

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

NATÁLIA HEIMANN

A INFLUÊNCIA DA ELEIÇÃO DE MULHERES NA PARTICIPAÇÃO POLÍTICA
FEMININA: UMA ANÁLISE NO CENÁRIO BRASILEIRO

SÃO PAULO
2016

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

NATÁLIA HEIMANN

A INFLUÊNCIA DA ELEIÇÃO DE MULHERES NA PARTICIPAÇÃO POLÍTICA
FEMININA: UMA ANÁLISE NO CENÁRIO BRASILEIRO

Dissertação apresentada à Escola de
Economia de São Paulo da Fundação
Getulio Vargas, como requisito para
obtenção do título de Mestre em Economia

Campo de conhecimento: Econometria

Orientador: Prof. Dr. Sérgio Pinheiro Firpo

SÃO PAULO
2016

Heimann, Natália.

A influência da eleição de mulheres na participação política feminina: uma análise no cenário brasileiro / Natália Heimann. - 2016.

87 f.

Orientador: Sergio Pinheiro Firpo

Dissertação (MPFE) - Escola de Economia de São Paulo.

1. Mulheres na política - Brasil. 2. Eleições locais. 3. Participação política. I. Firpo, Sergio Pinheiro. II. Dissertação (MPFE) - Escola de Economia de São Paulo. III. Título.

CDU 396.9(81)

A INFLUÊNCIA DA ELEIÇÃO DE MULHERES NA PARTICIPAÇÃO POLÍTICA FEMININA: UMA ANÁLISE NO CENÁRIO BRASILEIRO

Dissertação apresentada à Escola de Economia de São Paulo da Fundação Getúlio Vargas, como requisito para obtenção do título de Mestre em Economia

Campo de conhecimento: Econometria

Data de aprovação:

____/____/____

Banca examinadora:

Prof. PhD. Sergio Pinheiro Firpo (Orientador)
FGV – EESP

Prof. PhD. Vladimir Ponczek
FGV – EESP

Prof. PhD. Paulo Roberto Arvate
FGV – EAESP

AGRADECIMENTOS

Quero endereçar meus agradecimentos a Deus, a meus pais e irmão por serem meu porto seguro e nunca me deixarem desistir e por serem parâmetro indiscutível de determinação, orgulho e inspiração para atingir meus sonhos.

A meus amigos por entenderem minha ausência no período em que o mestrado foi prioridade e me darem todo o apoio que precisei por não morar em São Paulo.

A meu namorado pela excepcional paciência, dedicação e carinho que sempre me foi dado, por estar ao meu lado transmitindo calma e não pensar duas vezes em abdicar nossos finais de semana para estudarmos juntos e esmiuçar cada detalhe das inúmeras listas de exercícios que pareciam não ter fim. Por me mostrar que juntos podemos vencer qualquer desafio e que tudo vale a pena ao seu lado.

A meus professores pelos numerosos e-mails respondidos, pela disposição na enésima explicação dos mesmos exercícios e por todos os ensinamentos adquiridos nestes dois ricos anos de mestrado.

RESUMO

A participação política das mulheres é tema recorrente nos estudos sobre a desigualdade de gênero em diversas esferas da sociedade. Com uma literatura ainda incipiente, o estudo sobre o efeito que a eleição de mulheres tem sobre o aumento da participação política feminina em eleições subsequentes ainda não tem vertentes com foco no caso brasileiro. O objetivo deste trabalho é, mediante estudo dos dados do Tribunal Superior Eleitoral (TSE), mensurar a influência que a eleição de mulheres para o cargo de prefeito tem sobre a filiação de novas mulheres aos partidos políticos em pleitos subsequentes. O quadro metodológico se desenvolve ao redor das regressões descontínuas (RDD na abreviação em inglês), cuja funcionalidade é testar descontinuidades estruturais que seriam causadas pela eleição de prefeitas. Dentre os resultados, encontramos relações causais dúbias e a ausência de robustez nas análises estatísticas não nos permite tirar conclusões que corroborem a hipótese testada no trabalho.

Palavras-chave: eleições municipais no Brasil, mulheres na política, participação política, gênero, inferência causal, regressão descontínua.

ABSTRACT

Women participation in politics is a persistent subject found in studies and research on gender inequality across multiple social spheres. With a still incipient literature, the current studies of the effect of women's electoral victory over female political participation in subsequent elections do not focus on the Brazilian case. Through data gathered from the Brazilian Supreme Electoral Court, this study aims to measure the influence of women's municipal electoral victory over female political filiation in subsequent elections. The methodological framework walks through a regression discontinuity analysis (RDD), whose functionality is to test structural shifts caused by the election of women. The results are dubious and the lack of discernible causal effects fails to validate the main hypothesis.

Keywords: Brazilian municipal elections, women and politics, political participation, gender, causal inference, regression discontinuity designs.

LISTA DE ILUSTRAÇÕES

Gráfico 1 – Histograma da <i>running variable margin</i> : diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito, após tratar os valores acima do 95º percentil como <i>outliers</i>	58
Gráfico 2 – Histograma da variável resposta proporção de mulheres/homens filiados aos partidos políticos	59
Gráfico 3 – Histograma da variável resposta razão entre mulheres/homens filiados ao partido do prefeito eleito	60
Gráfico 4 – Histograma da variável resposta razão entre mulheres/homens filiados aos partidos que compõem a legenda do partido eleito	60
Gráfico 5 – Média da proporção de mulheres/homens filiados aos partidos políticos por faixa da margem de vitória (<i>running variable</i>)	61
Gráfico 6 – Média da razão entre mulheres/homens filiados ao partido do prefeito eleito por faixa da margem de vitória (<i>running variable</i>)	61
Gráfico 7 – Média da razão entre mulheres/homens filiados aos partidos que compõem a legenda do partido eleito por faixa da margem de vitória (<i>running variable</i>)	62
Gráfico 8 – Número de observações da variável resposta: proporção de mulheres/homens filiados aos partidos políticos por faixa da margem de vitória	63
Gráfico 9 – Número de observações da variável resposta: razão entre mulheres/homens filiados ao partido do prefeito eleito por faixa da margem de vitória	63
Gráfico 10 – Número de observações da variável resposta: razão entre mulheres/homens filiados aos partidos que compõem a legenda do partido eleito por faixa da margem de vitória	64
Gráfico 11 – Histograma da quantidade de homens filiados aos partidos políticos antes de 2012	70
Gráfico 12 – Média da quantidade de homens filiados aos partidos políticos antes de 2012 por faixa da margem de vitória (<i>running variable</i>)	71
Gráfico 13 – Histograma da quantidade de mulheres filiadas aos partidos políticos antes de 2012	71
Gráfico 14 – Média da quantidade de mulheres filiadas aos partidos políticos antes de 2012 por faixa da margem de vitória (<i>running variable</i>)	72

Figura 1 – Regressão descontínua na metodologia de Brookman (2013) 17

Figura 2 – Exemplo da macro utilizada para extrair as bases dos filiados ao PDT utilizando um arquivo executável 86

LISTA DE TABELAS

Tabela 1 – Descrição das variáveis contidas na base de dados dos candidatos nas eleições a partir de 2014	29
Tabela 2 – Relação das bases de dados dos candidatos nas eleições por ano e estado	30
Tabela 3 – Descrição das variáveis presentes na base dos eleitores filiados aos partidos políticos	31
Tabela 4 – Campos mantidos na base de candidatos	33
Tabela 5 – Exemplos de correções nas bases de candidatos e filiados	34
Tabela 6 – Ajustes nas siglas dos partidos políticos	34
Tabela 7 – Percentual de preenchimento da informação de gênero na base de filiados	36
Tabela 8 – Quantidade de gênero por primeiros nomes	37
Tabela 9 – Quantidade de gênero, exceto <i>missing</i> , por primeiros nomes	38
Tabela 10 – Quantidade de gênero ordenada por nome e maior volume	38
Tabela 11 – Gênero selecionado por nome de acordo com o maior volume encontrado	38
Tabela 12 – Novo percentual de preenchimento da informação de gênero na base de filiados	39
Tabela 13 – Descrição das variáveis presentes na base dos eleitores filiados aos partidos políticos após o processo de manipulação dos dados para trazer a informação de gênero da base de candidatos	39
Tabela 14 – Probabilidade de ser feminino/masculino por primeiro nome	40
Tabela 15 – Probabilidade ordenada de ser feminino/masculino por primeiro nome	40
Tabela 16 – Cálculo da diferença de probabilidade por nome	40
Tabela 17 – Quantidade de filiados masculino e feminino em cada partido por ano e município	42
Tabela 18 – Quantidade de filiados masculino e feminino em cada partido por ano e município e quantidade de homens e mulheres por ano e município, independente do partido	42

Tabela 19 – Exemplo da base criada contendo a variável resposta: proporção de mulheres sobre homens por município nos anos 2001, 2002, 2003 e 2004	43
Tabela 20 – Estatísticas descritivas da variável proporção de mulheres/homens que se filiaram entre 2001-2015	44
Tabela 21 – Estatísticas descritivas da variável: margem de vitória (Eleições de 2000, 2004, 2008 e 2012)	45
Tabela 22 – Frequência da variável: gênero do prefeito eleito considerando eleições de 2000, 2004, 2008 e 2012	45
Tabela 23 – Frequência da variável: Indicador das eleições em que os dois candidatos mais votados no município são necessariamente um homem e uma mulher (Período: eleições de 2000, 2004, 2008 e 2012)	46
Tabela 24 – Layout da base contendo a variável resposta considerando informações dos filiados, <i>running variable</i> e variável de tratamento	46
Tabela 25 – Ajustes nas siglas dos partidos políticos	49
Tabela 26 – Base de candidatos contendo as informações do partido e composição da legenda do prefeito eleito, ambas em formato de <i>dummies</i>	49
Tabela 27 – Campos provenientes da base de candidatos adicionados a base final de filiados	50
Tabela 28 – Variáveis criadas a partir da identificação do partido eleito e da composição da legenda	52
Tabela 29 – Volumetria da base final de acordo com o ano	55
Tabela 30 – Frequência da variável: gênero do prefeito eleito após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Período: 2000, 2004, 2008 e 2012)	57
Tabela 31 – Estatísticas descritivas das variáveis: margem de vitória após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Eleições de 2000, 2004, 2008 e 2012)	57
Tabela 32 – Estatísticas descritivas das variáveis: inter (mulher_venceu*margem de vitória) após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Eleições de 2000, 2004, 2008 e 2012)	58
Tabela 33 – Estatísticas descritivas das variáveis resposta após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Período: 2000, 2004, 2008 e 2012)	59

Tabela 34 – Resultados dos coeficientes da variável <i>mulher_venceu</i> para os modelos que utilizam: <i>prop_fem_masc</i> , <i>r_part</i> e <i>r_lege</i> como variáveis resposta	66
Tabela 35 – Média e desvio padrão para a variável <i>margin</i>	67
Tabela 36 – Resultados dos coeficientes da variável <i>mulher_venceu</i> para os modelos que utilizam: <i>prop_fem_masc</i> , <i>r_part</i> e <i>r_lege</i> como variáveis resposta considerando o intervalo1 = -0,1083 <= <i>margin</i> <= 0,1083 como corte	68
Tabela 37 – Resultados dos coeficientes da variável <i>mulher_venceu</i> para os modelos que utilizam: <i>prop_fem_masc</i> , <i>r_part</i> e <i>r_lege</i> como variáveis resposta considerando o intervalo2 = -0,0542 <= <i>margin</i> <= 0,0542 como corte	68
Tabela 38 – Resultados dos coeficientes da variável <i>mulher_venceu</i> para os modelos que utilizam: <i>prop_fem_masc</i> , <i>r_part</i> e <i>r_lege</i> como variáveis resposta considerando o intervalo3 = -0,0217 <= <i>margin</i> <= 0,0217 como corte	69
Tabela 39 – Resultados dos coeficientes da variável <i>mulher_venceu</i> em 2012 sobre as covariadas observadas nos 4 anos antes à eleição de 2012	74
Tabela 40 – Média e desvio padrão para a variável <i>margem de vitória</i> em 2012	74
Tabela 41 – Resultados dos coeficientes da variável <i>mulher_venceu</i> em 2012 sobre as covariadas observadas nos 4 anos antes à eleição de 2012 considerando o intervalo: -0,1022 <= <i>margin_2012</i> <= 0,1022 como corte	75
Tabela 42 – Descrição das variáveis contidas na base de dados dos candidatos nas eleições até 2010	84
Tabela 43 – Descrição das variáveis contidas na base de dados dos candidatos nas eleições em 2012	85
Tabela 44 – Siglas e nomes dos partidos políticos levados em consideração para a extração dos dados e análises presentes neste estudo	87

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Literatura sobre empoderamento das mulheres na política	16
2.2	Literatura sobre regressão descontínua	18
3	METODOLOGIA	20
3.1	Formulação da regressão descontínua	21
3.1.1	<i>Estimação do efeito causal do tratamento</i>	22
3.1.2	<i>Análise gráfica</i>	23
3.1.3	<i>Escolha da largura de banda para análise gráfica e regressões</i>	24
3.1.4	<i>Análise de covariadas</i>	26
3.2	Aplicação da regressão descontínua no presente estudo	26
4	DADOS	28
4.1	Extração dos dados	28
4.2	Manipulação dos dados	32
4.2.1	<i>Ajustes nas bases iniciais</i>	33
4.2.2	<i>Obter a informação de gênero para os filiados</i>	34
4.2.3	<i>Variável resposta considerando informações de filiados; running variable e variável de tratamento</i>	41
4.2.4	<i>Variável resposta considerando informações de candidatos</i>	48
5	RESULTADOS	56
5.1	Análises descritivas e análise gráfica das variáveis resposta	56
5.2	Estimação	64
5.3	Análise de covariadas	69

5.4	Discussão dos resultados	75
6	CONCLUSÃO	77
	REFERÊNCIAS	79
	APÊNDICES	82

1 INTRODUÇÃO

A literatura sobre desigualdade de gênero na política é tema constante na agenda de discussão sobre a participação das mulheres nas esferas decisórias do poder público. Estudos conduzidos pela Organização das Nações Unidas - ONU (2005) e a *Inter-Parliamentary Union* (2004) atestam a sub-representação política das mulheres como um estigma a ser combatido. Dados de 130 países coletados e analisados pelas instituições supracitadas ilustram que, dentro desta amostra, a participação das mulheres no universo dos eleitos não chega a 16%. Na mesma coletânea de estudos a ONU considera um percentual de 30% como patamar mínimo para que as mulheres exerçam algum tipo de influência política relevante. Não é objetivo deste estudo esmiuçar tais metodologias, mas voltar esta problemática para o caso brasileiro e, de forma mais aprofundada, entender se a eleição de mulheres tem algum tipo de influência no incremento da participação política feminina. Como estudo de caso, analisar-se-á esta relação no âmbito municipal, testando estatisticamente em que medida uma mulher no mais alto cargo político da cidade faz com que outras mulheres participem da vida pública, participação esta que, em nossa análise, traduz-se em número de filiações subsequentes a partidos políticos.

Segundo a ONU (*ibidem*), é curioso notar que a participação política das mulheres tem aumentado em alguns países com índice de desenvolvimento humano menor. Da amostra de 130 países citada acima, Ruanda, África do Sul e Moçambique têm, respectivamente, 48,8%, 32,8% e 30% de participação política feminina. A explicação para este desempenho acima da média é que políticas de inclusão eleitoral – como o sistema de cotas – contribuem para maior equalização participativa entre gêneros. Pande e Ford (2011) ponderam que há evidências corroborando a hipótese de que o desenho de cotas é importante para assegurar a representatividade feminina nas eleições. Na Índia, por exemplo, posições nos governos locais reservadas sob o sistema de cotas permitem que o líder (uma mulher a cada terço dos vilarejos) tenham jurisdição suficiente para alocar os recursos a ele (a) destinado da forma que achar mais proveitoso; esse tipo de rede de influência tem, segundo as autoras, um impacto positivo no aumento da participação política das mulheres no longo prazo.

A literatura sobre a participação política das mulheres na América Latina, Caribe e, especificamente, Brasil versam sobre as particularidades que este último tem dentre os demais. Cavenaghi e Alves (2012) lançam luz a uma peculiaridade: apesar de o empoderamento das mulheres no Brasil ter crescido nos últimos anos – inclusive pelo país ter eleito uma mulher para a Presidência da República -, a participação feminina no poder legislativo (8,8% em 2010) é muito menor que nas outras regiões citadas (23% em 2012). Inglehart (1997, *apud.* Cavenaghi e Alves, 2012) concluiu que no período pós-Segunda Guerra Mundial existia uma tendência de as mulheres votarem à direita do espectro político nos países ocidentais, mas isso mudou à medida que os padrões de consumo e bem-estar também cresceram. Inglehart e Norris (2000) perceberam que, ao longo dos anos, os votos das mulheres passou a migrar para a esquerda do espectro político. Desenvolvendo este fato, Cavenaghi e Alves (2012) analisam a eleição de Dilma Rousseff para a Presidência do Brasil em 2010, tentando entender a motivação dos votos que a presidente eleita recebeu. Em outras palavras, o que estaria por trás das intenções dos votos recebidos por Dilma?

