $F$  associada à covariável DRC no modelo reduzido é menor em relação ao seu análogo no modelo inicial, em que CRDAI também era especificada. Apesar disso, ele ainda está acima do nível nominal de significância (5%). Por outro lado, o valor-p da estatística  $t$  associado aos escores aumentou, evidenciando que o modelo reduzido produziu um banco pareado mais homogêneo nos escores do que o

obtido pelo modelo inicial no grupo das crianças. Assim, ele será utilizado para a avaliação de impacto do SWF em TSAT para o grupo das crianças, no caso em que consideramos o PAC.

Finalmente, tomando os escores preditos pelo modelo inicial para os adultos quando consideramos o PAC (com API; DRC e NPD especificadas como covariáveis), temos ao realizar o pareamento que todos os 1116 indivíduos tratados são pareados com algum controle. Ao todo, foram selecionados 917 controles distintos (82,17% do total) para o pareamento. Mesmo no banco pareado, não nenhuma província com um único indivíduo: todas as 19 províncias pesquisadas nas quatro rodadas estão representadas por dezenas ou mesmo centenas de indivíduos. Logo, todos os tratados e controles pareados são mantidos. Os escores preditos, quando separados por classe, possuem suporte comum, como exposto pela figura 19 (anexo C). Por fim, os resultados dos testes de balanceamento, adaptados para dados amostrais complexos,  $\chi^2$  de Pearson e MW nas três covariáveis do modelo inicial e nos escores preditos, respectivamente, são apresentados na tabela 18.

Tabela 18 – Resultados dos testes  $\chi^2$  de Pearson e MW para o balanceamento nos bancos original e pareado pelos escores do modelo inicial no grupo dos adultos, considerando o PAC da YNSPMS

Variável (Banco)	Estat. $F$ (Pearson)	GL	Valor-p (%)	Balanc.
API (Original)	14.14	(1, 543)	0.02*	Negativo
DRC (Original)	7.68	(1, 543)	0.58*	Negativo
NPD (Original)	13.02	(2.13, 1156.60)	<0.01*	Negativo
API (Pareado)	2.29	(1, 488)	13.07	Positivo
DRC (Pareado)	0.34	(1, 488)	55.99	Positivo
NPD (Pareado)	1.14	(2.35, 1144.70)	32.46	Positivo
Variável (Banco)	Estat. $t$ (MW)	GL	Valor-p (%)	Balanc.
Escore (Original)	4.63	542	<0.01*	Negativo
Escore (Pareado)	1.82	487	7.01	Positivo

Nenhum dos testes é rejeitado para o banco pareado pelo modelo inicial. Logo, este pode ser usado para a avaliação de impacto do SWF em TSAT nos adultos que trabalhavam na 1<sup>a</sup> e na 4<sup>a</sup> rodada. De fato, a presença de muitos pares no banco pareado dos adultos amenizou a influência dos pesos amostrais nos indivíduos pareados. Ainda assim, os valores-p dos testes adaptados para o PAC sequer se aproximam dos obtidos para o caso IID, cujo modelo inicial especifica uma covariável a mais (DRFP), mesmo estando acima do nível nominal de significância (5%). Entretanto, não é necessário retirar mais nenhuma covariável e refazer todo o processo de modelagem, pareamento e verificação do balanceamento como feito no grupo das crianças quando o PAC é considerado.

Concluído o pareamento e feitos os testes de balanceamento até que fosse encontrado um banco pareado satisfatório para cada grupo etário e abordagem, podemos finalmente passar para a última etapa, referente à avaliação de impacto propriamente dita. Também

dividida por grupo etário e abordagem, a avaliação tomará o banco pareado correspondente cujo balanceamento foi confirmado para as covariáveis do modelo que o gerou.

### 4.3 Avaliação do Impacto do SWF em TSAT sem e com PAC

Uma vez que obtemos bancos pareados para cada um dos quatro casos (de acordo com a abordagem e o grupo etário utilizados) com balanceamento confirmado pelos testes  $\chi^2$  de Pearson (em cada covariável) e de MW (no vetor de escores preditos) para as duas classes, podemos utilizar cada um destes bancos para a avaliação de impacto do SWF em uma resposta de interesse (TSAT). Como dito anteriormente, esta avaliação será feita através do método DD, combinando-o ao pareamento já realizado, pois o complementa ao permitir que características não observadas ou pesquisadas possam influenciar a resposta de interesse, desde que a influência destas características seja constante ao longo tempo. Logo, o método DD insere o componente da longitudinalidade.

Dado o grupo etário e o conhecimento sobre a incorporação (ou não) do PAC utilizado para extrair a amostra original, o modelo de regressão linear normal para a avaliação pelo método DD expresso na equação 3.23 incluirá como covariável de confundimento toda aquela que em tese pode influenciar na resposta de interesse (TSAT), mas não necessariamente na atribuição do tratamento (RSWF). Isso é válido também para aquelas com balanceamento confirmado dadas as observações no banco pareado pelos escores preditos do modelo de regressão logística que as especificava. A razão para isto é que o balanceamento nas classes de controles e tratados só é verificado para as observações destas covariáveis no período pré-tratamento (1ª rodada). Quando o PAC for ignorado, a estimação dos coeficientes do modelo para avaliação de impacto será feita por MV, e quando quisermos incorporá-lo, usaremos a estimação por MP, que utiliza o peso amostral (aqui, longitudinal) de cada indivíduo do banco pareado (como visto na subseção 2.3.3).

Seguindo a literatura discutida na seção 1.3, para cada grupo etário selecionaremos um conjunto com as potenciais variáveis de confundimento, dentre as pesquisadas na YNSPMS, para TSAT. Tanto no grupo das crianças quanto no dos adultos, serão incluídas as respectivas candidatas iniciais à especificação na modelagem do escore de propensão (mesmo aquelas excluídas posteriormente) e indicadores demográficos (como gênero e idade, por exemplo); geográficos e relacionados ao nível educacional. Evidentemente, a escolha das variáveis de confundimento a serem especificadas será distinta em cada grupo etário.

Para o grupo das crianças, temos as seguintes variáveis de confundimento, que podem ou não ser especificadas como covariáveis, para o modelo de regressão linear normal da avaliação do impacto pelo método DD, além das nove candidatas iniciais ao modelo de regressão logística para o escore de propensão. Todas elas estão acompanhadas de seus nomes originais na YNSPMS entre parênteses (Banco de Dados da YNSPMS, 2014c):

- CPI: Criança é Portadora de Invalidez (p2\_q\_11);
- EC: Escolaridade da Criança (yrschl);
- EMC: Escolaridade da Mãe da Criança (motheredu);
- EPC: Escolaridade do Pai da Criança (fatheredu);
- GC: Gênero da Criança, masculino ou feminino (p2\_q\_03);
- IC: Idade da Criança, em anos (p2\_q\_05\_y);
- LD: Localização do Domicílio, na zona urbana ou rural (f8);
- MCPE: Matrícula da Criança em Programa Educacional (p3\_q\_2);
- PCD: Pessoas por Cômodo no Domicílio (crowd2);
- TD: Topografia do Domicílio (topography).

Para algumas das potenciais variáveis de confundimento citadas acima faremos comentários adicionais a fim de facilitar a interpretação dos resultados. Nas variáveis EC; EMC e EPC, categorias correspondentes a valores próximos entre si de escolaridade (em anos) foram agrupadas em novas categorias devido ao fato de que algumas delas apresentavam uma quantidade bem reduzida de observações<sup>7</sup>. Além disso, uma das categorias se refere àqueles que jamais frequentaram a escola, mas que foram alfabetizados através da leitura do Alcorão<sup>8</sup>. No banco pareado para as crianças e somente para ele, nos casos IID e considerando o PAC, esta categoria (também com raras observações em EC, EMC e EPC) foi agrupada com a referente aos que não possuem nenhuma escolaridade. Ainda para EMC e EPC, há uma categoria para as crianças que não sabem a escolaridade da mãe ou do pai, respectivamente, por serem órfãs ou porque o adulto em questão não residia no domicílio. Logo, tais respostas serão renomeadas como “ausente” nas tabelas de resultados (note que não serão tratadas como valores faltantes). Com relação à variável CPI, esta não deve ser confundida com CRDAPI, pois nos referimos a uma condição da própria criança. Para a variável TD, categorias com poucas observações também foram agrupadas. Por fim, para as variáveis EC; EMC; EPC; GC; IC; LD e TD tomamos as mesmas respostas da 1ª rodada para a 4ª, pois a YNSPMS teve duração inferior a um ano e não seguiu entrevistados que se deslocaram do lugar em que residiam.

Ao todo, temos 17 candidatas a covariáveis de confundimento do modelo de regressão linear para as crianças. As variáveis EMC e EPC fornecem informação sobre a escolaridade da mãe ou do pai e sobre a sua ausência, cuja categoria correspondente será na prática perfeitamente correlacionada com as variáveis COM e COP. Desta forma, não especificaremos estas duas últimas independentemente da abordagem utilizada (considerando ou não o PAC), a fim de reduzir o erro-padrão das covariáveis restantes. Ainda, também eliminaremos da especificação do modelo para as crianças, dada a abordagem, todas as variáveis categóricas que possuírem poucas observações (até cinco indivíduos) em

<sup>7</sup> Os comandos para obtenção de frequências das potenciais variáveis de confundimento no banco pareado, em cada abordagem, são descritos nas subseções B.2.5 e B.2.6

<sup>8</sup> O texto religioso principal do islamismo.

uma ou mais categorias que não podem ser reunidas com outra mais frequente por não haver uma relação de proximidade semântica entre elas ou a variável possui apenas duas categorias. Com relação à taxa de valores faltantes, esta é praticamente nula para todas as candidatas, em ambos os bancos pareados para as crianças (ignorando e incorporando o PAC).

Por sua vez, no grupo dos adultos temos as seguintes variáveis de confundimento, das quais a grande maioria é análoga às existentes para o grupo das crianças, além das sete candidatas iniciais ao modelo de regressão logística para o escore de propensão:

- EA: Escolaridade do Adulto (*yrschl1*);
- GA: Gênero do Adulto, masculino ou feminino (*p2\_q\_03*);
- IA: Idade do Adulto, em anos (*p2\_q\_05\_y*);
- LD: Localização do Domicílio, na zona urbana ou rural (*f8*);
- PCD: Pessoas por Cômodo no Domicílio (*crowd2*);
- TD: Topografia do Domicílio (*topography*).

o que nos fornece um total de 13 candidatas a covariáveis de confundimento do modelo de regressão linear para os adultos. Nas variáveis análogas, o agrupamento pode ser desnecessário ou feito para uma quantidade menor de categorias em virtude do banco pareado ser bem maior com relação ao das crianças, para ambas as abordagens. Também nos dois bancos pareados para os adultos (sem e com PAC incorporado), a taxa de valores faltantes é praticamente nula para todas as candidatas correspondentes.

Além das covariáveis de confundimento, em cada modelo também serão incluídos parâmetros associados ao tratamento atribuído para o indivíduo (variável RSWF); à Rodada de Coleta (RC) das observações (uma variável binária) e à interação entre ambas (RSWF:RC), necessários para avaliação de impacto do SWF pelo método DD. O valor médio deste ao longo do tempo é fornecido pelas estimativas do coeficiente e erro-padrão associado à interação. Começando pela abordagem em que as observações são tratadas como IID (comandos em R na subseção B.2.5 do apêndice) e tomando o banco pareado para o grupo das crianças, excluimos das candidatas iniciais a covariáveis de confundimento do modelo de regressão linear normal as variáveis CRDCM; CPI e LD. As três têm cinco ou menos indivíduos com observações positivas (CPI e CRDCM) ou residentes na zona urbana (LD) para pelo menos uma das rodadas. Logo, teremos 14 covariáveis de confundimento a serem especificadas. As estimativas (pontual e erro-padrão), bem como as estatísticas *t* e valores-*p* correspondentes, para o modelo de regressão linear normal para a resposta de interesse TSAT (denotado por *m.dd.c*) são apresentadas na tabela 19. Quando significativas ao nível nominal de 5%, os respectivos valores-*p* são novamente marcados por \*. Sempre que analisarmos a variação em TSAT para um coeficiente associado a mudanças em uma dada variável categórica ou a variações unitárias em variáveis numéricas, vamos pressupor que as categorias ou valores numéricos das demais covariáveis são mantidos

constantes. No caso IID, os graus de liberdade para cada coeficiente são sempre dados pelo total de indivíduos na amostra pareada menos 1.

Tabela 19 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo das crianças, caso IID

Termo da Regr.	Est. Pontual	Erro-Pad.	Estat. $t$	Valor-p (%)
Intercepto	29.4713	6.0166	4.90	<0.01*
TD = “pl. des.”	6.1577	2.9711	2.07	3.88*
TD = “montan.”	3.8150	2.4203	1.58	11.57
PCD	-0.7448	0.4162	-1.79	7.42
GC = “masc.”	0.7895	1.6996	0.46	64.25
IC	0.1105	0.4219	0.26	79.35
MCPE = “sim, an. m.”	-5.5366	3.4871	-1.59	11.31
MCPE = “sim, at. m.”	-10.1589	3.3970	-2.99	0.29*
EC = “1 ano”	1.6680	3.9843	0.42	67.57
EC = “2 anos”	4.3532	3.5924	1.21	22.63
EC = “3 anos”	3.3270	3.5116	0.95	34.39
EC = “4 anos”	8.9561	3.5968	2.49	1.31*
EC = “5 anos”	2.5675	3.8138	0.67	50.12
EC = “6 anos”	12.7575	4.0740	3.13	0.19*
EC = “7 anos +”	-0.8026	4.3582	-0.18	85.40
EMC = “1 a 4 anos”	-0.4202	4.1369	-0.10	91.91
EMC = “5 a 8 anos”	3.2280	3.9878	0.81	41.87
EMC = “9 anos +”	-1.1447	4.2046	-0.27	78.56
EMC = “ausente”	0.6268	3.3632	0.19	85.22
EPC = “1 a 4 anos”	-6.4814	2.5532	-2.54	1.15*
EPC = “5 a 8 anos”	-0.2624	2.6470	-0.10	92.11
EPC = “9 anos +”	-2.2544	2.2365	-1.01	31.40
EPC = “ausente”	-3.2028	2.8592	-1.12	26.33
CRDAI = “sim”	-1.3656	2.9426	-0.46	64.28
CRDAPI = “sim”	-3.8855	2.8769	-1.35	17.75
DRFP = “sim”	6.0178	3.4074	1.77	7.81
DRC = “sim”	-3.3140	2.4184	-1.37	17.13
DRR = “sim”	-0.8540	2.2273	-0.38	70.16
NPD = “pob. mod.”	-0.7133	2.0901	-0.34	73.31
NPD = “vulner.”	-2.0814	2.5015	-0.83	40.58
NPD = “não-pob.”	-3.1853	2.2722	-1.40	16.17
RSWF	-1.1323	2.0557	-0.55	58.20
RC	0.3193	1.9982	0.16	87.31
RSWF:RC	3.0704	2.7893	1.10	27.16

Temos pela tabela 19 que os coeficientes associados ao intercepto; às crianças que moram em regiões planas desérticas; que estão atualmente matriculadas na escola ou em outro programa educacional; que possuem determinado nível de escolaridade (4 e 6 anos) e cujos pais têm de 1 a 4 anos de escolaridade são significativos. Como esperado, as crianças que disseram estar matriculadas em algum programa educacional no momento da entrevista

de fato alocam menos tempo para o trabalho durante a semana quando comparadas às que nunca estiveram matriculadas, a categoria de referência para MCPE (cerca de 10 horas e 10 minutos a menos), independentemente da rodada considerada. O modelo também confirma que crianças residentes em regiões desérticas alocam semanalmente mais tempo para o trabalho do que as residentes em regiões litorâneas, a categoria de referência para TD (aproximadamente 6 horas e 9 minutos a mais), algo também esperado. Um cuidado adicional deve ser tomado ao analisar unidades fracionárias de tempo, as quais são baseadas no sistema sexagesimal, e não decimal.

Por outro lado, o aumento da escolaridade para a criança implica em elevação significativa da jornada semanal de trabalho para aquelas que possuem 4 ou 6 anos de escolaridade (cerca de 8 horas e 57 minutos, e 12 horas e 45 minutos, respectivamente), comparadas às que não possuem nenhum ano de estudo (categoria de referência para EC). Mesmo supondo observações IID, coeficientes associados à variáveis relativas aos critérios de elegibilidade para o SWF, das quais se espera um efeito na redução do tempo alocado semanalmente para o trabalho nas crianças residentes de domicílios que têm maior acesso à renda (não só do SWF, mas também dos fundos de pensão, da caridade ou das remessas de parentes e amigos), não foram significativos. Isto pode ser reflexo do fato de que o banco pareado referente ao modelo contém apenas 236 indivíduos. Por fim, a estimativa do coeficiente para a interação de RSWF com RC foi positiva para a variável TSAT nas crianças. Porém, como este resultado não foi significativo, não se pode dizer que o SWF tem impacto positivo na jornada semanal de trabalho da criança residente no domicílio que o recebe. Quando a rodada é fixada, entretanto, as crianças beneficiárias trabalham menos do que as não beneficiárias (também um resultado não significativo, mas esperado). Quando a classe é fixada, as crianças trabalham mais na 4<sup>a</sup> rodada.

Tomando agora o banco pareado para o grupo dos adultos, ainda dentro do caso IID, nenhuma das candidatas iniciais a covariáveis de confundimento do modelo de regressão linear normal teve poucas observações em categorias que não pudessem ser aglutinadas. Logo, teremos 13 covariáveis de confundimento a serem especificadas. Apresentamos na tabela 20, para o modelo de regressão linear normal para a resposta de interesse TSAT (denotado por  $m.dd.a$ ), as estimativas (pontual e erro-padrão) de seus coeficientes, bem como as estatísticas  $t$  e valores- $p$  correspondentes.

Pela tabela 20, temos que o modelo de regressão para TSAT no grupo dos adultos possui mais coeficientes significativos do que o modelo análogo (com algumas pequenas diferenças nas covariáveis especificadas) no grupo das crianças, fato consequente de uma amostra pareada bem maior para os adultos (2222 indivíduos). São significativos ao nível nominal de 5% os coeficientes associados ao intercepto; à LD; aos adultos que moram no litoral Leste do Iêmen (banhado pelo Mar Árabe); à PCD; ao gênero; à idade; a algumas faixas de escolaridade (2; 11 e 13 anos ou mais); às variáveis AMCD; DRFP; DRR e

Tabela 20 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo dos adultos, caso IID

Termo da Regr.	Est. Pontual	Erro-Pad.	Estat. $t$	Valor-p (%)
Intercepto	36.5153	1.8895	19.33	<0.01*
LD = “urbano”	2.7203	0.9312	2.92	0.35*
TD = “lit., Mar Ar.”	3.2285	1.4680	2.20	2.79*
TD = “pl. des.”	-0.2398	1.2173	-0.20	84.39
TD = “montan.”	-0.2763	1.2128	-0.23	81.98
PCD	-0.4452	0.1780	-2.50	1.24*
GA = “masc.”	12.3906	0.7492	16.54	<0.01*
IA	-0.0813	0.0305	-2.67	0.77*
EA = “1 ano”	-2.2760	4.9926	-0.46	64.85
EA = “2 anos”	4.7553	2.3356	2.04	4.18*
EA = “3 anos”	2.0106	1.8117	1.11	26.72
EA = “4 anos”	-1.3977	1.6778	-0.83	40.49
EA = “5 anos”	0.2045	1.7648	0.12	90.78
EA = “6 anos”	0.0911	1.2470	0.07	94.18
EA = “7 anos”	-0.3476	1.6944	-0.21	83.75
EA = “8 anos”	-1.0725	1.6344	-0.66	51.17
EA = “9 anos”	1.6400	1.5143	1.08	27.89
EA = “10 anos”	-2.6834	1.9493	-1.38	16.87
EA = “11 anos”	-4.2199	1.8866	-2.24	2.54*
EA = “12 anos”	-0.8715	1.1346	-0.77	44.25
EA = “13 anos +”	-4.5999	1.3360	-3.44	0.06*
EA = “alf. no Alc.”	-0.3411	2.1805	-0.16	87.57
AMCD = “sim”	10.7302	2.4709	4.34	<0.01*
AI = “sim”	-0.0497	1.6136	-0.03	97.54
API = “sim”	-0.6682	2.0212	-0.33	74.10
DRFP = “sim”	-3.7394	1.0625	-3.52	0.04*
DRC = “sim”	2.1391	1.1385	1.88	6.03
DRR = “sim”	-4.4292	0.8566	-5.17	<0.01*
NPD = “pob. mod.”	-3.8390	0.8884	-4.32	<0.01*
NPD = “vulner.”	-2.6084	0.9825	-2.65	0.80*
NPD = “não-pob.”	-2.9699	0.8702	-3.41	0.06*
RSWF	-1.0721	0.8562	-1.25	21.06
RC	-2.3103	0.8459	-2.73	0.63*
RSWF:RC	1.7228	1.1937	1.44	14.90

aos três coeficientes para as quatro categorias de NPD (adotando a categoria referente aos domicílios extremamente pobres como referência do modelo). Por fim, também é significativo o coeficiente associado à variável RC.

Trabalham significativamente menos, ao nível nominal de 5%, os adultos que convivem com mais pessoas (cerca de 27 minutos a cada variação unitária positiva em PCD); mais velhos (em torno de 5 minutos a cada variação unitária positiva em IA); e que possuem elevada escolaridade (aproximadamente 4 horas e 13 minutos para quem tem 11

anos e 4 horas e 36 minutos para quem tem 13 anos ou mais de escolaridade), comparados aos adultos que nunca estudaram (categoria de referência para EA). Ainda, residir em domicílios com menores níveis de pobreza (comparados aos que são extremamente pobres, a categoria de referência para NPD); que recebem fundos de pensão ou remessas de parentes ou amigos também são características que influenciam negativamente e de forma significativa em TSAT (isoladamente, pelo menos 2 horas a menos em cada uma destas características).

Por sua vez, trabalham significativamente mais os adultos residentes na zona urbana (em torno de 2 horas e 43 minutos), comparados aos residentes na zona rural; do gênero masculino (cerca de 12 horas e 23 minutos); residentes no litoral Leste, banhado pelo Mar Árabe (aproximadamente 3 horas e 13 minutos), comparados aos que moram no litoral Oeste, banhado pelo Mar Vermelho; com apenas 2 anos de escolaridade (em torno de 4 horas e 45 minutos, comparados aos que nunca estudaram); e que são mulheres e chefes do domicílio (cerca de 10 horas e 44 minutos). Com exceção deste último resultado, todos os demais eram previstos dado o perfil demográfico e socioeconômico do Iêmen.

Finalmente, o período de coleta das informações, expresso na variável RC, influenciou de forma significativa em TSAT, com adultos trabalhando menos na 4ª rodada em relação à 1ª, uma provável consequência da crise decorrente dos eventos da Primavera Árabe em 2011. Por outro lado, a atribuição do tratamento (RSWF) e a sua interação com o período de coleta (RSWF:RC) não foram significativas. Em termos de sinal, seus resultados são similares aos obtidos para o grupo das crianças, com RSWF influenciando negativamente e a interação, que fornece o impacto do SWF, positivamente em TSAT. Logo, este último resultado contraria a existência de um suposto “efeito preguiça” do programa (impacto negativo) ao longo do tempo nos adultos, ao menos sob a suposição de observações IID.

A partir de agora, iremos reajustar modelos para a resposta de interesse TSAT tomando os bancos pareados, em cada um dos grupos etários definidos, na abordagem em que o PAC utilizado para extrair a amostra da YNSPMS é considerado (comandos em R na subseção B.2.6 do apêndice). Desta forma, a estimação dos coeficientes será feita por MP (e não mais por MV), de modo a incorporar os pesos amostrais de cada indivíduo. Assim, também será exposto o EPAA estimado para cada estimativa dos coeficientes de cada regressão, além dos resultados da estimação pontual; do erro padrão; das estatísticas  $t$  e do valor- $p$  associado. Agora, os graus de liberdade para a estatística  $t$  associada a cada coeficiente são dados pelo nº de UPA's menos o de estratos no 1º estágio, menos 1. Por outro lado, as candidatas a covariáveis de confundimento, dado o grupo etário, são as mesmas do caso IID. Começando pelo banco pareado no grupo das crianças (com 234 indivíduos, logo distinto e ligeiramente menor que o banco usado sob a abordagem IID), excluimos das candidatas iniciais a covariáveis de confundimento do modelo de regressão

linear normal as mesmas variáveis retiradas da especificação no caso IID (CRDCM; CPI e LD), permanecendo as outras 14 variáveis. No banco pareado em que consideramos o PAC, elas também têm raras observações (cinco ou menos) em categorias que não podem ser reunidas a outras da mesma variável. Deste modelo para a resposta de interesse TSAT (denotado por  $m.dd.cp$ ), apresentamos os resultados da estimação na tabela 21.

Tabela 21 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo das crianças, considerando o PAC da YNSPMS

Termo da Regr.	Est. Pontual	Erro-Pad.	EPAA	Estat. $t$	Valor-p (%)
Intercepto	28.5419	8.6424	1.95	3.30	0.14*
TD = “pl. des.”	4.4970	3.4837	1.75	1.29	20.05
TD = “montan.”	4.5795	2.6011	1.29	1.76	8.22
PCD	-0.4221	0.5786	2.06	-0.73	46.79
GC = “masc.”	-0.7274	2.3350	2.18	-0.31	75.62
IC	-0.2173	0.6499	2.15	-0.33	73.90
MCPE = “sim, an. m.”	5.0563	4.3288	1.74	1.17	24.63
MCPE = “sim, at. m.”	-4.7084	3.3676	1.17	-1.40	16.60
EC = “1 ano”	-1.7686	4.3633	1.43	-0.41	68.63
EC = “2 anos”	-6.3677	2.9225	0.82	-2.18	3.23*
EC = “3 anos”	-11.0790	3.7715	1.29	-2.94	0.43*
EC = “4 anos”	-2.0662	3.8455	1.22	-0.54	59.26
EC = “5 anos”	-5.4274	2.8383	0.62	-1.91	5.95
EC = “6 anos”	3.2558	2.7261	0.42	1.19	23.59
EC = “7 anos +”	6.4236	4.4487	1.15	1.44	15.27
EMC = “1 a 4 anos”	-11.9895	8.2994	4.68	-1.44	15.25
EMC = “5 a 8 anos”	-28.5428	4.8748	1.58	-5.86	<0.01*
EMC = “9 anos +”	5.6971	7.7172	2.84	0.74	46.26
EMC = “ausente”	2.9012	5.9643	2.85	0.49	62.80
EPC = “1 a 4 anos”	-11.8568	4.1368	2.33	-2.87	0.53*
EPC = “5 a 8 anos”	-0.5098	3.0074	1.71	-0.17	86.58
EPC = “9 anos +”	1.9959	3.4617	2.61	0.58	56.59
EPC = “ausente”	-10.1504	4.6244	2.59	-2.20	3.11*
CRDAI = “sim”	0.5719	3.6470	1.28	0.16	87.58
CRDAPI = “sim”	2.5835	5.0703	2.52	0.51	61.18
DRFP = “sim”	23.6002	5.2411	2.04	4.50	<0.01*
DRC = “sim”	7.0094	3.1093	1.65	2.25	2.69*
DRR = “sim”	1.3973	4.3242	3.99	0.32	74.75
NPD = “pob. mod.”	-7.9620	2.4660	1.43	-3.23	0.18*
NPD = “vulner.”	-10.5384	4.8627	3.73	-2.17	3.32*
NPD = “não-pob.”	-2.3089	4.1175	3.18	-0.56	57.66
RSWF	2.9277	3.2887	2.60	0.89	37.61
RC	6.4304	3.6286	3.58	1.77	8.02
RSWF:RC	0.7454	5.3653	4.01	0.14	88.99

Da tabela 21, podemos concluir que o EPAA estimado é superior ao valor unitário exceto para alguns valores específicos de escolaridade da criança (2; 5 e 6 anos) e é menor

do que 5 para todos os coeficientes. Logo, o PAC da YNSPMS, mesmo considerando a estratificação apenas no 1º estágio, não produziu erros-padrão muito elevados na estimação dos coeficientes do modelo de regressão especificado para avaliação de impacto do SWF em TSAT. Inclusive, o PAC foi mais eficiente do que a suposta adoção de um plano AAS para alguns dos coeficientes. Ressaltemos novamente que, assim como feito para os desvios de cada variável no modelo de regressão logística para o escore de propensão, a estimação do EPAA toma a razão entre a variância estimada (o quadrado do erro padrão) para a estimativa pontual considerando o PAC e a variância estimada para a estimativa pontual o mesmo modelo e com os mesmos dados (o banco pareado sob a existência de dados amostrais complexos), mas ignorando o PAC. Observe que esta última difere da variância estimada pela tabela 19, pois o modelo referente a esta foi ajustado para um banco pareado distinto, que ignorava o PAC desde a etapa de estimação dos escores de propensão.