Em seu estudo, Cavenaghi e Alves (*ibidem*) concluem que “o modelo logístico multinomial mostrou que a variável com maior poder de explicação das intenções de voto declaradas para a Presidência da República foi se o eleitor votaria em candidato indicado pelo presidente Lula. O carisma do ex-presidente e os bons índices da avaliação do governo reforçaram o desejo de continuidade da administração anterior. O resultado foi como uma reeleição do projeto político do governo, no qual a candidata Dilma era figura proeminente.” (p.77). Ou seja, neste caso específico o candidato da situação ganhou respaldo intrínseco ao governo e determinou a continuidade do partido. Avaliando tal situação de forma mais ampla, nos deparamos com um paradoxo: apesar da posse de uma presidente, o Brasil ainda caminha atrás de outros países em desenvolvimento na inclusão feminina na política. Mais do que isso, a própria eleição de Dilma parece estar mais ligada ao seu antecessor do que a características que a tornam um personagem político idiossincrático.

O exemplo acima correlaciona-se com um conceito largamente utilizado em Ciência Política: o efeito incumbência, segundo o qual uma pessoa que já ocupa um cargo político tem grandes chances (ou chances maiores em relação a candidatos de fora) de retê-lo. Entretanto, essa análise não fornece insumos para concluir se a eleição de

uma mulher aumenta a probabilidade de novas mulheres se candidatarem a eleições subsequentes. Dito isso, o objetivo deste estudo não é, estritamente, avaliar o efeito incumbência que as mulheres têm quando detêm um cargo político, mas testar uma relação conceitual diferente: uma mulher no mais alto cargo político da cidade faz com que outras mulheres participem da vida pública, em particular, se filiando a partidos?

2 REVISÃO BIBLIOGRÁFICA

2.1 Literatura sobre empoderamento das mulheres na política

Bhalotra *et al* (2013) analisam o impacto que a eleição de mulheres na Índia tem na diferenciação de gênero entre os eleitos nas votações subsequentes. O estudo se desenrola com base na literatura existente sobre as escolhas políticas favoráveis às mulheres em situações nas quais as representações políticas no poder são femininas. De maneira geral, os autores estudam o estigma de que as mulheres têm representação política inferior à dos homens, o que se dá por inúmeros motivos, quais sejam, discriminação de gênero, ausência de modelos de incentivo à participação feminina, menor ambição política por parte das mulheres, pouco apetite por situações muito competitivas, responsabilidades familiares, normas sociais e outros fatores. Como a literatura econômica (e em certa medida econométrica) sobre estes fatores é escassa e a vitória política de uma mulher pode ser atribuída ou correlacionada com características não observáveis – como preferências de voto -, os autores buscam testar as relações estatísticas que explicam ou comparam situações em que mulheres vencem os homens com reduzida margem de erro com situações em que o inverso ocorre, com o objetivo de encontrar descontinuidades nestes limiares. Para tal, utilizam um modelo de regressão descontínua (*regression discontinuity design*, doravante RDD), mesma ferramenta utilizada pela vasta literatura que testa os efeitos incumbência em Ciência Política.

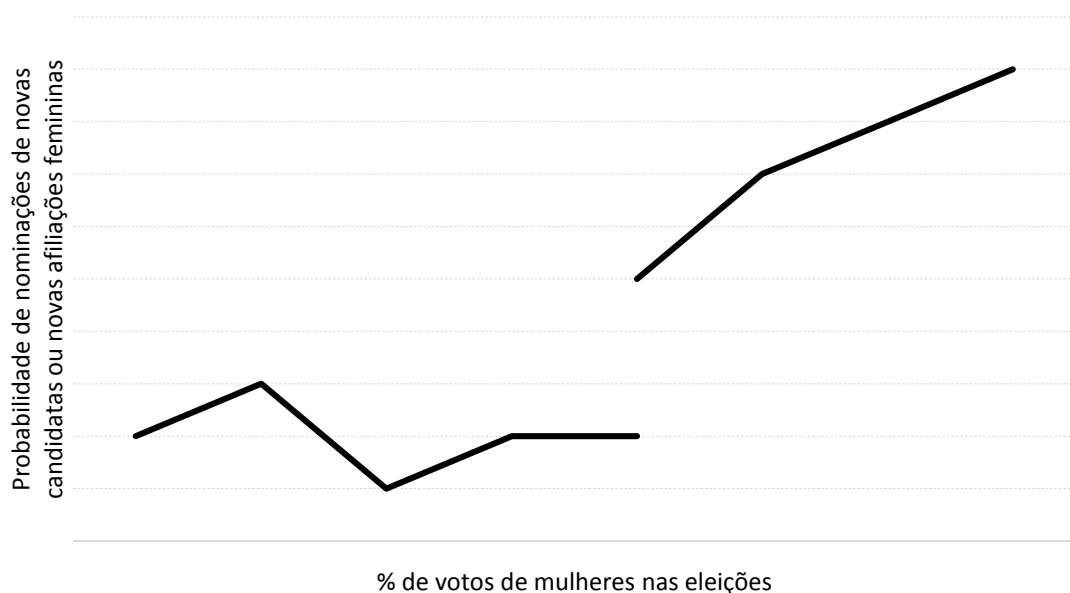
Os primeiros achados de Bhalotra *et al* (*ibidem*) são 1) a vitória de uma mulher numa eleição resulta num crescimento de 9,2% na fração de candidatas nos partidos maiores em eleições subsequentes e 2) não há evidência de *spillovers*, ou seja, o efeito tem limitação regional e não se alastra indefinidamente. O aumento das candidaturas femininas não ocorre por conta da redução do número de candidatos, ao contrário, há uma substituição de candidatos homens por mulheres.

Conclusões secundárias são igualmente interessantes: o número de mulheres concorrendo entre os candidatos mais competitivos (que recebem 5% ou mais dos

votos) também cresce sensivelmente. Contudo, essas mulheres não são novas entrantes nos partidos políticos, vale dizer, o percentual de mulheres dentro dos partidos políticos não muda, mas as candidatas preexistentes estimulam-se ou são estimuladas a participarem com mais ímpeto nas futuras eleições.

Análise semelhante é feita por Brookman (2013) ao estudar o cenário eleitoral e a divergência de gênero nos Estados Unidos. Novamente, o ponto de partida é a desigualdade, neste caso no que tange à lentidão no crescimento das afiliações políticas das mulheres norte-americanas. Dialogando com os trabalhos de Sonia Bhalotra, Brookman (*ibidem*) chega a conclusões diferentes das obtidas no caso indiano. Utilizando a mesma metodologia (RDD), o autor analisa 3813 eleições legislativas estaduais nos Estados Unidos para estimar os efeitos causais entre a eleição de mulheres e a presença de candidatas adicionais em novos pleitos. Dessa forma, testa-se 1) a probabilidade de, em casos nos quais a margem de votos recebidos por mulheres em relação a candidatos homens é pequena, as mulheres que venceram receberem mais votos em eleições subsequentes e 2) a probabilidade de, dadas as mesmas circunstâncias de margem de voto, mais mulheres se candidatarem em novos pleitos quando candidatas foram eleitas anteriormente. De forma gráfica, o autor busca testar a seguinte quebra estrutural:

Figura 1 - Regressão descontínua na metodologia de Brookman (2013)



Fonte: Elaboração própria com base em Brookman (2013)

A quebra estrutural percebida no gráfico acima ilustra – quando há significância estatística – a influência positiva que a eleição de uma mulher tem no incremento da participação política feminina em pleitos posteriores. Entretanto, não foi esse o resultado encontrado por Brookman (*ibidem*). As evidências nos EUA, diferente do que se observa na Índia, são mistas e insuficientes para corroborar a hipótese, vale dizer, dentro da metodologia proposta, a eleição de mulheres tem efeito estatístico nulo no aumento do número de candidatas futuras.

Este trabalho pretende analisar a mesma questão para o caso brasileiro. Utilizar-se-á a mesma metodologia de regressão descontínua para explorar os dados de filiados e candidatos por partidos políticos em cada município brasileiro, fornecidos pelo Tribunal Superior Eleitoral (TSE). A granularidade municipal nos permitirá avaliar se há ou não significância estatística na influência que a eleição (no âmbito municipal) de mulheres tem sobre o crescimento de filiações femininas em cada partido no ano eleitoral (também municipal) seguinte. A seguir, exploraremos a metodologia de regressão descontínua e como aplica-la no escopo deste trabalho.

2.2 Literatura sobre regressão descontínua

As Regressões Descontínuas objetivam, mediante manipulação quase-experimental dos dados observacionais, estabelecer relações de causalidade – e não somente de correlação – com imunidade à endogeneidade. Desde o final da década de 1990 inúmeros estudos no campo da economia tem sido publicados com o intuito de aplicar e estender métodos de regressão descontínua, incluindo Van der Klaauw (2002), Black (1999), Angrist e Lavy (1999), Chay e Greenstone (2005), DiNardo e Lee (2004), Chay, McEwan, e Urquiola (2005), McEwan e Shapiro (2007), e Card, Mas e Rothstein (2006). Porém, muito antes deste período, a introdução da regressão descontínua como metodologia de inferência causal para problemas envolvendo dados estritamente não-experimentais já pôde ser encontrada na literatura graças à Thistlewaite e Campbell (1960). Esses autores o fizeram em contraposição à técnica de pareamento de observáveis, e tiveram como objetivo analisar o efeito que o

reconhecimento acadêmico de certos estudantes teve sobre a continuidade de seus estudos e sua carreira.

Após algumas décadas, o emprego mais consistente da regressão descontínua (ou RDD, do inglês *Regression Discontinuity Design*) voltou a ganhar força, quando se tornou particularmente popular nos campos da Economia mais intimamente ligados às ciências políticas e sociais (LEE; LEMIEUX, 2010). Na verdade, não é incomum que programas governamentais em áreas como saúde e educação, bem como algumas leis, determinem políticas públicas para a concessão de benefícios e tratamentos específicos que eventualmente se enquadram nas condições de análise pelo RDD (van der KLAUW, 2008), condições essas que serão discutidas na seção Metodologia.

Na área político-eleitoral, o RDD foi utilizado para a verificação do que chamamos de “vantagem do incumbente”, em outras palavras, na vantagem competitiva que o candidato ou o partido que está no poder tem sobre aquele que é considerado desafiante, na perspectiva de novas eleições. Lee (2008) analisou o efeito da “vantagem” do partido Democrata, ou seja, analisou a probabilidade de reeleição do mesmo partido para a Câmara de Deputados americana. Atualmente, Cattaneo, Frandsen e Titiunik (2014) aplicaram o conceito de RDD, provida de certos aprofundamentos metodológicos, em uma análise semelhante aplicada às eleições para o Senado do mesmo país. Brambor e Ceneviva (2012) seguiram a ideia de Lee (2008) e mensuraram o efeito da incumbência no desempenho eleitoral de candidatos nas eleições municipais brasileiras. Por fim, Avelino, Biderman e Barone (2012) verificaram a influência do resultado das eleições para prefeito sobre as eleições estaduais subsequentes, em específico, exploraram a vantagem que um partido poderia ter, ao eleger o prefeito nas eleições de 2008, para as eleições proporcionais estaduais de 2010.

Além dos autores citados, em consequência da expansão de seu uso, a metodologia da regressão descontínua se tornou um objeto de interesse de alguns econometristas, que se esforçam por divulgá-la e aprofundá-la (IMBENS; LEMIEUX, 2008; LEE; CATONICO; CATTANEO; TITIUNIK, 2014a, 2014b e 2014c).

3 METODOLOGIA

Como descrito na sessão anterior, existem artigos que se dedicam a avaliar a metodologia de regressões descontínuas, nos quais se destacam os de van der Klaauw (2008), em análise generalista; os de Imbens e Lemieux (2008) e de Lee e Lemieux (2010), que se propõem a servir de guia prático para o uso desta técnica, apresentando-a de modo detalhado.

Os artigos citados nos apresentam a regressão descontínua (ou RDD, do inglês *Regression Discontinuity Design*) que, em sua formulação mais básica, se encarrega por estimar o efeito sobre uma variável resposta de um tratamento que é aplicado caso uma covariável observada e contínua assuma um valor superior a um ponto de corte especificado. A essa covariável denota-se na literatura *running*, *forcing*, *assigning* ou mesmo *score variable*.

Sob certas condições, argumenta-se que indivíduos ou elementos situados imediatamente acima do corte (que recebem tratamento) e imediatamente abaixo (que não o recebem) são iguais sob todos os aspectos, observáveis ou não, definidos anteriormente ao tratamento. Portanto, a diferença de comportamento ou de resposta que venham a apresentar posteriormente poderia ser imputada como efeito desse tratamento. Em outras palavras, trata-se de considerar que, na vizinhança do corte, a escolha de elementos tratados ou não-tratados é localmente aleatória. A formulação para análise dos dados observacionais aproxima-se assim da abordagem experimental, em que um processo efetivamente randômico é empregado para definir grupos de tratamento e de controle.

Ainda no contexto da inferência causal, a regressão descontínua apresenta-se como alternativa à técnica de pareamento em observáveis e ao próprio uso de variáveis instrumentais. Nas subseções a seguir, a metodologia de regressão descontínua será apresentada de modo mais formal, e em sequência será indicado de que forma o presente estudo fez uso da mesma.

3.1 Formulação da regressão descontínua

Imbens e Lemieux (2008) fazem a apresentação dos aspectos formais do RDD a partir de um enquadramento de resultados potenciais. Denotando $Y_i(1)$ e $Y_i(0)$ como um par de resultados da variável resposta para o i -ésimo elemento em uma população, respectivamente sob a incidência do tratamento ou não, os autores fazem notar que a comparação entre os dois resultados potenciais – indicando o efeito do tratamento sobre a resposta – não pode ser feita de maneira direta visto que, para cada elemento, apenas um dos resultados é efetivamente observado. De fato, se $W_i \in \{0,1\}$ denota o tratamento recebido, então o resultado observado é dado pela equação:

$$Y_i = (1 - W_i) \cdot Y_i(0) + W_i \cdot Y_i(1) \quad (1)$$

No caso básico de RDD, a aplicação do tratamento é determinística, ocorrendo em função do valor assumido pela *running variable* contínua, denotada X , e em função de um valor de corte especificado c , conforme a equação:

$$W_i = 1\{X_i \geq c\} \quad (2)$$

em que $1\{\cdot\}$ é a função indicadora, que vale 1 se seu argumento for verdadeiro e 0 caso contrário.

A identificação do efeito causal do tratamento passa então pela verificação da descontinuidade das expectativas condicionais do resultado observado no ponto de corte, o que se descreve conforme a equação:

$$\tau_{RD} = \lim_{x \downarrow c} E[Y_i|X_i = x] - \lim_{x \uparrow c} E[Y_i|X_i = x] = E[Y_i(1) - Y_i(0)|X_i = c] \quad (3)$$

em que se assume a continuidade das expectativas condicionais dos resultados potenciais no ponto $X = c$.

3.1.1 Estimação do efeito causal do tratamento

Em uma aplicação prática, e dado o que foi apresentado, uma forma intuitiva de se estimar o efeito do tratamento em uma população seria a de se tomar a diferença entre as médias da variável resposta para os dois subconjuntos compostos dos elementos situados a uma distância h do ponto de corte, acima ou abaixo do mesmo. A isso corresponde a abordagem não-paramétrica com *kernel* retangular, cujo viés é discutido por Imbens e Lemieux (2008) e por van der Klaauw (2008).

Uma elaboração desse conceito é permitir que a resposta varie linearmente como função de X nos subconjuntos, e daí a abordagem de regressões lineares locais, em que o efeito τ é dado pela solução do problema de minimização:

$$\min_{\alpha, \beta, \tau, \gamma} \sum_{i=1}^N 1\{c - h \leq X_i \leq c + h\} \cdot (Y_i - \alpha - \beta \cdot (X_i - c) - \tau \cdot W_i - \gamma \cdot (X_i - c) \cdot W_i)^2 \quad (4)$$

em que N é o número total de observações, referenciadas pela letra i . O processo equivale então a se ajustar uma função linear para cada subconjunto e se tomar como efeito do tratamento a diferença entre o valor assumido por cada reta no ponto de corte.

Conforme Lee e Lemieux (2010), é natural que se verifique a sensibilidade da estimativa obtida à inclusão de termos de maior ordem nas regressões locais, ao passo que Gelman e Imbens (2014) discutem razões para que não se vá além dos termos de segunda ordem (regressões quadráticas locais). Além disso, outro aspecto importante do problema é a questão de como se definir a distância h do ponto de corte que limita os elementos participantes da estimação, chamada na literatura de *bandwidth*, ou largura de banda – o que será tratado na subseção 3.1.3 a seguir.

3.1.2 Análise gráfica

Os artigos citados nesta seção Metodologia e demais artigos relacionados reforçam a importância da análise gráfica na aplicação prática do RDD. O exame visual da variável resposta em função da *running variable*, assim como da densidade desta última, é tido por Imbens e Lemieux (2008) como uma forma simples e poderosa de identificação do efeito do tratamento, além de conferir credibilidade e transparência às análises. Não obstante, Lee e Lemieux (2010) alertam para o fato de que tais análises não podem ser conclusivas isoladamente, e que métodos gráficos criteriosos são recomendados para que se evitem vieses de construção.