Com relação às estimativas pontuais, há várias diferenças em significância, magnitude e sinal quando os resultados com PAC são comparados aos obtidos sem considerá-lo desde o princípio (tabela 19). Por exemplo, os coeficientes associados às crianças que residem em regiões desérticas; estão atualmente matriculadas em algum programa educacional e possuem 4 ou 6 anos de escolaridade não são mais significativos ao nível nominal de 5%. Por outro lado, os coeficientes associados às crianças que possuem 2 ou 3 anos de escolaridade; cujas mães têm de 5 a 8 anos de escolaridade; cujos pais estão ausentes do domicílio (sem informação sobre a escolaridade dele); residentes em domicílios que recebem fundos de pensão ou renda através de caridade e em domicílios de pobreza moderada ou vulneráveis passaram a ser significativos ao incorporar o PAC. Os coeficientes associados ao intercepto e às crianças cujos pais têm de 1 a 4 anos de escolaridade são significativos em ambos os modelos.

Dentre os coeficientes significativos da tabela 21, houve mudanças de sinal em relação à tabela 19 para as estimativas pontuais daqueles associados às crianças que possuem 2 ou 3 anos de escolaridade e cujas mães têm de 5 a 8 anos de escolaridade (de positivo para negativo) e para o coeficiente associado às crianças residentes em domicílios que recebem renda através de caridade (de negativo para positivo). Assim, incorporar o PAC da YNSPMS na modelagem para avaliação de impacto do SWF em TSAT para o grupo das crianças produziu uma análise bem distinta da realizada no caso IID, dadas as mesmas categorias de referência em cada variável.

Para começar, na maioria das faixas de escolaridade, não apenas as referentes às crianças com 2 ou 3 anos de estudo (ainda que somente estas tenham sido significativas), a influência da escolaridade da criança em TSAT muda de sentido, passando a ser negativa para as faixas mais baixas (de 1 a 5 anos) e positiva para as mais altas (7 anos ou mais). Ou seja, trabalham mais as crianças que nunca frequentaram a escola (categoria de referência para EC) ou que possuem pelo menos seis anos de escolaridade. O coeficiente associado às

crianças cujas mães têm de 5 a 8 anos de escolaridade também passa a ser significativo e a estimativa pontual associada, negativa e bem maior em magnitude. Logo, crianças com esta característica trabalham significativamente menos do que aquelas cuja mãe nunca frequentou a escola (cerca de 28 horas e 33 minutos).

Por outro lado, para crianças cuja mãe tem pelo menos 9 anos de escolaridade, a influência ocorre no sentido contrário, sendo positiva, mas não significativa. Finalmente, residir em um domicílio que recebe renda através de caridade influencia em maior magnitude, positivamente (não mais negativamente) e de forma significativa no tempo alocado semanalmente para o trabalho pela criança (em torno de 7 horas e 1 minuto). Mesmo que o domicílio tenha fontes adicionais de renda, não apenas da caridade, mas também advinda dos fundos de pensão ou das remessas de parentes ou amigos, isto não é suficiente para que o valor de TSAT seja menor, em média, para a criança (pelo contrário). Este resultado difere do obtido caso o PAC fosse completamente ignorado (exceto para o coeficiente associado às crianças residentes em domicílios beneficiários de fundos de pensão, também positivo no caso IID).

Para os coeficientes significativos nos quais não há mudança de sinal, mas apenas em magnitude, das estimativas pontuais correspondentes na abordagem que incorpora o PAC, temos os seguintes resultados. O intercepto do modelo continua significativo, positivo e elevado em magnitude, embora ligeiramente menor (estimado aproximadamente em 28 horas e 33 minutos). As crianças cujos pais estão ausentes ou, quando presentes, têm de 1 a 4 anos de escolaridade, trabalham significativamente menos do que aquelas cujo pai nunca estudou (a categoria de referência para EPC), com estimativas pontuais dos coeficientes associados bem maiores em magnitude com relação ao caso IID (somente a segunda característica era significativa no caso IID). Por fim, residir em domicílios em situação de pobreza moderada ou vulnerabilidade passou a ser um fator significativo, com coeficientes ainda negativos, mas bem maiores em magnitude. Logo, as crianças residentes nestes domicílios trabalham significativamente menos com relação àquelas que residem em domicílios extremamente pobres (variando em torno de 2 horas e 19 minutos a 10 horas e 32 minutos).

Quanto aos coeficientes associados às covariáveis RSWF; RC e à interação entre ambas (que fornece o impacto da política), todos são positivos, porém não significativos. Exceto pela mudança de sinal no coeficiente associado à RSWF e por pequenas variações em magnitude, os resultados são semelhantes aos obtidos no caso IID. Como o coeficiente associado à interação não é significativo, não podemos dizer que o SWF tem qualquer impacto no valor de TSAT para as crianças. Logo, o SWF não contribui para a redução ou aumento na jornada semanal de trabalho da criança residente no domicílio que o recebe, tanto para o caso IID quanto para o caso em que o PAC é considerado.

Resta somente a avaliação de impacto para o SWF em TSAT, considerando o

PAC desde a etapa de modelagem do escore de propensão, no grupo dos adultos. Assim, para o banco pareado correspondente (com 2332 indivíduos, dez a mais do que no caso IID), nenhuma das candidatas iniciais a covariáveis de confundimento do modelo de regressão linear normal para TSAT teve quantidade muito reduzida de observações em categorias que não pudessem ser reunidas. Logo, teremos 13 covariáveis de confundimento para a especificação deste modelo (denotado por *m.dd.ap*), para o qual apresentamos as estimativas (pontual e erro-padrão) dos seus coeficientes; o EPAA associado, calculado por procedimento análogo ao usado para o EPAA apresentado na tabela 21, e as estatísticas *t* correspondentes com seus valores-p na tabela 22.

Pela tabela 22, o EPAA estimado é superior ao valor unitário, exceto para o coeficiente associado aos adultos que possuem 2 anos de escolaridade, e seus valores são em geral maiores do que os mostrados na tabela 21 para as covariáveis análogas no modelo do grupo das crianças. Como o banco pareado para os adultos é bem maior, a influência da conglomeração e do uso de pesos amostrais desiguais do PAC da YNSPMS nas variâncias estimadas é mais evidente para este grupo etário. Assim como verificado no grupo das crianças, há muitas diferenças na significância, magnitude e sinal das estimativas pontuais entre as tabelas 20 (ignorando completamente o PAC) e 22 (incorporando-o desde o início da modelagem do escore de propensão).

Não houve nenhum coeficiente do modelo de regressão linear que não foi significativo ao nível nominal de 5% no caso IID que passou a ter significância na abordagem em que o PAC é incorporado. Por outro lado, muitos coeficientes significativos da tabela 20 deixaram de sê-lo ao tratar as observações como dados amostrais complexos. Como exemplos, temos o coeficiente associado aos adultos residentes no litoral Leste; à PCD; à IA; aos adultos que possuem 11 anos de escolaridade; residentes em domicílios que recebem fundos de pensão ou vulneráveis ou não-pobres. Por fim, o coeficiente associado à RC também deixou de ser significativo.

Olhando agora para os coeficientes que também são significativos na abordagem em que o PAC é considerado desde o princípio, somente a estimativa pontual para o coeficiente associado aos adultos que possuem 2 anos de escolaridade tem sinal distinto da obtida no caso IID, passando a ser negativa (embora a magnitude seja praticamente a mesma). Adultos com este nível de escolaridade trabalham significativamente menos do que quem não possui nenhum ano de estudo (cerca de 4 horas e 48 minutos a menos). Para os demais coeficientes significativos, mudanças são observadas apenas na magnitude das estimativas pontuais, em maior ou menor grau, quando comparadas às expostas na tabela 20.

Como resultado, temos que a estimativa pontual para o coeficiente associado ao intercepto é positiva (aproximadamente 30 horas e 57 minutos, quase 6 horas a menos do que no caso IID) para os adultos nas categorias de referência do modelo (em todas as covariáveis). Além disso, trabalham significativamente mais os adultos residentes na zona

Tabela 22 – Estimativas para os coeficientes no modelo de regressão linear normal para TSAT no grupo dos adultos, considerando o PAC da YNSPMS

Termo da Regr.	Est. Pontual	Erro-Pad.	EPAA	Estat. $t$	Valor-p (%)
Intercepto	30.9573	4.3233	5.21	7.16	<0.01*
LD = “urbano”	7.1081	2.5382	7.65	2.80	0.53*
TD = “lit., Mar Ar.”	0.9308	3.6112	5.67	0.26	79.67
TD = “pl. des.”	-0.0959	2.4713	3.97	-0.04	96.91
TD = “montan.”	4.0998	2.6993	4.78	1.52	12.95
PCD	-0.2295	0.3812	4.87	-0.60	54.75
GA = “masc.”	15.8230	1.9998	7.31	7.91	<0.01*
IA	0.0240	0.0667	5.11	0.36	71.95
EA = “1 ano”	-0.5629	5.8809	2.45	-0.10	92.38
EA = “2 anos”	-4.8071	2.1244	0.86	-2.26	2.41*
EA = “3 anos”	0.4513	4.9526	6.24	0.09	92.74
EA = “4 anos”	-1.9278	4.3656	8.19	-0.44	65.90
EA = “5 anos”	3.2715	4.2596	5.59	0.77	44.29
EA = “6 anos”	5.1336	3.8266	9.91	1.34	18.04
EA = “7 anos”	-2.6268	3.6809	4.49	-0.71	47.58
EA = “8 anos”	9.6348	8.4443	28.30	1.14	25.45
EA = “9 anos”	-0.7555	3.6264	5.47	-0.21	83.51
EA = “10 anos”	-2.6296	3.7789	3.83	-0.70	48.69
EA = “11 anos”	-4.0309	6.4743	10.77	-0.62	53.39
EA = “12 anos”	-1.8684	2.9821	7.03	-0.63	53.13
EA = “13 anos +”	-7.0137	2.7004	4.84	-2.60	0.97*
EA = “alf. no Alc.”	2.7986	3.4994	2.51	0.80	42.43
AMCD = “sim”	12.0953	3.4426	1.68	3.51	0.05*
AI = “sim”	-6.7982	3.5026	5.12	-1.94	5.29
API = “sim”	-4.3704	2.7982	1.83	-1.56	11.90
DRFP = “sim”	-3.0373	2.1204	3.75	-1.43	15.27
DRC = “sim”	0.9541	1.5887	2.12	0.60	54.84
DRR = “sim”	-6.2977	1.7677	4.46	-3.56	0.04*
NPD = “pob. mod.”	-8.6149	2.7553	9.94	-3.13	0.19*
NPD = “vulner.”	-5.0182	2.8369	8.30	-1.77	7.76
NPD = “não-pob.”	-3.8702	3.1566	13.22	-1.23	22.08
RSWF	-1.4680	2.7103	10.35	-0.54	58.83
RC	-4.0447	2.7497	10.81	-1.47	14.20
RSWF:RC	-0.2190	4.5096	14.66	-0.05	96.13

urbana (em torno de 7 horas e 6 minutos), comparados aos residentes na zona rural, e do gênero masculino (aproximadamente 15 horas e 49 minutos a mais do que as mulheres). Por sua vez, trabalham significativamente menos os adultos que possuem 13 anos ou mais de escolaridade (cerca de 7 horas e 1 minuto), comparados aos que nunca estudaram; que são mulheres e chefes do domicílio (aproximadamente 12 horas e 6 minutos); e residentes em domicílios que recebem remessas (em torno de 6 horas e 18 minutos) ou em situação de pobreza moderada (em torno de 8 horas e 37 minutos), comparados aos que estão em

situação de extrema pobreza.

Finalmente, os coeficientes associados às covariáveis RSWF; RC e à interação entre ambas, embora todas as estimativas pontuais correspondentes sejam negativas, não são significativos ao nível nominal de 5%. Logo, também no grupo dos adultos, não se pode dizer que o SWF teve impacto negativo em TSAT quando o PAC é considerado. De forma similar, nada se pode concluir para os efeitos isolados da atribuição do tratamento (RSWF) e do período de coleta das informações (RC), ao contrário da análise feita no caso IID, quando o coeficiente associado à RC foi significativo. Portanto, ao considerar o PAC efetivamente utilizado na YNSPMS, foi possível perceber dentre as potenciais variáveis de confundimento quais delas de fato influenciavam significativamente na resposta de interesse (TSAT) para os adultos. Mais importante, foi possível confirmar se a política sob estudo (o SWF) teve algum impacto nesta mesma resposta ao longo do tempo, o que não ocorreu, assim como na abordagem supondo observações IID.

Comparando as conclusões da avaliação de impacto do SWF em TSAT quando consideramos o PAC da YNSPMS na estimação dos coeficientes com o que já foi encontrado na literatura e descrito na seção 1.3 para outras políticas de transferência direta de renda, em cada grupo etário, há mais diferenças do que semelhanças. No grupo das crianças, a estimativa do coeficiente associado à interação é positiva, mas não é significativa ao nível nominal de 5% (não há impacto), divergindo dos resultados encontrados por Skoufias & Parker (2001) e Cacciamali *et al.* (2010) para a participação da força de trabalho e por Ferro & Kasouf (2005) para a jornada semanal de trabalho (equivalente à variável TSAT). Por outro lado, no grupo dos adultos, a estimativa do coeficiente associado à interação é negativa, porém também não é significativa, indo ao encontro apenas do resultado encontrado por Skoufias & di Maro (2006).

A avaliação de impacto do SWF em TSAT indica, portanto, que o programa de fato não contribui para um suposto “efeito preguiça” nos adultos, mas está longe de reduzir o tempo semanal alocado para o trabalho pelas crianças que ainda trabalham. De fato, o próprio relatório final da YNSPMS aconselha ao governo iemenita uma possível revisão nos critérios de elegibilidade do SWF, de modo que este favoreça a entrada de domicílios extremamente pobres e em que residam mais crianças e exija contrapartidas do domicílio beneficiário, as chamadas condicionalidades, existentes nos demais programas de transferência direta de renda mencionados (YNSPMS Final Report, 2014a).

Caso ignorássemos o PAC da YNSPMS bem como a EP da amostra selecionada desde a etapa de modelagem do score de propensão, a conclusão que encontramos para a estimativa que fornece o impacto, pelo menos em termos de sinal, já não seria a mesma. Para as estimativas dos coeficientes associados às covariáveis de confundimento, em cada modelo, haveria ainda mais diferenças em termos de significância; sinal e magnitude, especialmente no grupo das crianças. Logo, incorporar as características do PAC permitiu

---

o cálculo de estimativas com o nível verdadeiro de precisão fornecido pela amostra, em geral menor do que o obtido quando se supõe, erroneamente, que os dados amostrais são IID.

## 5 Considerações Finais

O presente trabalho teve como objetivo principal analisar a influência de um PAC na avaliação do impacto de uma política de transferência direta de renda (o SWF) na variável TSAT (a resposta de interesse, suposta aproximadamente contínua e normal). A análise foi feita tomando os indivíduos entrevistados nas quatro rodadas da pesquisa YNSPMS, os quais foram divididos em dois grupos ou subpopulações (crianças e adultos). Para a avaliação de impacto, foi utilizado o método DD, de modo que consideramos somente as observações dos indivíduos que responderam à TSAT na 1<sup>a</sup> e 4<sup>a</sup> rodadas, sob duas abordagens para estimar os coeficientes da regressão linear normal correspondente a cada grupo etário. Na primeira delas, tratamos todas as observações amostrais como IID, ignorando completamente o PAC utilizado na YNSPMS. Por outro lado, na segunda abordagem o incorporamos à estimação, através dos pesos amostrais associados a cada indivíduo pesquisado. Antes da avaliação de impacto, para cada grupo etário pareamos os indivíduos tratados (aqueles que passaram a receber o SWF a partir da 2<sup>a</sup> rodada) com os controles (aqueles que não recebiam o benefício em nenhuma das quatro entrevistas).

O pareamento dos indivíduos tratados e controles foi feito através do escore de propensão para a variável de atribuição do tratamento (RSWF). Para estimar o escore, foi ajustado um modelo de regressão logística para todos os indivíduos de um dado grupo etário que pudessem ser classificados como tratados ou controles. Assim, desde a etapa de modelagem dos escores, a análise foi dividida nas duas abordagens citadas anteriormente. O uso do modelo de regressão logística não é obrigatório, ainda que seja o de uso mais corrente na literatura sobre modelagem do escore de propensão. Em cada abordagem, a especificação do modelo de regressão logística considerou apenas covariáveis que influenciassem tanto na atribuição do tratamento quanto na resposta de interesse para avaliação do impacto (TSAT).

Desta forma, foram especificadas como covariáveis apenas aquelas relacionadas diretamente aos critérios de elegibilidade do SWF ou que fossem pelo menos fortemente correlacionadas. Ainda, covariáveis cujos desvios não fossem significativos ao nível nominal de 5% pelo teste RV (ou RP, quando o PAC é considerado) eram eliminadas da especificação. Neste caso, um novo ajuste era feito apenas com as covariáveis remanescentes, até que fosse encontrado um modelo em que todas as covariáveis fossem significativas. Como todos os escores preditos devem estar estritamente entre 0 e 1, é necessário encontrar o equilíbrio entre a informação fornecida pelo modelo e a sua parcimônia. Este equilíbrio é atingido com mais facilidade quanto maior for o tamanho amostral. Por esta razão, modelos ajustados para os adultos possuíam mais covariáveis especificadas, independentemente da abordagem.

Após estimar os escores de propensão para todos os indivíduos, dado o grupo etário da amostra, tomamos todos os tratados e selecionamos controles que sejam o mais semelhantes possível com base nas observações das covariáveis ou do escore predito. Assim, tais controles são vistos como os contrafatuais do indivíduo tratado. Neste trabalho, utilizamos o PDE com reposição; sem sobreamostragem e impondo um nível máximo de tolerância para pareamento dos indivíduos controles com tratados. O indivíduo tratado que não pudesse ser pareado com nenhum controle dada a restrição do nível de tolerância era eliminado do banco pareado a ser formado. Constituído o banco com os pares de tratados e controles, são realizados testes de balanceamento para cada covariável especificada no modelo que gerou os escores ( $\chi^2$  de Pearson para independência) e para os escores preditos (teste MW para homogeneidade). Tais testes também são adaptados para a abordagem que incorpora o PAC.

A finalidade de tais testes é garantir que não há evidência para dizer que os indivíduos tratados e controles do banco a ser utilizado na avaliação de impacto não são similares. Ou seja, que o banco pareado é também balanceado dadas as classes formadas por tratados e controles. Caso um dos testes de balanceamento seja significativo para as observações de alguma covariável ou para os escores preditos, voltamos à etapa de modelagem do escore de propensão. Agora, sem especificar uma dada covariável (se o teste  $\chi^2$  de Pearson foi significativo) ou especificando o modelo encaixado com a maior taxa de acerto na previsão da classificação de tratados e controles (se somente o teste MW foi significativo). Isto é feito até que encontremos um modelo em que o pareamento de tratados e controles feito através dos escores preditos forneça um banco pareado que esteja balanceado dada qualquer covariável especificada e dados os escores preditos. Por esta razão, quando consideramos o PAC, a rejeição dos testes de balanceamento (adaptados para dados amostrais complexos) ocorria com frequência, já que indivíduos pareados não necessariamente têm pesos amostrais iguais. Ainda assim, ao refazer o processo de modelagem dos escores de propensão, foram encontradas especificações que cumprissem com a exigência de balanceamento nas covariáveis especificadas e nos escores preditos. Evidentemente, tais modelos eram menos informativos do que no caso IID.

Com a garantia de um banco pareado e ao mesmo tempo balanceado nas classes de indivíduos tratados e controles, finalmente prosseguimos para a avaliação de impacto do SWF em TSAT pelo método DD. Esta é feita através da especificação de um modelo de regressão linear normal contendo indicadores da atribuição do tratamento (RSWF); da rodada de coleta das observações (RC) e da interação entre ambas (RSWF:RC). A estimativa do coeficiente associado à interação fornece o impacto do programa em TSAT, dado o grupo etário e a abordagem utilizada. Ainda, também são especificadas como covariáveis qualquer potencial variável de confundimento para TSAT, incluindo as que foram candidatas à especificação do modelo para o escore de propensão. Logo, ajustamos no total quatro modelos para avaliação de impacto, tabelando os resultados da estimação

(incluindo o EPAA na abordagem em que incorporamos o PAC).

Apesar de uma especificação bem extensa das potenciais covariáveis de confundimento, em acordo com o que a literatura sugere, para os quatro modelos ajustados não foi encontrado impacto significativo do SWF na média de TSAT nos adultos nem nas crianças. Portanto, ainda que seja uma programa de transferência direta de renda incondicional, o SWF não contribui para estimular um suposto “efeito preguiça” nos adultos em atividade laboral e que são beneficiários da política. Por outro lado, o programa também não contribui para reduzir o trabalho infantil nas crianças que ainda exercem atividade laboral. É possível que isto ocorra devido ao componente da ambição familiar, também observado em outros resultados da literatura, e do fato de que o Iêmen passou por uma crise interna durante a realização da YNSPMS. Por estas razões, os responsáveis pelas crianças são mais relutantes em dispensar a participação destas na força de trabalho e conseqüentemente do ganho adicional gerado pela mão-de-obra infantil na renda familiar.

Embora não haja diferenças para a avaliação de impacto do SWF dada qualquer abordagem, do ponto de vista metodológico, considerar o PAC na obtenção das estimativas dos coeficientes (através da MP, e não da tradicional MV) forneceu uma estimativa menos enviesada do erro padrão verdadeiro associado a cada coeficiente. Na abordagem em que as observações eram tratadas como IID, em geral o erro padrão era subestimado (menor do que o real), induzindo a conclusões incorretas sobre a significância de determinada covariável ou coeficiente dado o modelo. Como consequência, o EPAA estimado era quase sempre superior ao valor unitário. Assim, adotar tal abordagem em detrimento da existência de um PAC e de uma EP como geradores da amostra utilizada na avaliação de impacto acarreta na escolha de modelos com menos parcimônia, ou seja, com mais parâmetros do que de fato deveriam ter. Neste estudo, isto ficou evidente tanto na modelagem dos escores quanto na avaliação de impacto do SWF em TSAT. Quando o PAC era ignorado, mais variáveis apresentavam significância, independentemente do grupo etário sob estudo. Por esta razão, é possível concluir que o PAC de uma pesquisa longitudinal possui informações importantes, também quando se avalia o impacto de uma política, sobre a incerteza de qualquer estimativa para um parâmetro populacional a ser investigado. Logo, não é recomendável ignorá-lo.

Não obstante, a avaliação de impacto feita neste estudo pode ser estendida de várias formas. A primeira delas, mais clara, é através do uso de outras funções de distribuição para ligação entre o preditor linear e a resposta binária para a atribuição do tratamento, além da logística e da normal. Além disso, o pareamento dos escores preditos pode ser melhorado através de um critério que também leve em consideração a proximidade com relação às características do PAC. Por exemplo, pareando indivíduos com pesos amostrais similares ou que pertençam ao mesmo estrato no 1º estágio. Além disso, outros testes para verificação do balanceamento também poderiam ser utilizados, muito embora os

---

resultados praticamente não diverjam com relação ao valor-p quando a amostra pareada é grande. Para o teste MW, Lumley & Scott (2013) citam extensões já existentes e como elas poderiam ser adaptadas para dados amostrais complexos. Por fim, é suficiente que a resposta de interesse para a qual avaliaremos o impacto da política seja pelo menos aproximadamente contínua na reta real. No momento, apenas a distribuição normal (com a função de ligação identidade ou outras usualmente empregadas) está disponível no pacote `survey` do R e na maioria das bibliotecas análogas dos demais *software* estatísticos para análises envolvendo dados amostrais complexos. Entretanto, qualquer distribuição com suporte contínuo real pode ser utilizada para modelar a resposta de interesse.

## Referências

- ABNTEX2. *AbnTeX2: Documentos técnicos e científicos em L<sup>A</sup>T<sub>E</sub>X*. 2016. Disponível em: <<http://www.abntex.net.br/>>. Acesso em: 20 de março de 2016. Citado na página 46.
- Akaike, H. Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*. Budapest: [s.n.], 1973. p. 267–281. Citado na página 64.
- Anderson, T. W. On the distribution of the two-sample Cramér–von Mises criterion. *Annals of Mathematical Statistics*, v. 33, n. 3, p. 1148–1159, 1962. Citado na página 82.
- Ashenfelter, O. Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, v. 60, n. 1, p. 47–57, 1978. Citado na página 89.
- Athey, S. e Imbens, G. W. Identification and inference in nonlinear difference-in-difference models. *Econometrica*, v. 74, n. 2, p. 431–497, 2006. Citado na página 92.
- Austin, P. C. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, v. 10, p. 150–161, 2011. Citado 2 vezes nas páginas 75 e 77.
- Bakri, N. e Goodman, J. D. *Thousands in Yemen Protest Against the Government*. 2011. Disponível em: <[http://www.nytimes.com/2011/01/28/world/middleeast/28yemen.html?\\_r=0](http://www.nytimes.com/2011/01/28/world/middleeast/28yemen.html?_r=0)>. Acesso em: 7 de março de 2016. Citado na página 40.
- Baser, O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value in Health*, v. 9, n. 6, p. 377–385, 2006. Citado na página 77.
- Beaumont, J. F. e Bocci, C. A practical bootstrap method for testing hypotheses from survey data. *Survey Methodology*, v. 35, p. 25–35, 2009. Citado na página 63.
- Berry, K. J.; Mielke, P. W e Johnston, J. E. The two-sample rank-sum test: Early development. *Electronic Journal for History of Probability and Statistics*, v. 8, p. 1–26, 2012. Citado na página 84.
- Binder, D. A. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, v. 51, n. 3, p. 279–292, 1983. Citado na página 28.
- Blair, R. C. e Higgins, J. J. A comparison of the power of Wilcoxon’s rank-sum statistic to that of Student’s *t* statistic under various nonnormal distributions. *Journal of Educational Statistics*, v. 5, n. 4, p. 309–335, 1980. Citado na página 77.
- Blundell, R. e Dias, M. C. Evaluation Methods for Non-Experimental Data. *Fiscal Studies*, v. 21, n. 4, p. 427–468, 2000. Citado na página 96.
- Bowley, A. L. Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science. *Journal of the Royal Statistical Society*, v. 69, p. 548–557, 1906. Citado na página 26.

- Breslow, N. E. e Clayton, D. G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, v. 88, n. 421, p. 9–25, 1993. Citado na página 54.
- Cacciamali, M. C.; Tatei, F. e Batista, N. F. Impactos do Programa Bolsa Família federal sobre o trabalho infantil e a frequência escolar. *Revista de Economia Contemporânea*, v. 14, n. 2, p. 269–301, 2010. Citado 2 vezes nas páginas 34 e 129.
- Caliendo, M. e Kopeinig, S. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, v. 22, n. 1, p. 31–72, 2008. Citado 5 vezes nas páginas 53, 59, 60, 74 e 75.
- Card, D. The impact of the *Mariel* boatlift on the Miami labor market. *Industrial and Labor Relations Review*, v. 33, p. 245–257, 1990. Citado na página 89.
- CIA. *The World Factbook: Yemen*. 2016. Disponível em: <<https://www.cia.gov/library/publications/the-world-factbook/geos/ym.html>>. Acesso em: 4 de março de 2016. Citado 2 vezes nas páginas 39 e 40.
- Claeskens, G. e Hjort, N. L. *Model Selection and Model Averaging*. Cambridge: Cambridge University Press, 2008. Citado 2 vezes nas páginas 57 e 64.
- Cochran, W. G. *Sampling Techniques*. 3. ed. New York: John Wiley & Sons, 1977. Citado 5 vezes nas páginas 20, 27, 64, 69 e 146.
- Cochran, W. G. e Chambers, S. P. The planning of observational studies of human populations. *Journal of the Royal Statistical Society, Series A*, v. 128, n. 2, p. 234–266, 1965. Citado na página 73.
- Cohen, J. E. The distribution of the chi-squared statistic under cluster sampling from contingency tables. *Journal of American Statistician Association*, v. 71, n. 355, p. 665–670, 1976. Citado na página 79.
- Cordeiro, G. M. e Paula, G. A. Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika*, v. 76, n. 1, p. 93–100, 1989. Citado na página 54.
- Corder, G. W. e Foreman, D. I. *Nonparametric Statistics for Non-Statisticians*. Hoboken: John Wiley & Sons, 2009. Citado 2 vezes nas páginas 83 e 84.
- Cox, D. R. e Hinkley, D. V. *Theoretical Statistics*. London: Chapman & Hall, 1974. Citado na página 54.
- Dahl, D. B. *Package 'xtable': Export tables to L<sup>A</sup>T<sub>E</sub>X or HTML*. [S.l.], 2016. Disponível em: <<http://CRAN.R-project.org/package=xtable>>. Acesso em: 26 de abril de 2016. Citado na página 150.
- Davies, R. B. Algorithm AS 155: the distribution of a linear combination of  $\chi^2$  Random Variables. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, v. 29, n. 3, p. 323–333, 1980. Citado na página 62.
- Deming, W. E. *Some Theory of Sampling*. New York: John Wiley & Sons, 1950. Citado na página 26.