Nesse sentido, e conforme os artigos citados anteriormente, a análise gráfica se baseia na definição de bandas ou faixas ao longo da *running variable* X , dentro das quais são calculados os valores médios das variáveis de interesse. Em outras palavras, trata-se de um processo de categorização da *running variable*, ou de binagem (do inglês, *binning*). Dada uma largura de banda h , e para um certo número de bandas (em inglês, *bins*) K_0 e K_1 à esquerda e à direita do ponto de corte em $X = c$, são definidos intervalos do tipo $(b_k, b_{k+1}]$, com k variando de 1 a $K = K_0 + K_1$ e de tal modo que b_k seja definido pela equação:

$$b_k = c - (k_0 + k + 1)/h \quad (5)$$

Para cada intervalo, é calculado o ponto médio \tilde{b}_k ; o número de observações N_k ; e a média \bar{Y}_k da variável resposta Y (ou de qualquer outra variável de interesse Z , dita covariada), conforme:

$$\tilde{b}_k = (b_k + b_{k+1})/2 \quad (6)$$

$$N_k = \sum_{i=1}^N 1\{b_k < X_i \leq b_{k+1}\} \quad (7)$$

$$\bar{Y}_k = \frac{1}{N_k} \sum_{i=1}^N Y_i \cdot 1\{b_k < x_i \leq b_{k+1}\} \quad (8)$$

Os gráficos de maior relevância para a análise são os das médias e o do número de observações contra os pontos médios. É natural que o efeito do tratamento, em existindo, se faça notar pelo salto da média da variável resposta no ponto de corte. Por outro lado, normalmente não é esperada a ocorrência de saltos de magnitude similar em outros pontos ao longo da *running variable*, tanto quanto saltos em covariadas determinadas anteriormente ao tratamento denotam fragilidade à interpretação causal. Por fim, uma descontinuidade da distribuição das observações (ou seja, de sua densidade) pode ser indício de manipulação do processo de auferição do tratamento por parte dos agentes, o que fere diretamente o argumento de quase-aleatoriedade (IMBENS; LEMIEUX, 2008 e LEE; LEMIEUX, 2010).

3.1.3 Escolha da largura de banda para análise gráfica e regressões

Tanto a análise gráfica quanto as regressões para determinação do efeito do tratamento requerem, como se viu, a definição de uma largura de banda, seja para a discretização dos dados em face de sua exposição gráfica, seja para a seleção dos limites das regressões.

No enquadramento das regressões lineares locais, a escolha da largura de banda se traduz num balanço entre precisão – favorecida com o uso de bandas maiores, pelo maior número de observações incluídas nas regressões – e acurácia – sendo o menor viés obtido com o uso de bandas menores, quando apenas observações mais próximas do ponto de corte são empregadas para estimação do efeito do tratamento nesse ponto (LEE; LEMIEUX, 2010).

Tem-se, portanto, um contexto de otimização, e o procedimento preferido no presente trabalho para verificação da largura de banda das regressões de estimação do efeito do tratamento, por seu apelo intuitivo, é o de *cross-validação*.

O processo de *cross-validação* consiste em se estimar, para cada observação i , o valor Y_i assumido pela variável resposta com base em regressões lineares computadas sobre observações estritamente à direita ou à esquerda de i , conforme a

própria observação i esteja à direita ou à esquerda do ponto de corte. Assim, se $X_i > c$, estima-se para Y_i o valor \hat{Y}_i como uma extrapolação linear a partir das observações situadas no intervalo $(X_i, X_i + h]$. Analogamente, se $X_i < c$, o intervalo considerado é $(X_i - h, X_i]$.

Uma função de *cross-validação* é definida pela soma dos erros quadráticos (comparação entre os valores observados Y_i e estimados \hat{Y}_i), e depende da escolha de h :

$$CV(h) = \frac{1}{N^*} \sum_{i=1}^{N^*} (Y_i - \hat{Y}_i)^2 \quad (9)$$

em que N^* é o número de observações considerado para escolha da largura de banda por esse processo.

A largura de banda h_{CV}^* escolhida por meio desse processo de *cross-validação* é dada pelo valor de h que minimiza a função de *cross-validação* (LEE; LEMIEUX, 2010):

$$h_{CV}^* = \arg_h \min CV(h) \quad (10)$$

Ligeira modificação desse procedimento tendo em vista a escolha da largura dos *bins* na análise gráfica requer apenas que os Y_i sejam estimados com base em regressões sobre uma constante – ou seja, anula-se o coeficiente do termo linear das regressões originais, o que tem sua razão de ser no fato de que a análise gráfica é baseada na inspeção do valor médio das variáveis de interesse no interior dos *bins*.

Para uma maior formalização do procedimento de *cross-validação*, refere-se a Imbens e Lemieux (2008). Para procedimentos alternativos de escolha da largura de banda para as regressões de estimativa do efeito do tratamento, sugere-se a consulta a Imbens e Lemieux (2008); Lee e Lemieux (2010); e Imbens e Kalyanaraman (2012). Em particular para a análise gráfica, Lee e Lemieux (2010) também apresentam duas outras propostas intuitivas, denominadas pelos autores de *bin* e *regression test*. Essas últimas técnicas serão apresentadas na seção Resultados como complemento das análises.

3.1.4 Análise de covariadas

A inspeção de covariadas, definidas como variáveis determinadas (ou seja, cujo conteúdo é conhecido) antes da especificação do tratamento, é dada na literatura como uma etapa importante da verificação da validade da regressão descontínua.

Particularmente, se as covariadas são anteriores ao tratamento, não pode haver relação de causalidade deste em direção àquelas. Assim, examinada uma covariada segundo a mesma metodologia aplicada à variável resposta, não é esperado que se encontre uma descontinuidade em seu valor no ponto de corte que define o tratamento – caso contrário, tal fato poderia indicar problemas de especificação.

Lee e Lemieux (2010) propõem que se examine as covariadas de modo conjunto, por meio de *Seemingly Unrelated Regressions* (regressão SUR) – em que se estima o sistema de equações das covariadas em função da variável de tratamento e da *running variable* – para em seguida se verificar se os coeficientes associados ao tratamento para cada covariada são conjuntamente nulos, conforme seria desejado.

Os autores sugerem, por fim, que as covariadas sejam incorporadas às regressões para estimação do efeito do tratamento sobre a variável resposta, seja por meio da residualização desta última em regressão exclusiva contra as primeiras, seja por meio da incorporação direta das covariadas nas regressões originais. O objetivo de tal inclusão seria o de se reduzir a variabilidade amostral nas estimativas (LEE; LEMIEUX, 2010).

3.2 Aplicação da regressão descontínua no presente estudo

Conforme dito na Introdução, o objetivo do presente estudo é verificar o efeito que as mulheres no mais alto cargo político da cidade têm sob a participação de outras mulheres na vida pública. Mais especificamente, analisar o empoderamento das

mulheres respondendo à seguinte pergunta: uma mulher no mais alto cargo político da cidade faz com que outras mulheres se filiem aos partidos políticos?

No processo eleitoral, está claro que o tratamento “eleição do candidato do gênero feminino” é aplicado caso este candidato tenha mais votos do que um candidato do gênero masculino. Dentro do enquadramento da metodologia de regressão descontínua, então, a *running variable*, nesse caso, é a margem de vitória do(a) candidato(a) eleito(a), ou seja, a diferença do percentual de votos entre os candidatos femininos e masculinos considerando o primeiro e segundo lugares nas eleições municipais. Se essa diferença for positiva (acima do ponto de corte valendo 0, implicando que a candidata eleita venceu com margem de vitória positiva), o tratamento é aplicado, e caso contrário, não.

A aplicação de RDD ao contexto eleitoral, tendo por *running variable* uma variável de percentual de votos nas eleições, já foi levada a cabo por Brookman (2013), Lee (2008) e por Cattaneo, Frandsen e Titiunik (2014) no cenário americano, por Bhalotra *et al* (2013) no cenário indiano e por Brambor e Ceneviva (2012) no cenário brasileiro, e se apoia no fato de que a regra que define o resultado das eleições, a de maioria dos votos, em geral é clara e irremissível. Além disso, ante a quantidade de agentes envolvidos no processo (eleitores, candidatos etc), também fica patente a incapacidade desses agentes, em seu conjunto, de influenciarem de modo preciso o resultado das eleições.

4 DADOS

Como citado anteriormente, não foi encontrado na literatura brasileira algum estudo que analisasse o impacto que a eleição de mulheres tem na diferenciação de gênero entre os filiados aos partidos políticos nas eleições subsequentes. Com isso, a novidade da informação foi aproveitada como motivação e utilizou-se uma fonte de dados incomum na literatura.

A principal fonte de dados para o presente estudo foi o repositório de Dados Eleitorais do Superior Tribunal Eleitoral, em que se encontram informações detalhadas sobre os resultados eleitorais no Brasil. A fonte de dados, assim como o tratamento e o uso que foi feito dos dados serão detalhados nas subseções a seguir.

4.1 Extração dos dados

A extração dos dados foi separada em duas partes: bases com informações acerca do perfil dos candidatos nas eleições e bases dos eleitores filiados aos partidos políticos. As bases com informações dos candidatos foram obtidas do repositório de dados eleitorais do site do Tribunal Superior Eleitoral: <http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais>. Neste endereço é possível selecionar o tópico “Candidatos” e encontrar os dados a partir de 1945, contudo utilizamos dados dos seguintes anos: 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012 e 2014.

Para fazer o *download* dos arquivos, basta clicar no ano desejado e então os dados referentes aos candidatos daquele ano (de todos os estados) serão carregados em formato .zip (que facilmente são descompactados e transformados para .txt). O *layout* das tabelas existentes no repositório de dados eleitorais e suas respectivas características podem ser encontrados no apêndice.

Utilizou-se a sexta versão dos arquivos do repositório de dados eleitorais. Os arquivos com data de geração a partir de 12/01/2014 contemplam as seguintes correções:

- ✓ Inclusão da variável CPF_CANDIDATO nas tabelas de candidatos, para 2012;
- ✓ Inclusão da variável DS_EMAIL nas tabelas CONSULTA_CAND_<ANO ELEIÇÃO>_<SIGLA UF>, a partir do ano de 2012.

As tabelas que contém o *layout* e a descrição das variáveis presentes na base de dados dos candidatos nas versões anteriores (eleições até 2010 e 2012) podem ser encontradas no apêndice. A seguir, temos a descrição das variáveis presentes na base dos candidatos na eleição de 2014:

Tabela 1 - Descrição das variáveis contidas na base de dados dos candidatos nas eleições a partir de 2014

VARIÁVEL	DESCRIÇÃO
DATA_GERACAO	Data de geração do arquivo (data da extração)
HORA_GERACAO	Hora de geração do arquivo (hora da extração) - Horário de Brasília
ANO_ELEICAO	Ano da eleição
NUM_TURNO (*)	Número do turno
DESCRICAO_ELEICAO (*)	Descrição da eleição
SIGLA_UF	Sigla da Unidade da Federação em que ocorreu a eleição
SIGLA_UE (*)	Sigla da Unidade Eleitoral (Em caso de eleição majoritária é a sigla da UF que o candidato concorre (texto) e em caso de eleição municipal é o código TSE do município (número)). Assume os valores especiais BR, ZZ e VT para designar, respectivamente, o Brasil, Exterior e Voto em Trânsito
DESCRICAO_UE	Descrição da Unidade Eleitoral
CODIGO_CARGO (*)	Código do cargo a que o candidato concorre
DESCRICAO_CARGO	Descrição do cargo a que o candidato concorre
NOME_CANDIDATO	Nome completo do candidato
SEQUENCIAL_CANDIDATO (*)	Número sequencial do candidato gerado internamente pelos sistemas eleitorais. Não é o número de campanha do candidato.
NUMERO_CANDIDATO	Número do candidato na urna
CPF_CANDIDATO	CPF do candidato
NOME_URNA_CANDIDATO	Nome de urna do candidato
COD_SITUACAO_CANDIDATURA	Código da situação de candidatura
DES_SITUACAO_CANDIDATURA	Descrição da situação de candidatura
NUMERO_PARTIDO	Número do partido
SIGLA_PARTIDO	Sigla do partido
NOME_PARTIDO	Nome do partido
CODIGO_LEGENDA	Código sequencial da legenda gerado pela Justiça Eleitoral
SIGLA_LEGENDA	Sigla da legenda
COMPOSICAO_LEGENDA	Composição da legenda
NOME_LEGENDA	Nome da legenda
CODIGO_OCUPACAO	Código da ocupação do candidato
DESCRICAO_OCUPACAO	Descrição da ocupação do candidato
DATA_NASCIMENTO	Data de nascimento do candidato
NUM_TITULO_ELEITORAL_CANDIDATO	Número do título eleitoral do candidato
IDADE_DATA_ELEICAO	Idade do candidato da data da eleição
CODIGO_SEXO	Código do sexo do candidato
DESCRICAO_SEXO	Descrição do sexo do candidato
COD_GRAU_INSTRUCAO	Código do grau de instrução do candidato. Gerado internamente pelos sistemas eleitorais
DESCRICAO_GRAU_INSTRUCAO	Descrição do grau de instrução do candidato
CODIGO_ESTADO_CIVIL	Código do estado civil do candidato
DESCRICAO_ESTADO_CIVIL	Descrição do estado civil do candidato
CODIGO_COR_RACA	Código da cor/raça do candidato (auto declaração)

DESCRICAO_COR_RACA	Descrição da cor/raça do candidato (auto declaração)
CODIGO_NACIONALIDADE	Código da nacionalidade do candidato
DESCRICAO_NACIONALIDADE	Descrição da nacionalidade do candidato
SIGLA_UF_NASCIMENTO	Sigla da UF de nascimento do candidato
CODIGO_MUNICIPIO_NASCIMENTO	Código TSE do município de nascimento do candidato
NOME_MUNICIPIO_NASCIMENTO	Nome do município de nascimento do candidato
DESPESA_MAX_CAMPANHA	Despesa máxima de campanha declarada pelo partido para aquele cargo. Valores em Reais.
COD_SIT_TOT_TURNO	Código da situação de totalização do candidato naquele turno
DESC_SIT_TOT_TURNO	Descrição da situação de totalização do candidato naquele turno
NM_EMAIL	E-mail para comunicação com o candidato

As bases dos candidatos nas eleições com as variáveis descritas acima estão distribuídas conforme os anos e estados da seguinte forma:

Tabela 2 - Relação das bases de dados dos candidatos nas eleições por ano e estado

Ano								
1998	2000	2002	2004	2006	2008	2010	2012	2014
AC	AC	AC	AC	AC	AC	AC	AC	AC
AL	AL	AL	AL	AL	AL	AL	AL	AL
AM	AM	AM	AM	AM	AM	AM	AM	AM
AP	AP	AP	AP	AP	AP	AP	AP	AP
BA	BA	BA	BA	BA	BA	BA	BA	BA
BR	CE	BR	CE	BR	CE	BR	CE	BR
CE	ES	CE	ES	CE	ES	CE	ES	CE
DF	GO	DF	GO	DF	GO	DF	GO	DF
ES	MA	ES	MA	ES	MA	ES	MA	ES
GO	MG	GO	MG	GO	MG	GO	MG	GO
MA	MS	MA	MS	MA	MS	MA	MS	MA
MG	MT	MG	MT	MG	MT	MG	MT	MG
MS	PA	MS	PA	MS	PA	MS	PA	MS
MT	PB	MT	PB	MT	PB	MT	PB	MT
PA	PE	PA	PE	PA	PE	PA	PE	PA
PB	PI	PB	PI	PB	PI	PB	PI	PB
PE	PR	PE	PR	PE	PR	PE	PR	PE
PI	RJ	PI	RJ	PI	RJ	PI	RJ	PI
PR	RN	PR	RN	PR	RN	PR	RN	PR
RJ	RO	RJ	RO	RJ	RO	RJ	RO	RJ
RN	RR	RN	RR	RN	RR	RN	RR	RN
RO	RS	RO	RS	RO	RS	RO	RS	RO
RR	SC	RR	SC	RR	SC	RR	SC	RR
RS	SE	RS	SE	RS	SE	RS	SE	RS
SC	SP	SC	SP	SC	SP	SC	SP	SC
SE	TO	SE	TO	SE	TO	SE	TO	SE
SP		SP		SP		SP		SP
TO		TO		TO		TO		TO
28 bases	26 bases	28 bases	26 bases	28 bases	26 bases	28 bases	26 bases	28 bases

Existem 28 bases para os anos de 1998, 2002, 2006, 2010 e 2014, 2 a mais que os restantes, pois aqueles são anos em que ocorrem eleições para cargos federais e são consideradas as bases referentes ao DF e BR. Ao todo, são 244 bases com a seguinte nomenclatura `consulta_cand_AAAA_ZZ` em que AAAA representa o ano e ZZ a sigla do estado (ou DF/BR, quando aplicável).

A listagem de eleitores filiados aos partidos políticos foi encontrada no seguinte endereço: <http://www.tse.jus.br/partidos/filiacao-partidaria/relacao-de-filiados>. Nesta página é possível acessar a relação de filiados de cada partido político. Contudo, é necessário selecionar o partido político e o estado que desejamos consultar e depois clicar no *link* “Baixar lista”. A composição do nome do arquivo é definida como: `filiados_<sigla do partido>_<sigla da uf>.csv` e a formatação dos arquivos estão na seguinte forma:

- ✓ A primeira linha contém o cabeçalho das informações;
- ✓ As informações estão entre aspas duplas e separadas pelo caractere “,”.