- Draper, N. R. e Smith, H. *Applied Regression Analysis*. 3. ed. New York: John Wiley & Sons, 1998. Citado na página 100.
- Dudewicz, E. J.; Ma, Y.; Mai, E. e Su, H. Exact solutions to the Behrens–Fisher Problem: Asymptotically optimal and finite sample efficient choice among. *Biometrika*, v. 137, p. 1584–1605, 2007. Citado na página 76.
- Farebrother, R. W. Algorithm AS 204: the distribution of a positive linear combination of  $\chi^2$  random variables. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, v. 33, n. 3, p. 332–339, 1984. Citado na página 62.
- Faria, A. L. C. Aplicação do teste de elegibilidade multidimensional na definição do público-alvo beneficiário de políticas públicas. *Revista de Economia Contemporânea*, v. 11, n. 2, p. 253–278, 2007. Citado na página 42.
- Fellegi, I. P. Approximate tests of independence and goodness of fit based on stratified multistage samples. *Journal of the American Statistical Association*, v. 75, n. 370, p. 261–268, 1980. Citado na página 79.
- Ferraz, C. e Vieira, M. D. T. Sample design for impact evaluation of welfare programs: The Yemen case. In: *In Proceedings 59th ISI World Statistics Congress*. Hong Kong: [s.n.], 2013. Citado 3 vezes nas páginas 11, 38 e 43.
- Ferraz, C. e Vieira, M. D. T. *Technical Report on the Cross-Sectional Sampling Weights for the Yemen National Social Protection and Monitoring Survey*. Brasília: International Policy Centre for Inclusive Growth, 2014. Technical Report. Citado na página 147.
- Ferro, A. R. e Kasouf, A. L. Avaliação do impacto dos programas Bolsa-Escola sobre o trabalho infantil no Brasil. *Pesquisa e Planejamento Econômico*, v. 35, n. 3, p. 417–444, 2005. Citado 3 vezes nas páginas 32, 35 e 129.
- Fisher, R. A. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd, 1925. Citado na página 26.
- Fiszbein, A. e Schady, N. *Conditional cash transfers: reducing present and future poverty*. Washington, D. C.: The International Bank for Reconstruction and Development, 2009. Citado na página 31.
- Foguel, M. N. e Barros, R. P. The effects of conditional cash transfer programmes on adult labour supply: an empirical analysis using a time-series-cross-section sample of Brazilian municipalities. *Estudos Econômicos*, v. 40, n. 2, p. 259–293, 2010. Citado na página 36.
- Fuller, W. A. *Sampling Statistics*. New York: John Wiley and Sons, 2009. Citado na página 58.
- Gertler, P. J.; Martinez, S.; Premand, P.; Rawlings, L. B. e Vermeersch, C. M. J. *Impact Evaluation in Practice*. Washington, D. C.: The World Bank, 2011. Citado na página 96.
- Gujarati, D. N. e Porter, D. C. *Basic Econometrics*. 5. ed. New York: McGraw-Hill/Irwin, 2009. Citado na página 52.
- Hansen, M. H.; Hurwitz, W. N. e Madow, W. G. *Sample Survey Methods and Theory, Volumes I and II*. New York: John Wiley & Sons, 1953. Citado na página 27.

Harville, D. A. *Matrix Algebra from a Statistician's Perspective*. New York: Springer, 2001. Citado na página 81.

Heckman, J.; Ichimura, H.; e Todd, P. Matching as an econometric evaluation estimator. *Review of Economic Studies*, v. 65, p. 261–294, 1998. Citado na página 60.

Heckman, J. J. e Hotz, V. J. Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, v. 84, n. 408, p. 862–874, 1989. Citado na página 89.

Heeringa, S. G.; West, B. T. e Belgrund, P. A. *Applied Survey Data Analysis*. Florida: Chapman and Hall/CRC, 2010. Citado 6 vezes nas páginas 21, 22, 27, 67, 71 e 148.

Hollander, M.; Wolfe, D.A. e Chicken, E. *Nonparametric Statistical Methods*. Hoboken: John Wiley & Sons, 2014. Citado na página 82.

Holt, D.; Scott, A. J. e Ewings, P. D. Chi-squared tests with survey data. *Journal of the Royal Statistical Society. Series A (General)*, v. 143, n. 3, p. 303–320, 1980. Citado 3 vezes nas páginas 78, 79 e 80.

Hornik, K. *R FAQ: Frequently Asked Questions on R*. 2016. Disponível em: <[https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R\\_003f](https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R_003f)>. Acesso em: 7 de abril de 2016. Citado na página 46.

Horvitz, D. G. e Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, v. 47, p. 663–685, 1952. Citado 3 vezes nas páginas 26, 64 e 147.

IFPRI. *Conditional Cash Transfers in Latin America*. Baltimore: The John Hopkins University Press, 2010. Citado na página 30.

IPC-IG e UNICEF. *Yemen National Social Protection Monitoring Survey (NSPMS): 2012-2013 Final Report*. Brasília: International Policy Centre for Inclusive Growth, 2014. Citado 8 vezes nas páginas 23, 40, 41, 42, 43, 48, 100 e 129.

IPC-IG e UNICEF. *Yemen National Social Protection Monitoring Survey (NSPMS): 2012-2013 Final Report - Technical Note No. 1: Sample and Survey Methodology*. Brasília: International Policy Centre for Inclusive Growth, 2014. Citado 2 vezes nas páginas 145 e 147.

IPC-IG e UNICEF. *Yemen National Social Protection Monitoring Survey: Version 1 [Machine-readable database]*. Brasília: International Policy Centre for Inclusive Growth, 2014. Citado 5 vezes nas páginas 23, 48, 94, 98 e 117.

Ipea. *Programa Bolsa Família: Uma Década de Inclusão e Cidadania*. Brasília: Instituto de Pesquisa Econômica Aplicada, 2013. Citado 3 vezes nas páginas 30, 31 e 32.

Isaki, C. T. e Fuller, W. A. Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, v. 77, n. 377, p. 89–96, 1982. Citado na página 148.

Jørgensen, B. Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika*, v. 70, n. 1, p. 19–28, 1983. Citado na página 54.

- Jørgensen, B. Exponential dispersion models. *Journal of the Royal Statistical Society, Series B (Methodological)*, v. 49, n. 2, p. 127–162, 1987. Citado na página 55.
- Johnson, N.L.; Kotz, S. e Balakrishnan, N. *Continuous Univariate Distributions*. New York: Wiley & Sons, 1995. Citado na página 63.
- Judkins, D. R.; Morganstein, D.; Zador, P.; Piesse, A.; Barrett, B. e Mukhopadhyay, P. Variable selection and raking in propensity scoring. *Statistics in Medicine*, v. 26, p. 1022–1033, 2007. Citado na página 77.
- Khandker, S. R.; Koolwal, G. B. e Samad, H. A. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, D. C.: The World Bank, 2010. Citado na página 93.
- Kiaer, A. N. Observations et expériences concernant des dénombrements représentatifs. *Bulletin of the International Statistical Institute*, v. 9, p. 176–183, 1895. Citado na página 26.
- Kish, L. *Survey Sampling*. New York: John Wiley & Sons, 1965. Citado 2 vezes nas páginas 27 e 68.
- Kish, L. e Frankel, M. R. Inference from complex samples. *Journal of the Royal Statistical Society, Series B*, v. 36, n. 1, p. 1–37, 1974. Citado na página 28.
- Koch, G. G. e Lemeshow, S. An application of multivariate analysis to complex sample survey data. *Journal of the Royal Statistical Society, Series B*, v. 67, n. 340, p. 780–782, 1972. Citado na página 28.
- Korn, E. L. e Graubard, B. I. *Analysis of Health Surveys*. Hoboken: John Wiley & Sons, 1999. Citado 2 vezes nas páginas 62 e 107.
- Krewski, D. e Rao, J. N. K. Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, v. 9, n. 5, p. 1010–1019, 1981. Citado na página 68.
- Kruskal, W. H. e Mosteller, F. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, v. 47, n. 260, p. 583–621, 1952. Citado na página 85.
- Kruskal, W. H. e Mosteller, F. Representative sampling, IV: The history of the concept in Statistics, 1895–1939. *International Statistical Review*, v. 48, n. 2, p. 169–195, 1980. Citado na página 26.
- Kuonen, D. Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, v. 86, n. 4, p. 929–935, 1999. Citado na página 62.
- Lechner, M. The estimation of causal effects by Difference-in-Difference methods. *Foundations and Trends in Econometrics*, v. 4, n. 3, p. 165–224, 2010. Citado 6 vezes nas páginas 88, 90, 92, 93, 94 e 95.
- Lemonte, A. J. *The Gradient Test: Another Likelihood-Based Test*. London: Academic Press, 2016. Citado na página 56.
- Lohr, S. *Sampling: Design and Analysis*. 2. ed. Boston: Cengage Learning, 2010. Citado 5 vezes nas páginas 22, 27, 29, 43 e 94.

- Lumley, T. *Package 'survey': Analysis of complex survey samples*. [S.l.], 2016. Disponível em: <<https://cran.r-project.org/web/packages/survey/index.html>>. Acesso em: 21 de abril de 2016. Citado 5 vezes nas páginas 63, 68, 82, 95 e 157.
- Lumley, T. e Scott, A. Two-sample rank tests under complex sampling. *Biometrika*, v. 100, n. 4, p. 831–842, 2013. Citado 5 vezes nas páginas 82, 85, 86, 87 e 134.
- Lumley, T. e Scott, A. Tests for regression models fitted to survey data. *Australian & New Zealand Journal of Statistics*, v. 56, n. 1, p. 1–14, 2014. Citado 6 vezes nas páginas 57, 58, 60, 61, 62 e 63.
- Luo, Z.; Gardiner, J. C. e Bradley, C. J. Applying propensity score methods in medical research: Pitfalls and prospects. *Medical Care Research and Review*, v. 67, n. 5, p. 528–554, 2010. Citado na página 73.
- Mahalanobis, P. C. Recent experiments in statistical sampling in the Indian Statistical Institute. *Sankhya: The Indian Journal of Statistics*, v. 20, n. 3, p. 329–398, 1958. Citado na página 67.
- Mann, H. B. e Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, v. 18, n. 1, p. 50–60, 1947. Citado 3 vezes nas páginas 77, 82 e 84.
- Matsumoto, M. e Nishimura, T. Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, v. 8, n. 1, p. 3–30, 1998. Citado na página 112.
- Medeiros, M.; Britto, T. e Soares, F. *Programas Focalizados de Transferência de Renda no Brasil: Contribuições para o Debate*. Brasília: Instituto de Pesquisa Econômica Aplicada, 2007. Citado na página 31.
- Meyer, B. D.; Viscusi, W. K. e Durbin, D. L. Workers' compensation and injury duration: Evidence from a natural experiment. *The American Economic Review*, v. 85, n. 3, p. 322–340, 1995. Citado na página 89.
- Narain, R. D. On sampling without replacement with varying probabilities. *Journal of the Indian Society for Agricultural Statistics*, v. 4, p. 169–175, 1951. Citado na página 26.
- Nelder, J. A. e Wedderburn, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, v. 135, n. 3, p. 370–384, 1972. Citado 2 vezes nas páginas 53 e 56.
- Neyman, J. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, v. 97, p. 558–606, 1934. Citado na página 26.
- Obenauer, M. e von der Nienburg, B. Effect of minimum-wage determinations in oregon. *Monthly Review of the U.S. Bureau of Labor Statistics*, v. 1, n. 3, p. 5–8, 1915. Citado na página 89.
- Oetiker, T.; Partl, H.; Hyna, I. e Schlegl, E. *The Not So Short Introduction to L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>*. 5.05. ed. [S.l.], 2015. Published under GNU General Public License by the Free Software Foundation. Citado na página 45.

- Paula, G. A. *Modelos de Regressão com Apoio Computacional*. São Paulo, 2013. Citado 3 vezes nas páginas 54, 55 e 56.
- Pessoa, D. G. C. e Silva, P. L. D. N. *Análise de Dados Amostrais Complexos*. Rio de Janeiro: IBGE, 1998. Citado 13 vezes nas páginas 21, 29, 58, 59, 68, 69, 70, 71, 78, 79, 80, 81 e 94.
- Pfeffermann, D.; Skinner, C. J.; Holmes, D. J.; Goldstein, H. e Rasbash, J. Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society, Series B*, v. 60, n. 1, p. 23–40, 1998. Citado na página 29.
- Putter, J. The treatment of ties in some nonparametric tests. *The Annals of Mathematical Statistics*, v. 26, n. 3, p. 368–386, 1955. Citado na página 83.
- Quatember, A. *Pseudo-Populations: A Basic Concept in Statistical Surveys*. Switzerland: Springer International Publishing, 2015. Citado na página 79.
- R Core Team. *foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...* [S.l.], 2016. R package version 0.8-66. Disponível em: <<https://CRAN.R-project.org/package=foreign>>. Acesso em: 21 de abril de 2016. Citado na página 150.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <<http://www.R-project.org/>>. Citado 4 vezes nas páginas 24, 45, 102 e 150.
- Rao, J. N. K. and Wu, C. F. J. Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association*, v. 80, n. 391, p. 620–630, 1985. Citado na página 68.
- Rao, J. N. K. e Scott, A. J. The analysis of categorical data from complex sample surveys. In: *American Statistical Association Conference*. Washington, D. C.: [s.n.], 1979. Citado na página 78.
- Rao, J. N. K. e Scott, A. J. The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, v. 76, n. 374, p. 221–230, 1981. Citado 2 vezes nas páginas 28 e 80.
- Rao, J. N. K. e Scott, A. J. On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics*, v. 12, n. 1, p. 46–60, 1984. Citado na página 79.
- Rao, J. N. K.; Scott, A. J. e Skinner, C. J. Quasi-score tests with survey data. *Statistica Sinica*, v. 8, n. 4, p. 1059–1070, 1998. Citado na página 63.
- Roberts, G; Rao, J. N. K. e Kumar, S. Logistic regression analysis of sample survey data. *Biometrika*, v. 74, n. 1, p. 1–12, 1987. Citado na página 59.
- Rosenbaum, P. R. e Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, v. 70, n. 1, p. 41–55, 1983. Citado 4 vezes nas páginas 48, 49, 50 e 51.

- Rosenbaum, P. R. e Rubin, D. B. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, v. 39, n. 1, p. 33–38, 1985. Citado 2 vezes nas páginas 73 e 77.
- Rubin, D. B. Inference and missing data. *Biometrika*, v. 63, n. 3, p. 581–592, 1976. Citado na página 147.
- Rubin, D. B. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, v. 2, n. 1, p. 1–26, 1977. Citado na página 90.
- Särndal, C. E.; Swensson, B. e Wretman, J. *Model Assisted Survey Sampling*. New York: Springer, 1992. Citado 3 vezes nas páginas 21, 27 e 29.
- Satterthwaite, F. E. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, v. 2, n. 6, p. 110–114, 1946. Citado 2 vezes nas páginas 62 e 81.
- Scholz, F. W. e Stephens, M. A. K-Sample Anderson-Darling tests. *Journal of the American Statistical Association*, v. 82, n. 399, p. 918–924, 1987. Citado na página 82.
- Sekhon, J. S. Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software*, v. 42, n. 7, p. 1–52, 2011. Citado 3 vezes nas páginas 53, 111 e 172.
- Sen, P. K. e Singer, J. M. *Large Sample Methods in Statistics: An Introduction with Applications*. London: Chapman & Hall, 1993. Citado na página 56.
- Skinner, C. J.; Holt, D. e Smith, T. M. F. *Analysis of Complex Surveys*. New York: John Wiley & Sons, 1989. Citado 4 vezes nas páginas 27, 28, 29 e 69.
- Skoufias, E. e di Maro, V. Conditional cash transfers, adult work incentives, and poverty. *Impact Evaluation Series*, v. 5, p. 1–37, 2006. Citado 3 vezes nas páginas 35, 36 e 129.
- Skoufias, E. e Parker, S. W. Conditional cash transfers and their impact on child work and schooling: Evidence from the PROGRESA program in Mexico. *Economía*, v. 2, n. 1, p. 45–96, 2001. Citado 2 vezes nas páginas 31 e 129.
- Smith, H. L. Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology*, v. 27, n. 1, p. 325–353, 1997. Citado na página 74.
- Snow, J. *On the Mode of Communication of Cholera*. London: John Churchill, 1855. Citado na página 88.
- Sukhatme, P. V. *Sampling Theory of Surveys, with Applications*. Ames: Iowa State College Press, 1954. Citado na página 27.
- Tavares, P. A. Efeito do Programa Bolsa Família sobre a oferta de trabalho das mães. *Economia e Sociedade*, v. 19, n. 3, p. 613–635, 2010. Citado 2 vezes nas páginas 35 e 36.
- Thomas, D. R. e Rao, J. N. K. Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. *Journal of the American Statistical Association*, v. 82, n. 398, p. 630–636, 1987. Citado 4 vezes nas páginas 60, 62, 78 e 82.
- Thomas, D. R.; Singh, A. C. e Roberts, G. R. Tests of independence on two-way tables under cluster sampling: An evaluation. *International Statistical Review*, v. 64, n. 3, p. 295–311, 1996. Citado 2 vezes nas páginas 78 e 81.

- UNICEF. *Female Genital Mutilation/Cutting: A statistical Overview and Exploration of the Dynamics of Change*. New York: UNICEF, 2013. Citado na página 42.
- UNICEF Brasil. *Convenção n.º 138 – OIT – Idade Mínima para Admissão em Emprego*. 2016. Disponível em: <[http://www.unicef.org/brazil/pt/resources\\_10231.htm](http://www.unicef.org/brazil/pt/resources_10231.htm)>. Acesso em: 19 de maio de 2016. Citado na página 24.
- van der Vaart, A. W. *Asymptotic Statistics*. Cambridge: Cambridge University Press, 1998. Citado na página 86.
- van der Vaart, A. W. e Wellner, J. *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer, 1996. Citado na página 86.
- Venables, W. N.; Smith, D. M. e R Core Team. *An Introduction to R*. Vienna, Austria, 2016. Versão 3.2.4. Disponível em: <<http://www.R-project.org/>>. Acesso em: 21 de abril de 2016. Citado na página 46.
- Vieira, M. D. T. e Ferraz, C. *Technical Report on the Longitudinal Sampling Weights for the Yemen National Social Protection and Monitoring Survey*. Brasília: International Policy Centre for Inclusive Growth, 2014. Technical Report. Citado 2 vezes nas páginas 148 e 149.
- Wang, J. C. Sample distribution function based goodness-of-fit test for complex surveys. *Computational Statistics and Data Analysis*, v. 56, p. 664–679, 2012. Citado na página 82.
- Wang, Y.; Cai, H.; Li, C.; Jiang, Z.; Wang, L.; Song, J. e Xia, J. Optimal caliper width for propensity score matching of three treatment groups: A monte carlo study. *PLoS ONE*, v. 8, n. 12, p. 1–7, 2013. Citado na página 75.
- Wedderburn, R. W. M. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, v. 61, n. 3, p. 439–447, 1974. Citado na página 54.
- Welch, B. L. The generalization of “Student’s” problem when several different population variances are involved. *Biometrika*, v. 34, p. 28–35, 1947. Citado na página 76.
- Wickham, H. The Split-Apply-Combine strategy for data analysis. *Journal of Statistical Software*, v. 40, n. 1, p. 1–29, 2011. Citado na página 150.
- Wolter, K. M. *Introduction to Variance Estimation*. 2. ed. New York: Springer-Verlag, 2007. Citado 3 vezes nas páginas 65, 66 e 67.
- Yates, F. *Sampling Methods for Censuses and Surveys*. 3. ed. London: Griffin, 1960. Citado na página 27.

# Apêndices

# APÊNDICE A – Cálculo dos Pesos Amostrais

## A.1 Cálculo dos Pesos Amostrais Básicos

Nesta seção, será mostrada a forma de cálculo das probabilidades de seleção das UPA's (as Áreas de Enumeração - AE's) e das USA's (os domicílios) na YNSPMS. Para estas últimas, o inverso da respectiva probabilidade de seleção fornecerá o peso amostral básico do domicílio (observação da pesquisa amostral).

Considerando o plano amostral descrito na subseção 1.4.3, temos que no 1º estágio de seleção cada AE  $i$  foi geograficamente estratificada pela província  $h$  a qual pertencia e selecionada através do esquema PPT com APS e a variável cujo tamanho foi utilizado é o número médio de pessoas pobres por AE, considerando o total no distrito ao qual a AE pertence. Denotemos então por  $\mu_{hk}$  o nº médio de pessoas pobres por AE no distrito  $k$  dentro da província  $h$ , com  $\mu_{hk} = \tau_{hk}/M_{hk}$ , em que  $M_{hk}$  e  $\tau_{hk}$  representam o nº de AE's e o total de pessoas pobres no distrito  $k$  dentro da província  $h$ , respectivamente, obtidos dos dados da pesquisa HBS. A seguir, são listados os passos do algoritmo para seleção da AE por APS (YNSPMS Final Report, 2014b, p. 3):

- Para cada AE  $i$  dentro da província  $h$ , gere um valor  $u_i$  a partir de uma variável aleatória com distribuição uniforme padrão;
- Calcule  $p_{hi} = \mu_{hk} / \sum_{k=1}^{M_h} \mu_{hk}$  para cada AE  $i$ ;
- Calcule  $Z_{hi} = u_i / p_{hi}$  para cada AE  $i$ ;
- Após ordenar as AE's cadastradas de acordo com valores crescentes em  $Z_{hi}$ , selecione as 30 primeiras para a amostra dentro da província  $h$ .

Note que devido ao último passo as AE's com maiores valores para  $p_{hi}$  (ou seja, com maiores médias de pessoas pobres) terão maior probabilidade de seleção, o qual não é determinístico devido ao fato de que  $u_i$  foi gerado a partir de uma variável aleatória. Ainda, se duas AE's estão dentro do mesmo distrito  $k$ , elas terão o mesmo valor para  $\mu_{hk}$  e portanto a mesma probabilidade de seleção. A análise prévia dos valores de  $\mu_{hk}$ , antes da implementação do algoritmo acima, garantiu que não houve nenhum exemplo de distrito com valor extremo.

Sejam então  $\pi_{hi}$  a probabilidade de seleção da AE  $i$  dentro da província  $h$  no 1º estágio;  $\gamma_{hi} = \mu_{hk}$  o valor da variável tamanho (o nº médio de pessoas pobres) nesta mesma

AE e  $m_h$  o n<sup>o</sup> de AE's selecionadas na província  $h$  (sempre será igual a 30). Temos que

$$\pi_{hi} = m_h \frac{\gamma_{hi}}{\sum_{i \in U_{ih}} \gamma_{hi}} \quad (\text{A.1})$$

em que  $U_{ih}$  denota o n<sup>o</sup> total (universo) de AE's na província  $h$ . É esperado que a adoção do método de amostragem com PPT, em que o tamanho é uma medida de pobreza, aumente a eficiência da inferência estatística realizada e da política pública (no caso, o SWF) em atingir quem de fato precisa dela.

Amostras planejadas para avaliação de programas de bem-estar social exigem dos estatísticos amostristas o foco na estimação de parâmetros para estudos comparativos. Assim, os tamanhos amostrais em cada estrato, no 1<sup>o</sup> estágio, foram alocados com base apenas na produção de estimativas de contraste, por estrato, com níveis similares de precisão (após a análise da pesquisa). Quando o 1<sup>o</sup> estágio foi desenhado, o principal objetivo era a estimação de um contraste entre subpopulações de domicílios beneficiários ( $b$ ) e não beneficiários ( $nb$ ) do SWF. Desprezando os fatores de correção para populações finitas, as regras de alocação ótima do tamanho amostral para amostras aleatórias simples em cada estrato podem ser derivadas minimizando (Cochran, 1977, p. 140)

$$\text{Var}(\bar{y}_b - \bar{y}_{nb}) = \frac{\sigma_b^2}{n_b} + \frac{\sigma_{nb}^2}{n_{nb}} \quad (\text{A.2})$$

sujeito à restrição de custo

$$C = c_0 + c_b n_b + c_{nb} n_{nb} \quad (\text{A.3})$$

A minimização da variância calculada em (A.2) é atingida quando

$$n_b = n \frac{\sigma_b c_b}{\sqrt{\frac{\sigma_b^2}{c_b} + \frac{\sigma_{nb}^2}{c_{nb}}}} \quad (\text{A.4})$$

$$n_{nb} = n \frac{\sigma_{nb} c_{nb}}{\sqrt{\frac{\sigma_b^2}{c_b} + \frac{\sigma_{nb}^2}{c_{nb}}}} \quad (\text{A.5})$$

Como não temos informação prévia das variâncias populacionais  $\sigma_b^2$  e  $\sigma_{nb}^2$  e assumindo que  $\sigma_b^2 = \sigma_{nb}^2$  e  $c_b = c_{nb}$ , teremos igualdade entre as expressões (A.4) e (A.5), ou seja,  $n_b = n_{nb}$  e portanto a alocação será uniforme (explicando a seleção de 30 AE's em cada estrato no 1<sup>o</sup> estágio). É interessante observar que, quando lidamos com estimação de contrastes (contra-factual) para avaliar o impacto da política pública, os tamanhos populacionais dos grupos de contraste (no caso, beneficiários e não beneficiários) não influenciam na decisão de alocação do tamanho amostral por estrato.

Por sua vez, no 2<sup>o</sup> estágio de seleção, foram selecionados 12 domicílios em cada AE  $i$  por AES, onde o estrato  $g$  deste estágio é dado pelo grupo (ET; EC1 ou EC2) ao qual o domicílio  $j$  pertence, com alocação definida na subseção 1.4.3. Sejam  $\pi_{gj|hi}$

a probabilidade condicional de seleção do domicílio  $j$  dentro do estrato  $g$  dada a AE  $i$ , selecionada previamente, dentro da província  $h$  no estágio anterior;  $n_{g|hi}$  o tamanho amostral dos domicílios classificados no estrato  $g$  dada a AE  $i$  dentro da província  $h$  e  $N_{g|hi}$  o respectivo total populacional de domicílios. Temos que,  $\forall j \in U_{g|hi}$ , onde  $U_{g|hi}$  representa o total de domicílios listados no estrato  $g$  dada a AE  $i$  dentro da província  $h$  (YNSPMS Final Report, 2014b, p. 6)

$$\pi_{gj|hi} = \frac{n_{g|hi}}{N_{g|hi}} \quad (\text{A.6})$$

Finalmente, a probabilidade de seleção do domicílio  $j$  da AE  $i$  dentro da província  $h$  é fornecida por

$$\pi_{hij} = \pi_{hi}\pi_{gj|hi} \quad (\text{A.7})$$

com o respectivo peso amostral básico  $d_{hij}$  dado pelo seu inverso, ou seja,  $d_{hij} = 1/\pi_{hij}$ .

## A.2 Cálculo dos Pesos Transversais

Os pesos amostrais transversais descritos a seguir refletem não apenas o PAC utilizado para a YNSPMS, mas também a introdução de termos de ajuste para não-resposta unitária (um domicílio selecionado não respondeu ou não teve acesso a nenhuma variável do questionário) encontrados na primeira rodada da pesquisa. Considere as quantidades  $n_{g|hi}$  e  $d_{hij}$  definidas na seção A.1. O ponto de partida para a construção dos pesos amostrais transversais é o estimador ENHT para o total populacional, dado por (Horvitz & Thompson, 1952, p. 668)

$$\hat{Y} = \sum_h \sum_i \sum_g \sum_j d_{hij} y_{hij} \quad (\text{A.8})$$

em que  $y_{hij}$  é o valor da variável de interesse neste mesmo domicílio. Desta forma, uma vez que o 1º estágio do plano amostral seleciona AE's com probabilidades proporcionais a uma medida de pobreza (o nº de pessoas pobres no distrito ao qual a AE pertence), pesos amostrais maiores serão atribuídos às AE's com menores níveis de pobreza, e pesos amostrais menores, às AE's com maiores níveis de pobreza.