O *layout* das bases segue o seguinte formato:

Tabela 3 - Descrição das variáveis presentes na base dos eleitores filiados aos partidos políticos

Variável	Descrição
DATA_DA_EXTRACAO	Data da extração das informações do banco de dados
HORA_DA_EXTRACAO	Hora da extração das informações do banco de dados
NUMERO_DA_INSCRICAO	Número da inscrição eleitoral
NOME_DO_FILIADO	Nome do filiado
SIGLA_DO_PARTIDO	Sigla do partido político
NOME_DO_PARTIDO	Nome do partido político
UF	Unidade da federação
CODIGO_DO_MUNICIPIO	Código do município
NOME_DO_MUNICIPIO	Nome do município
ZONA_ELEITORAL	Zona eleitoral
SECAO_ELEITORAL	Seção eleitoral
DATA_DA_FILIACAO	Data da filiação
SITUACAO_DO_REGISTRO	Situação do registro de filiação
TIPO_DO_REGISTRO	Tipo do registro de filiação
DATA_DO_PROCESSAMENTO	Data do processamento do registro de filiação
DATA_DA_DESFILIACAO	Data da desfiliação
DATA_DO_CANCELAMENTO	Data do cancelamento do registro de filiação
DATA_DA_REGULARIZACAO	Data da regularização do registro de filiação
MOTIVO_DO_CANCELAMENTO	Motivo do cancelamento do registro de filiação

Ao todo são 32 partidos políticos e 27 estados (contando DF), porém não foram encontradas informações para o partido PCB (Partido Comunista Brasileiro) no estado de Roraima e PCO (Partido da Causa Operária) no estado de Tocantins, portanto são 862 arquivos no total.

Para facilitar o *download* dos 862 arquivos, criou-se uma macro com o intuito de automatizar a extração dos dados. O processo consiste em acessar a página e fazer o *download* dos arquivos sem a necessidade de escolher manualmente todos os partidos e estados. Para fazer isto, basta abrir um bloco de notas, digitar o seguinte comando: `"C:\Program Files (x86)\Google\Chrome\Application\chrome.exe"` `http://agencia.tse.jus.br/estatistica/sead/eleitorado/filiados/uf/filiados_XXXX_YY.zip` (onde XXXX representa a sigla do partido e YY representa a sigla do estado) e salvar em formato .bat (para se tornar um arquivo executável). Ao clicar, este comando fará com que o navegador (neste caso Google Chrome) abra a página da internet, acesse o endereço e faça o download do arquivo para o partido XXXX no estado YY em formato .zip. É necessário criar um comando para cada partido e estado, ou seja, haverá 862 comandos, que podem ser executados no mesmo arquivo de uma só vez (exemplo da extração de dados para o PDT – Partido Democrático Trabalhista pode ser encontrado no apêndice). Basta lembrar que ao clicar no ícone, o navegador abrirá 862 páginas e fará o *download* dos 862 arquivos. Em resumo, cada partido terá 27 comandos (consequentemente 27 bases de dados, com exceção do PCB e PCO para os quais não há informação referentes a RR e TO respectivamente).

4.2 Manipulação dos dados

As bases de dados dos candidatos e eleitores filiados aos partidos políticos, extraídas conforme descrição acima, foram exportadas para o software estatístico SAS e toda a descrição referente à manipulação dos dados que será detalhada a seguir ocorreu na mesma plataforma.

4.2.1 Ajustes nas bases iniciais

Para facilitar as análises, todas as 862 bases dos filiados foram empilhadas em uma base única denominada “filiados”. As variáveis DATA_DA_EXTRACAO e HORA_DA_EXTRACAO foram retiradas da base, pois não serão utilizadas nas análises. Portanto, a base “filiados” contém 19.198.083 linhas e 17 colunas, conforme descrito na tabela 3.

As bases que contém os candidatos nas eleições também foram empilhadas em uma base única e mantivemos apenas as variáveis que seriam utilizadas nas análises para que a base ficasse mais leve e o processamento mais rápido. Com isso, a base de candidatos contém 1.758.550 linhas e 20 colunas, sendo elas:

Tabela 4 - Campos mantidos na base de candidatos

Variável	Descrição
ano_eleicao	Ano da eleição
num_turno	Número do turno
sigla_uf	Sigla da Unidade da Federação em que ocorreu a eleição
sigla_ue	Sigla da Unidade Eleitoral (Em caso de eleição majoritária é a sigla da UF que o candidato concorre (texto) e em caso de eleição municipal é o código TSE do município (número)). Assume os valores especiais BR, ZZ e VT para designar, respectivamente, o Brasil, Exterior e Voto em Trânsito.
cargo	Descrição do cargo a que o candidato concorre
nome_candidato	Nome completo do candidato
sigla_partido	Sigla do partido
nome_partido	Nome do partido
codigo_legenda	Código sequencial da legenda gerado pela Justiça Eleitoral
sigla_legenda	Sigla da legenda
composicao_legenda	Composição da legenda
nome_legenda	Nome da legenda
ocupacao	Descrição da ocupação do candidato
data_nasc	Data de nascimento do candidato
num_tit_eleit_cand	Número do título eleitoral do candidato
genero	Descrição do sexo do candidato
grau_instrucao	Descrição do grau de instrução do candidato
estado_civil	Descrição do estado civil do candidato
cod_sit_tot_turno	Código da situação de totalização do candidato naquele turno
desc_sit_tot_turno	Descrição da situação de totalização do candidato naquele turno

4.2.2 Obter a informação de gênero para os filiados

Como o intuito é verificar o efeito que as mulheres eleitas têm sob a filiação de outras mulheres aos partidos políticos, precisamos da informação de gênero dos filiados na base de dados. Porém, esta informação está presente apenas na base de candidatos. Com isso, os passos descritos a seguir serão destinados ao processo de obter a informação “feminino”/”masculino” para cada filiado presente na base de dados.

O primeiro ajuste foi corrigir tanto na base de candidatos como na base de filiados os nomes que estavam escritos de maneira incorreta, como exemplos abaixo:

Tabela 5 - Exemplos de correções nas bases de candidatos e filiados

Nome incorreto	Nome ajustado
"CARLOS FERNANDO D"AGUIAR SILVA PALÁCIO"	CARLOS FERNANDO D"AGUIAR SILVA PALÁCIO
'REGINA ROCHA FIGUEIREDO NOGUEIRA	REGINA ROCHA FIGUEIREDO NOGUEIRA
,ADVALDO DE FREITAS SOUSA	ADVALDO DE FREITAS SOUSA
.JOSE MILTON PEREIRA	JOSE MILTON PEREIRA
ODEMIR OLELIS DA SILVA PORTO	ODEMIR OLELIS DA SILVA PORTO
A NTONIO BASÍLIO DE FREITAS	ANTONIO BASÍLIO DE FREITAS
# ROSIMAR FIRMINIANO CAVALCANTE	ROSIMAR FIRMINIANO CAVALCANTE
"ZELIA PEREIRA BESSA	ZELIA PEREIRA BESSA
-NILDA ANA VICENTE	NILDA ANA VICENTE
. ALEXANDRE GOMES LIMA	ALEXANDRE GOMES LIMA
.RITA PEREIRA DA SILVA.	RITA PEREIRA DA SILVA
: LUIS EDUARDO EDWIRGES PINTO	LUIS EDUARDO EDWIRGES PINTO
:IVONE SILVA AZEVEDO	IVONE SILVA AZEVEDO
A CLAUDIA LIMA DE OLIVEIRA	CLAUDIA LIMA DE OLIVEIRA
A'VILA DE SANTANA BRITO DOS SANTOS	ÁVILA DE SANTANA BRITO DOS SANTOS
AAB JOSEPH PABLO VIANA VELOSO	JOSEPH PABLO VIANA VELOSO
AALBERTO VALDENIR DE OLIVEIRA GUEDES	ALBERTO VALDENIR DE OLIVEIRA GUEDES

Ao todo foram 206 correções, sendo 18 na base de candidatos e 188 na base dos filiados. Da mesma forma que foi feito para os nomes, ajustou-se também os casos em que as siglas dos partidos políticos estavam incorretas nas bases de candidatos:

Tabela 6 - Ajustes nas siglas dos partidos políticos

Sigla do partido incorreta	Sigla do partido ajustada
PC do	PC do B
PT do	PT do B

Os passos seguintes foram:

1. Criação de uma base de candidatos apenas com as seguintes variáveis: nome_candidato, num_tit_eleit_cand e gênero. Estas são as variáveis necessárias que precisamos manter na base para futuramente cruzar com a base de filiados e obter a informação do gênero.

Nome da base: candidatos_genero

Quantidade de linhas: 1.758.550 Quantidade de colunas: 3

2. Tirar duplicidade para garantir que haverá apenas um candidato por título de eleitor e gênero. Retirar a duplicidade é essencial, pois no passo anterior poderia acontecer, por exemplo, do mesmo indivíduo se candidatar para diferentes cargos, em diferentes anos e partidos (para governador em 2002 no PT e depois se candidatar para senador em 2010 no PMDB) e então ele apareceria na base duplicado. Como estamos interessados apenas em buscar o nome e gênero independente do período, e não capturar a diferença de candidatura no decorrer dos anos, retiramos a duplicidade.

Nome da base: candidatos_genero

Quantidade de linhas: 1.313.371 Quantidade de colunas: 3

3. Repetimos o passo 1 para a base de filiados mantendo apenas: nome_do_filiado e numero_da_inscricao.

Nome da base: filiados

Quantidade de linhas: 19.198.083 Quantidade de colunas: 2

4. Repetimos o passo 2 para a base de filiados com o objetivo de retirar a duplicidade por nome do filiador e título de eleitor. Retirar a duplicidade na base dos filiados também foi essencial pois no passo 3 encontramos casos em que houve cancelamento ou desfiliação em um ano e algum tempo depois o indivíduo decidiu filiar-se novamente ao mesmo partido político, ou situações em que o mesmo indivíduo, em anos diferentes, filiou-se a partidos políticos diferentes, ambos casos gerando duplicidade na base, uma vez que só mantivemos o nome e título de eleitor.

Nome da base: filiados

Quantidade de linhas: 16.930.886 Quantidade de colunas: 2

5. Cruzar as bases obtidas nos passos 2 e 4, para obter a informação de gênero dos filiados. A ideia é fixar a base de filiados e trazer a informação de gênero contida na base de candidatos. O cruzamento é feito por nome completo e número do título de eleitor.

Nome da base: filiados_com_genero

Quantidade de linhas: 16.931.651 Quantidade de colunas: 3

A base gerada aumentou 765 linhas, isso ocorreu pois para alguns candidatos encontramos a informação de gênero em um ano e em outro ano, o mesmo candidato, não tem a informação de gênero preenchida, gerando duas linhas para o mesmo candidato neste caso, uma com informação de gênero preenchida e outra linha com o campo de gênero vazio.

Através dos itens anteriores encontramos o seguinte preenchimento para a informação de gênero na base filiados_com_genero:

Tabela 7 - Percentual de preenchimento da informação de gênero na base de filiados

Gênero	% Preenchimento
Feminino	1.3%
Masculino	4.0%
Missing	94.7%

É importante lembrar que estamos buscando a informação de gênero para uma base de filiados que contém aproximadamente 17 milhões de observações em uma base de candidatos que contém aproximadamente 1,3 milhões. Além disso, cruzamos as bases para trazermos o gênero se encontramos o nome completo do filiado na base de candidatos. Portanto, este baixo preenchimento da informação de gênero ocorre pois nem todos os filiados aos partidos políticos são candidatos nas eleições.

Para aumentarmos o preenchimento do campo “gênero” na base dos filiados, decidiu-se cruzar as bases apenas pelo primeiro nome e não pelo nome completo, como feito anteriormente. Para exemplificar este benefício no preenchimento da informação, temos 11.157 filiados que têm “Augusto” como primeiro nome e, pela forma anterior de trazer o gênero para a base dos filiados – cruzando pelo nome completo – encontramos o gênero para apenas 773 filiados (que possuem “Augusto” como

primeiro nome e estão na base de candidatos). Ou seja, encontramos a informação de gênero para apenas 7% dos filiados que têm o primeiro nome “Augusto”, pois estes também são candidatos. Isso implica que os demais 10.384 (93%) filiados que têm como primeiro nome “Augusto” e que não foram encontrados na base de candidatos estariam sem a informação de gênero preenchida.

Portanto, os passos a seguir referem-se a busca da informação de gênero na base dos candidatos utilizando apenas o primeiro nome como chave de cruzamento, para que por exemplo, todos os filiados que têm “Augusto” como primeiro nome tenham a informação de gênero preenchida.

6. Criou-se um campo denominado “first_name” na base filiados_com_genero (desenvolvida no passo 5) que contém o primeiro nome de cada filiado.

Nome da base: filiados_com_genero

Quantidade de linhas: 16.931.651 Quantidade de colunas: 4

7. Gerou-se uma base apartada com o objetivo de contar a quantidade de cada gênero encontrado (inclusive “missing”) para cada primeiro nome, por exemplo:

Tabela 8 - Quantidade de gênero por primeiros nomes

first_name	genero	Volume
ADALCI		0
ADALCI	FEMININO	3
ADALCI	MASCULINO	8
NUBIA		0
NUBIA	FEMININO	124
NUBIA	MASCULINO	2

Nome da base: volumetria

Quantidade de linhas: 439.038 Quantidade de colunas: 3

8. Retirou-se os casos em que não há preenchimento para o gênero, conforme abaixo:

Tabela 9 - Quantidade de gênero, exceto missing, por primeiros nomes

first_name	genero	Volume
ADALCI	FEMININO	3
ADALCI	MASCULINO	8
NUBIA	FEMININO	124
NUBIA	MASCULINO	2

Nome da base: volumetria

Quantidade de linhas: 62.867

Quantidade de colunas: 3

9. Ordenou-se o campo first_name por ordem alfabética e o volume de forma decrescente:

Tabela 10 - Quantidade de gênero ordenada por nome e maior volume

first_name	genero	Volume
ADALCI	MASCULINO	8
ADALCI	FEMININO	3
NUBIA	FEMININO	124
NUBIA	MASCULINO	2

Nome da base: volumetria

Quantidade de linhas: 62.867

Quantidade de colunas: 3

10. Retirou-se duplicidade por first_name, desta forma mantemos apenas a primeira linha de cada nome, garantindo que o gênero selecionado para cada nome é o que possui maior volume.

Tabela 11 - Gênero selecionado por nome de acordo com o maior volume encontrado

first_name	genero	Volume
ADALCI	MASCULINO	8
NUBIA	FEMININO	124

Nome da base: gêneros

Quantidade de linhas: 59.743

Quantidade de colunas: 3

11. Neste passo, cruzamos a base `filiados_com_genero` descrita no passo 6 com a base “gêneros” descrita no item anterior. A chave de cruzamento foi apenas o primeiro nome (`first_name`) e não mais o nome completo, como feito anteriormente, com o intuito de aumentar o *match* de preenchimento do gênero para os filiados. A nova informação de gênero foi substituída pela informação trazida anteriormente e obtivemos o seguinte resultado:

Tabela 12 - Novo percentual de preenchimento da informação de gênero na base de filiados

Gênero	% Preenchimento
Feminino	40.8%
Masculino	55.0%
Missing	4.2%

Nome da base: `filiados_com_genero2`

Quantidade de linhas: 16.931.651 Quantidade de colunas: 4

12. O último passo foi levar a informação de gênero obtida no item anterior para a base original de filiados.

Nome da base: `base_filiados_completa`

Quantidade de linhas: 19.198.083 Quantidade de colunas: 18

Tabela 13 - Descrição das variáveis presentes na base dos eleitores filiados aos partidos políticos após o processo de manipulação dos dados para trazer a informação de gênero da base de candidatos

Variável	Descrição
NUMERO_DA_INSCRICAO	Número da inscrição eleitoral
NOME_DO_FILIADO	Nome do filiado
SIGLA_DO_PARTIDO	Sigla do partido político
NOME_DO_PARTIDO	Nome do partido político
UF	Unidade da federação
CODIGO_DO_MUNICIPIO	Código do município
NOME_DO_MUNICIPIO	Nome do município
ZONA_ELEITORAL	Zona eleitoral
SECAO_ELEITORAL	Seção eleitoral
DATA_DA_FILIACAO	Data da filiação
SITUACAO_DO_REGISTRO	Situação do registro de filiação
TIPO_DO_REGISTRO	Tipo do registro de filiação
DATA_DO_PROCESSAMENTO	Data do processamento do registro de filiação
DATA_DA_DESFILIACAO	Data da desfiliação
DATA_DO_CANCELAMENTO	Data do cancelamento do registro de filiação
DATA_DA_REGULARIZACAO	Data da regularização do registro de filiação
MOTIVO_DO_CANCELAMENTO	Motivo do cancelamento do registro de filiação
GENERO	Informação de gênero trazida da base de candidatos

Para garantir que os procedimentos descritos acima (atribuir o gênero para os filiados) foram assertivos, fizemos o seguinte estudo:

1. Através da quantidade de feminino/masculino encontrada para cada primeiro nome na base 'volumetria', descrita no passo 9, criou-se a probabilidade de ser feminino/masculino por primeiro nome, da seguinte forma:

Tabela 14 - Probabilidade de ser feminino/masculino por primeiro nome

first_name	genero	volume	probabilidade
CLAUDECIR	MASCULINO	95	0.96
CLAUDECIR	FEMININO	4	0.04
ELISMAR	FEMININO	5	0.06
ELISMAR	MASCULINO	74	0.94
JOSEANE	MASCULINO	3	0.03
JOSEANE	FEMININO	88	0.97
MARILIA	FEMININO	161	1.00

2. Ordenou-se as probabilidades de ser feminino/masculino de forma decrescente:

Tabela 15 - Probabilidade ordenada de ser feminino/masculino por primeiro nome

first_name	genero	volume	probabilidade
CLAUDECIR	MASCULINO	95	0.96
CLAUDECIR	FEMININO	4	0.04
ELISMAR	MASCULINO	74	0.94
ELISMAR	FEMININO	5	0.06
JOSEANE	FEMININO	88	0.97
JOSEANE	MASCULINO	3	0.03
MARILIA	FEMININO	161	1.00

3. Calculou-se a diferença de probabilidade (nos casos em que há mais de um gênero por nome), subtraindo a primeira linha (maior probabilidade) da segunda linha (menor probabilidade) de cada nome. Para os casos em que existe apenas um gênero por nome, a diferença será 1.00, conforme exemplo abaixo:

Tabela 16 - Cálculo da diferença de probabilidade por nome

first_name	Diferença de probabilidades
CLAUDECIR	0.92
ELISMAR	0.87
JOSEANE	0.93
MARILIA	1.00

4. Calculou-se a diferença média de probabilidades entre todos os primeiros nomes dos filiados na base. Se a diferença média for próxima de 0, significa que existem muitos casos em que o mesmo nome é classificado 50% como homem e 50% como mulher. Mas, se a diferença média for próxima de 1, significa que não houve dúvida para atribuir o gênero do filiado, pois a maioria foi classificada apenas com um gênero.

No exemplo acima encontramos que a diferença média é 0.93. Na base completa encontramos que a diferença média é 0.9760, o que é bastante assertivo, pois podemos garantir que houve muita diferença entre a quantidade de gêneros atribuídos para o mesmo nome (quando aplicável, pois existem muitos casos em que encontramos apenas um gênero por nome), não deixando dúvida de qual seria o gênero mais apropriado para se atribuir ao primeiro nome.

Portanto, a maneira que atribuímos o gênero na base dos filiados aos partidos políticos fez-se satisfatória, podendo então seguir com a manipulação dos dados para as futuras análises.

4.2.3 Variável resposta considerando informações de filiados; running variable e variável de tratamento

Os ajustes a seguir destinam-se a adaptação do *layout* e manipulação dos dados na base de filiados (com informação de gênero) para construção de uma das variáveis resposta, captura da *running variable* e da variável de tratamento que utilizaremos nos futuros modelos.

O primeiro passo foi manter na base apenas os filiados que não tinham data de cancelamento do registro de filiação preenchida ou, se por acaso tivessem data de cancelamento preenchida, que fosse no ano seguinte ao ano de filiação, para garantir que no ano em questão o indivíduo realmente estava filiado ao partido. Após este ajuste, contou-se a quantidade de filiados masculino e feminino em cada partido político por ano e município, conforme exemplo abaixo:

Tabela 17 - Quantidade de filiados masculino e feminino em cada partido por ano e município

ano_filiacao	nome_do_municipio	uf	qtd_masc_PMDB	qtd_masc_PDT	...	qtd_masc_PT	qtd_fem_PMDB	qtd_fem_PDT	...	qtd_fem_PT
2000	SÃO PAULO	SP	65	104	...	1444	99	89	...	1389
2000	BOA VISTA	RR	3	1	...	35	.	1	...	41
2014	CURITIBA	PR	272	43	...	11	268	14	...	5
2014	GOIÂNIA	GO	62	5	...	45	54	2	...	39

A base contém 68 colunas, sendo 4 referentes às informações chave da base (ano de filiação, nome do município, código do município e unidade federativa) e as demais 64 colunas referem-se à quantidade de filiados do gênero masculino em cada um dos 32 partidos políticos e a quantidade de filiadas do gênero feminino em cada um dos mesmos 32 partidos políticos.

O próximo passo foi selecionar apenas os indivíduos que se filiaram aos partidos políticos entre os anos 2001 a 2015 e separar em quatro bases, sendo elas:

- ✓ Eleição 2000: Indivíduos que se filiaram em 2001, 2002, 2003 e 2004;
- ✓ Eleição 2004: Indivíduos que se filiaram em 2005, 2006, 2007 e 2008;
- ✓ Eleição 2008: Indivíduos que se filiaram em 2009, 2010, 2011 e 2012;
- ✓ Eleição 2012: Indivíduos que se filiaram em 2013, 2014 e 2015.

Nestas quatro bases acima, criamos também a quantidade de homens (qtd_homens) e mulheres (qtd_mulheres) por ano e município (somando todos os partidos políticos), conforme tabela abaixo para a base Eleição 2008:

Tabela 18 - Quantidade de filiados masculino e feminino em cada partido por ano e município e quantidade de homens e mulheres por ano e município, independente do partido

ano_filiacao	nome_do_municipio	uf	qtd_masc_PMDB	...	qtd_masc_PT	qtd_fem_PMDB	...	qtd_fem_PT	qtd_homens	qtd_mulheres
2009	CRUZEIRO DO SUL	AC	4	...	6	6	...	3	95	32
2010	ALFENAS	MG	3	...	29	4	...	45	117	112
2011	NOVO AIRÃO	AM	18	...	58	11	...	36	327	241
2012	OSASCO	SP	4	...	35	1	...	40	860	1014

A partir das quatro bases citadas acima, calculou-se as mesmas quantidades de filiados do gênero masculino e feminino por partido e município (mas agora independente do ano, ou seja, somando as informações dos quatro anos presentes em cada base) e calculou-se também a quantidade de homens e mulheres por município, independente do partido político e também do ano de filiação. Ou seja, para a base Eleição 2000, calculou-se a quantidade de filiados do gênero masculino e feminino em cada partido político para cada município, independente se foram filiados em 2001, 2002, 2003 ou 2004 e calculou-se a quantidade de homens (qtd_homens) e mulheres (qtd_mulheres) por município, somando todos os partidos políticos e os quatro anos: 2001, 2002, 2003 e 2004. Além disso, criou-se um campo que será utilizado como variável dependente (ou variável resposta) nos modelos futuros, que indica a proporção entre mulheres/homens (prop_fem_masc) por município, através dos campos qtd_mulheres e qtd_homens, levando em consideração todos os partidos políticos e os quatro anos de filiação presentes na base. Na tabela abaixo é possível identificar um exemplo dos campos criados:

Tabela 19 – Exemplo da base criada contendo a variável resposta: proporção de mulheres sobre homens por município nos anos 2001, 2002, 2003 e 2004

nome_do_municipio	uf	qtd_masc_PMDB	...	qtd_masc_PT	qtd_fem_PMDB	...	qtd_fem_PT	qtd_homens	qtd_mulheres	prop_fem_masc
TERESINA	PI	133	...	1630	113	...	1813	4990	5086	1.02
OLINDA	PE	103	...	994	103	...	989	3654	3684	1.01
CAMAÇARI	BA	269	...	825	244	...	767	2938	2598	0.88
ANGRA DOS REIS	RJ	412	...	116	366	...	62	2366	1841	0.78

As quatro bases: Eleição 2000, Eleição 2004, Eleição 2008 e Eleição 2012 foram empilhadas em uma base única. Foi criado um campo de identificação para cada base denominado “eleição”, em que eleição = 2000 representa os indivíduos que se filiaram entre 2001-2004, eleição = 2004 representa os indivíduos que se filiaram entre 2005-2008, eleição = 2008 representa os indivíduos que se filiaram entre 2009-2012 e eleição = 2012 representa os indivíduos que se filiaram entre 2013-2015. Isso foi feito apenas para que fosse possível realizar análises tanto em conjunto, quanto segregar as quatro bases novamente, dependendo da análise.

Utilizando a base empilhada, é possível verificar as estatísticas descritivas para a variável: proporção entre mulheres/homens por município, que utilizaremos como variável resposta nos modelos futuros.

Tabela 20 - Estatísticas descritivas da variável proporção de mulheres/homens que se filiaram entre 2001-2015

Estatística	Variável: proporção de mulheres/homens filiados por município considerando todos os partidos políticos
Número total de observações	21602
Número de obs. não- <i>missing</i>	21039
Número de obs. <i>missing</i>	563
Média	0.59
Desvio-padrão	0.30
Mínimo	0.04
1° percentil	0.18
5° percentil	0.29
10° percentil	0.34
25° percentil	0.44
Mediana	0.55
75° percentil	0.69
90° percentil	0.85
95° percentil	1.00
99° percentil	1.53
Máximo	8.00

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

O último passo foi trazer para a base a *running variable*, a variável de tratamento e um campo que nos permitirá filtrar a base com o intuito de selecionar o público alvo para as regressões. Tais variáveis foram extraídas a partir de uma base de dados utilizada para o trabalho de conclusão do curso de Doutorado em Economia: “Ensaio sobre economia política aplicada” desenvolvida por Pieri (2015), apresentado à Escola de Economia de São Paulo da Fundação Getulio Vargas. As informações trazidas foram:

- ✓ *‘margin’ (running variable)*: margem de vitória, definida como a diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito considerando o primeiro e segundo lugares nas eleições de primeiro turno para eleições municipais.

Tabela 21 - Estatísticas descritivas da variável: margem de vitória (Eleições de 2000, 2004, 2008 e 2012)

Estatística	Variável: margem de vitória
Número total de observações	21602
Número de obs. não-missing	17742
Número de obs. missing	3860
Média	-0.13
Desvio-padrão	0.17
Mínimo	-1.00
1° percentil	-0.72
5° percentil	-0.43
10° percentil	-0.32
25° percentil	-0.19
Mediana	-0.10
75° percentil	-0.04
90° percentil	0.00
95° percentil	0.07
99° percentil	0.28
Máximo	0.97

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

- ✓ 'mulher_venceu' (variável de tratamento): recebe o valor 0 quando o prefeito eleito na cidade é do gênero masculino e 1 quando a prefeita eleita na cidade é do gênero feminino.

Tabela 22 – Frequência da variável: gênero do prefeito eleito considerando eleições de 2000, 2004, 2008 e 2012

Variável: mulher_venceu. Indica o gênero do prefeito eleito.	Frequência	%
Prefeito eleito na cidade é do gênero masculino	16847	91.67
Prefeita eleita na cidade é do gênero feminino	1531	8.33
<hr/>		
Número total de observações	21602	
Número de obs. não-missing	18378	
Número de obs. Missing	3224	

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

- ✓ 'D_BASE_HM' (variável binária) que ao receber o valor 1 nos restringe às eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher.

Tabela 23 – Frequência da variável: Indicador das eleições em que os dois candidatos mais votados no município são necessariamente um homem e uma mulher (Período: eleições de 2000, 2004, 2008 e 2012)

Variável: D_BASE_HM Indica as eleições em que os dois candidatos mais votados no município são necessariamente um homem e uma mulher.	Frequência	%
Os dois candidatos mais votados NÃO são necessariamente um homem e uma mulher	14665	79.8
Os dois candidatos mais votados são necessariamente um homem e uma mulher	3713	20.2
Número total de observações	21602	
Número de obs. não-missing	18378	
Número de obs. missing	3224	

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

Considerando os ajustes e manipulação de dados descritos neste tópico, o layout da base é dado por:

Tabela 24 – Layout da base contendo a variável resposta considerando informações dos filiados, running variable e variável de tratamento

Campos da base	Descrição
Eleição	<p>Campo de identificação para separar a base de acordo com o ano de filiação.</p> <p>Recebe os valores:</p> <p>2000 = indivíduos que se filiaram entre 2001-2004</p> <p>2004 = indivíduos que se filiaram entre 2005-2008</p> <p>2008 = indivíduos que se filiaram entre 2009-2012</p> <p>2012 = indivíduos que se filiaram entre 2013-2015</p>
codigo_do_municipio	Código do município
nome_do_municipio	Nome do município
UF	Unidade da federação
qtd_masc_DEM	Qtd de filiados ao partido DEM do gênero masculino no período definido pelo campo eleição.
qtd_masc_PC_DO_B	Qtd de filiados ao partido PC do B do gênero masculino no período definido pelo campo eleição.
qtd_masc_PCB	Qtd de filiados ao partido PCB do gênero masculino no período definido pelo campo eleição.
qtd_masc_PCO	Qtd de filiados ao partido PCO do gênero masculino no período definido pelo campo eleição.
qtd_masc_PDT	Qtd de filiados ao partido PDT do gênero masculino no período definido pelo campo eleição.
qtd_masc_PEN	Qtd de filiados ao partido PEN do gênero masculino no período definido pelo campo eleição.
qtd_masc_PHS	Qtd de filiados ao partido PHS do gênero masculino no período definido pelo campo eleição.
qtd_masc_PMDB	Qtd de filiados ao partido PMDB do gênero masculino no período definido pelo campo eleição.
qtd_masc_PMN	Qtd de filiados ao partido PMN do gênero masculino no período definido pelo campo eleição.
qtd_masc_PP	Qtd de filiados ao partido PP do gênero masculino no período definido pelo campo eleição.
qtd_masc_PPL	Qtd de filiados ao partido PPL do gênero masculino no período definido pelo campo eleição.
qtd_masc_PPS	Qtd de filiados ao partido PPS do gênero masculino no período definido pelo campo eleição.
qtd_masc_PR	Qtd de filiados ao partido PR do gênero masculino no período definido pelo campo eleição.
qtd_masc_PRB	Qtd de filiados ao partido PRB do gênero masculino no período definido pelo campo eleição.
qtd_masc_PROS	Qtd de filiados ao partido PROS do gênero masculino no período definido pelo campo eleição.
qtd_masc_PRP	Qtd de filiados ao partido PRP do gênero masculino no período definido pelo campo eleição.

qtd_fem_PTB	Qtd de filiados ao partido PTB do gênero feminino no período definido pelo campo eleição.
qtd_fem_PTC	Qtd de filiados ao partido PTC do gênero feminino no período definido pelo campo eleição.
qtd_fem_PTN	Qtd de filiados ao partido PTN do gênero feminino no período definido pelo campo eleição.
qtd_fem_PV	Qtd de filiados ao partido PV do gênero feminino no período definido pelo campo eleição.
qtd_fem_SD	Qtd de filiados ao partido SD do gênero feminino no período definido pelo campo eleição.
qtd_homens	Qtd de filiados do gênero masculino no período definido pelo campo eleição.
qtd_mulheres	Qtd de filiados do gênero feminino no período definido pelo campo eleição.
prop_fem_masc	Proporção entre mulheres/homens filiadas no período definido pelo campo eleição.
mulher_venceu	Indicador binário que recebe o valor 1 quando a prefeita eleita na cidade é do gênero feminino, 0 caso contrário.
Margin	Margem de vitória (diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito considerando o 1° e 2° lugares nas eleições de 1° turno para eleições municipais).
D_BASE_HM	Indicador binário que ao receber o valor 1 nos restringe às eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher.

4.2.4 Variável resposta considerando informações de candidatos

Os ajustes a seguir destinam-se a adaptação do *layout* na base de candidatos a fim de construir duas variáveis resposta que serão levadas para a base anterior que contém a variável resposta considerando informações dos filiados, a *running variable* e a variável de tratamento, descritas no item anterior.

O primeiro passo foi manter na base de candidatos descrita na tabela 4 apenas os candidatos a prefeito nos anos 2000, 2004, 2008 e 2012 que foram eleitos (período utilizado no item anterior quando analisamos informações dos filiados). Este passo é importante pois considerar apenas os candidatos eleitos nos permitirá medir o efeito da eleição nas informações dos filiados nos anos seguintes.

As informações contidas na base de candidatos que estamos interessados em adicionar nas análises são: sigla do partido eleito e composição da legenda, por isso vamos manter na base apenas os campos *ano_eleição*, *sigla_ue*, *sigla_partido* e *composição_legenda*. O intuito é identificar quais são os partidos dos prefeitos eleitos e os partidos presentes na composição da legenda do partido vinculado ao prefeito(a) eleito(a) e com isso construir variáveis ligadas à esses partidos políticos.

Da mesma forma que foi feito na base de filiados, ajustou-se também na base de candidatos os casos em que as siglas dos partidos políticos estavam incorretas, conforme exemplo abaixo:

Tabela 25 - Ajustes nas siglas dos partidos políticos

Sigla do partido incorreta	Sigla do partido ajustada
"PC do" e "PC d"	PC do B
"PT do" e "PT d"	PT do B

Uma vez identificado qual é o partido político do prefeito eleito, criou-se uma variável *dummie* para cada partido político (*partido_prefeito*) que recebe o valor 1 quando o prefeito eleito for do partido em questão e 0 quando o prefeito eleito for de outro partido. Da mesma forma foi feito para os partidos contidos na composição da legenda do partido eleito, ou seja, uma vez identificados quais são os partidos que fazem parte da composição da legenda do partido eleito, criou-se uma variável *dummie* para cada partido político (*partido_legenda*) que recebe o valor 1 quando o partido pertencer a composição da legenda do partido eleito e 0 quando o partido em questão não pertencer a composição da legenda do partido que foi eleito. Tal construção pode ser exemplificada conforme tabela a seguir:

Tabela 26 – Base de candidatos contendo as informações do partido e composição da legenda do prefeito eleito, ambas em formato de dummies

ano_ eleicao	cod_ municipio	PMDB_ prefeito	PDT_ prefeito	PC_do_B_ prefeito	...	PSDB_ prefeito	PSB_ legenda	PPS_ legenda	PT_ legenda	...	PV_ legenda
2000	1120	0	0	0	...	1	1	1	1	...	1
2004	39870	1	0	0	...	0	0	0	1	...	0
2012	56308	0	1	0	...	0	0	1	0	...	1

De acordo com o exemplo acima, podemos concluir que no ano 2004, na cidade que recebe o *cod_municipio* = 39870 (neste caso Santa Luiza - BA), o prefeito eleito era do partido PMDB e a composição da legenda do PMDB era PT.