Assumiu-se que as não-respostas unitárias foram geradas por um mecanismo MR (*Missing at Random*, em português: Não-resposta Aleatória, introduzido por Rubin em 1976), no qual a não-resposta dependerá apenas das variáveis utilizadas para estratificação no plano amostral (ou seja, a província no 1º estágio e os grupos classificados com relação ao SWF no 2º estágio). Para atenuar este problema, foi feito um procedimento de ajuste para corrigir o peso amostral básico. Seja (Ferraz & Vieira, 2014, p. 3)

$$\hat{\varphi}_{hig} = \frac{m_{h(r)} n_{g|hi(r)}}{m_h n_{g|hi}} \quad (\text{A.9})$$

o escore de propensão estimado para resposta dentro do estrato  $g$  na AE  $i$  dentro da província  $h$ . Na expressão acima,  $m_h$  é o nº de AE's selecionadas na província  $h$ ;  $m_{h(r)}$  é o

nº de AE's que foram de fato pesquisadas e  $n_{g|hi(r)}$  é o nº de domicílios (subconjunto de  $n_{g|hi}$ ) que de fato responderam ao questionário. Logo, temos que

$$\tilde{w}_{higj} = \frac{d_{higj}}{\hat{\varphi}_{hig}} \quad (\text{A.10})$$

é o peso amostral ajustado por não-resposta unitária do domicílio  $j$  dentro do estrato  $g$  na AE  $i$  dentro da província  $h$ . Uma vez que a não-resposta é suposta seguir um mecanismo MR, o estimador ENHT continuará não enviesado e consistente mesmo com o ajuste dos pesos amostrais (Isaki & Fuller, 1982). Logo,  $\tilde{w}_{higj}$  é o peso amostral transversal para a pesquisa YNSPMS.

É possível ainda melhorar a precisão das estimativas obtidas para os parâmetros de interesse incorporando a informação de variáveis auxiliares. Preferencialmente, estas devem estar fortemente correlacionadas com as variáveis sob estudo. Tal informação é incorporada através de novo ajuste nos pesos amostrais transversais, que pode ser feito por pós-estratificação ou calibração (Heeringa *et al.*, 2010). Na YNSPMS, não é possível aplicá-las devido à ausência de dados censitários atualizados para as variáveis pesquisadas.

### A.3 Cálculo dos Pesos Longitudinais

Em pesquisas amostrais longitudinais, dois tipos de pesos amostrais precisam ser calculados: os pesos transversais (cuja metodologia de cálculo foi discutida na seção anterior), específicos para uma rodada  $t$  e para uso em análises envolvendo dados apenas desta rodada; e os pesos longitudinais, também específicos para uma rodada  $t$ , mas para uso em análises longitudinais considerando todas as rodadas até  $t$ .

Assim, os pesos transversais são reajustados pela não-resposta na rodada  $t$  e, quando disponíveis, também pela incorporação de informações auxiliares referentes à mesma rodada. Para a construção de pesos longitudinais, abordagens semelhantes também podem ser adotadas. De fato, é possível calcular até  $2^T - 1$  (onde  $T$  é o nº total de rodadas) conjuntos de pesos longitudinais, permitindo a análise de todas as combinações dos padrões de não-resposta. Por razões de simplicidade, Vieira & Ferraz (2014) adotaram uma abordagem que lida apenas com a não-resposta por exclusão de unidades amostrais primárias ou secundárias (o domicílio não pôde ser entrevistado ou a AE a qual pertence sequer pôde ser visitada), resultando na necessidade de se calcular apenas  $t$  conjuntos de pesos longitudinais.

Sob esta abordagem, somente os domicílios que responderam ao questionário em todas as  $t$  rodadas terão pesos longitudinais positivos na rodada  $t$ . Se tomamos, por exemplo, uma pesquisa amostral com duas rodadas, então os pesos longitudinais para a 2ª rodada vão considerar: as probabilidades desiguais de seleção na 1ª rodada; o ajuste para não-resposta unitária, o qual pode ocorrer nas duas rodadas, ou seja, um domicílio que

previamente respondeu ao questionário na 1ª rodada pode não ter participado da pesquisa na 2ª rodada por algum motivo (principal diferença em relação aos pesos transversais); e o ajuste utilizando possíveis informações de variáveis auxiliares referentes à rodada atual.

Em seguida, os pesos longitudinais serão calculados de maneira que seja garantida a consistência dos momentos amostrais ponderados, como estimativas dos momentos populacionais, com respeito à distribuição de probabilidade conjunta de não-resposta (pois temos  $t$  rodadas) na amostra. Mantendo a mesma notação utilizada na seção anterior, definimos (Vieira & Ferraz (2014, p. 3)

$$\hat{\varphi}_{hig,t} = \frac{m_{h(r),t} n_{g|hi(r),t}}{m_h n_{g|hi}} \quad (\text{A.11})$$

como o escore de propensão estimado para resposta dentro do estrato  $g$  na AE  $i$  dentro da província  $h$ , a ser considerado para o ajuste dos pesos longitudinais na rodada  $t$ . Os termos  $m_h$  e  $n_{g|hi}$  são os mesmos presentes em (A.9);  $m_{h(r),t}$  é o n.º de AE's que foram pesquisadas até a rodada  $t$  (inclusive) e  $n_{g|hi(r),t}$  é o n.º de domicílios classificados no estrato  $g$  da AE  $i$  dentro da província  $h$  e que de fato responderam ao questionário até a rodada  $t$  (inclusive). Assim

$$\tilde{w}_{higj,t} = \frac{d_{higj}}{\hat{\varphi}_{hig,t}} \quad (\text{A.12})$$

é o peso amostral longitudinal na rodada  $t$  ajustado por não-resposta unitária, nas  $t$  primeiras rodadas, do domicílio  $j$  classificado no grupo  $g$  da AE  $i$  dentro da província  $h$ . Ao contrário do que foi feito para os pesos transversais, não é realizada nenhuma análise que possibilite aparar qualquer dos  $t$  conjuntos de pesos amostrais no contexto da YNSPMS (Vieira & Ferraz, 2014).

Uma vez calculados, os pesos longitudinais serão úteis para avaliações de mudanças ao longo do tempo nas estimativas das variáveis pesquisadas e do impacto de determinada política (aqui, o programa SWF) no seu alvo (os domicílios que recebem o benefício).

## APÊNDICE B – Códigos Computacionais para o R

Os códigos abaixo foram integralmente executados no *software* R versão 3.2.4 (R Core Team, 2016b) e divididos em duas partes. A primeira versa sobre estatísticas descritivas produzidas na dissertação e a segunda, sobre estimativas obtidas para os escores de propensão (considerando o plano amostral); o EMT e o impacto do SWF no nº de horas trabalhadas para os dois grupos etários. Antes de executar os códigos, o diretório do R deve ser atribuído a uma pasta que contenha o(s) banco(s) de dado(s) necessário(s) durante a execução. Isto pode ser feito indo em *File* → *Change Dir...* e escolhendo a pasta onde o(s) banco(s) de dado(s) se localiza(m), clicando finalmente em *OK*.

Após isso, instale e carregue no R os pacotes descritos a seguir, procedimento também obrigatório para a execução do código. Nem todas as funções apresentadas no código serão nativas do R. Para instalação de pacotes, use o comando `install.packages()`

```
install.packages("plyr"); install.packages("xtable")
```

Para carregar os pacotes previamente instalados, use o comando `library()`

```
library(foreign); library(plyr)
```

Embora presente na instalação básica do R, o pacote `foreign` (R Core Team, 2016a) precisa ser carregado se quisermos usar suas funções, úteis para leitura e escrita de arquivos em diversos formatos de outros *software* estatísticos. Por sua vez, o pacote `plyr` (Wickham, 2011) possui funções bem simplificadas (com relação a sintaxes análogas da base do R) para manipulação de bancos de dados (no R, representados pela classe *data frame*). Já o pacote `xtable` (Dahl, 2016) possui funções que permitem exibir no R tabelas de objetos diretamente na tipografia L<sup>A</sup>T<sub>E</sub>X.

Os dados da YNSPMS estão salvos no formato `.dta`, significando que o banco foi originalmente montado no *software* Stata. Para ler o banco de dados, use a função `read.dta()`, do pacote `foreign`

```
yenspms = read.dta("yenspms.dta")
```

Onde `yenspms.dta` é o nome do arquivo que contém o banco de dados da YNSPMS. Os dados foram lidos e salvos no objeto `yenspms`, agora um *data frame*.

Antes de extrair qualquer medida do banco de dados, podemos dividi-lo em quatro de acordo com a rodada e em seguida reorganizá-lo indexando cada linha por indivíduo pesquisado. Note que serão necessários os códigos do domicílio e do indivíduo, sendo as demais variáveis replicadas por rodada. É recomendável salvar sempre qualquer banco criado e, quando a execução for encerrada, removê-lo para recuperar espaço na memória antes de encerrar o R

```
yd.1r = ynspps[which(ynspms$f2 == "first"),]
write.csv(yd.1r, "ynspms1rod.csv")

yd.2r = ynspps[which(ynspms$f2 == "second"),]
write.csv(yd.2r, "ynspms2rod.csv")

yd.3r = ynspps[which(ynspms$f2 == "third"),]
write.csv(yd.3r, "ynspms3rod.csv")

yd.4r = ynspps[which(ynspms$f2 == "fourth"),]
write.csv(yd.4r, "ynspms4rod.csv")

ym12 = merge(yd.1r, yd.2r, by = c("f1", "line_no"))
ym34 = merge(yd.3r, yd.4r, by = c("f1", "line_no"))

rm(yd.1r, yd.2r, yd.3r, yd.4r)

ym1234 = merge(ym12, ym34, by = c("f1", "line_no"))
write.csv(ym1234, "ynspmsMerDI.csv")

rm(ym12, ym34); rm(ynspms)
```

Para ler novamente os dados, simplificando a nomenclatura dos objetos, basta fazer

```
yd = read.dta("ynspms.dta")

yd.1r = read.csv("ynspms1rod.csv"); yd.2r = read.csv("ynspms2rod.csv")
yd.3r = read.csv("ynspms3rod.csv"); yd.4r = read.csv("ynspms4rod.csv")

ym1234 = read.csv("ynspmsMerDI.csv")
```

Para mais detalhes sobre as funções utilizadas, digite `help(<nome.da.função>)` no console do R (janela executável).

## B.1 Código para o Cálculo de Estatísticas Descritivas

### B.1.1 Frequência de Controles e Tratados pela Variável RSWF

A variável RSWF corresponde ao nome `p6b_q_15` no dicionário de variáveis da YNSPMS. Sendo uma variável binária, as duas respostas possíveis são “yes” (afirmativa) e “no” (negativa). Antes de analisá-la, vamos retirar do banco completo os indivíduos cujos pesos longitudinais da 1ª à 4ª rodada (nome `adj_weight_w1ew2ew3ew4`) são faltantes, indicando que o indivíduo não foi entrevistado a partir de uma das três rodadas após a 1ª, logo com certeza ele não o foi na 4ª rodada

```
names(ym1234)[names(ym1234) == "adj_weight_w1ew2ew3ew4.x.x"] <- "p1"
ym1234 = ym1234[complete.cases(ym1234$p1),]
```

Note que a terminação da variável (`.x.x`) após a reorganização por domicílio e indivíduo não interfere nesta variável, pois os mesmos dados estavam disponíveis para todas as rodadas. Isto permite que as análises considerando e ignorando o PAC sejam comparáveis, dado que nos basearemos em informações de todas as rodadas para definir os controles e tratados. Considerando esta variável para classificação, os comandos que geram o total de controles e tratados são apresentados abaixo

```
count(ym1234$p6b_q_15.x.x == "no" & ym1234$p6b_q_15.y.x == "no" &
      ym1234$p6b_q_15.x.y == "no" & ym1234$p6b_q_15.y.y == "no")

count(ym1234$p6b_q_15.x.x == "no" & (ym1234$p6b_q_15.y.x == "yes" |
      ym1234$p6b_q_15.x.y == "yes" |
      ym1234$p6b_q_15.y.y == "yes"))
```

O total de controles é cerca de três vezes maior do que o de tratados. É comum, em pesquisas amostrais, que haja mais controles do que tratados, o que é adequado para o pareamento a ser feito mais adiante. Isso se explica pela maior facilidade em encontrar um controle que possa ser pareado a cada indivíduo tratado. Apesar disso, o total de tratados não pode ser muito baixo, ou as estimativas do EMT serão enviesadas. Esta recomendação também é atendida pela pesquisa YNSPMS.

Para garantir a validade da classificação, é verificado primeiramente se a mesma é uniforme com relação ao domicílio

```
crit = rep(NA, nrow(ym1234))

con = which(ym1234$p6b_q_15.x.x == "no" & ym1234$p6b_q_15.y.x == "no" &
           ym1234$p6b_q_15.x.y == "no" & ym1234$p6b_q_15.y.y == "no")
```

```

tra = which(ym1234$p6b_q_15.x.x == "no" & (ym1234$p6b_q_15.y.x == "yes" |
      ym1234$p6b_q_15.x.y == "yes" | ym1234$p6b_q_15.y.y == "yes"))

crit = replace(crit, contr, 0); crit = replace(crit, trata, 1)
count(crit)

ym1234 = cbind(ym1234, crit)

unif.dom = table(ym1234$f1, ym1234$crit)

which(is.na(unif.dom[,1:2])) #Note que os indivíduos NA terão valores
unif.dom[1:30,]             #nulos em ambas as categorias da tabela.

conta.erro = 0
for(i in 1:nrow(unif.dom)){
  if(unif.dom[i,1] != 0 & unif.dom[i,2] != 0) conta.erro = conta.erro + 1}

conta.erro #0 contador continua igual a zero, logo não há indivíduos
           #dentro de um mesmo domicílio com classificações distintas.

e em seguida, se não há indivíduos residentes em domicílios que já possuíram um beneficiário
residente, mas não o possuíam nas quatro rodadas da pesquisa (falsos controles)

count(ym1234$p6b_q_15.x.x == "no" & ym1234$p6b_q_15.y.x == "no" &
      ym1234$p6b_q_15.x.y == "no" & ym1234$p6b_q_15.y.y == "no" &
      ym1234$bitreat.x.x == "Control")

```

Como ambas as condições são garantidas, temos que o critério usado para classificação em controles e tratados é válido. Desta forma, podemos reduzir o banco excluindo os indivíduos não classificados como controle ou tratados, salvando esta nova versão em seguida

```

ym1234.red = ym1234[complete.cases(crit),]
write.csv(ym1234.red, "ynspmsMerDIRed.csv"); rm(ym1234)

```

Separando os indivíduos presentes no banco acima pelos grupos etários considerados (de 6 a 14 anos, inclusive, e acima de 14 anos na 1ª rodada), temos os respectivos totais de controles e tratados a seguir

```

ym.chi = ym1234.red[(ym1234.red$p2_q_05_y.x.x >= 6 &

```

```

                                ym1234.red$p2_q_05_y.x.x <= 14),]
ym.adu = ym1234.red[(ym1234.red$p2_q_05_y.x.x >= 15),]

count(ym.chi$crit); count(ym.adu$crit)

```

É importante notar que não há respostas faltantes para a idade, correspondente ao nome `p2_q_05_y` no dicionário de variáveis, na primeira rodada

```
count(ym1234.red$p2_q_05_y.x.x); rm(ym1234.red)
```

Devido ao período curto entre as quatro rodadas da YNSPMS (de outubro de 2012 a setembro de 2013), foi suposto que a idade respondida na 1ª rodada se manteve constante nas demais rodadas.

### B.1.2 Frequência de Respostas para a Variável TSAT

São apresentados abaixo os comandos que geram o total de respostas válidas para os dois grupos etários considerados (entre 6 e 14 anos e acima de 14 anos) para a variável TSAT, nossa resposta de interesse. No dicionário de variáveis da YNSPMS, ela corresponde ao nome `p6a_q_10` e qualquer valor entre 0 e 97 (inclusive) indica uma resposta válida. Foram considerados apenas os indivíduos que responderam “sim” à variável indicadora de trabalho pelo indivíduo nos últimos 30 dias (nomeada `p6a_q_2`) na 1ª e 4ª rodadas da YNSPMS. Isto se explica pelo fato de que somente quem está de fato trabalhando está apto a responder quantas horas são dedicadas ao trabalho, conforme explicitado pelo questionário, e o objetivo é avaliar o impacto do SWF neste período. Logo, foram retirados quem respondeu "não" ou teve respostas faltantes nesta variável.

A frequência foi calculada para duas situações: considerando primeiramente todos os indivíduos da YNSPMS que se encaixam na restrição enunciada no parágrafo anterior e depois somente para aqueles também classificados como controles ou tratados e com pesos longitudinais da 1ª à 4ª rodada. Tais estatísticas descritivas estão separadas para as duas rodadas consideradas e pelos dois grupos etários sob estudo, com código para execução mostrado logo abaixo

```
ym1234 = read.csv("ynspmsMerDI.csv") #Recarrega o banco completo.
```

```
ym1234.c = ym1234[(ym1234$p2_q_05_y.x.x >= 6 &
                  ym1234$p2_q_05_y.x.x <= 14),]
ym1234.a = ym1234[ym1234$p2_q_05_y.x.x >= 15,]
```

```
ym1234.c = ym1234.c[complete.cases(ym1234.c$p6a_q_2.x.x,
```

```

                                ym1234.c$p6a_q_2.y.y),]
ym1234.a = ym1234.a[complete.cases(ym1234.a$p6a_q_2.x.x,
                                ym1234.a$p6a_q_2.y.y),]

ym1234.c = ym1234.c[-which(ym1234.c$p6a_q_2.x.x == "no" |
                            ym1234.c$p6a_q_2.y.y == "no"),]
ym1234.a = ym1234.a[-which(ym1234.a$p6a_q_2.x.x == "no" |
                            ym1234.a$p6a_q_2.y.y == "no"),]

ym.chi = ym.chi[complete.cases(ym.chi$p6a_q_2.x.x, ym.chi$p6a_q_2.y.y),]
ym.adu = ym.adu[complete.cases(ym.adu$p6a_q_2.x.x, ym.adu$p6a_q_2.y.y),]

ym.chi = ym.chi[-which(ym.adu$p6a_q_2.x.x == "no" |
                        ym.chi$p6a_q_2.y.y == "no"),]
ym.adu = ym.adu[-which(ym.adu$p6a_q_2.x.x == "no" |
                        ym.adu$p6a_q_2.y.y == "no"),]

count(ym.chi$crit) #A quantidade de controles e tratados ainda é grande
count(ym.adu$crit) #para a avaliação de impacto na resposta considerada.

rm(ym1234) #Temos agora 4 bancos de dados no ambiente de trabalho.

count(ym1234.c$p6a_q_10.x.x); count(ym1234.a$p6a_q_10.x.x)

count(ym.chi$p6a_q_10.x.x); count(ym.adu$p6a_q_10.x.x)

count(ym1234.c$p6a_q_10.y.y); count(ym1234.a$p6a_q_10.y.y)

count(ym.chi$p6a_q_10.y.y); count(ym.adu$p6a_q_10.y.y)

```

Como a frequência de TSAT está restrita aos indivíduos com respostas positivas na 1ª e 4ª rodadas para a variável `p6a_q_2`, não há respostas que não se aplicam em TSAT. Entretanto, pelo dicionário da YNSPMS, ainda temos valores faltantes nesta variável, representados pelas respostas “98” (Do not remember) e “99” (No response). Logo, qualquer análise envolvendo TSAT deve registrar estas respostas como faltantes (no R, atribuindo o código NA)

```

for(i in 1:nrow(ym1234.c)){
  if(ym1234.c$p6a_q_10.x.x[i] > 97) ym1234.c$p6a_q_10.x.x[i] <- NA
  if(ym1234.c$p6a_q_10.y.y[i] > 97) ym1234.c$p6a_q_10.y.y[i] <- NA}

```

```

for(i in 1:nrow(ym1234.a)){
  if(ym1234.a$p6a_q_10.x.x[i] > 97) ym1234.a$p6a_q_10.x.x[i] <- NA
  if(ym1234.a$p6a_q_10.y.y[i] > 97) ym1234.a$p6a_q_10.y.y[i] <- NA}

for(i in 1:nrow(ym.chi)){
  if(ym.chi$p6a_q_10.x.x[i] > 97) ym.chi$p6a_q_10.x.x[i] <- NA
  if(ym.chi$p6a_q_10.y.y[i] > 97) ym.chi$p6a_q_10.y.y[i] <- NA}

for(i in 1:nrow(ym.adu)){
  if(ym.adu$p6a_q_10.x.x[i] > 97) ym.adu$p6a_q_10.x.x[i] <- NA
  if(ym.adu$p6a_q_10.y.y[i] > 97) ym.adu$p6a_q_10.y.y[i] <- NA}

```

O percentual de respostas faltantes em cada situação é dado por

```

length(which(is.na(ym1234.c$p6a_q_10.x.x)) == T)/nrow(ym1234.c)
length(which(is.na(ym1234.c$p6a_q_10.y.y)) == T)/nrow(ym1234.c)

length(which(is.na(ym1234.a$p6a_q_10.x.x)) == T)/nrow(ym1234.a)
length(which(is.na(ym1234.a$p6a_q_10.y.y)) == T)/nrow(ym1234.a)

length(which(is.na(ym.chi$p6a_q_10.x.x)) == T)/nrow(ym.chi)
length(which(is.na(ym.chi$p6a_q_10.y.y)) == T)/nrow(ym.chi)

length(which(is.na(ym.adu$p6a_q_10.x.x)) == T)/nrow(ym.adu)
length(which(is.na(ym.adu$p6a_q_10.y.y)) == T)/nrow(ym.adu)

```

A proporção de respostas faltantes para a variável TSAT é bem residual e não prejudicará a avaliação de impacto do SWF pelo método DD. Para mostrar que a variável resposta de interesse pode ser tratada como aproximadamente contínua nos grupos etários e nas rodadas consideradas, basta fazer um histograma simples da densidade nesta variável. Tomando agora somente os indivíduos que responderam afirmativamente à p6a\_q\_2 nas duas rodadas consideradas, em cada grupo etário, temos os seguintes histogramas para a variável TSAT quando ignoramos o PAC

```

hist(ym.adu$p6a_q_10.x.x, breaks = 10, main = "", xlab = "TSAT",
     ylab = "Densidade", probability = T, col = "gray",
     border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

```

```
hist(ym.adu$p6a_q_10.y.y, breaks = 10, main = "", xlab = "TSAT",
     ylab = "Densidade", probability = T, col = "gray",
     border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

```
hist(ym.chi$p6a_q_10.x.x, breaks = 10, main = "", xlab = "TSAT",
     ylab = "Densidade", probability = T, col = "gray",
     border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

```
hist(ym.chi$p6a_q_10.y.y, breaks = 10, main = "", xlab = "TSAT",
     ylab = "Densidade", probability = T, col = "gray",
     border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

Caso consideremos o PAC, é necessário antes remover indivíduos que são únicos na sua província (estrato no 1º estágio) ao qual pertencem, do contrário não é possível estimar os totais populacionais em cada intervalo do histograma. Para fazer isso, é suficiente fazer

```
count(ym.adu$f5) #Todos os estratos possuem ao menos dois indivíduos.
count(ym.chi$f5) #Em "shabwa", há apenas uma observação. Vamos retirá-la.
```

```
ym.c.p = ym.chi[-which(ym.chi$f5=="shabwa"),]
```

Isto será feito sempre que algum parâmetro for estimado incorporando os pesos amostrais dos indivíduos. Para criar objetos relativos ao PAC utilizado, necessários para considerá-lo na produção de estimativas, instale e carregue o pacote `survey` (Lumley, 2016)

```
install.packages("survey"); library(survey)

d.re.c = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.c.p)
d.re.a = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.adu)
```

Finalmente, temos os histogramas abaixo sem ignorar o PAC

```
svyhist(~p6a_q_10.x.x, design = d.re.a, breaks = 10, main = "",
       xlab = "TSAT", ylab = "Densidade", probability = T,
       col = "gray", border = "black", axes = F)
```

```
axis(side = 1); axis(side = 2, las = 1, line = -.4)

svyhist(~p6a_q_10.y.y, design = d.re.a, breaks = 10, main = "",
        xlab = "TSAT", ylab = "Densidade", probability = T,
        col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

svyhist(~p6a_q_10.x.x, design = d.re.c, breaks = 10, main = "",
        xlab = "TSAT", ylab = "Densidade", probability = T,
        col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

svyhist(~p6a_q_10.y.y, design = d.re.c, breaks = 10, main = "",
        xlab = "TSAT", ylab = "Densidade", probability = T,
        col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

Temos que a variável TSAT pode ser tratada como aproximadamente contínua em ambos os grupos etários, ainda que originalmente discreta. Esta conclusão é válida tanto na situação em que o PAC é ignorado quanto na em que ele é incorporado. As figuras 2 a 9 representam os histogramas produzidos.

## B.2 Código para o Cálculo de Estimativas

### B.2.1 Ajuste de Modelos para o Escore de Propensão sem PAC

Tomando a variável de atribuição do tratamento como dependente (`crit`), obteremos os coeficientes do respectivo modelo de regressão logística para o escore de propensão e como os valores preditos, dado um conjunto de covariáveis especificadas. A partir de agora, utilizaremos somente os dados dos indivíduos classificados como controle ou tratamento e que trabalhavam na 1<sup>a</sup> e 4<sup>a</sup> rodadas da pesquisa YNSPMS, com pesos amostrais longitudinais para as quatro rodadas. Tanto na situação em que o modelo para o escore de propensão é ajustado ignorando o PAC quanto na em que este é considerado, temos como candidatas ao modelo especificado as seguintes variáveis observadas (com nomes entre parênteses quando originalmente presentes no dicionário de variáveis), no grupo das crianças:

- COM: Criança é Órfã de Mãe;
- COP: Criança é Órfã de Pai;

- CRDCM: Criança Reside em Domicílio Chefiado por Mulher, desde que esta seja adulta e solteira, divorciada ou viúva;
- CRDAI: Criança Reside em Domicílio com Adulto Idoso;
- CRDAPI: Criança Reside em Domicílio com Adulto Portador de Invalidez;
- DRR (*remit*): Domicílio Recebe Remessas (de parentes ou amigos);
- DRC (*charit*): Domicílio Recebe Caridade
- DRFP (*pension*): Domicílio Recebe Fundos de Pensão;
- NPD (*poor*): Nível de Pobreza do Domicílio.

e no grupo dos adultos:

- AMCD: Adulto é Mulher e Chefe do Domicílio, desde que seja solteira, divorciada ou viúva;
- AI: Adulto é Idoso;
- API: Adulto é Portador de Invalidez;
- DRR (*remit*): Domicílio Recebe Remessas (de parentes ou amigos);
- DRC (*charit*): Domicílio Recebe Caridade
- DRFP (*pension*): Domicílio Recebe Fundos de Pensão;
- NPD (*poor*): Nível de Pobreza do Domicílio.