As informações do partido do prefeito eleito e composição da legenda, exemplificadas pela Tabela 26, foram levadas para a base de filiados, que já contém as variáveis explicativas, representada pela Tabela 24. Com isso, a base final que utilizaremos para gerarmos os modelos contém os 74 campos presentes na Tabela 24 mais os campos descritos na tabela abaixo:

Tabela 27 - Campos provenientes da base de candidatos adicionados a base final de filiados

[illegible]

PSB_legenda	Campo binário: recebe o valor 1 se o partido PSB fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PSC_legenda	Campo binário: recebe o valor 1 se o partido PSC fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PSD_legenda	Campo binário: recebe o valor 1 se o partido PSD fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PSDB_legenda	Campo binário: recebe o valor 1 se o partido PSDB fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PSDC_legenda	Campo binário: recebe o valor 1 se o partido PSDC fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PSL_legenda	Campo binário: recebe o valor 1 se o partido PSL fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PSOL_legenda	Campo binário: recebe o valor 1 se o partido PSOL fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PSTU_legenda	Campo binário: recebe o valor 1 se o partido PSTU fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PT_legenda	Campo binário: recebe o valor 1 se o partido PT fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PT_do_B_legenda	Campo binário: recebe o valor 1 se o partido PT do B fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PTB_legenda	Campo binário: recebe o valor 1 se o partido PTB fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PTC_legenda	Campo binário: recebe o valor 1 se o partido PTC fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PTN_legenda	Campo binário: recebe o valor 1 se o partido PTN fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
PV_legenda	Campo binário: recebe o valor 1 se o partido PV fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.
SD_legenda	Campo binário: recebe o valor 1 se o partido SD fizer parte da composição da legenda do partido eleito na cidade em questão no período definido pelo campo eleição e 0, caso contrário.

Através dos campos que identificam os partidos dos prefeitos eleitos e os partidos presentes na composição da legenda deste mesmo partido apresentados na tabela acima, criou-se as variáveis que darão origem a uma das variáveis resposta dos modelos:

Tabela 28 - Variáveis criadas a partir da identificação do partido eleito e da composição da legenda

Campos da base	Descrição
num_mul_part_pref	Quantidade de mulheres filiadas ao partido do(a) prefeito(a) eleito(a) na cidade em questão e no período definido pelo campo eleição.
num_hom_part_pref	Quantidade de homens filiados ao partido do(a) prefeito(a) eleito(a) na cidade em questão e no período definido pelo campo eleição.
num_mul_part_leg_pref	Quantidade de mulheres filiadas ao(s) partido(s) que compõem a legenda do partido eleito na cidade em questão e no período definido pelo campo eleição.
num_hom_part_leg_pref	Quantidade de homens filiados ao(s) partido(s) que compõem a legenda do partido eleito na cidade em questão e no período definido pelo campo eleição.

O valor das variáveis `num_mul_part_pref` e `num_hom_part_pref` será o valor das variáveis provenientes das bases de filiados (Tabela 24) de acordo com o partido em questão, conforme exemplo abaixo.

Se o(a) prefeito(a) eleito(a) na cidade de Anápolis (GO) em 2004 for do partido PSOL, a variável `num_mul_part_pref` (Quantidade de mulheres filiadas ao partido PSOL em Anápolis entre 2005-2008) = `qtd_fem_PSOL` (Quantidade de filiadas em Anápolis entre 2005-2008 ao partido PSOL). O mesmo acontece para `num_hom_part_pref`, ou seja:

Se o(a) prefeito(a) eleito(a) na cidade de Uberlândia (MG) em 2008 for do partido PDT, a variável `num_hom_part_pref` (Quantidade de homens filiados ao partido PDT em Uberlândia entre 2009-2012) = `qtd_masc_PDT` (Quantidade de filiados em Uberlândia entre 2009-2012 ao partido PDT).

Para as variáveis `num_mul_part_leg_pref` e `num_hom_part_leg_pref` a ideia foi somar o valor das variáveis provenientes das bases de filiados (Tabela 27) de acordo com todos os partidos que fazem parte da composição da legenda, ou seja:

Se o partido do(a) prefeito(a) eleito(a) em 2004 na cidade de Penápolis (SP) for PMDB e a composição da legenda do PMDB for composta pelos partidos PHS, PP, PSC, PSDB e PV, a variável `num_mul_part_leg_pref` (Quantidade de mulheres filiadas ao(s) partido(s) que compõem a legenda do PMDB em Penápolis entre 2005-2008) = `qtd_fem_PMDB` (Quantidade de filiadas em Penápolis entre 2005-2008 ao partido PMDB) + `qtd_fem_PHS` (Quantidade de filiadas em Penápolis entre 2005-2008 ao partido PHS) + `qtd_fem_PP` (Quantidade de filiadas em Penápolis entre 2005-2008 ao partido PP) + `qtd_fem_PSC` (Quantidade de filiadas em Penápolis entre 2005-2008 ao partido PSC) + `qtd_fem_PSDB` (Quantidade de filiadas em Penápolis entre 2005-2008 ao partido PSDB) + `qtd_fem_PV` (Quantidade de filiadas em Penápolis entre 2005-2008 ao partido PV). O mesmo se aplica para a variável `num_hom_part_leg_pref` (Quantidade de homens filiados ao(s) partido(s) que compõem a legenda do partido eleito na cidade em questão e no período definido pelo campo eleição), ou seja:

Se o partido do(a) prefeito(a) eleito(a) em 2000 na cidade de Fortaleza (CE) for PT e a composição da legenda do PT for composta apenas pelo partido PC do B, a variável `num_hom_part_leg_pref` (Quantidade de homens filiados ao(s) partido(s) que compõem a legenda do PT em Fortaleza entre 2001-2004) = `qtd_masc_PT` (Quantidade de filiados em Fortaleza entre 2001-2004 ao partido PT) + `qtd_masc_PC_do_B` (Quantidade de filiados em Fortaleza entre 2001-2004 ao partido PC do B).

Vale ressaltar que o partido eleito também será identificado como partido que compõe a legenda, ou seja, se o partido do prefeito eleito for PV, este também fará parte da composição da legenda. Com isso, somaremos a quantidade de filiados(as) do PV na cidade e anos em questão com as demais quantidades de filiados(as) dos outros partidos na mesma cidade e anos em questão que fazem parte da composição da legenda. Caso não tenha nenhum partido na composição da legenda (além do partido do prefeito eleito), as variáveis `num_mul_part_leg_pref` e `num_hom_part_leg_pref` receberão apenas a quantidade de filiados(as) do partido eleito na cidade nos anos analisados, neste exemplo `qtd_fem_PV` (Quantidade de filiadas ao PV na cidade e período analisado) e `qtd_masc_PV` (Quantidade de filiados ao PV na cidade e período analisado), respectivamente.

Através das variáveis descritas acima, podemos então construir as variáveis resposta:

✓ $r_{part} = \frac{\text{num_mul_part_pref}}{\text{num_hom_part_pref}}$, ou seja, razão entre o número de mulheres filiadas ao partido do prefeito(a) eleito(a) sobre o número de homens filiados ao partido do prefeito(a) eleito(a), ambos na cidade em questão e no período definido pelo campo eleição.

✓ $r_{lege} = \frac{\text{num_mul_part_leg_pref}}{\text{num_hom_part_leg_pref}}$, ou seja, razão entre o número de mulheres filiadas ao(s) partido(s) que compõem a legenda do partido eleito sobre o número de homens filiados ao(s) partido(s) que compõem a legenda do partido eleito, ambos na cidade em questão e no período definido pelo campo eleição.

Além das variáveis dependentes citadas acima, criou-se variáveis auxiliares, tais como:

- ✓ $inter = mulher_venceu \times margin$
- ✓ $inter^2 = mulher_venceu \times margin^2$
- ✓ $margin^2 = margin \times margin$

mulher_venceu (variável de tratamento): Indicador binário que recebe o valor 0 quando o prefeito eleito na cidade é do gênero masculino e 1 quando a prefeita eleita na cidade é do gênero feminino.

margin (running variable) = margem de vitória, definida como a diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito considerando o primeiro e segundo lugares nas eleições de primeiro turno para eleições municipais.

Com isso, a base final que será utilizada para desenvolver os modelos possui 138 campos (junção das tabelas 24 e 27) + 4 campos (tabela 28) + 2 variáveis resposta considerando informações de candidatos (*r_part* e *r_lege*) + 3 variáveis auxiliares descritas anteriormente (*inter*, *inter²* e *margin²*), totalizando 147 variáveis. A volumetria da base é de 21.771 linhas (aproximadamente 5.500 cidades por ano) de acordo com a tabela a seguir:

Tabela 29 - Volumetria da base final de acordo com o ano

Ano	Volumetria
2000 (indivíduos que se filiaram entre 2001-2004)	5554
2004 (indivíduos que se filiaram entre 2005-2008)	5521
2008 (indivíduos que se filiaram entre 2009-2012)	5705
2012 (indivíduos que se filiaram entre 2013-2015)	4991

A volumetria referente ao ano de 2012 é inferior às demais pois considera apenas 3 anos (indivíduos que se filiaram em 2013, 2014 e 2015), enquanto os demais consideram quatro anos.

5 RESULTADOS

Apresentam-se nesta seção os resultados obtidos, e faz-se uso de subseções para marcar as etapas sucessivas do processo de análise. Neste tópico comenta-se sobre análise gráfica das variáveis resposta, as larguras de banda utilizadas para a *running variable margin*, estimação do efeito do tratamento (ou seja, eleição de mulheres) sobre as respostas e análise de covariadas.

5.1 Análises descritivas e análise gráfica das variáveis resposta

A partir da base final com 147 campos detalhados anteriormente, restringimos apenas as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (utilizamos a variável $D_BASE_HM = 1$). Com isso, ficamos com 3.793 observações (17%) do volume da base para gerar as regressões.

Como dito anteriormente, os modelos que serão testados utilizam as seguintes informações:

- ✓ Variáveis explicativas: ***mulher_venceu*** (variável de tratamento): Indicador binário que recebe o valor 1 quando a prefeita eleita na cidade é do gênero feminino, 0 caso contrário, ***margin*** (*running variable*): margem de vitória que é a diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito, ***margin²*** (*margin* * *margin*), ***inter*** (*mulher_venceu* * margem de vitória) e ***inter²*** (*mulher_venceu* * margem de vitória ao quadrado).
- ✓ Variáveis dependentes: ***prop_fem_masc*** (proporção de mulheres/homens filiados aos partidos políticos), ***r_part*** (razão entre o número de mulheres filiadas ao partido do prefeito eleito sobre o número de homens filiados ao partido do prefeito eleito), ***r_lege*** (razão entre o número de mulheres filiadas aos partidos que compõem a legenda do partido eleito sobre o número de homens filiados aos partidos que compõem a legenda do partido eleito).

A seguir encontram-se estatísticas descritivas das variáveis mencionadas acima considerando o público selecionado:

Tabela 30 - Frequência da variável: gênero do prefeito eleito após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Período: 2000, 2004, 2008 e 2012)

Variável: mulher_venceu Indica o gênero do prefeito eleito.	Frequência	%
Prefeito eleito na cidade é do gênero masculino	2477	65.3
Prefeita eleita na cidade é do gênero feminino	1316	34.7
<hr/>		
Número total de observações	3793	
Número de obs. não-missing	3793	
Número de obs. missing	0	

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

Tabela 31 - Estatísticas descritivas das variáveis: margem de vitória após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Eleições de 2000, 2004, 2008 e 2012)

Estatística	Variável: margem de vitória	Variável: (margem de vitória) ²
Número total de observações	3793	3793
Número de obs. não-missing	3168	3168
Número de obs. missing	625	625
Média	-0.0367	0.0483
Desvio-padrão	0.2167	0.1111
Mínimo	-0.9996	0.0000
1° percentil	-0.6995	0.0000
5° percentil	-0.3792	0.0001
10° percentil	-0.2752	0.0004
25° percentil	-0.1430	0.0026
Mediana	-0.0312	0.0118
75° percentil	0.0725	0.0428
90° percentil	0.1928	0.1132
95° percentil	0.2832	0.2118
99° percentil	0.6046	0.6687
Máximo	0.9661	0.9992

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

Além das estatísticas descritivas, a seguir apresentamos o histograma da variável margem de vitória, restringindo as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher e considerando os valores acima do 95º percentil como *outliers*.

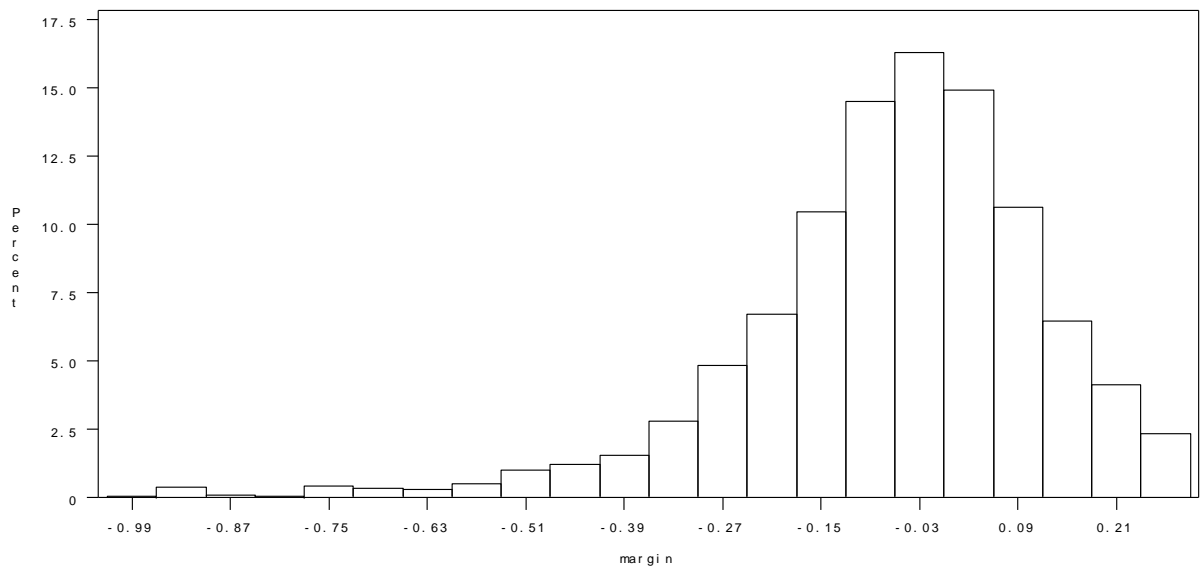


Gráfico 1 – Histograma da running variable margin: diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito, após tratar os valores acima do 95º percentil como outliers.

Fonte: Elaboração própria.

Tabela 32 - Estatísticas descritivas das variáveis: *inter* (*mulher_venceu***margem de vitória*) após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Eleições de 2000, 2004, 2008 e 2012)

Estatística	Variável: <i>inter</i> (<i>mulher_venceu</i> * <i>margem de vitória</i>)	Variável: <i>inter</i> ² (<i>mulher_venceu</i> * <i>margem de vitória</i> ²)
Número total de observações	3793	3793
Número de obs. não- <i>missing</i>	3168	3168
Número de obs. <i>missing</i>	625	625
Média	0.0593	0.0181
Desvio-padrão	0.1208	0.0746
Mínimo	0.0000	0.0000
1º percentil	0.0000	0.0000
5º percentil	0.0000	0.0000
10º percentil	0.0000	0.0000
25º percentil	0.0000	0.0000
Mediana	0.0000	0.0000
75º percentil	0.0725	0.0053
90º percentil	0.1928	0.0372
95º percentil	0.2832	0.0802
99º percentil	0.6046	0.3655
Máximo	0.9661	0.9333

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

Tabela 33 – Estatísticas descritivas das variáveis resposta após restringir as eleições em que os dois candidatos mais votados são necessariamente um homem e uma mulher (Período: 2000, 2004, 2008 e 2012)

Estatística	Variável: prop_fem_masc (proporção de mulheres/homens filiados aos partidos políticos)	Variável: r_part (número de mulheres/homens filiados ao partido do prefeito eleito)	Variável: r_lege (número de mulheres/homens filiados aos partidos que compõem a legenda do partido eleito)
Número total de observações	3793	3793	3793
Número de obs. não-missing	3636	2127	1966
Número de obs. missing	157	1666	1827
Média	0.64	0.63	0.65
Desvio-padrão	0.41	0.58	0.84
Mínimo	0.05	0.00	0.00
1º percentil	0.16	0.00	0.00
5º percentil	0.29	0.00	0.00
10º percentil	0.34	0.09	0.10
25º percentil	0.46	0.33	0.33
Mediana	0.58	0.54	0.55
75º percentil	0.73	0.83	0.80
90º percentil	0.92	1.11	1.01
95º percentil	1.05	1.43	1.35
99º percentil	2.00	2.50	3.00
Máximo	8.00	12.25	21.40

Fonte: Repositório de Dados Eleitorais do Superior Tribunal Eleitoral.

Nota: Dados trabalhados pelo autor.

Para análise gráfica, foram construídos histogramas e gráficos da média das 3 variáveis resposta em cada faixa da margem de vitória. É importante ressaltar que os valores acima do 95º percentil de cada variável foram considerados como *outliers* e então retirados da base.

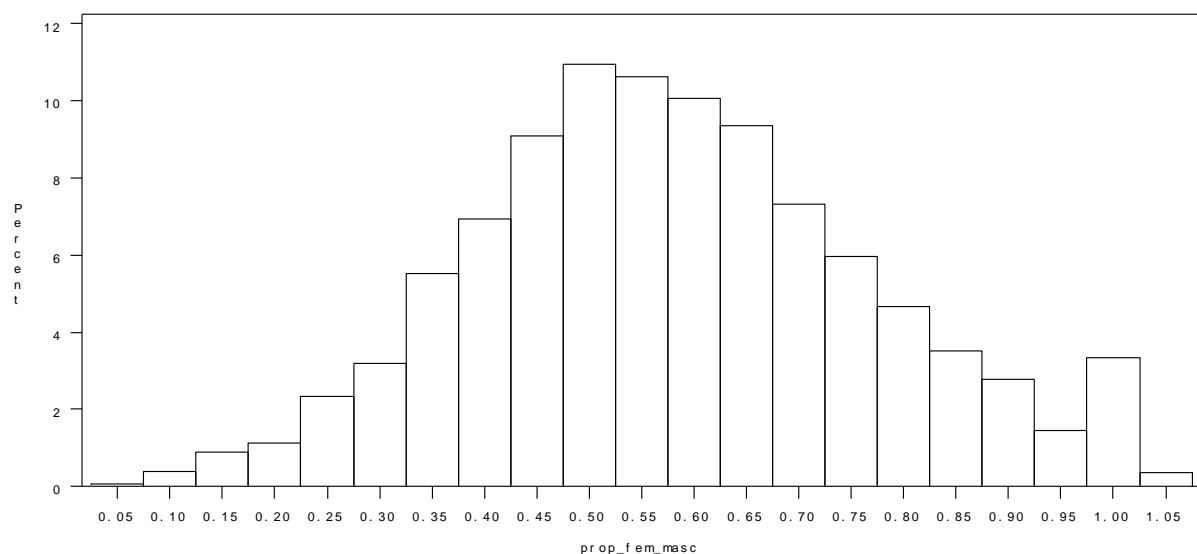


Gráfico 2 – Histograma da variável resposta proporção de mulheres/homens filiados aos partidos políticos.

Fonte: Elaboração própria.

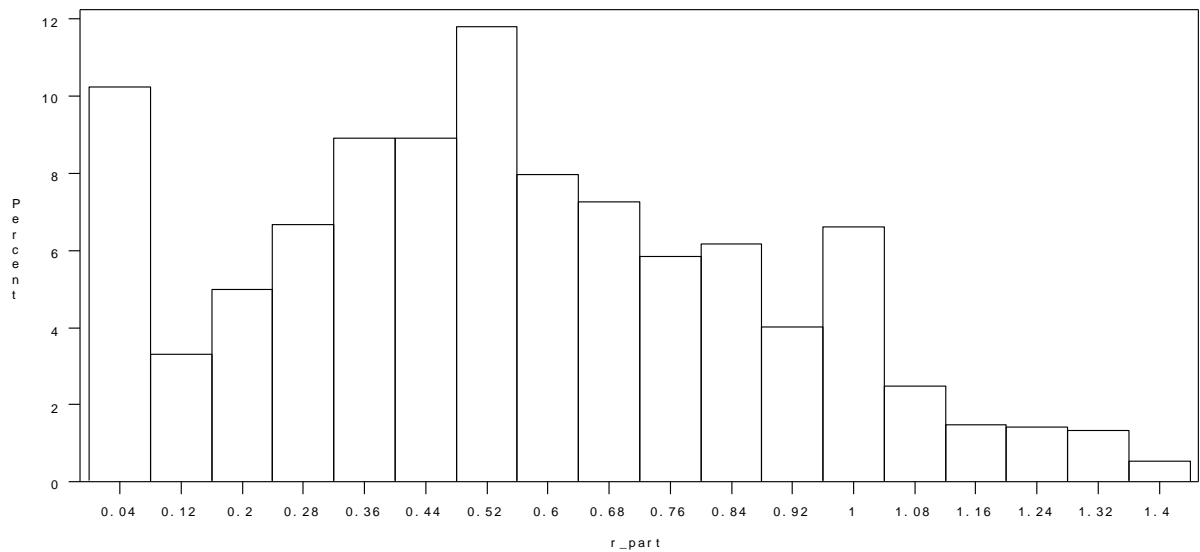


Gráfico 3 – Histograma da variável resposta razão entre mulheres/homens filiados ao partido do prefeito eleito.
Fonte: Elaboração própria.

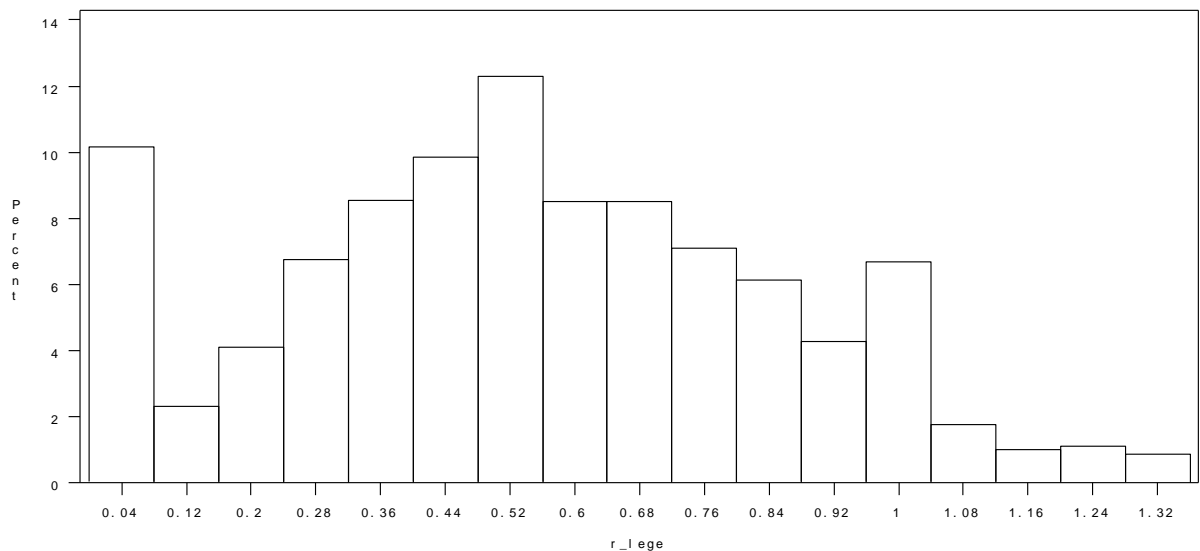


Gráfico 4 – Histograma da variável resposta razão entre mulheres/homens filiados aos partidos que compõem a legenda do partido eleito.
Fonte: Elaboração própria.

Os gráficos 5, 6 e 7 a seguir retratam o comportamento das três variáveis resposta “proporção de mulheres/homens filiados aos partidos políticos”, “razão entre mulheres/homens filiados ao partido do prefeito eleito” e “razão entre mulheres/homens filiados aos partidos que compõem a legenda do partido eleito” entre os limites da margem de vitória.

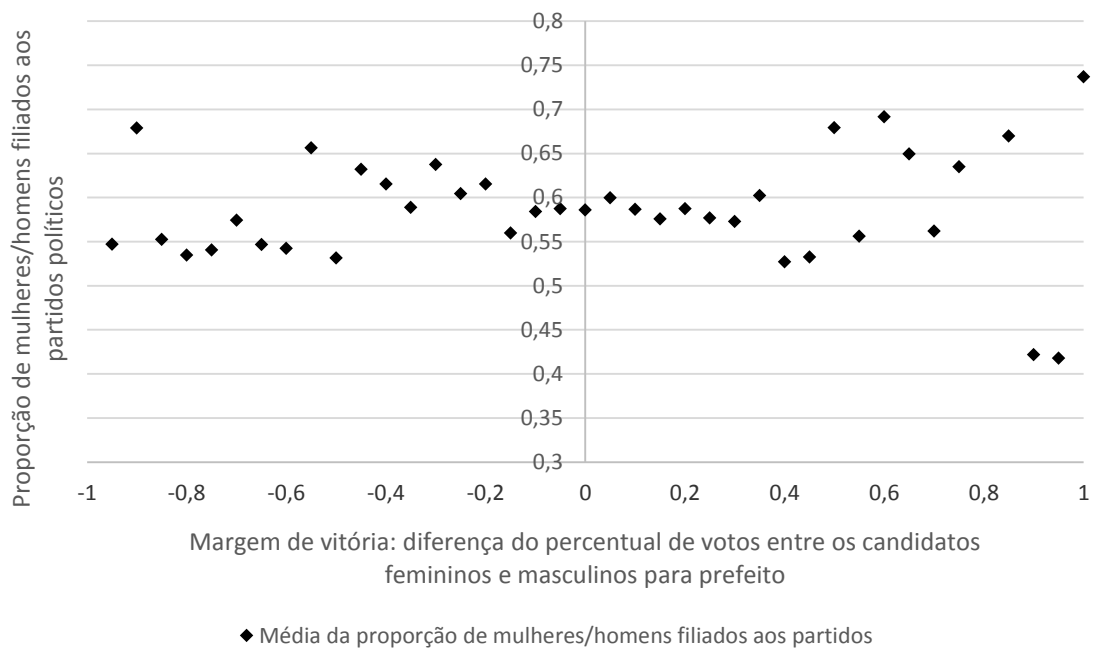


Gráfico 5 - Média da proporção de mulheres/homens filiados aos partidos políticos por faixa da margem de vitória (running variable).

Fonte: Elaboração própria.

Nota: Os pontos representam a média da variável resposta, mensurada no eixo y contra o ponto médio em cada faixa da margem de vitória.

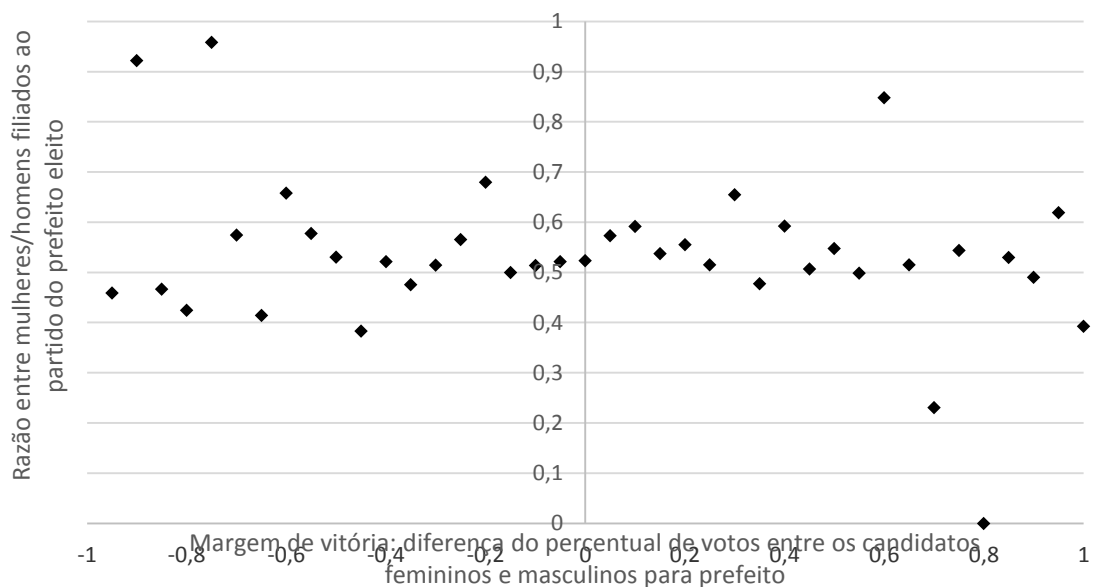


Gráfico 6 - Média da razão entre mulheres/homens filiados ao partido do prefeito eleito por faixa da margem de vitória (running variable).

Fonte: Elaboração própria.

Nota: Os pontos representam a média da variável resposta, mensurada no eixo y contra o ponto médio em cada faixa da margem de vitória.

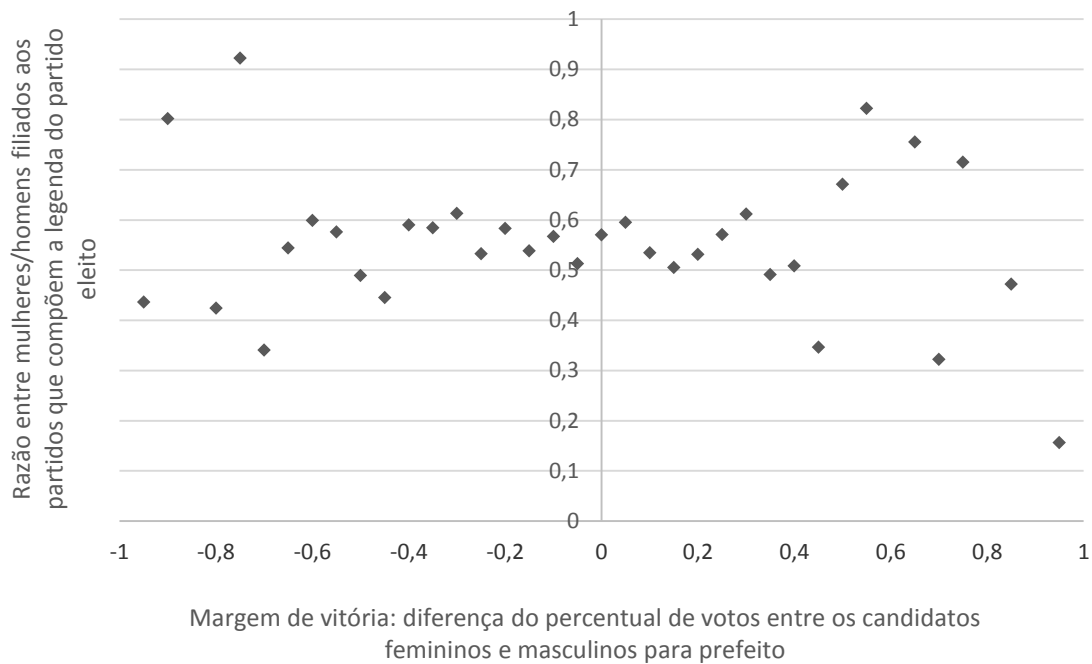


Gráfico 7 - Média da razão entre mulheres/homens filiados aos partidos que compõem a legenda do partido eleito por faixa da margem de vitória (*running variable*).

Fonte: Elaboração própria.

Nota: Os pontos representam a média da variável resposta, mensurada no eixo y contra o ponto médio em cada faixa da margem de vitória.

A observação dos gráficos acima nos permite uma inspeção visual da variabilidade dos dados e da sutil relação das variáveis resposta com a *running variable*. Nota-se pequenos saltos das variáveis resposta nos respectivos pontos de corte, como por exemplo: tem-se a razão entre mulheres/homens filiados ao partido do prefeito eleito ligeiramente maior a direita do corte (ou seja, para municípios em que houve vitória de uma mulher por uma pequena margem de votos) do que à esquerda (para municípios em que houve vitória de um homem). Entretanto, cumpre notar que saltos de dimensões análogas são observados em outros pontos da *running variable* ao analisar as três variáveis resposta, de tal modo que a análise gráfica por si só não pode ser conclusiva para apreciação do efeito do tratamento sobre a resposta.

Diferentes larguras de banda e limites mais estreitados para a margem de vitória foram testados a fim de que a análise gráfica fosse direcionada na região mais próxima do corte da *running variable*, porém os resultados não se mostraram diferentes dos descritos anteriormente.

No que diz respeito à densidade das observações conforme a margem de vitória, pode-se afirmar que não se notam descontinuidades significativas da mesma, o que corrobora a tese de que não há manipulação precisa da running variable por parte dos agentes (candidatos) e, portanto, não invalida a aplicação da metodologia de regressão descontínua para este conjunto de dados. De modo a complementar esta constatação, os gráficos 8, 9 e 10 a seguir trazem apenas a distribuição das observações para as três variáveis resposta ao longo de todo o espectro da margem de vitória (ou seja, de -100% a +100%), com uma largura de banda de 0,05 ou 5%.