Para todas as variáveis acima, foram consideradas observações apenas na 1<sup>a</sup> rodada (.x.x) para ajuste do modelo para o escore de propensão. Boa parte é derivada de transformações em variáveis originalmente observadas na YNSPMS. Qualquer transformação necessária será feita para as observações na 1<sup>a</sup> e também na 4<sup>a</sup> rodada, pois as variáveis transformadas serão especificadas posteriormente no modelo de regressão linear para o método DD. As variáveis AMCD; AI e API são derivadas de transformações em variáveis originalmente observadas nos adultos da YNSPMS

```
count(ym.adu$p2_q_03.x.x) #Não há valores faltantes para gênero; ...
count(ym.adu$p2_q_04.x.x) #... relação com o chefe domiciliar; ...
count(ym.adu$p2_q_05_y.x.x) #... idade; ...
count(ym.adu$p2_q_11.x.x) #... invalidez na vida diária ...
count(ym.adu$p2_q_16.x.x) #... ou para o estado civil na 1ª rodada.
```

```
count(ym.adu$p2_q_03.y.y) #Não há valores faltantes para gênero; ...
count(ym.adu$p2_q_04.y.y) #... relação com o chefe domiciliar; ...
count(ym.adu$p2_q_05_y.y.y) #... idade; ...
count(ym.adu$p2_q_11.y.y) #... invalidez na vida diária ...
count(ym.adu$p2_q_16.y.y) #... ou para o estado civil na 4ª rodada.
```

```
mun = rep("no", nrow(ym.adu))
mul.nac = which((ym.adu$p2_q_03.x.x == "female" &
```

```
ym.adu$p2_q_04.x.x == " head " & ym.adu$p2_q_05_y.x.x >= 18 &
ym.adu$p2_q_16.x.x == "single") |
(ym.adu$p2_q_03.x.x == "female" & ym.adu$p2_q_04.x.x == " head " &
ym.adu$p2_q_05_y.x.x >= 18 & ym.adu$p2_q_16.x.x == "divorced") |
(ym.adu$p2_q_03.x.x == "female" & ym.adu$p2_q_04.x.x == " head " &
ym.adu$p2_q_05_y.x.x >= 18 & ym.adu$p2_q_16.x.x == "widowed"))
mun = replace(mun, mul.nac, "yes")

mun.4 = rep("no", nrow(ym.adu))
mul.nac.4 = which((ym.adu$p2_q_03.y.y == "female" &
ym.adu$p2_q_04.y.y == " head " & ym.adu$p2_q_05_y.y.y >= 18 &
ym.adu$p2_q_16.y.y == "single") |
(ym.adu$p2_q_03.y.y == "female" & ym.adu$p2_q_04.y.y == " head " &
ym.adu$p2_q_05_y.y.y >= 18 & ym.adu$p2_q_16.y.y == "divorced") |
(ym.adu$p2_q_03.y.y == "female" & ym.adu$p2_q_04.y.y == " head " &
ym.adu$p2_q_05_y.y.y >= 18 & ym.adu$p2_q_16.y.y == "widowed"))
mun.4 = replace(mun.4, mul.nac.4, "yes")

avi = rep("no", nrow(ym.adu))
ava.ida = which((ym.adu$p2_q_03.x.x == "male" &
ym.adu$p2_q_05_y.x.x >= 60) |
(ym.adu$p2_q_03.x.x == "female" & ym.adu$p2_q_05_y.x.x >= 55))
avi = replace(avi, ava.ida, "yes")

avi.4 = rep("no", nrow(ym.adu))
ava.ida.4 = which((ym.adu$p2_q_03.y.y == "male" &
ym.adu$p2_q_05_y.y.y >= 60) |
(ym.adu$p2_q_03.y.y == "female" & ym.adu$p2_q_05_y.y.y >= 55))
avi.4 = replace(avi.4, ava.ida.4, "yes")

inc = rep("no", nrow(ym.adu))
adu.inc = which(ym.adu$p2_q_11.x.x == "yes")
inc = replace(inc, adu.inc, "yes")

inc.4 = rep("no", nrow(ym.adu))
adu.inc.4 = which(ym.adu$p2_q_11.y.y == "yes")
inc.4 = replace(inc.4, adu.inc.4, "yes")

ym.adu = cbind(ym.adu, mun, avi, inc, mun.4, avi.4, inc.4)
```

```
rm(mul.nac, ava.ida, adu.inc, mul.nac.4, ava.ida.4, adu.inc.4)
```

assim como as variáveis CRDCM; CRDAI e CRDAPI, embora estas também utilizem como auxiliares dados da variável indicadora de domicílio (nome `f1`) no banco das crianças para identificar as que residem junto a uma mulher adulta e solteira, divorciada ou viúva, desde que chefe o domicílio; a um aposentado por idade ou a um adulto inválido

```
ym.mun = ym.adu[ym.adu$mun == "yes",]
ym.avi = ym.adu[ym.adu$avi == "yes",]
ym.inc = ym.adu[ym.adu$inc == "yes",]
```

```
ym.mun.4 = ym.adu[ym.adu$mun.4 == "yes",]
ym.avi.4 = ym.adu[ym.adu$avi.4 == "yes",]
ym.inc.4 = ym.adu[ym.adu$inc.4 == "yes",]
```

```
d.mun = ym.mun$f1[!duplicated(ym.mun$f1)]
d.avi = ym.avi$f1[!duplicated(ym.avi$f1)]
d.inc = ym.inc$f1[!duplicated(ym.inc$f1)]
```

```
d.mun.4 = ym.mun.4$f1[!duplicated(ym.mun.4$f1)]
d.avi.4 = ym.avi.4$f1[!duplicated(ym.avi.4$f1)]
d.inc.4 = ym.inc.4$f1[!duplicated(ym.inc.4$f1)]
```

```
cri.mun = match(ym.chi$f1, d.mun)
cri.mun.0 = which(is.na(cri.mun) == T)
cri.mun.1 = which(is.na(cri.mun) == F)
```

```
cri.d.mun = rep("no", nrow(ym.chi))
cri.d.mun = replace(cri.d.mun, cri.mun.1, "yes")
```

```
cri.mun.4 = match(ym.chi$f1, d.mun.4)
cri.mun.0.4 = which(is.na(cri.mun.4) == T)
cri.mun.1.4 = which(is.na(cri.mun.4) == F)
```

```
cri.d.mun.4 = rep("no", nrow(ym.chi))
cri.d.mun.4 = replace(cri.d.mun.4, cri.mun.1.4, "yes")
```

```
cri.avi = match(ym.chi$f1, d.avi)
cri.avi.0 = which(is.na(cri.avi) == T)
cri.avi.1 = which(is.na(cri.avi) == F)
```

```
cri.d.avi = rep("no", nrow(ym.chi))
cri.d.avi = replace(cri.d.avi, cri.avi.1, "yes")

cri.avi.4 = match(ym.chi$f1, d.avi.4)
cri.avi.0.4 = which(is.na(cri.avi.4) == T)
cri.avi.1.4 = which(is.na(cri.avi.4) == F)

cri.d.avi.4 = rep("no", nrow(ym.chi))
cri.d.avi.4 = replace(cri.d.avi.4, cri.avi.1.4, "yes")

cri.inc = match(ym.chi$f1, d.inc)
cri.inc.0 = which(is.na(cri.inc) == T)
cri.inc.1 = which(is.na(cri.inc) == F)

cri.d.inc = rep("no", nrow(ym.chi))
cri.d.inc = replace(cri.d.inc, cri.inc.1, "yes")

cri.inc.4 = match(ym.chi$f1, d.inc.4)
cri.inc.0.4 = which(is.na(cri.inc.4) == T)
cri.inc.1.4 = which(is.na(cri.inc.4) == F)

cri.d.inc.4 = rep("no", nrow(ym.chi))
cri.d.inc.4 = replace(cri.d.inc.4, cri.inc.1.4, "yes")
```

Por sua vez, as variáveis COM e COP são o exato complementar de p2\_q\_07 e p2\_q\_09, que são indicadoras da mãe e do pai do indivíduo (criança) estarem vivos, respectivamente. Como é nosso desejo testar se ser órfão de mãe ou pai explica a variabilidade do escore a ser predito, pois haver crianças órfãs de mãe ou pai no domicílio é um dos critérios de elegibilidade do SWF, temos

```
count(ym.chi$p2_q_07.x.x); count(ym.chi$p2_q_09.x.x)
count(ym.chi$p2_q_07.y.y); count(ym.chi$p2_q_09.y.y)

aus.mae = rep(1, nrow(ym.chi))
c.mae = which(ym.chi$p2_q_07.x.x == "yes")
aus.mae = replace(aus.mae, c.mae, 0)

aus.mae.4 = rep(1, nrow(ym.chi))
c.mae.4 = which(ym.chi$p2_q_07.y.y == "yes")
```

```
aus.mae.4 = replace(aus.mae.4, c.mae.4, 0)
```

```
aus.pai = rep(1, nrow(ym.chi))
c.pai = which(ym.chi$p2_q_09.x.x == "yes")
aus.pai = replace(aus.pai, c.pai, 0)
```

```
aus.pai.4 = rep(1, nrow(ym.chi))
c.pai.4 = which(ym.chi$p2_q_09.y.y == "yes")
aus.pai.4 = replace(aus.pai.4, c.pai.4, 0)
```

Finalmente, temos as variáveis necessárias no ajuste do modelo para as crianças

```
ym.chi = cbind(ym.chi, cri.d.mun, cri.d.avi, cri.d.inc, aus.mae,
              aus.pai, cri.d.mun.4, cri.d.avi.4, cri.d.inc.4,
              aus.mae.4, aus.pai.4)

ym.c.p = ym.chi[-which(ym.chi$f5=="shabwa"),] #Incluimos novas variáveis
                                             #no banco original.

rm(avi, mun, inc, avi.4, mun.4, inc.4)
rm(d.avi, d.mun, d.inc, d.avi.4, d.mun.4, d.inc.4)
rm(c.mae, c.pai, c.mae.4, c.pai.4)
rm(ym.mun, ym.avi, ym.inc, ym.mun.4, ym.avi.4, ym.inc.4)
rm(mul.nac, ava.ida, adu.inc, mul.nac.4, ava.ida.4, adu.inc.4)
rm(cri.avi.0, cri.avi.1, cri.mun.0, cri.mun.1, cri.inc.0, cri.inc.1,
   cri.avi.0.4, cri.avi.1.4, cri.mun.0.4, cri.mun.1.4,
   cri.inc.0.4, cri.inc.1,.4)
rm(cri.avi, cri.mun, cri.inc, cri.avi.4, cri.mun.4, cri.inc.4)
rm(cri.d.avi, cri.d.mun, cri.d.inc, cri.d.avi.4,
   cri.d.mun.4, cri.d.inc.4)
```

Nas demais covariáveis candidatas, em ambos os grupos etários e para a 1ª rodada da YNSPMS, também não há valores faltantes. Para a 4ª rodada, são muito raros os indivíduos que não responderam a algumas delas

```
count(ym.chi$poor.x.x); count(ym.chi$remit.x.x)
count(ym.chi$charit.x.x); count(ym.chi$pension.x.x)

count(ym.chi$poor.y.y); count(ym.chi$remit.y.y)
count(ym.chi$charit.y.y); count(ym.chi$pension.y.y)
```

```
count(ym.adu$remit.x.x); count(ym.adu$charit.x.x)
count(ym.adu$pension.x.x); count(ym.adu$poor.x.x)
```

```
count(ym.adu$remit.y.y); count(ym.adu$charit.y.y)
count(ym.adu$pension.y.y); count(ym.adu$poor.y.y)
```

Por outro lado, as covariáveis `p6_a_12` e `p6_a_13`, embora meçam diretamente a renda do indivíduo pesquisado (pagamento em dinheiro e em espécie, respectivamente), possuem uma alta taxa de valores faltantes em ambos os grupos e rodadas. Por esta razão, foram ignoradas para a modelagem

```
count(ym.chi$p6a_q_12.x.x); count(ym.chi$p6a_q_13.x.x)
count(ym.chi$p6a_q_12.y.y); count(ym.chi$p6a_q_13.y.y)
```

```
count(ym.adu$p6a_q_12.x.x); count(ym.adu$p6a_q_13.x.x)
count(ym.adu$p6a_q_12.y.y); count(ym.adu$p6a_q_13.y.y)
```

Após estas modificações preliminares, podemos selecionar quais covariáveis devem ser especificadas a partir de todas as candidatas. Começando pelo grupo das crianças na situação em que as observações são supostas IID, ao incluir todas as covariáveis candidatas em ordem crescente de variabilidade interna, temos que

```
count(ym.c$aus.mae); count(ym.c$aus.pai); count(ym.c$cri.d.mun)
count(ym.c$cri.d.avi); count(ym.c$cri.d.inc); count(ym.c$remit.x.x)
count(ym.c$charit.x.x); count(ym.c$pension.x.x); count(ym.c$poor.x.x)
```

```
m1.c = glm(crit ~ cri.d.mun + aus.mae + pension.x.x + aus.pai +
           charit.x.x + cri.d.avi + cri.d.inc + remit.x.x + poor.x.x,
           family = binomial(link = logit), data = ym.chi)
```

```
m1.c.rv = anova(m1.c, test = "LRT"); xtable(m1.c.rv, digits = 4)
```

Dado o nível nominal  $\alpha = 5\%$ , apenas DRC e NPD são significativas neste primeiro modelo, com a segunda variável tendo valor-p apenas um pouco menor do que o nível  $\alpha$ . Desta forma, vamos ajustar um modelo que especifique somente as duas covariáveis

```
m2.c = glm(crit ~ charit.x.x + poor.x.x,
           family = binomial(link = logit), data = ym.chi)
```

```
anova(m2.c, test = "LRT")
```

Agora, apenas a variável DRC é significativa ao nível de 5%. Isto quer dizer que NPD só será significativa quando incluirmos mais covariáveis (não significativas) no modelo, tornando-o menos parcimonioso. Neste caso, será mais fácil rejeitar a hipótese de balanceamento posteriormente no banco pareado. Por esta razão, somente a permanência da variável DRC é dada como certa no modelo para o escore de propensão para as crianças quando ignoramos o PAC, pois ela continua significativa mesmo se isolada das demais

```
m3.c = glm(crit ~ charit.x.x, family = binomial(link = logit),
           data = ym.chi)

m3.c.rv = anova(m3.c, test = "LRT"); xtable(m3.c.rv, digits = 4)

m3.c.sm = summary(m3.c); xtable(m3.c.sm, digits = 4)
```

Para verificar se as demais covariáveis candidatas, não significativas no primeiro modelo, o serão se cada uma for especificada de forma isolada com DRC, considere os seguintes modelos especificados

```
m4.c.1 = glm(crit ~ cri.d.mun + charit.x.x,
             family = binomial(link = logit), data = ym.chi)

m4.c.2 = glm(crit ~ aus.mae + charit.x.x,
             family = binomial(link = logit), data = ym.chi)

m4.c.3 = glm(crit ~ pension.x.x + charit.x.x,
             family = binomial(link = logit), data = ym.chi)

m4.c.4 = glm(crit ~ aus.pai + charit.x.x,
             family = binomial(link = logit), data = ym.chi)

m4.c.5 = glm(crit ~ charit.x.x + cri.d.avi,
             family = binomial(link = logit), data = ym.chi)

m4.c.6 = glm(crit ~ charit.x.x + cri.d.inc,
             family = binomial(link = logit), data = ym.chi)

m4.c.7 = glm(crit ~ charit.x.x + remit.x.x,
             family = binomial(link = logit), data = ym.chi)

anova(m4.c.1, test = "LRT"); anova(m4.c.2, test = "LRT")
```

```
anova(m4.c.3, test = "LRT"); anova(m4.c.4, test = "LRT")
anova(m4.c.5, test = "LRT"); anova(m4.c.6, test = "LRT")
anova(m4.c.7, test = "LRT")
```

Como nenhuma das sete covariáveis foi significativa, mantemos o modelo `m3.c`, especificando apenas DRC (além do intercepto), para produzir os escores preditos das crianças sem considerar o PAC. Antes de lidar com os escores, vamos remover os demais modelos

```
rm(m1.c, m2.c, m4.c.1, m4.c.2, m4.c.3, m4.c.4, m4.c.5, m4.c.6, m4.c.7)
```

Vamos verificar se os escores preditos pelo modelo especificado possuem suporte estritamente entre 0 e 1

```
esc.c = fitted.values(m3.c); count(esc.c)

hist(esc.c, breaks = 10, main = "", xlab = "Escores Preditos",
     ylab = "Frequência", col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1)

ym.chi = cbind(ym.chi, esc.c)
```

Há apenas dois valores possíveis para os escores, mas ambos estão distantes de zero e um (em verdade, entre 0.1 e 0.5), atendendo aos dois pressupostos citados. Estes são os escores a serem usados no pareamento das crianças que afirmaram estar trabalhando na 1ª e 4ª rodada da YNSPMS, ignorando completamente o PAC.

Para os adultos, ainda sem considerar o PAC, temos ao incluir todas as covariáveis candidatas que

```
count(ym.a$mun); count(ym.a$avi); count(ym.a$inc)
count(ym.a$remit.x.x); count(ym.a$charit.x.x)
count(ym.a$pension.x.x); count(ym.a$poor.x.x)

m1.a = glm(crit ~ mun + inc + pension.x.x + avi + charit.x.x +
          remit.x.x + poor.x.x, family = binomial(link = logit),
          data = ym.adu)

m1.a.rv = anova(m1.a, test = "LRT"); xtable(m1.a.rv, digits = 4)
```

Temos quatro covariáveis significativas ao nível nominal de 5%: API; DRC; DRFP e NPD. Assim como feito para as crianças, vamos isolá-las para saber se todas continuam significativas

```
m2.a = glm(crit ~ inc + pension.x.x + charit.x.x + poor.x.x,
           family = binomial(link = logit), data = ym.adu)
```

```
m2.a.rv = anova(m2.a, test = "LRT"); xtable(m2.a.rv, digits = 4)
```

```
m2.a.sm = summary(m2.a); xtable(m2.a.sm, digits = 4)
```

Ao contrário do que aconteceu no modelo para as crianças, não é necessário retirar mais nenhuma covariável. Agora, vamos incluir neste segundo modelo, uma por vez, as covariáveis AMCD; AI e DRR, que não foram significativas no primeiro modelo para os adultos

```
m3.a.1 = glm(crit ~ mun + inc + pension.x.x + charit.x.x + poor.x.x,
             family = binomial(link = logit), data = ym.adu)
```

```
m3.a.2 = glm(crit ~ inc + pension.x.x + avi + charit.x.x + poor.x.x,
             family = binomial(link = logit), data = ym.adu)
```

```
m3.a.3 = glm(crit ~ inc + pension.x.x + charit.x.x + remit.x.x + poor.x.x,
             family = binomial(link = logit), data = ym.adu)
```

```
anova(m3.a.1, test = "LRT"); anova(m3.a.2, test = "LRT")
```

```
anova(m3.a.3, test = "LRT")
```

Mesmo nas especificações acima, AMCD; AI e DRR continuam não sendo significativas a 5%. Logo, mantemos o modelo `m2.a` para produzir os escores preditos dos adultos sem incorporar qualquer característica do PAC. Novamente, antes de verificar os escores preditos, vamos remover os demais modelos para os adultos

```
rm(m1.a, m3.a.1, m3.a.2, m3.a.3)
```

Tomando os escores preditos nos adultos (sem PAC), temos que o PATFI é respeitado e o SWF é fortemente ignorável nas quatro covariáveis especificadas

```
esc.a = fitted.values(m2.a); count(esc.a)
```

```
hist(esc.a, breaks = 10, main = "", xlab = "Escores Preditos",
```

```
ylab = "Frequência", col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

```
ym.adu = cbind(ym.adu, esc.a)
```

Com mais valores distintos nos escores preditos (26 ao todo), há uma tendência maior de aparecer valores extremos no modelo para os adultos, considerando que uma das covariáveis especificadas (NPD) tem quatro categorias. Porém, os valores dos escores ainda estão entre 0.1 e 0.5, portanto bem longe dos limites para sua distribuição.

Finalizados os modelos para o escore de propensão em ambos os grupos etários no caso IID, prosseguiremos para a situação em que o PAC é incorporado na modelagem.

## B.2.2 Ajuste de Modelos para o Escore de Propensão com PAC

Sempre que formos incorporar o PAC, faremos uso do objeto relativo a este criado na subseção B.1.2. Nos modelos para o escore de propensão e na avaliação de impacto do SWF, temos o interesse de estimar o EPAA para saber o quanto se perde ao ignorar (indevidamente) o PAC. Desta forma, ele será impresso junto com o erro-padrão das estimativas dos coeficientes, o qual será aproximado através da LST (opção padrão no pacote `survey` do R).

Como inserimos novas variáveis no banco original (para crianças e adultos), devemos recriar os objetos relativos ao PAC de cada grupo (apenas o argumento `data` de `svydesign` se altera na prática). Tomando novamente as mesmas candidatas ao modelo definitivo para o escore no caso IID e na mesma ordem de especificação, inicialmente para o grupo das crianças (apenas um indivíduo foi removido), temos

```
d.re.c = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.c.p)
```

```
m1.cp = svyglm(crit ~ cri.d.mun + aus.mae + pension.x.x + aus.pai +
              charit.x.x + cri.d.avi + cri.d.inc + remit.x.x + poor.x.x,
              family = binomial(link = logit), design = d.re.c)
```

```
names(m1.cp); anova(m1.cp, deff = T); m1.cp$null.deviance
```

No pacote `survey`, a função `svyglm` é a responsável pela estimação por MP dos coeficientes do modelo para o escore de propensão. No momento, o pacote `xtable` não tem suporte para a classe dos objetos gerados pelas funções do pacote `survey`. Por padrão, a análise de desvio é feita pelo teste da RP, a versão do teste da RV adaptada para dados amostrais complexos. Neste primeiro ajuste, temos duas covariáveis significativas ao nível

nominal de 5%: DRC e CRDAI. Além disso, em quatro destas covariáveis (CRDCM; COM; COP e CRDAPI) o EPAA estimado é menor do que o valor unitário. Isto indica que o PAC utilizado na YNSPMS produziu ganhos de eficiência na estimação dos respectivos coeficientes (com relação a um plano AAS), mesmo desconsiderando a informação de estratificação no 2º estágio. Mantendo apenas as covariáveis significativas, temos

```
m2.cp = svyglm(crit ~ charit.x.x + cri.d.avi, design = d.re.c,  
              family = binomial (link = logit))
```

```
names(m2.cp); anova(m2.cp, deff = T)  
m2.cp$null.deviance; summary(m2.cp)
```

Novamente, as duas covariáveis continuam significativas e permanecerão na especificação do modelo final. Testando as demais candidatas, uma por vez, obtemos

```
m3.cp.1 = svyglm(crit ~ cri.d.mun + charit.x.x + cri.d.avi,  
                design = d.c, family = binomial (link = logit))
```

```
m3.cp.2 = svyglm(crit ~ aus.mae + charit.x.x + cri.d.avi,  
                design = d.c, family = binomial (link = logit))
```

```
m3.cp.3 = svyglm(crit ~ pension.x.x + charit.x.x + cri.d.avi,  
                design = d.c, family = binomial (link = logit))
```

```
m3.cp.4 = svyglm(crit ~ aus.pai + charit.x.x + cri.d.avi,  
                design = d.c, family = binomial (link = logit))
```

```
m3.cp.5 = svyglm(crit ~ charit.x.x + cri.d.avi + cri.d.inc,  
                design = d.c, family = binomial (link = logit))
```

```
m3.cp.6 = svyglm(crit ~ charit.x.x + cri.d.avi + remit.x.x,  
                design = d.c, family = binomial (link = logit))
```

```
m3.cp.7 = svyglm(crit ~ charit.x.x + cri.d.avi + poor.x.x,  
                design = d.c, family = binomial (link = logit))
```

```
anova(m3.cp.1, deff = T); anova(m3.cp.2, deff = T);  
anova(m3.cp.3, deff = T); anova(m3.cp.4, deff = T);  
anova(m3.cp.5, deff = T); anova(m3.cp.6, deff = T);  
anova(m3.cp.7, deff = T)
```

Quando adicionadas uma por vez, nenhuma das demais variáveis é significativa ao nível nominal de 5% quando especificada conjuntamente às covariáveis do modelo `m2.cp`. Portanto, para modelar o escore de propensão das crianças, incorporando devidamente o PAC, adotamos o modelo `m2.cp` e removemos os demais

```
rm(m1.cp, m3.cp.1, m3.cp.2, m3.cp.3, m3.cp.4, m3.cp.5, m3.cp.6, m3.cp.7)
```

Para obter o EPAA estimado dos coeficientes da regressão, também precisamos das estimativas de suas variâncias (quadrado do erro-padrão) quando ignoramos o PAC no mesmo modelo, com os mesmos dados

```
m2.cp.s = glm(crit ~ charit.x.x + cri.d.avi, data = ym.c.p,
              family = binomial (link = logit))
```

```
summary(m2.cp.s)
```

Tomando os escores preditos do modelo escolhido, para saber se o PATFI é atendido, fazemos

```
esc.cp = fitted.values(m2.cp); count(esc.cp)
```

```
hist(esc.cp, breaks = 10, main = "", xlab = "Escore Preditos",
     ylab = "Frequência", col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

Todos os escores preditos estão acima de 0.01 e abaixo de 0.99, logo o PATFI é respeitado. Tomamos os seus escores preditos e removemos os demais modelos

```
esc.cp = fitted.values(m2.cp); ym.c.p = cbind(ym.c.p, esc.cp)
```

```
rm(m1.cp, m3.cp.1, m3.cp.2, m3.cp.3, m3.cp.4, m3.cp.5, m3.cp.6, m3.cp.7)
```

Por fim, falta apenas modelar os escores preditos para os adultos incorporando o plano amostral, com as respectivas covariáveis candidatas

```
d.re.a = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                  nest = TRUE, data = ym.adu)
```

```
m1.ap = svyglm(crit ~ mun + inc + pension.x.x + avi + charit.x.x +
               remit.x.x + poor.x.x, design = d.re.a,
               family = binomial(link = logit))
```

```
anova(m1.ap, deff = T); m1.ap$null.deviance; summary(m1.ap)
```

No modelo com todas as covariáveis candidatas, somente API; DRC e NPD são significativas ao nível nominal de 5%. Dentre estas três covariáveis, apenas API tem EPAA estimado abaixo de 1. Mantendo as covariáveis significativas e retirando as demais, temos

```
m2.ap = svyglm(crit ~ inc + charit.x.x + poor.x.x, design = d.re.a,  
              family = binomial(link = logit))
```

```
anova(m2.ap, deff = T); m2.ap$null.deviance; summary(m2.ap)
```

As três covariáveis, isoladas, permanecem significativas. Inserindo uma por vez as demais covariáveis, temos os modelos abaixo

```
m3.ap.1 = svyglm(crit ~ mun + inc + charit.x.x + poor.x.x,  
                design = d.re.a, family = binomial(link = logit))
```

```
m3.ap.2 = svyglm(crit ~ inc + pension.x.x + charit.x.x + poor.x.x,  
                design = d.re.a, family = binomial(link = logit))
```

```
m3.ap.3 = svyglm(crit ~ inc + avi + charit.x.x + poor.x.x,  
                design = d.re.a, family = binomial(link = logit))
```

```
m3.ap.4 = svyglm(crit ~ inc + charit.x.x + remit.x.x + poor.x.x,  
                design = d.re.a, family = binomial(link = logit))
```

```
anova(m3.ap.1, deff = T); anova(m3.ap.2, deff = T)
```

```
anova(m3.ap.3, deff = T); anova(m3.ap.4, deff = T)
```

Mesmo incluídas uma por vez, nenhuma das quatro variáveis passará a ser significativa ao nível nominal de 5%. Ficamos então com o modelo `m2.ap`. Novamente, para obter o EPAA estimado dos coeficientes da regressão precisaremos das estimativas de suas variâncias (quadrado do erro-padrão) quando ignoramos o PAC no mesmo modelo, com os mesmos dados

```
m2.ap.s = glm(crit ~ inc + charit.x.x + poor.x.x, data = ym.a,  
              family = binomial(link = logit))
```

```
summary(m2.ap.s)
```

Note que o intercepto deixou de ser significativo quando incorporamos o PAC, logo ele não pode ser considerado para a interpretação dos coeficientes. Tomando os escores preditos por este para fazer o pareamento e balanceamento no grupo dos adultos, considerando o PAC

```
esc.ap = fitted.values(m2.ap); count(esc.ap)

hist(esc.ap, breaks = 10, main = "", xlab = "Escores Preditos",
     ylab = "Frequência", col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4);

ym.adu = cbind(ym.adu, esc.ap)

rm(m1.ap, m3.ap.1, m3.ap.2, m3.ap.3, m3.ap.4)
```

Antes de parear cada grupo etário em cada perspectiva (com e sem PAC), o que será feito nas subseções [B.2.3](#) e [B.2.4](#), vamos por praticidade renomear os quatro modelos escolhidos

```
mc = m3.c; ma = m2.a; mcp = m2.cp; map = m2.ap
rm(m3.c, m2.a, m2.cp, m2.ap)
```

### B.2.3 Pareamento e Balanceamento dos Modelos Ajustados sem PAC

Escolhidos os modelos para crianças e adultos na subseção [B.2.1](#), vamos pará-los usando o PDE com reposição de indivíduos controles selecionados; um controle pareado por tratado e com nível máximo de tolerância igual a 0.2 vezes o desvio padrão dos escores preditos pelo modelo de regressão logística. No R, usaremos o pacote `Matching` (Sekhon, 2011), que possui uma ampla variedade de configurações, para pareamento dos dados pelos escores preditos. Se houver empates entre dois ou mais controles, o pareamento será aleatório, motivando a implantação de uma semente para tornar os resultados reproduzíveis a cada vez que parearmos

```
install.packages("Matching"); library(Matching);

set.seed(2017)
par.c = Match(Tr = ym.chi$crit, X = esc.c, M = 1, caliper = 0.2,
             replace = T, ties = F)

par.c$index.control; which(duplicated(par.c$index.control) == F)
```

```
set.seed(2017)
par.a = Match(Tr = ym.adu$crit, X = esc.a, M = 1, caliper = 0.2,
              replace = T, ties = F)

par.a$index.control; which(duplicated(par.a$index.control) == F)

par.c$ndrops.matches #Todos os indivíduos tratados são pareados.
par.a$ndrops.matches #Cinco tratados não são pareados.

ym.par.c = ym.chi[c(par.c$index.treated, par.c$index.control),]
ym.par.a = ym.adu[c(par.a$index.treated, par.a$index.control),]
```

Tendo o banco pareado para ambos os grupos etários (sem considerar o PAC), é necessário checar antes do balanceamento se o suporte dos escores preditos é comum para indivíduos tratados e controles

```
par(mfrow = c(1,2))

hist(ym.chi[c(par.c$index.treated),]$esc.c, breaks = 10, main = "",
     xlab = "Escore dos Tratados Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1)

hist(ym.chi[c(par.c$index.control),]$esc.c, breaks = 10, main = "",
     xlab = "Escore dos Controles Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1)

par(mfrow = c(1,2))

hist(ym.adu[c(par.a$index.treated),]$esc.a, breaks = 10, main = "",
     xlab = "Escore dos Tratados Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

hist(ym.adu[c(par.a$index.control),]$esc.a, breaks = 10, main = "",
     xlab = "Escore dos Controles Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

Como controles e tratados têm suporte comum para os escores preditos dentro de cada grupo etário, podemos prosseguir para os testes de balanceamento no banco pareado:  $\chi^2$  com correção de 2ª ordem de Rao-Scott  $F$ -ajustada (nas covariáveis categóricas) e MW (nos escores preditos). A título de comparação, os mesmos testes são realizados nos bancos originais, para mostrar que estes eram desbalanceados antes do pareamento. Também foram realizados testes  $t$  para duas amostras nos escores preditos, ainda que não fossem usados para verificar o balanceamento por serem mais liberais do que os testes MW. Começando pelo grupo das crianças, o qual possui uma única covariável (DRC), temos

```
chisq.test(ym.chi$crit, ym.chi$charit.x.x, correct = F)$statistic
chisq.test(ym.chi$crit, ym.chi$charit.x.x, correct = F)$p.value
```

```
kruskal.test(esc.c ~ crit, data = ym.chi)$statistic
kruskal.test(esc.c ~ crit, data = ym.chi)$p.value
```

```
t.test(esc.c ~ crit, data = ym.chi)
```

```
chisq.test(ym.par.c$crit, ym.par.c$charit.x.x, correct = F)$statistic
chisq.test(ym.par.c$crit, ym.par.c$charit.x.x, correct = F)$p.value
```

```
kruskal.test(esc.c ~ crit, data = ym.par.c)$statistic
kruskal.test(esc.c ~ crit, data = ym.par.c)$p.value
```

```
t.test(esc.c ~ crit, data = ym.par.c)
```

Como os testes  $\chi^2$  e MW (caso particular do teste  $H$  de Kruskal-Wallis para duas amostras) não rejeitam as hipóteses nula de independência entre as proporções de DRC por nível de RSWF e homogeneidade nos escores preditos, respectivamente, temos que o banco pareado pelos escores preditos do modelo com DRC como única covariável é balanceado. Observe que a conclusão não é a mesma para o banco completo das crianças.

De forma análoga para os adultos, agora com quatro covariáveis categóricas a serem testadas, temos

```
chisq.test(ym.adu$crit, ym.adu$inc, correct = F)$statistic
chisq.test(ym.adu$crit, ym.adu$pension.x.x, correct = F)$statistic
chisq.test(ym.adu$crit, ym.adu$charit.x.x, correct = F)$statistic
chisq.test(ym.adu$crit, ym.adu$poor.x.x, correct = F)$statistic
```

```
chisq.test(ym.adu$crit, ym.adu$inc, correct = F)$p.value
chisq.test(ym.adu$crit, ym.adu$pension.x.x, correct = F)$p.value
```

```

chisq.test(ym.adu$crit, ym.adu$charit.x.x, correct = F)$p.value
chisq.test(ym.adu$crit, ym.adu$poor.x.x, correct = F)$p.value

kruskal.test(esc.a ~ crit, data = ym.adu)$statistic
kruskal.test(esc.a ~ crit, data = ym.adu)$p.value

t.test(esc.a ~ crit, data = ym.adu)

chisq.test(ym.par.a$crit, ym.par.a$inc, correct = F)$statistic
chisq.test(ym.par.a$crit, ym.par.a$charit.x.x, correct = F)$statistic
chisq.test(ym.par.a$crit, ym.par.a$pension.x.x, correct = F)$statistic
chisq.test(ym.par.a$crit, ym.par.a$poor.x.x, correct = F)$statistic

chisq.test(ym.par.a$crit, ym.par.a$inc, correct = F)$p.value
chisq.test(ym.par.a$crit, ym.par.a$charit.x.x, correct = F)$p.value
chisq.test(ym.par.a$crit, ym.par.a$pension.x.x, correct = F)$p.value
chisq.test(ym.par.a$crit, ym.par.a$poor.x.x, correct = F)$p.value

kruskal.test(esc.a ~ crit, data = ym.par.a)$statistic
kruskal.test(esc.a ~ crit, data = ym.par.a)$p.value

t.test(esc.a ~ crit, data = ym.par.a)

```

Também para o banco pareado dos adultos, sem considerar o PAC, o balanceamento é confirmado pelos testes  $\chi^2$  (nas quatro covariáveis) e MW (nos escores preditos), o que não ocorria no banco completo. Com o balanceamento verificado em ambos os grupos etários (ignorando o PAC), é recomendável salvar os bancos pareados e fazer uma pausa antes de recarregá-los novamente para a avaliação o impacto do SWF em TSAT nos respectivos grupos etários

```

write.csv(ym.par.c, "ynspmsParCriSemPA.csv")
write.csv(ym.par.a, "ynspmsParAduSemPA.csv")

```

## B.2.4 Pareamento e Balanceamento dos Modelos Ajustados com PAC

De forma análoga ao que foi feito acima, tomemos agora os modelos ajustados na subseção B.2.2 para pareamento e verificação do balanceamento no banco pareado, mas incorporando o PAC. Note que o banco das crianças terá uma observação a menos em relação ao caso IID. O banco dos adultos ainda é o mesmo, mas o vetor de escores preditos usado para pareamento é outro

```

set.seed(2017)
par.cp = Match(Tr = ym.c.p$crit, X = esc.cp, M = 1, caliper = 0.2,
               replace = T, ties = F)

set.seed(2017)
par.ap = Match(Tr = ym.adu$crit, X = esc.ap, M = 1, caliper = 0.2,
               replace = T, ties = F)

par.cp$index.control; which(duplicated(par.cp$index.control) == F)
par.ap$index.control; which(duplicated(par.ap$index.control) == F)

par.cp$ndrops.matches #Todos os indivíduos tratados são pareados.
par.ap$ndrops.matches #Todos os indivíduos tratados são pareados.

ym.par.cp = ym.c.p[c(par.cp$index.treated, par.cp$index.control),]
ym.par.ap = ym.adu[c(par.ap$index.treated, par.ap$index.control),]

```

Como estamos considerando o PAC, um objeto alternativo que contenha as suas informações apenas para os dados pareados deve ser criado, para ambos os grupos etários. Antes de criá-lo, devemos remover as observações que são únicas em suas províncias (1º nível de estratificação) no banco pareado, como feito para o banco completo

```

count(ym.par.cp$f5) #Apenas um indivíduo representa a província de Aden.
count(ym.par.ap$f5) #No grupo dos adultos, todas as províncias têm dois
                   #ou mais indivíduos representando cada uma.

which(ym.par.cp$f5 == "aden") #Indivíduo da linha 144 do banco pareado,
                              #que é controle.

ym.par.cp = ym.par.cp[-c(28, 144),] #Como o banco é pareado, eliminamos
                                   #também os contrafatuais.

count(ym.par.cp$f5) #Todas as 14 províncias restantes têm pelo menos 2
                   #indivíduos que as representam cada uma.

ym.par.cp$crit #Os indivíduos tratados agora estão entre as linhas 1 e
               #117 (inclusive).

```

Criando o objeto que contém as informações do PAC para os bancos completos (pois inserimos vetores de escores preditos) e pareados de crianças e adultos

```
d.re.c = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.c.p)
d.re.a = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.adu)

d.par.c = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.par.cp)
d.par.a = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.par.ap)
```

Verificando que o suporte dos escores preditos é comum para indivíduos tratados e controles

```
par(mfrow = c(1,2))

hist(ym.par.cp[c(1:117),]$esc.cp, breaks = 10, main = "",
     xlab = "Escore dos Tratados Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

hist(ym.par.cp[c(118:234),]$esc.cp, breaks = 10, main = "",
     xlab = "Escore dos Controles Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

par(mfrow = c(1,2))

hist(ym.adu[c(par.ap$index.treated),]$esc.ap, breaks = 10, main = "",
     xlab = "Escore dos Tratados Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

hist(ym.adu[c(par.ap$index.control),]$esc.ap, breaks = 10, main = "",
     xlab = "Escore dos Controles Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

Podemos enfim fazer os testes de balanceamento considerando o PAC da pesquisa YNSPMS. Iniciando novamente pelo grupo das crianças, temos que

```

svychisq(~crit + charit.x.x, design = d.re.c)$statistic
svychisq(~crit + cri.d.avi, design = d.re.c)$statistic
svychisq(~crit + charit.x.x, design = d.re.c)$p.value
svychisq(~crit + cri.d.avi, design = d.re.c)$p.value

svyranktest(esc ~ crit, design = d.re.c, test = "wilcoxon")$statistic
svyranktest(esc ~ crit, design = d.re.c, test = "wilcoxon")$p.value

svychisq(~crit + charit.x.x, design = d.par.c)$statistic
svychisq(~crit + cri.d.avi, design = d.par.c)$statistic
svychisq(~crit + charit.x.x, design = d.par.c)$p.value
svychisq(~crit + cri.d.avi, design = d.par.c)$p.value

svyranktest(esc ~ crit, design = d.par.c, test = "wilcoxon")$statistic
svyranktest(esc ~ crit, design = d.par.c, test = "wilcoxon")$p.value

svyttest(esc ~ crit, design = d.re.c)

svyttest(esc ~ crit, design = d.par.c)

```

Para o banco pareado, a covariável CRDAI não está balanceada nas classes de indivíduos tratados e controles pareados pelo score de propensão, pois o teste  $\chi^2$  rejeita a hipótese nula de independência ao nível nominal de 5%. Neste caso, teremos que refazer a análise desde a modelagem para o score de propensão, mas agora especificando somente DRC como covariável dentre as presentes no modelo `m2.cp`. Temos então que

```

mcp.1 = svyglm(crit ~ charit.x.x, family = binomial(link = logit),
              design = d.re.c)

summary(mcp.1); esc.cp.1 = fitted.values(mcp.1); count(esc.cp.1)

hist(esc.cp.1, breaks = 10, main = "", xlab = "Escores Preditos",
     ylab = "Frequência", col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

ym.c.p = cbind(ym.c.p, esc.cp.1)

```

O intercepto e a covariável DRC continuam significativos e o PATFI é respeitado. É muito provável que DRC esteja balanceada ao final dos testes de balanceamento no banco pareado. Para confirmar isto, fazemos

```
rm(mcp.1)

set.seed(2017)
par.cp.1 = Match(Tr = ym.c.p$crit, X = esc.cp.1, M = 1, caliper = 0.2,
                replace = T, ties = F)

par.cp.1$ndrops.matches #Todos os indivíduos tratados são pareados.

par.cp.1$index.control; which(duplicated(par.cp.1$index.control) == F);

ym.par.cp.1 = ym.c.p[c(par.cp.1$index.treated, par.cp.1$index.control),]

count(ym.par.cp.1$f5) #No novo banco pareado, temos um único in-
                    #divíduo pertencente à província da Cidade
                    #de Sana'a.

which(ym.par.cp.f5 == "sana'a city") #Indivíduo da linha 6 do banco
                                    #pareado, que é tratado.

ym.par.cp = ym.par.cp[-c(6, 124),] #Como o banco é pareado, eliminamos
                                   #também os contrafatuais.

count(ym.par.cp.f5) #Todas as 13 províncias restantes têm pelo menos 2
                   #indivíduos que as representam cada uma.

ym.par.cp$crit #Os indivíduos tratados agora estão entre as linhas 1 e
               #117, inclusive.

par(mfrow = c(1,2))

hist(ym.par.cp[c(1:117),]$esc.cp.1, breaks = 10, main = "",
     xlab = "Escores dos Tratados Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)

hist(ym.par.cp[c(118:234),]$esc.cp.1, breaks = 10, main = "",
     xlab = "Escores dos Controles Pareados", ylab = "Frequência",
     col = "gray", border = "black", axes = F)
axis(side = 1); axis(side = 2, las = 1, line = -.4)
```

Recriando os objetos do plano amostral para cada banco e fazendo os testes de balanceamento com o novo modelo, temos

```
d.re.c = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.c.p)

d.par.c = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                  nest = TRUE, data = ym.par.cp.1)

svychisq(~crit + charit.x.x, design = d.re.c)$statistic
svychisq(~crit + charit.x.x, design = d.re.c)$p.value

svyranktest(esc ~ crit, design = d.re.c, test = "wilcoxon")$statistic
svyranktest(esc ~ crit, design = d.re.c, test = "wilcoxon")$p.value

svychisq(~crit + charit.x.x, design = d.par.c)$statistic
svychisq(~crit + charit.x.x, design = d.par.c)$p.value

svyranktest(esc ~ crit, design = d.par.c, test = "wilcoxon")$statistic
svyranktest(esc ~ crit, design = d.par.c, test = "wilcoxon")$p.value

svytttest(esc ~ crit, design = d.re.c)

svytttest(esc ~ crit, design = d.par.c)
```

Agora, temos que a covariável DRC está balanceada, o que também ocorria no banco pareado pelo modelo inicial (em que CRDAI estava especificada, mas não balanceada). Com relação aos escores preditos, estes também estão balanceados. Logo, o banco pareado pelos escores do modelo reduzido será utilizado na avaliação de impacto do SWF em TSAT no grupo das crianças, quando consideramos o PAC.

Para o grupo dos adultos, considerando o PAC, temos pelos testes  $\chi^2$  e MW que

```
d.re.a = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
                 nest = TRUE, data = ym.adu)

d.par.a = svydesign(id = ~f15 + f1, strat = ~f5, #Apenas
                  weight = ~pl, nest = TRUE, #relembrando.
                  data = ym.par.ap)

svychisq(~crit + inc, design = d.re.a)$statistic
```

```
svychisq(~crit + charit.x.x, design = d.re.a)$statistic
svychisq(~crit + poor.x.x, design = d.re.a)$statistic

svychisq(~crit + inc, design = d.re.a)$p.value
svychisq(~crit + charit.x.x, design = d.re.a)$p.value
svychisq(~crit + poor.x.x, design = d.re.a)$p.value

svyranktest(esc ~ crit, design = d.re.a, test = "wilcoxon")$statistic
svyranktest(esc ~ crit, design = d.re.a, test = "wilcoxon")$p.value

svychisq(~crit + inc, design = d.par.a)$statistic
svychisq(~crit + charit.x.x, design = d.par.a)$statistic
svychisq(~crit + poor.x.x, design = d.par.a)$statistic

svychisq(~crit + inc, design = d.par.a)$p.value
svychisq(~crit + charit.x.x, design = d.par.a)$p.value
svychisq(~crit + poor.x.x, design = d.par.a)$p.value

svyranktest(esc ~ crit, design = d.par.a, test = "wilcoxon")$statistic
svyranktest(esc ~ crit, design = d.par.a, test = "wilcoxon")$p.value

svyttest(esc ~ crit, design = d.re.a)

svyttest(esc ~ crit, design = d.par.a)
```

todas as três covariáveis, bem como os escores preditos, estão balanceados. Logo, o modelo `map`, em sua especificação original, será usado para a avaliação de impacto do SWF em TSAT neste grupo, quando incorporamos o PAC na análise. Podemos salvar os bancos pareados que de fato estão balanceados

```
write.csv(ym.par.cp.1, "ynspmsParCriComPA.csv")
write.csv(ym.par.ap, "ynspmsParAduComPA.csv")
```

e enfim prosseguir para os comandos da avaliação de impacto do SWF.

### B.2.5 Avaliação de Impacto do SWF em TSAT sem PAC

Para os dois bancos pareados obtidos na subseção [B.2.3](#) (caso IID), ajustaremos modelos de regressão linear normal para TSAT, a resposta de interesse, a fim de avaliarmos o impacto do SWF, dado pela interação entre RSWF (a variável de atribuição do tratamento)

e RC (a rodada de coleta das observações). Também é relevante quantificar a influência de RSWF e RC isoladamente, bem como de variáveis em tese fortemente associadas ao tempo que a criança ou adulto aloca semanalmente para o trabalho (potenciais covariáveis de confundimento do modelo). Começando pelo grupo das crianças, temos as seguintes candidatas a covariáveis de confundimento, além das utilizadas para a modelagem do escore de propensão, com seus nomes originais na YNSPMS entre parênteses:

- CPI: Criança é Portadora de Invalidez (p2\_q\_11);
- EC: Escolaridade da Criança (yrschl);
- EMC: Escolaridade da Mãe da Criança (motheredu);
- EPC: Escolaridade do Pai da Criança (fatheredu);
- GC: Gênero da Criança, masculino ou feminino (p2\_q\_03);
- IC: Idade da Criança, em anos (p2\_q\_05\_y);
- LD: Localização do Domicílio, na zona urbana ou rural (f8);
- MCPE: Matrícula da Criança em Programa Educacional (p3\_q\_2);
- PCD: Pessoas por Cômodo no Domicílio (crowd2);
- TD: Topografia do Domicílio (topography).

Para todas as variáveis categóricas (IC é discreta e PCD é contínua) candidatas a covariáveis de confundimento na 1<sup>a</sup> e 4<sup>a</sup> rodadas, exceto EC; EMC; EPC; GC; IC; LD e TD (para as quais consideramos observações apenas na 1<sup>a</sup> rodada), a frequência nas categorias é dada por

```
count(ym.par.c$f8); count(ym.par.c$topography.x.x)
count(ym.par.c$p2_q_03.x.x)
```

```
count(ym.par.c$p2_q_11.x.x); count(ym.par.c$p2_q_11.y.y)
```

```
count(ym.par.c$p3_q_2.x.x); count(ym.par.c$p3_q_2.y.y)
```

```
count(ym.par.c$yrschl.x.x); count(ym.par.c$motheredu.x.x)
count(ym.par.c$fatheredu.x.x)
```

```
count(ym.par.c$cri.d.mun); count(ym.par.c$cri.d.mun.4)
count(ym.par.c$cri.d.avi); count(ym.par.c$cri.d.avi.4)
count(ym.par.c$cri.d.inc); count(ym.par.c$cri.d.inc.4)
count(ym.par.c$pension.x.x); count(ym.par.c$pension.y.y)
count(ym.par.c$charit.x.x); count(ym.par.c$charit.y.y)
count(ym.par.c$remit.x.x); count(ym.par.c$remit.y.y)
count(ym.par.c$poor.x.x); count(ym.par.c$poor.y.y)
```

Por possuírem cinco ou menos observações em categorias que não podem ser agrupadas com outras da mesma variável para pelo menos uma das rodadas, excluimos da especificação para o modelo de regressão linear normal as variáveis CPI; LD e CRDCM. Logo, especificaremos 14 covariáveis de confundimento. Antes do ajuste, reuniremos categorias, alterando também a de referência quando necessário (para que corresponda a uma das categorias extremas), nas covariáveis EC; EMC; EPC e TD

```
ym.par.c$yrschl.x.x = mapvalues(ym.par.c$yrschl.x.x,
  from = c("7 years of schooling", "8 years of schooling",
           "9 years of schooling", "non-standard: Quran & Literacy"),
  to = c("7 years or more", "7 years or more",
         "7 years or more", "no schooling"))
```

```
count(ym.par.c$yrschl.x.x)
```

```
ym.par.c$motheredu.x.x = mapvalues(ym.par.c$motheredu.x.x,
  from = c("2 years of schooling", "3 years of schooling",
           "4 years of schooling", "5 years of schooling",
           "6 years of schooling", "8 years of schooling",
           "9 years of schooling", "10 years of schooling",
           "12 years of schooling", "non-standard: Quran & Literacy",
           "non-app: orphan or mother does not live HH"),
  to = c("1 to 4 years of schooling", "1 to 4 years of schooling",
         "1 to 4 years of schooling", "5 to 8 years of schooling",
         "5 to 8 years of schooling", "5 to 8 years of schooling",
         "9 years or more", "9 years or more", "9 years or more",
         "no schooling", "orphan or mother does not live HH"))
```

```
count(ym.par.c$motheredu.x.x)
```

```
ym.par.c$fatheredu.x.x = mapvalues(ym.par.c$fatheredu.x.x,
  from = c("2 years of schooling", "3 years of schooling",
           "4 years of schooling", "5 years of schooling",
           "6 years of schooling", "7 years of schooling",
           "8 years of schooling", "9 years of schooling",
           "10 years of schooling", "11 years of schooling",
           "12 years of schooling", "13 years of schooling",
           "14 years of schooling", "15 years of schooling",
           "16+ years of schooling", "non-standard: Quran & Literacy"),
```

```

      "non-app: orphan or father does not live HH"),
to = c("1 to 4 years of schooling", "1 to 4 years of schooling",
      "1 to 4 years of schooling", "5 to 8 years of schooling",
      "5 to 8 years of schooling", "5 to 8 years of schooling",
      "5 to 8 years of schooling", "9 years or more",
      "9 years or more", "9 years or more", "9 years or more",
      "9 years or more", "9 years or more", "9 years or more",
      "9 years or more", "no schooling",
      "orphan or father does not live HH"))

count(ym.par.c$fatheredu.x.x)

ym.par.c$topography.x.x = mapvalues(ym.par.c$topography.x.x,
  from = c("CoastalArea-ArabianSea", "CoastalArea-RedSea"),
  to = c("CoastalArea", "CoastalArea"))

count(ym.par.c$topography.x.x)

ym.par.c$yrschl.x.x = relevel(ym.par.c$yrschl.x.x,
  ref = "no schooling")
ym.par.c$motheredu.x.x = relevel(ym.par.c$motheredu.x.x,
  ref = "no schooling")
ym.par.c$fatheredu.x.x = relevel(ym.par.c$fatheredu.x.x,
  ref = "no schooling")

```

Agora os dados estão prontos para que façamos o ajuste do modelo

```

r.pre.c = data.frame(ym.par.c$p6a_q_10.x.x, ym.par.c$crit,
  ym.par.c$topography.x.x, ym.par.c$crowd2.x.x,
  ym.par.c$p2_q_03.x.x, ym.par.c$p2_q_05_y.x.x,
  ym.par.c$p3_q_2.x.x, ym.par.c$yrschl.x.x,
  ym.par.c$motheredu.x.x, ym.par.c$fatheredu.x.x,
  ym.par.c$cri.d.avi, ym.par.c$cri.d.inc,
  ym.par.c$pension.x.x, ym.par.c$charit.x.x,
  ym.par.c$remit.x.x, ym.par.c$poor.x.x)

r.pos.c = data.frame(ym.par.c$p6a_q_10.y.y, ym.par.c$crit,
  ym.par.c$topography.x.x, ym.par.c$crowd2.y.y,
  ym.par.c$p2_q_03.x.x, ym.par.c$p2_q_05_y.x.x,
  ym.par.c$p3_q_2.y.y, ym.par.c$yrschl.x.x,

```

```

ym.par.c$motheredu.x.x, ym.par.c$fatheredu.x.x,
ym.par.c$cri.d.avi.4, ym.par.c$cri.d.inc.4,
ym.par.c$pension.y.y, ym.par.c$charit.y.y,
ym.par.c$remit.y.y, ym.par.c$poor.y.y)

names(r.pre.c); names(r.pos.c) #Note que os nomes são distintos
                                #para as mesmas variáveis.

colnames(r.pre.c) = c("p6a_q_10", "crit", "topography", "crowd2",
  "p2_q_03", "p2_q_05_y", "p3_q_2", "yrschl", "motheredu", "fatheredu",
  "cri.d.avi", "cri.d.inc", "pension", "charit", "remit", "poor")

colnames(r.pos.c) = c("p6a_q_10", "crit", "topography", "crowd2",
  "p2_q_03", "p2_q_05_y", "p3_q_2", "yrschl", "motheredu", "fatheredu",
  "cri.d.avi", "cri.d.inc", "pension", "charit", "remit", "poor")

dd.av1.c = rbind.data.frame(r.pre.c, r.pos.c)

indper.c = c(rep(0, nrow(ym.par.c)), rep(1, nrow(ym.par.c)))

dd.av2.c = cbind.data.frame(dd.av1.c, indper.c)

m.dd.c = glm(p6a_q_10 ~ topography + crowd2 + p2_q_03 + p2_q_05_y +
  p3_q_2 + yrschl + motheredu + fatheredu +
  cri.d.avi + cri.d.inc + pension + charit + remit + poor +
  crit + indper + crit:indper, data = dd.av2.c,
  family = gaussian(link = "identity"))

summary(m.dd.c); xtable(summary(m.dd.c), digits = 4)

```

Pelos resultados do modelo acima, o SWF não tem impacto significativo ao longo do tempo, dado o nível nominal de 5%, na resposta de interesse TSAT para o grupo das crianças. Este impacto é dado pela estimativa pontual do coeficiente associado à interação entre RSWF e RC. Isoladamente, os coeficientes associados à RSWF e RC também não são significativos. Para os demais coeficientes, somente os associados ao intercepto; às crianças residentes em regiões desérticas; atualmente matriculadas em algum programa educacional; com 4 ou 6 anos de escolaridade ou cujos pais têm de 1 a 4 anos de estudo são significativos.

De maneira similar e ainda dentro do caso IID, a avaliação de impacto pode ser feita para os adultos. Neste grupo etário, temos as seguintes candidatas a covariáveis de confundimento, além das utilizadas para a modelagem do escore de propensão, com seus nomes originais na YNSPMS entre parênteses:

- EA: Escolaridade do Adulto (`yrschl`);
- GA: Gênero do Adulto, masculino ou feminino (`p2_q_03`);
- IA: Idade do Adulto, em anos (`p2_q_05_y`);
- LD: Localização do Domicílio, na zona urbana ou rural (`f8`);
- PCD: Pessoas por Cômodo no Domicílio (`crowd2`);
- TD: Topografia do Domicílio (`topography`).

Novamente, para todas as variáveis categóricas (IA é discreta e PCD é contínua) candidatas a covariáveis de confundimento na 1<sup>a</sup> e 4<sup>a</sup> rodadas, exceto EA; GA; IA; LD e TD (para as quais consideramos observações apenas na 1<sup>a</sup> rodada), a frequência nas categorias pode ser obtida por

```
count(ym.par.a$f8); count(ym.par.a$topography.x.x)
count(ym.par.a$p2_q_03.x.x); count(ym.par.a$yrschl.x.x)

count(ym.par.a$mun); count(ym.par.a$mun.4)
count(ym.par.a$avi); count(ym.par.a$avi.4)
count(ym.par.a$inc); count(ym.par.a$inc.4)
count(ym.par.a$pension.x.x); count(ym.par.a$pension.y.y)
count(ym.par.a$charit.x.x); count(ym.par.a$charit.y.y)
count(ym.par.a$remit.x.x); count(ym.par.a$remit.y.y)
count(ym.par.a$poor.x.x); count(ym.par.a$poor.y.y)
```

Não há nenhuma variável com cinco ou menos observações em categorias que não podem ser agrupadas com outras da mesma para pelo menos uma das rodadas. Logo, especificaremos todas as 13 covariáveis. Antes do ajuste, reuniremos categorias, alterando também a de referência quando necessário (para que corresponda a uma das categorias extremas), nas covariáveis EA e TD

```
ym.par.a$yrschl.x.x = mapvalues(ym.par.a$yrschl.x.x,
  from = c("13 years of schooling", "14 years of schooling",
           "15 years of schooling", "16+ years of schooling"),
  to = c("13 years or more", "13 years or more",
         "13 years or more", "13 years or more"))

count(ym.par.a$yrschl.x.x)
```

```
ym.par.a$topography.x.x = relevel(ym.par.a$topography.x.x,
                                  ref = "CoastalArea-RedSea")
ym.par.a$yrschl.x.x = relevel(ym.par.a$yrschl.x.x,
                              ref = "no schooling")
```

Por fim, prosseguimos para o ajuste do modelo correspondente

```
r.pre.a = data.frame(ym.par.a$p6a_q_10.x.x, ym.par.a$crit,
                    ym.par.a$f8, ym.par.a$topography.x.x,
                    ym.par.a$crowd2.x.x, ym.par.a$p2_q_03.x.x,
                    ym.par.a$p2_q_05_y.x.x, ym.par.a$yrschl.x.x,
                    ym.par.a$mun, ym.par.a$avi, ym.par.a$inc,
                    ym.par.a$pension.x.x, ym.par.a$charit.x.x,
                    ym.par.a$remit.x.x, ym.par.a$poor.x.x)

r.pos.a = data.frame(ym.par.a$p6a_q_10.y.y, ym.par.a$crit,
                    ym.par.a$f8, ym.par.a$topography.x.x,
                    ym.par.a$crowd2.y.y, ym.par.a$p2_q_03.x.x,
                    ym.par.a$p2_q_05_y.x.x, ym.par.a$yrschl.x.x,
                    ym.par.a$mun.4, ym.par.a$avi.4, ym.par.a$inc.4,
                    ym.par.a$pension.y.y, ym.par.a$charit.y.y,
                    ym.par.a$remit.y.y, ym.par.a$poor.y.y)

names(r.pre.a); names(r.pos.a) #Note que os nomes são distintos
                              #para as mesmas variáveis.

colnames(r.pre.a) = c("p6a_q_10", "crit", "f8", "topography",
                    "crowd2", "p2_q_03", "p2_q_05_y", "yrschl", "mun", "avi", "inc",
                    "pension", "charit", "remit", "poor")

colnames(r.pos.a) = c("p6a_q_10", "crit", "f8", "topography",
                    "crowd2", "p2_q_03", "p2_q_05_y", "yrschl", "mun", "avi", "inc",
                    "pension", "charit", "remit", "poor")

dd.av1.a = rbind.data.frame(r.pre.a, r.pos.a)

indper.a = c(rep(0, nrow(ym.par.a)), rep(1, nrow(ym.par.a)))

dd.av2.a = cbind.data.frame(dd.av1.a, indper.a)
```

```
m.dd.a = glm(p6a_q_10 ~ f8 + topography + crowd2 + p2_q_03 +
             p2_q_05_y + yrschl + mun + avi + inc +
             pension + charit + remit + poor +
             crit + indper + crit:indper, data = dd.av2.a,
             family = gaussian(link = "identity"))

summary(m.dd.a); xtable(summary(m.dd.a), digits = 4)
```

Pelos resultados do modelo acima, o SWF não tem impacto significativo ao longo do tempo, dado o nível nominal de 5%, na resposta de interesse TSAT para o grupo dos adultos. Isoladamente, o coeficiente associado à RC é negativo e significativo, indicando que os adultos trabalham menos na 4ª rodada do que na 1ª. Dentre os demais coeficientes, são significativos os associados ao intercepto; aos adultos residentes na zona urbana ou no litoral Leste (banhado pelo Mar Árábico); à PCD; aos adultos do gênero masculino; à IA; que possuem 2, 11 ou pelo menos 13 anos de escolaridade; que são mulheres e chefiam o domicílio; residem em domicílios que recebem fundos de pensão ou remessas de parentes ou amigos. Ainda, todos os três coeficientes associados às categorias de NPD (exceto a de referência) são significativos.

## B.2.6 Avaliação de Impacto do SWF em TSAT com PAC

Agora considerando o PAC, para os dois bancos pareados obtidos na subseção B.2.4, ajustaremos modelos de regressão linear normal para TSAT a fim de avaliarmos o impacto do SWF nesta resposta de interesse, dado pela interação entre RSWF e RC. Para cada modelo, também quantificaremos a influência de RSWF e RC isoladamente, bem como de potenciais covariáveis de confundimento. Começando novamente pelo grupo das crianças, temos que as candidatas a covariáveis de confundimento são as mesmas do caso IID. Dentre estas, para todas as variáveis categóricas na 1ª e 4ª rodadas, exceto EC; EMC; EPC; GC; IC; LD e TD (para as quais consideramos observações apenas na 1ª rodada), a frequência nas categorias é dada por

```
count(ym.par.cp.1$f8); count(ym.par.cp.1$topography.x.x)
count(ym.par.cp.1$p2_q_03.x.x)

count(ym.par.cp.1$p2_q_11.x.x); count(ym.par.cp.1$p2_q_11.y.y)

count(ym.par.cp.1$p3_q_2.x.x); count(ym.par.cp.1$p3_q_2.y.y)

count(ym.par.cp.1$yrschl.x.x); count(ym.par.cp.1$motheredu.x.x)
```

```
count(ym.par.cp.1$fatheredu.x.x)
```

```
count(ym.par.cp.1$cri.d.mun); count(ym.par.cp.1$cri.d.mun.4)
count(ym.par.cp.1$cri.d.avi); count(ym.par.cp.1$cri.d.avi.4)
count(ym.par.cp.1$cri.d.inc); count(ym.par.cp.1$cri.d.inc.4)
count(ym.par.cp.1$pension.x.x); count(ym.par.cp.1$pension.y.y)
count(ym.par.cp.1$charit.x.x); count(ym.par.cp.1$charit.y.y)
count(ym.par.cp.1$remit.x.x); count(ym.par.cp.1$remit.y.y)
count(ym.par.cp.1$poor.x.x); count(ym.par.cp.1$poor.y.y)
```

Por possuírem cinco ou menos observações em categorias que não podem ser agrupadas com outras da mesma variável para pelo menos uma das rodadas, também no banco pareado considerando o PAC, excluímos da especificação para o modelo de regressão linear normal as variáveis CPI; LD e CRDCM. Logo, especificaremos as mesmas 14 covariáveis de confundimento do caso IID. Antes do ajuste, novamente reuniremos categorias, alterando também a de referência quando necessário, nas covariáveis EC; EMC; EPC e TD

```
ym.par.cp.1$yrschl.x.x = mapvalues(ym.par.cp.1$yrschl.x.x,
  from = c("7 years of schooling", "8 years of schooling",
           "9 years of schooling", "non-standard: Quran & Literacy"),
  to = c("7 years or more", "7 years or more",
         "7 years or more", "no schooling"))
```

```
count(ym.par.cp.1$yrschl.x.x)
```

```
ym.par.cp.1$motheredu.x.x = mapvalues(ym.par.cp.1$motheredu.x.x,
  from = c("2 years of schooling", "3 years of schooling",
           "4 years of schooling", "5 years of schooling",
           "6 years of schooling", "8 years of schooling",
           "9 years of schooling", "10 years of schooling",
           "12 years of schooling", "non-standard: Quran & Literacy",
           "non-app: orphan or mother does not live HH"),
  to = c("1 to 4 years of schooling", "1 to 4 years of schooling",
         "1 to 4 years of schooling", "5 to 8 years of schooling",
         "5 to 8 years of schooling", "5 to 8 years of schooling",
         "9 years or more", "9 years or more", "9 years or more",
         "no schooling", "orphan or mother does not live HH"))
```

```
count(ym.par.cp.1$motheredu.x.x)
```

```

ym.par.cp.1$fatheredu.x.x = mapvalues(ym.par.cp.1$fatheredu.x.x,
  from = c("2 years of schooling", "3 years of schooling",
    "4 years of schooling", "5 years of schooling",
    "6 years of schooling", "7 years of schooling",
    "8 years of schooling", "9 years of schooling",
    "10 years of schooling", "11 years of schooling",
    "12 years of schooling", "13 years of schooling",
    "14 years of schooling", "15 years of schooling",
    "16+ years of schooling", "non-standard: Quran & Literacy",
    "non-app: orphan or father does not live HH"),
  to = c("1 to 4 years of schooling", "1 to 4 years of schooling",
    "1 to 4 years of schooling", "5 to 8 years of schooling",
    "5 to 8 years of schooling", "5 to 8 years of schooling",
    "5 to 8 years of schooling", "9 years or more",
    "9 years or more", "9 years or more", "9 years or more",
    "9 years or more", "9 years or more", "9 years or more",
    "9 years or more", "no schooling",
    "orphan or father does not live HH"))

count(ym.par.cp.1$fatheredu.x.x)

ym.par.cp.1$topography.x.x = mapvalues(ym.par.cp.1$topography.x.x,
  from = c("CoastalArea-ArabianSea", "CoastalArea-RedSea"),
  to = c("CoastalArea", "CoastalArea"))

count(ym.par.cp.1$topography.x.x)

ym.par.cp.1$yrschl.x.x = relevel(ym.par.cp.1$yrschl.x.x,
  ref = "no schooling")
ym.par.cp.1$motheredu.x.x = relevel(ym.par.cp.1$motheredu.x.x,
  ref = "no schooling")
ym.par.cp.1$fatheredu.x.x = relevel(ym.par.cp.1$fatheredu.x.x,
  ref = "no schooling")

```

Os dados estão prontos para que façamos o ajuste do modelo. Note que agora incorporaremos o PAC da YNSPMS na estimação dos coeficientes da regressão

```

r.pre.cp = data.frame(ym.par.cp.1$p6a_q_10.x.x, ym.par.cp.1$crit,
  ym.par.cp.1$topography.x.x, ym.par.cp.1$crowd2.x.x,

```

```

ym.par.cp.1$p2_q_03.x.x, ym.par.cp.1$p2_q_05_y.x.x,
ym.par.cp.1$p3_q_2.x.x, ym.par.cp.1$yrsl.x.x,
ym.par.cp.1$motheredu.x.x, ym.par.cp.1$fatheredu.x.x,
ym.par.cp.1$cri.d.avi, ym.par.cp.1$cri.d.inc,
ym.par.cp.1$pension.x.x, ym.par.cp.1$charit.x.x,
ym.par.cp.1$remit.x.x, ym.par.cp.1$poor.x.x)

r.pos.cp = data.frame(ym.par.cp.1$p6a_q_10.y.y, ym.par.cp.1$crit,
ym.par.cp.1$topography.x.x, ym.par.cp.1$crowd2.y.y,
ym.par.cp.1$p2_q_03.x.x, ym.par.cp.1$p2_q_05_y.x.x,
ym.par.cp.1$p3_q_2.y.y, ym.par.cp.1$yrsl.x.x,
ym.par.cp.1$motheredu.x.x, ym.par.cp.1$fatheredu.x.x,
ym.par.cp.1$cri.d.avi.4, ym.par.cp.1$cri.d.inc.4,
ym.par.cp.1$pension.y.y, ym.par.cp.1$charit.y.y,
ym.par.cp.1$remit.y.y, ym.par.cp.1$poor.y.y)

names(r.pre.cp); names(r.pos.cp) #Note que os nomes são distintos
                                #para as mesmas variáveis.

colnames(r.pre.cp) = c("p6a_q_10", "crit", "topography", "crowd2",
"p2_q_03", "p2_q_05_y", "p3_q_2", "yrsl", "motheredu", "fatheredu",
"cri.d.avi", "cri.d.inc", "pension", "charit", "remit", "poor")

colnames(r.pos.cp) = c("p6a_q_10", "crit", "topography", "crowd2",
"p2_q_03", "p2_q_05_y", "p3_q_2", "yrsl", "motheredu", "fatheredu",
"cri.d.avi", "cri.d.inc", "pension", "charit", "remit", "poor")

dd.av1.cp = rbind.data.frame(r.pre.cp, r.pos.cp)

indper.cp = c(rep(0, nrow(ym.par.cp.1)), rep(1, nrow(ym.par.cp.1)))

dd.av2.cp = cbind.data.frame(dd.av1.cp, indper.cp)

d.c = svydesign(id = ~f15 + f1, strat = ~f5, weight = ~pl,
nest = TRUE, data = dd.av2)

d.c

m.dd.cp = svyglm(p6a_q_10 ~ topography + crowd2 + p2_q_03 + p2_q_05_y +
p3_q_2 + yrsl + motheredu + fatheredu +

```

```

cri.d.avi + cri.d.inc + pension + charit +
remit + poor + crit + indper + crit:indper,
design = d.c, family = gaussian(link = "identity"))

summary(m.dd.cp); xtable(summary(m.dd.cp), digits = 4)

```

Pelos resultados do modelo acima, também no caso em que consideramos o PAC, o SWF não tem impacto significativo ao longo do tempo, dado o nível nominal de 5%, na resposta de interesse TSAT para o grupo das crianças. Isoladamente, os coeficientes associados à RSWF e RC também não são significativos. Para os demais coeficientes, somente os associados ao intercepto; às crianças que possuem 2 ou 3 anos de escolaridade; cujas mães têm de 5 a 8 anos de estudo; cujos pais têm de 1 a 4 anos de estudo ou não residem no domicílio (ausente); residentes em domicílios que recebem fundos de pensão ou renda através de caridade são significativos. Finalmente, residir em um domicílio em situação de pobreza moderada ou de vulnerabilidade também influencia significativamente em TSAT.

Para estimar o EPAA, um modelo adicional com os mesmos dados e especificação deve ser ajustado, mas partindo da premissa incorreta de que o banco foi pareado supondo observações IID

```

m.dd.cp.s = glm(p6a_q_10 ~ topography + crowd2 + p2_q_03 +
p2_q_05_y + p3_q_2 + yrschl + motheredu + fatheredu +
cri.d.avi + cri.d.inc + pension + charit +
remit + poor + crit + indper + crit:indper,
data = dd.av2.cp, family = gaussian(link = "identity"))

summary(m.dd.cp.s, digits = 4); xtable(summary(m.dd.cp.s), digits = 4)

```

De forma análoga, mas agora considerando o PAC, a avaliação de impacto pode ser feita para os adultos. Neste grupo etário, temos as mesmas candidatas a covariáveis de confundimento do caso IID. Para todas as variáveis categóricas na 1ª e 4ª rodadas, exceto EA; GA; IA; LD e TD (para as quais consideramos observações apenas na 1ª rodada), a frequência nas categorias pode ser obtida por

```

count(ym.par.ap$f8); count(ym.par.ap$topography.x.x)
count(ym.par.ap$p2_q_03.x.x); count(ym.par.ap$yrschl.x.x)

count(ym.par.ap$mun); count(ym.par.ap$mun.4)
count(ym.par.ap$avi); count(ym.par.ap$avi.4)
count(ym.par.ap$inc); count(ym.par.ap$inc.4)

```

```
count(ym.par.ap$pension.x.x); count(ym.par.ap$pension.y.y)
count(ym.par.ap$charit.x.x); count(ym.par.ap$charit.y.y)
count(ym.par.ap$remit.x.x); count(ym.par.ap$remit.y.y)
count(ym.par.ap$poor.x.x); count(ym.par.ap$poor.y.y)
```

Também aqui, não há nenhuma variável com cinco ou menos observações em categorias que não podem ser agrupadas com outras da mesma para pelo menos uma das rodadas. Logo, especificaremos todas as 13 covariáveis. Antes do ajuste, reuniremos categorias, alterando também a de referência quando necessário (para que corresponda a uma das categorias extremas), nas covariáveis EA e TD

```
ym.par.ap$yrschl.x.x = mapvalues(ym.par.ap$yrschl.x.x,
  from = c("13 years of schooling", "14 years of schooling",
           "15 years of schooling", "16+ years of schooling"),
  to = c("13 years or more", "13 years or more",
         "13 years or more", "13 years or more"))
```

```
count(ym.par.ap$yrschl.x.x)
```

```
ym.par.ap$topography.x.x = relevel(ym.par.ap$topography.x.x,
  ref = "CoastalArea-RedSea")
ym.par.ap$yrschl.x.x = relevel(ym.par.ap$yrschl.x.x,
  ref = "no schooling")
```

Por fim, prosseguimos para o ajuste do modelo correspondente

```
r.pre.ap = data.frame(ym.par.ap$p6a_q_10.x.x, ym.par.ap$crit,
  ym.par.ap$f8, ym.par.ap$topography.x.x,
  ym.par.ap$crowd2.x.x, ym.par.ap$p2_q_03.x.x,
  ym.par.ap$p2_q_05_y.x.x, ym.par.ap$yrschl.x.x,
  ym.par.ap$mun, ym.par.ap$avi, ym.par.ap$inc,
  ym.par.ap$pension.x.x, ym.par.ap$charit.x.x,
  ym.par.ap$remit.x.x, ym.par.ap$poor.x.x)
```

```
r.pos.ap = data.frame(ym.par.ap$p6a_q_10.y.y, ym.par.ap$crit,
  ym.par.ap$f8, ym.par.ap$topography.x.x,
  ym.par.ap$crowd2.y.y, ym.par.ap$p2_q_03.x.x,
  ym.par.ap$p2_q_05_y.x.x, ym.par.ap$yrschl.x.x,
  ym.par.ap$mun.4, ym.par.ap$avi.4, ym.par.ap$inc.4,
  ym.par.ap$pension.y.y, ym.par.ap$charit.y.y,
```

```
ym.par.ap$remit.y.y, ym.par.ap$poor.y.y)

names(r.pre.ap); names(r.pos.ap) #Note que os nomes são distintos
                                #para as mesmas variáveis.

colnames(r.pre.ap) = c("p6a_q_10", "crit", "f8", "topography",
  "crowd2", "p2_q_03", "p2_q_05_y", "yrschl", "mun", "avi", "inc",
  "pension", "charit", "remit", "poor")

colnames(r.pos.ap) = c("p6a_q_10", "crit", "f8", "topography",
  "crowd2", "p2_q_03", "p2_q_05_y", "yrschl", "mun", "avi", "inc",
  "pension", "charit", "remit", "poor")

dd.av1.ap = rbind.data.frame(r.pre.ap, r.pos.ap)

indper.ap = c(rep(0, nrow(ym.par.ap)), rep(1, nrow(ym.par.ap)))

dd.av2.ap = cbind.data.frame(dd.av1.ap, indper.ap)

d.a = svydesign(id = ~f15 + f1, strat = ~f5,
  weight = ~pl, nest = TRUE, data = dd.av2)

d.a

m.dd.ap = svyglm(p6a_q_10 ~ f8 + topography + crowd2 + p2_q_03 +
  p2_q_05_y + yrschl + mun + avi + inc +
  pension + charit + remit + poor +
  crit + indper + crit:indper, design = d.a,
  family = gaussian(link = "identity"))

summary(m.dd.ap); xtable(summary(m.dd.ap), digits = 4)
```

Pelos resultados do modelo acima, o SWF não tem impacto significativo ao longo do tempo, dado o nível nominal de 5%, na resposta de interesse TSAT para o grupo dos adultos. Isoladamente, o coeficiente associado à RC não é mais significativo, ao contrário do que ocorre no caso IID. Dentre os demais coeficientes, são significativos os associados ao intercepto; aos adultos residentes na zona urbana; do gênero masculino; que possuem 2 ou pelo menos 13 anos de escolaridade; que são mulheres e chefiam o domicílio; e que residem em domicílios que recebem remessas de parentes ou amigos. Por fim, residir em um domicílio em situação de pobreza moderada também influencia significativamente em

TSAT.

Para estimar o EPAA, um modelo adicional com os mesmos dados e especificação deve ser ajustado, mas partindo da premissa incorreta de que o banco foi pareado supondo observações IID

```
m.dd.ap.s = glm(p6a_q_10 ~ f8 + topography + crowd2 + p2_q_03 +  
                p2_q_05_y + yrschl + mun + avi + inc +  
                pension + charit + remit + poor +  
                crit + indper + crit:indper,  
                data = dd.av2.ap, family = gaussian(link = "identity"))  
  
summary(m.dd.ap.s, digits = 4); xtable(summary(m.dd.ap.s), digits = 4)
```

# Anexos

## ANEXO A – Histogramas da Densidade da Variável TSAT

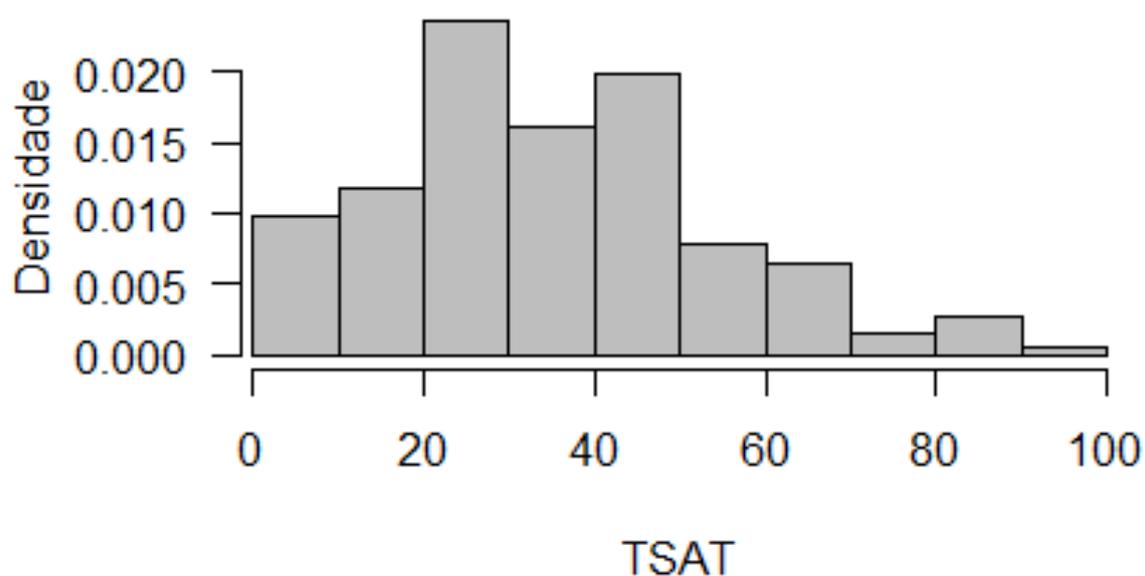


Figura 2 – Densidade da variável TSAT nos adultos, caso IID, 1ª rodada.

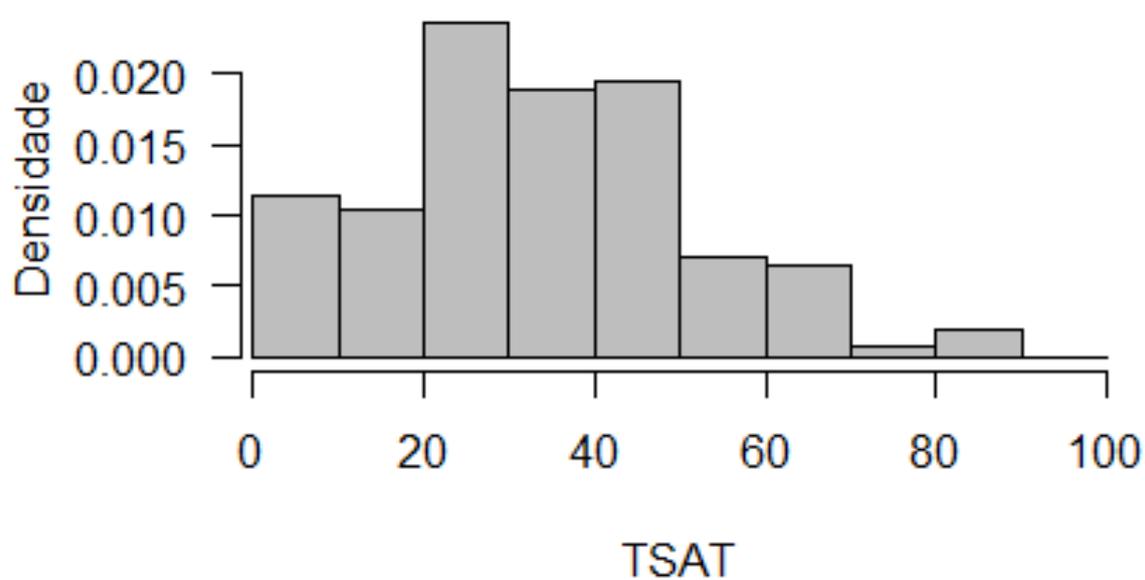


Figura 3 – Densidade da variável TSAT nos adultos, caso IID, 4ª rodada.

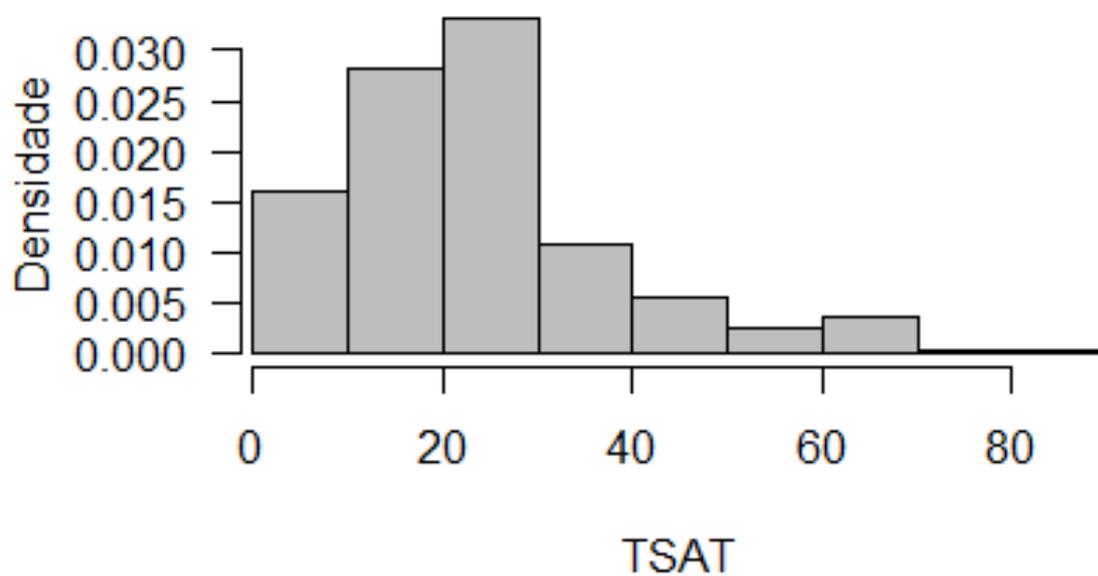


Figura 4 – Densidade da variável TSAT nas crianças, caso IID, 1ª rodada.

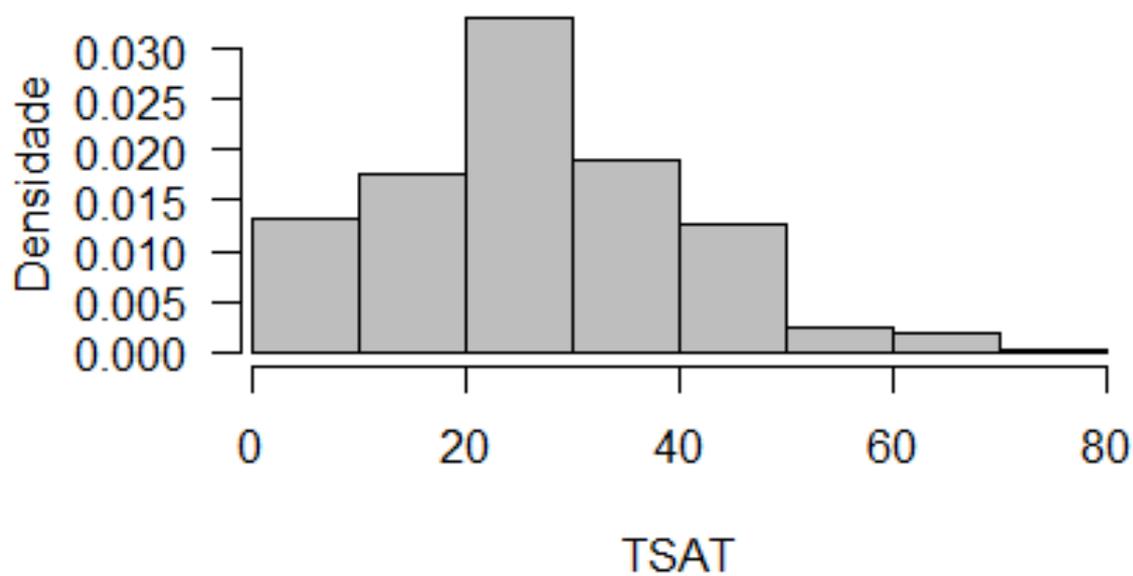


Figura 5 – Densidade da variável TSAT nas crianças, caso IID, 4ª rodada.

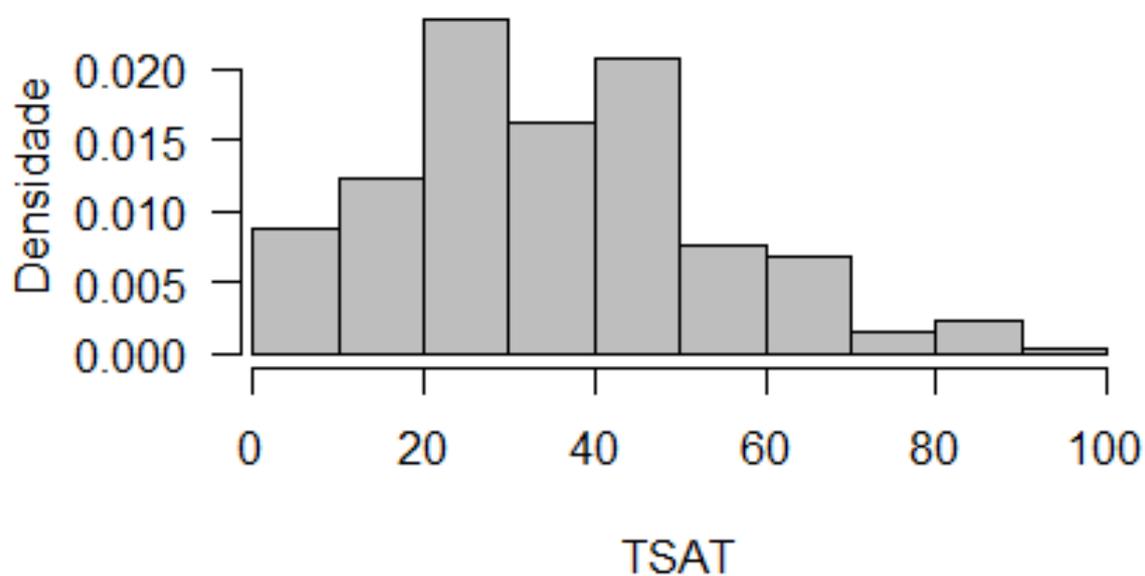


Figura 6 – Densidade da variável TSAT nos adultos, considerando o PAC da YNSPMS, 1ª rodada.

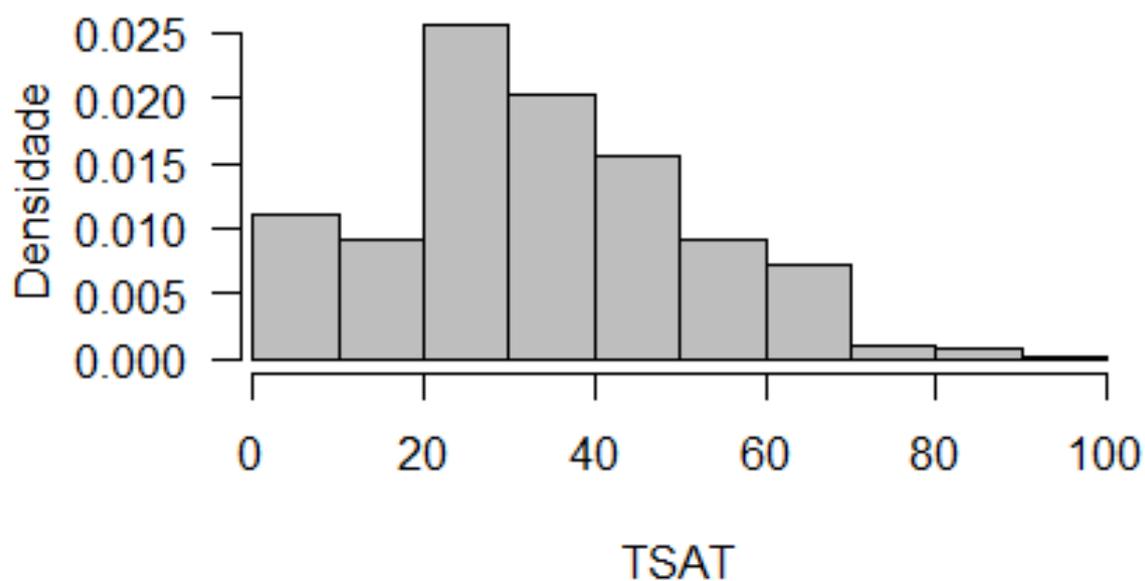


Figura 7 – Densidade da variável TSAT nos adultos, considerando o PAC da YNSPMS, 4ª rodada.

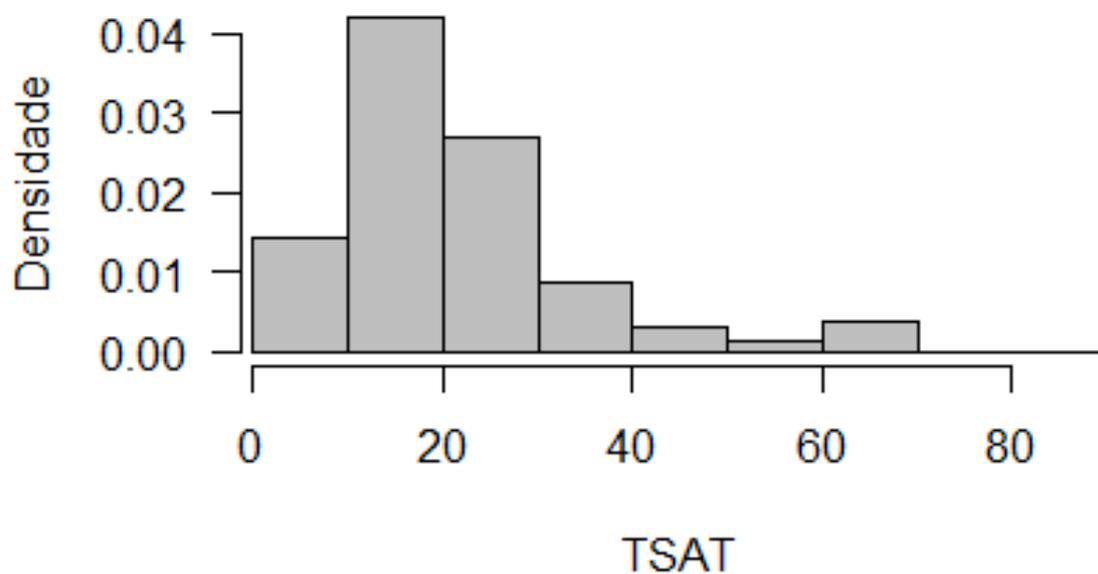


Figura 8 – Densidade da variável TSAT nas crianças, considerando o PAC da YNSPMS, 1ª rodada.

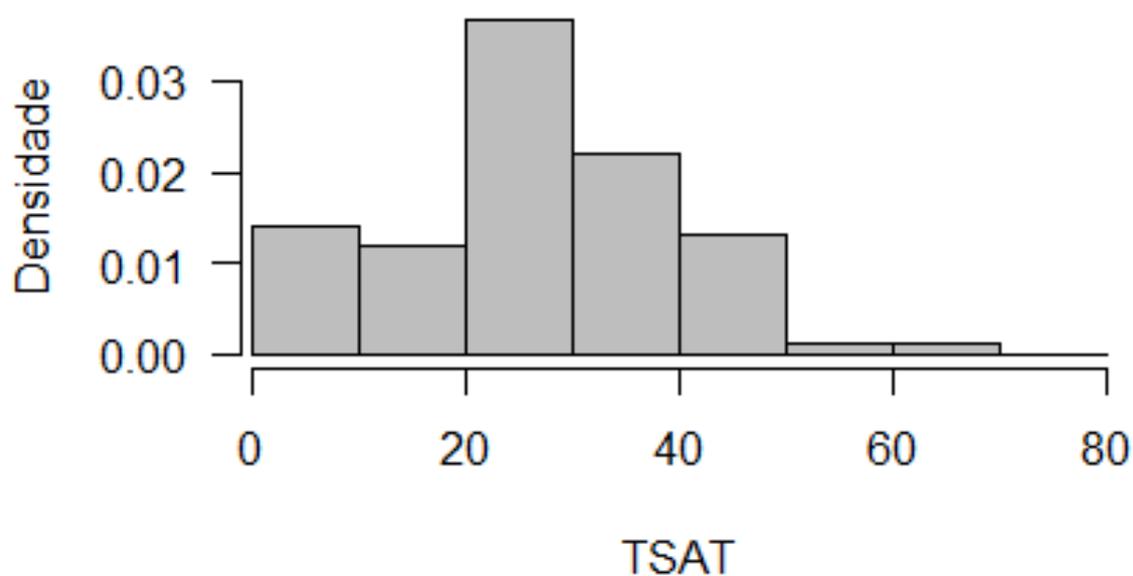


Figura 9 – Densidade da variável TSAT nas crianças, considerando o PAC da YNSPMS, 4ª rodada.

## ANEXO B – Histogramas da Frequência dos Escores Preditos para o Banco Original

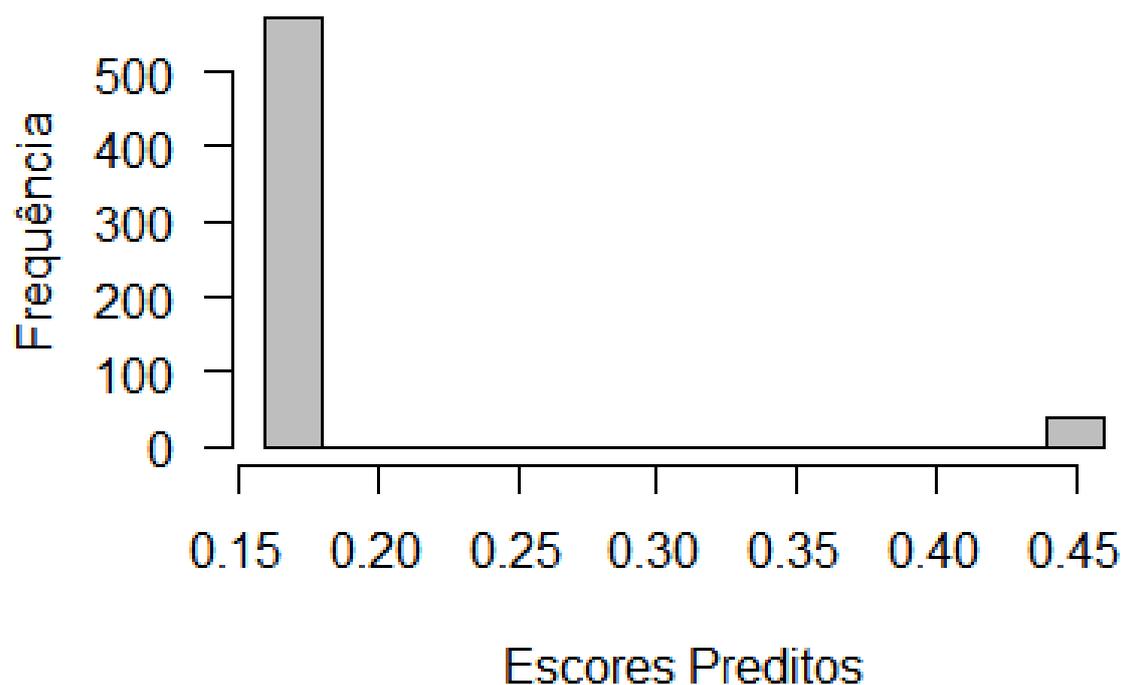


Figura 10 – Frequência dos escores preditos pelo modelo inicial para o banco original das crianças, caso IID.

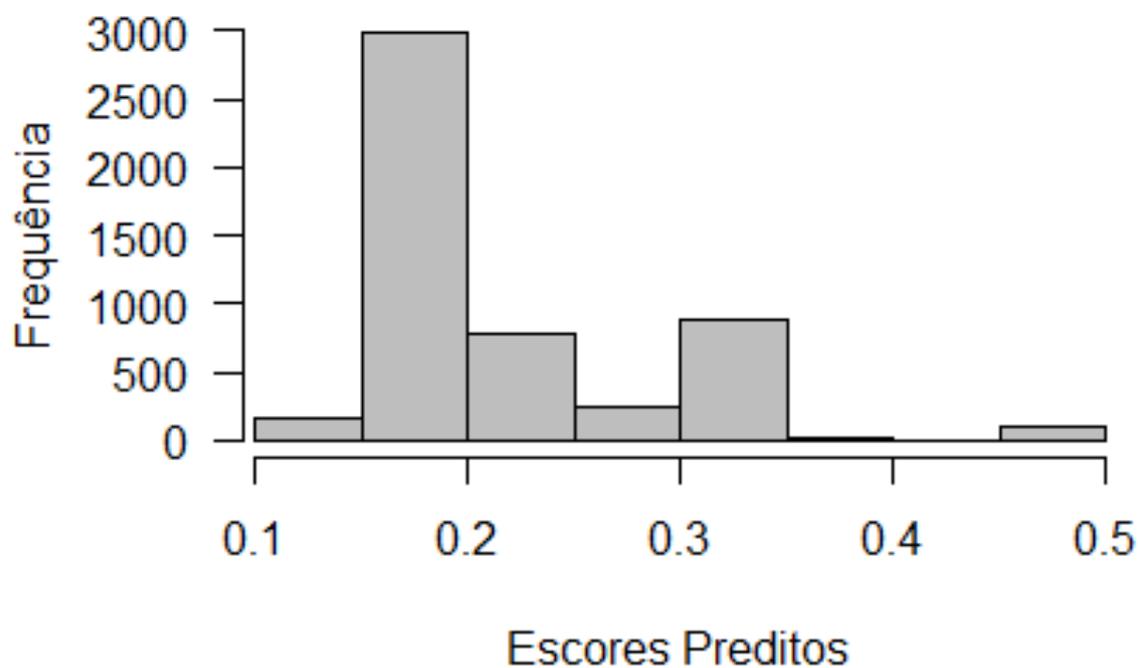


Figura 11 – Frequência dos escores preditos pelo modelo inicial para o banco original dos adultos, caso IID.

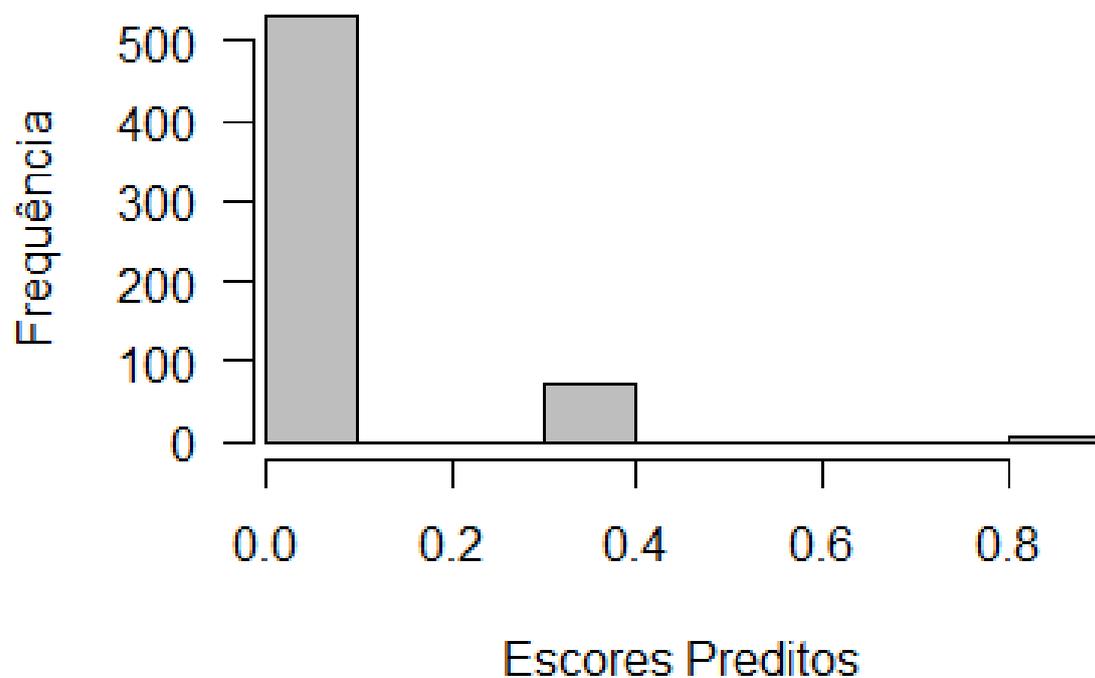


Figura 12 – Frequência dos escores preditos pelo modelo inicial para o banco original das crianças, considerando o PAC da YNSPMS.

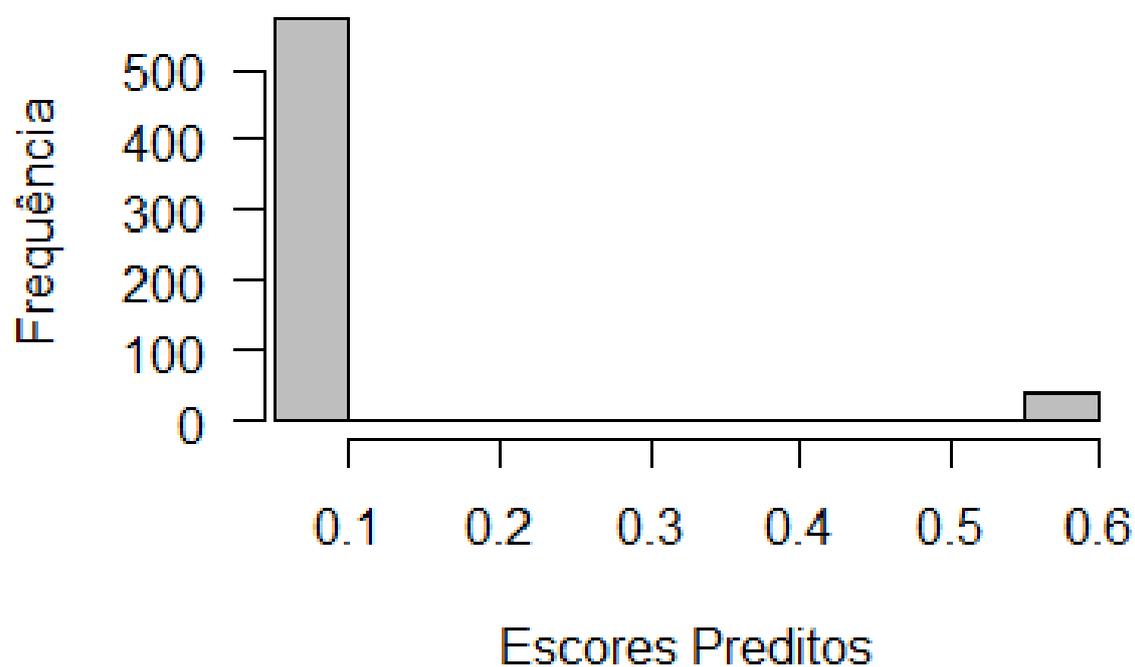


Figura 13 – Frequência dos escores preditos pelo modelo reduzido para o banco original das crianças, considerando o PAC da YNSPMS.

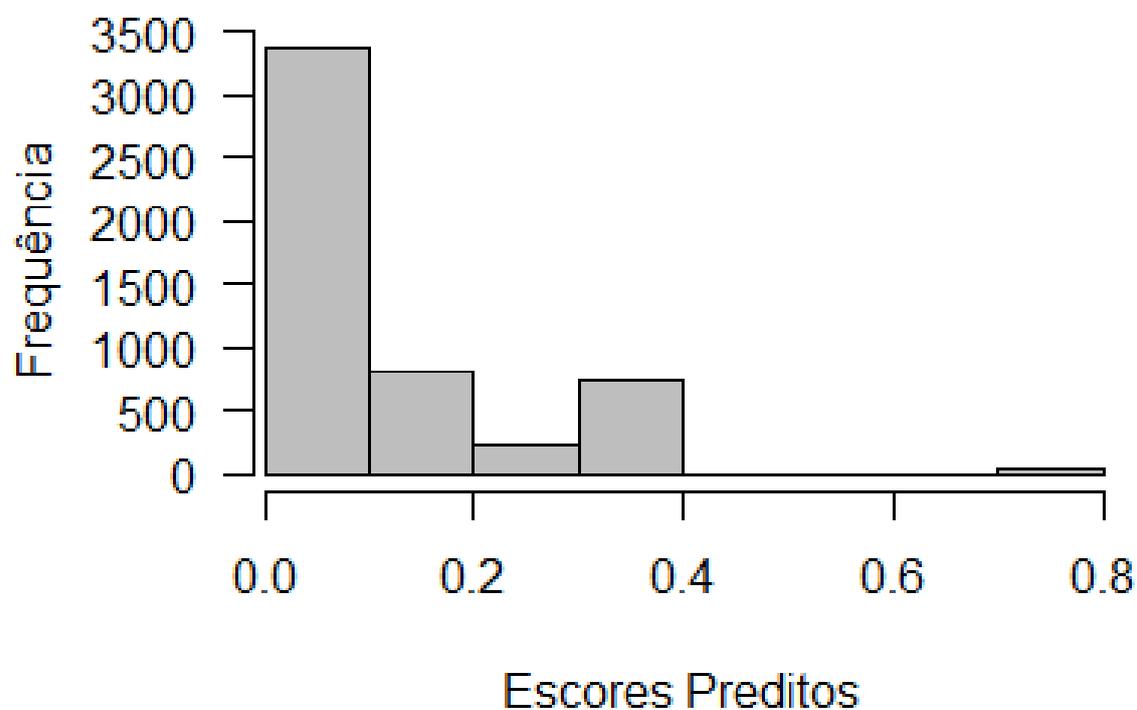


Figura 14 – Frequência dos escores preditos pelo modelo inicial para o banco original dos adultos, considerando o PAC da YNSPMS.

## ANEXO C – Histogramas da Frequência dos Escores Preditos para o Banco Pareado

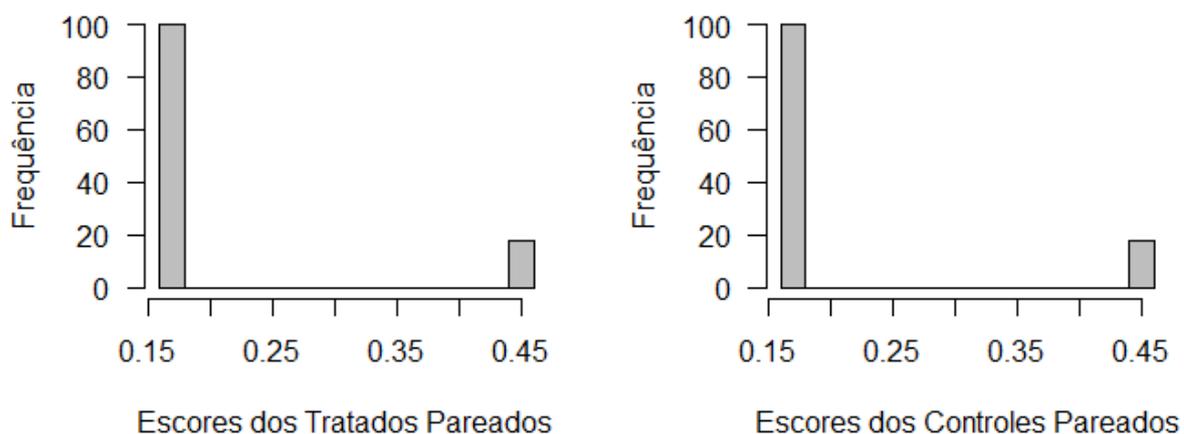


Figura 15 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado das crianças, caso IID.

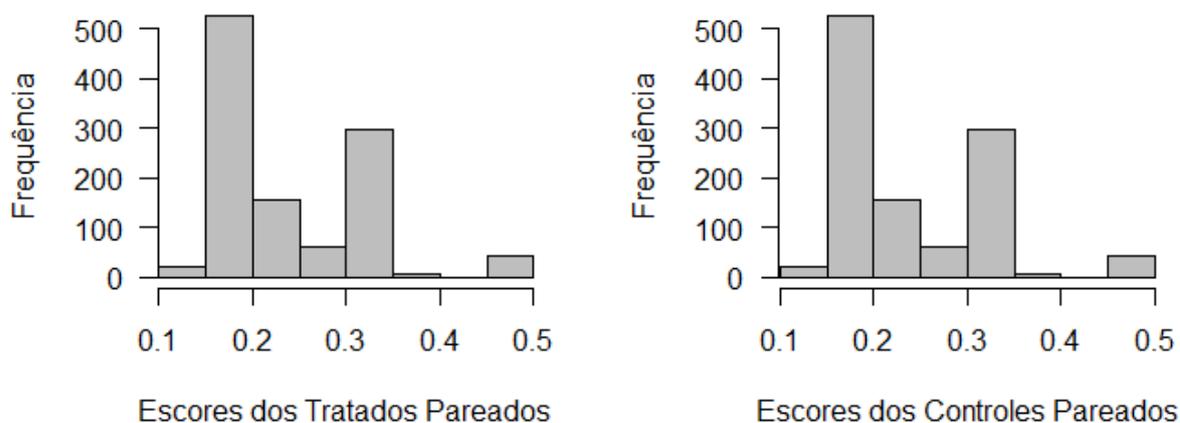


Figura 16 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado dos adultos, caso IID.

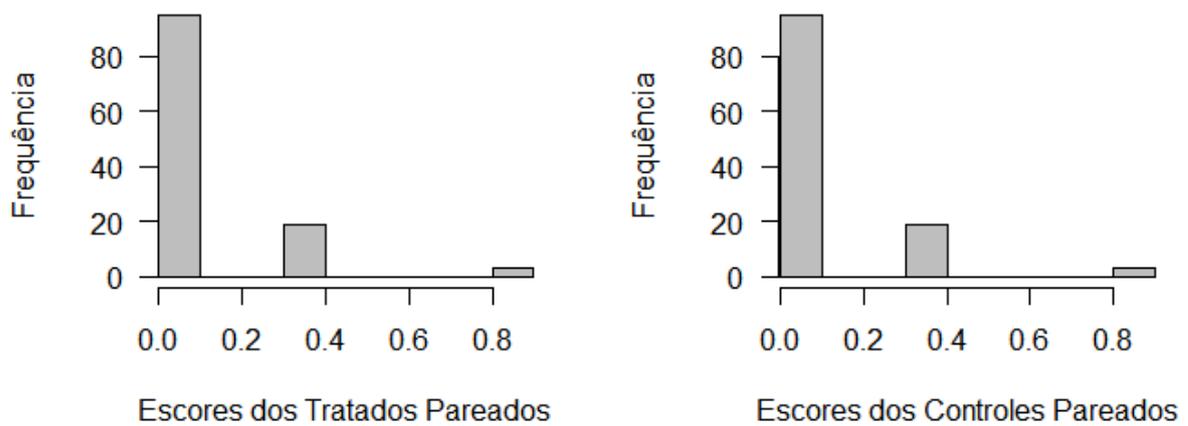


Figura 17 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado das crianças, considerando o PAC da YNSPMS.

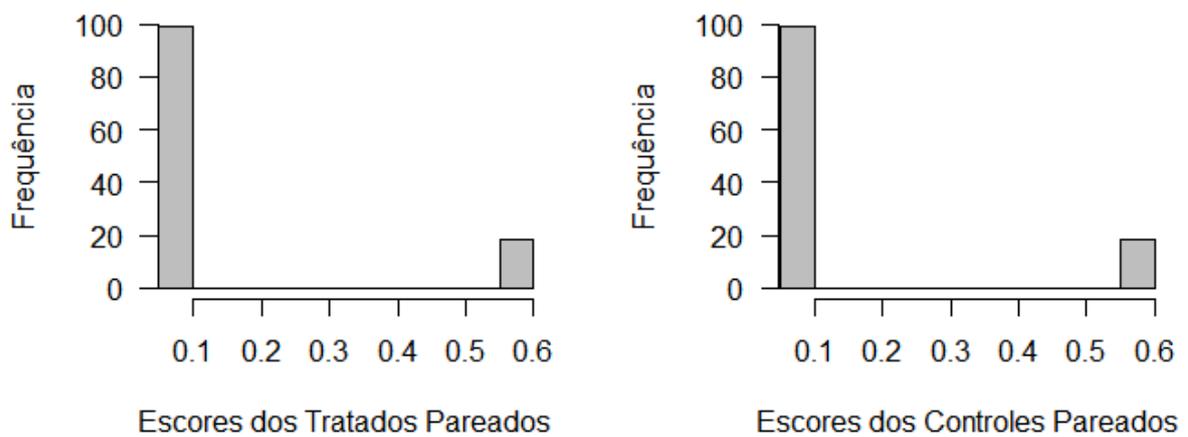


Figura 18 – Frequência por classe dos escores preditos pelo modelo reduzido para o banco pareado das crianças, considerando o PAC da YNSPMS.

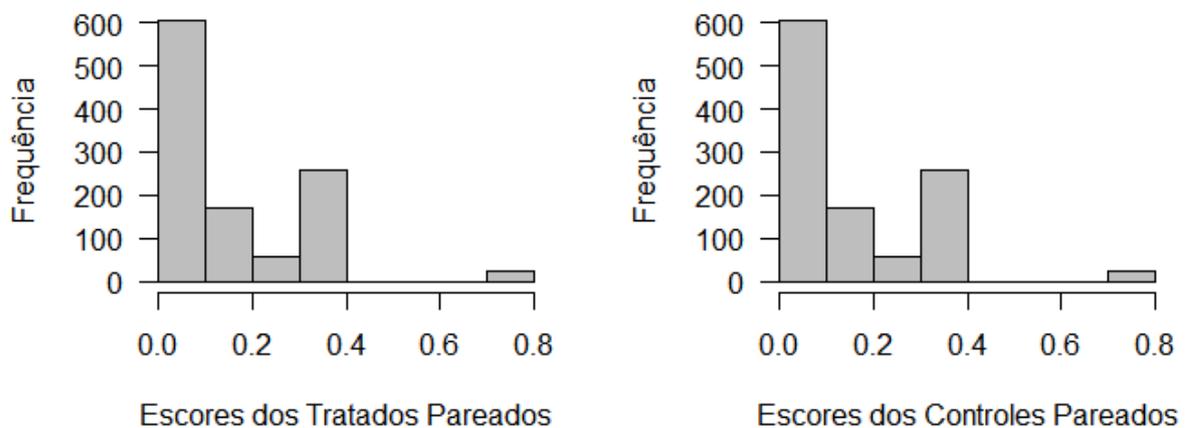


Figura 19 – Frequência por classe dos escores preditos pelo modelo inicial para o banco pareado dos adultos, considerando o PAC da YNSPMS.