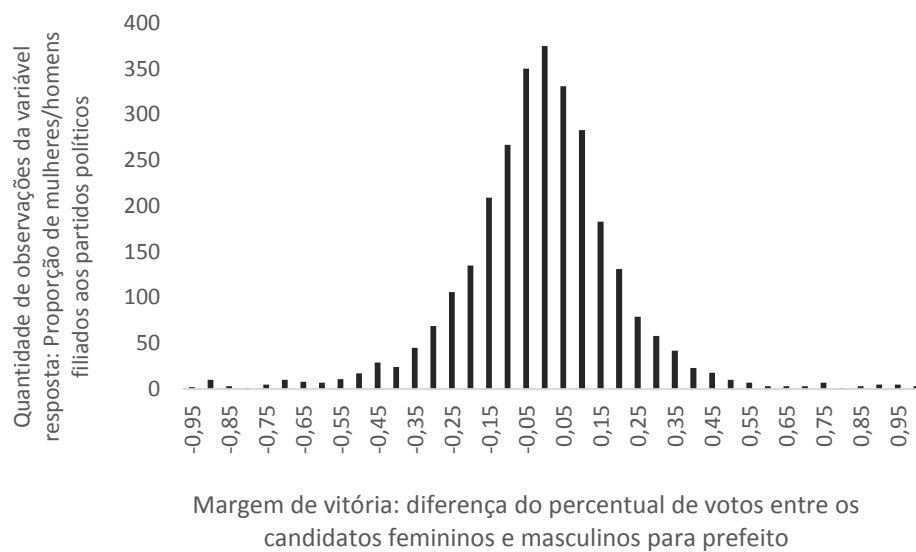


Gráfico 8 - Número de observações da variável resposta: proporção de mulheres/homens filiados aos partidos políticos por faixa da margem de vitória.

Fonte: Elaboração própria.

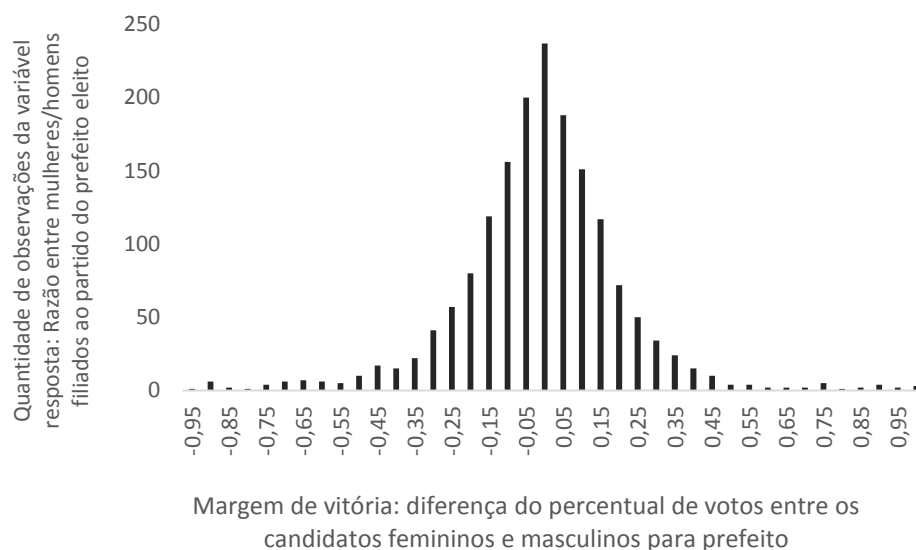


Gráfico 9 - Número de observações da variável resposta: razão entre mulheres/homens filiados ao partido do prefeito eleito por faixa da margem de vitória.

Fonte: Elaboração própria.

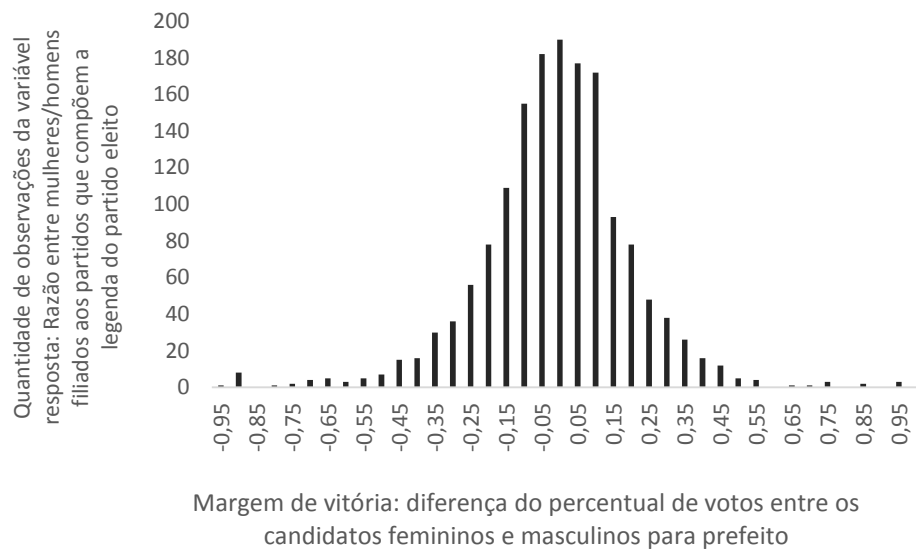


Gráfico 10 - Número de observações da variável resposta: razão entre mulheres/homens filiados aos partidos que compõem a legenda do partido eleito por faixa da margem de vitória.

Fonte: Elaboração própria.

Concluindo a análise gráfica descrita nesta sessão, encontramos alta variabilidade dos dados e pequenos saltos das variáveis resposta também são observados em outros pontos da *running variable* (além do ponto de corte). Com isso, as estimações deverão ser analisadas antes que se chegue a uma conclusão a respeito do efeito do tratamento sobre a resposta.

5.2 Estimação

A estimação segundo a metodologia indicada na seção 3.1.1 pode ser traduzida no ajuste dos dados ao modelo com a especificação dada a seguir:

$$Y_i = \alpha + \tau \cdot W_i + \beta_1 \cdot X_i + \gamma_1 \cdot X_i \cdot W_i + \varepsilon_i, \text{ para } X_i \in [-h, h] \quad (11)$$

em que Y é a variável resposta, W é a variável indicadora do tratamento, X é a *running variable*, ε é o termo de erro, e h é a largura da banda selecionada para a estimação. Aqui, já se considera que o ponto de corte do tratamento, no presente caso, dá-se em $c = 0$.

Além dessa especificação linear, a especificação quadrática também foi testada no presente trabalho, conforme equação a seguir:

$$Y_i = \alpha + \tau \cdot W_i + \beta_1 \cdot X_i + \gamma_1 \cdot X_i \cdot W_i + \beta_2 \cdot X_i^2 + \gamma_2 \cdot X_i^2 \cdot W_i + \varepsilon_i \quad (12)$$

Com o objetivo de se obter várias estimativas para o efeito do tratamento sobre as respostas, os seguintes modelos foram testados:

Modelo1: $prop_fem_masc = \alpha + \tau \text{mulher_venceu} + \varepsilon$

Modelo2: $prop_fem_masc = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \varepsilon$

Modelo3: $prop_fem_masc = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{inter} + \varepsilon$

Modelo4: $prop_fem_masc = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{margin}^2 + \varepsilon$

Modelo5: $prop_fem_masc = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{margin}^2 + \delta \text{inter} + \theta \text{inter}^2 + \varepsilon$

Modelo6: $r_part = \alpha + \tau \text{mulher_venceu} + \varepsilon$

Modelo7: $r_part = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \varepsilon$

Modelo8: $r_part = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{inter} + \varepsilon$

Modelo9: $r_part = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{margin}^2 + \varepsilon$

Modelo10: $r_part = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{margin}^2 + \delta \text{inter} + \theta \text{inter}^2 + \varepsilon$

Modelo11: $r_lege = \alpha + \tau \text{mulher_venceu} + \varepsilon$

Modelo12: $r_lege = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \varepsilon$

Modelo13: $r_lege = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{inter} + \varepsilon$

Modelo14: $r_lege = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{margin}^2 + \varepsilon$

Modelo15: $r_lege = \alpha + \tau \text{mulher_venceu} + \beta \text{margin} + \gamma \text{margin}^2 + \delta \text{inter} + \theta \text{inter}^2 + \varepsilon$

mulher_venceu (variável de tratamento): Indicador binário que recebe o valor 1 quando a prefeita eleita na cidade é do gênero feminino, 0 caso contrário;

margin (*running variable*): margem de vitória que é a diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito;

*margin*²: $\text{margin} * \text{margin}$;

inter: $\text{mulher_venceu} * \text{margem de vitória}$;

*inter*²: $\text{mulher_venceu} * \text{margem de vitória ao quadrado}$;

prop_fem_masc: proporção de mulheres/homens filiados aos partidos políticos;

r_part: razão entre o número de mulheres filiadas ao partido do prefeito eleito sobre o número de homens filiados ao partido do prefeito eleito;

r_lege: razão entre o número de mulheres filiadas aos partidos que compõem a legenda do partido eleito sobre o número de homens filiados aos partidos que compõem a legenda do partido eleito.

A partir dos modelos estruturados anteriormente, vamos direcionar as análises para o parâmetro τ , com o intuito de examinar o efeito do tratamento sobre as respostas. A tabela 34 a seguir demonstra os resultados das estimativas para:

- I. O efeito da vitória de uma mulher sobre a proporção de mulheres/homens filiados aos partidos políticos nos quatro anos seguintes ao ano da eleição (Modelo 1 ao Modelo5);
- II. O efeito da vitória de uma mulher sobre a razão entre o número de mulheres/homens filiados ao partido do prefeito eleito nos quatro anos seguintes ao ano da eleição (Modelo 6 ao Modelo10);
- III. O efeito da vitória de uma mulher sobre a razão entre o número de mulheres/homens filiados aos partidos que compõem a legenda do partido eleito nos quatro anos seguintes ao ano da eleição (Modelo 10 ao Modelo15).

Tabela 34 - Resultados dos coeficientes da variável *mulher_venceu* para os modelos que utilizam:

prop_fem_masc, *r_part* e *r_lege* como variáveis resposta

Variável resposta: <i>prop_fem_masc</i>										
Variável	Modelo1		Modelo2		Modelo3		Modelo4		Modelo5	
	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
<i>mulher_venceu</i>	0.0076	0.2646	0.0087	0.3883	0.0089	0.3754	0.0090	0.3701	0.0123	0.3479

Variável resposta: <i>r_part</i>										
Variável	Modelo6		Modelo7		Modelo8		Modelo9		Modelo10	
	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
<i>mulher_venceu</i>	0.0251	0.1102	0.0622	0.0063	0.0622	0.0066	0.0616	0.0071	0.0443	0.1395

Variável resposta: <i>r_lege</i>										
Variável	Modelo11		Modelo12		Modelo13		Modelo14		Modelo15	
	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
<i>mulher_venceu</i>	0.0109	0.4686	0.0201	0.3770	0.0217	0.3452	0.0209	0.3661	0.0038	0.8984

Conforme apresentado na tabela 34, para os modelos que levam em consideração a proporção de mulheres/homens filiados aos partidos políticos (*prop_fem_masc*) e a razão entre o número de mulheres/homens filiados aos partidos que compõem a legenda do partido eleito (*r_lege*) como variáveis resposta, o efeito do tratamento é indistinguível de zero, de modo que se poderia afirmar, com base apenas nessas estimações, que a vitória de uma mulher não afeta tanto a proporção de mulheres/homens filiados aos partidos políticos quanto razão entre o número de mulheres/homens filiados aos partidos que compõem a legenda do partido eleito nos quatro anos seguintes ao ano da eleição.

Por outro lado, é possível encontrar resultados com significância moderada, em que a vitória de uma mulher afeta a razão entre o número de mulheres/homens filiados ao partido do prefeito eleito nos quatro anos seguintes ao ano da eleição, conforme verificamos nos modelos 7, 8 e 9.

Além dos resultados apresentados nas tabelas acima, foram testados diferentes cortes para a *running variable margin*. Para isso, calculou-se a média e o desvio padrão da variável conforme apresentado a seguir:

Tabela 35 - Média e desvio padrão para a variável *margin*

	Variável: <i>margin</i>
Média	-0.0367
Desvio Padrão	0.2167

A partir do desvio padrão, construiu-se intervalos restringindo as observações que estivessem num raio de $0,5 \times DP$ do zero ou seja, $0,5 \times DP$ para cima e $0,5 \times DP$ para baixo do zero. Repetimos este processo também para $0,25 \times DP$ e $0,10 \times DP$, conforme calculado abaixo:

$$Intervalo1 = 0,5 \times DP = 0,5 \times 0,2167 = 0,1083 \therefore Intervalo1 = [-0,1083; 0,1083]$$

$$Intervalo2 = 0,25 \times DP = 0,25 \times 0,2167 = 0,0542 \therefore Intervalo2 = [-0,0542; 0,0542]$$

$$Intervalo3 = 0,1 \times DP = 0,1 \times 0,2167 = 0,0217 \therefore Intervalo3 = [-0,0217; 0,0217]$$

Utilizando os intervalos acima, geramos os 15 modelos apresentados anteriormente e obtivemos os seguintes resultados:

Resultados utilizando o Intervalo1 = [-0,1083;0,1083]

Tabela 36 - Resultados dos coeficientes da variável mulher_venceu para os modelos que utilizam: prop_fem_masc, r_part e r_lege como variáveis resposta considerando o intervalo1 = -0,1083 <= margin <= 0,1083 como corte

Variável resposta: prop_fem_masc										
	Modelo1		Modelo2		Modelo3		Modelo4		Modelo5	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	0.0052	0.6041	0.0128	0.5076	0.0135	0.4872	0.0134	0.4899	-0.0149	0.6039

Variável resposta: r_part										
	Modelo6		Modelo7		Modelo8		Modelo9		Modelo10	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	0.0520	0.0250	0.0324	0.4644	0.0336	0.4495	0.0354	0.4262	-0.0602	0.3530

Variável resposta: r_lege										
	Modelo11		Modelo12		Modelo13		Modelo14		Modelo15	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	0.0133	0.5461	0.0256	0.5630	0.0282	0.5246	0.0280	0.5286	-0.0248	0.7111

Resultados utilizando o Intervalo2 = [-0,0542;0,0542]

Tabela 37 - Resultados dos coeficientes da variável mulher_venceu para os modelos que utilizam: prop_fem_masc, r_part e r_lege como variáveis resposta considerando o intervalo2 = -0,0542 <= margin <= 0,0542 como corte

Variável resposta: prop_fem_masc										
	Modelo1		Modelo2		Modelo3		Modelo4		Modelo5	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	0.0135	0.3251	-0.0137	0.6115	-0.0141	0.6011	-0.0139	0.6049	-0.0441	0.2618

Variável resposta: r_part										
	Modelo6		Modelo7		Modelo8		Modelo9		Modelo10	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	0.0537	0.0874	-0.0429	0.4846	-0.0439	0.4742	-0.0433	0.4807	-0.1695	0.0547

Variável resposta: r_lege										
	Modelo11		Modelo12		Modelo13		Modelo14		Modelo15	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	0.0221	0.4570	0.0004	0.9941	0.0007	0.9913	0.0006	0.9921	-0.0809	0.3734

Resultados utilizando o Intervalo3 = [-0,0217;0,0217]

Tabela 38 - Resultados dos coeficientes da variável mulher_venceu para os modelos que utilizam: prop_fem_masc, r_part e r_lege como variáveis resposta considerando o intervalo3 = -0,0217 <= margin <= 0,0217 como corte

Variável resposta: prop_fem_masc										
	Modelo1		Modelo2		Modelo3		Modelo4		Modelo5	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	-0.0119	0.5851	-0.0292	0.4774	-0.0290	0.4808	-0.0294	0.4739	0.0349	0.5784

Variável resposta: r_part										
	Modelo6		Modelo7		Modelo8		Modelo9		Modelo10	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	-0.0334	0.5295	-0.1477	0.1331	-0.1524	0.1248	-0.1498	0.1314	-0.0464	0.7561

Variável resposta: r_lege										
	Modelo11		Modelo12		Modelo13		Modelo14		Modelo15	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	0.0066	0.8924	-0.0843	0.3742	-0.0860	0.3664	-0.0858	0.3675	-0.0733	0.6155

Considerando os resultados obtidos nas tabelas 36, 37 e 38, em que utiliza-se diferentes intervalos de corte para a *running variable margin*, o efeito do tratamento sobre as três variáveis resposta nos traz resultados também indistinguíveis de zero, ou seja, a vitória de uma mulher não afeta: 1) proporção de mulheres/homens filiados aos partidos políticos, 2) razão entre o número de mulheres/homens filiados ao partido do prefeito eleito nem 3) razão entre o número de mulheres/homens filiados aos partidos que compõem a legenda do partido eleito, todas considerando os quatro anos seguintes ao ano da eleição.

5.3 Análise de covariadas

Conforme indicado na seção 3.1.4, a análise de covariadas é uma importante etapa da aplicação da metodologia de regressão descontínua, podendo indicar falhas de especificação caso se constate a existência de um suposto impacto do tratamento sobre variáveis que foram definidas anteriormente a ele – e que, portanto, não poderiam ser interpretadas como consequência dele. Para essa análise, no presente

trabalho, escolheram-se duas covariadas: quantidade de mulheres e quantidade de homens filiados aos partidos políticos antes de 2012, cuja descrição e tratamento foram detalhados na seção 4.

Ressalta-se que as variáveis escolhidas foram selecionadas em razão a sua relação direta ou indireta com a dinâmica político-eleitoral nos municípios, além do fato de, como se falou, serem todas conhecidas anteriormente ao pleito.

Inicialmente, a distribuição e o valor médio das covariadas em função da *running variable margin* foram examinadas graficamente – em análise semelhante à da seção 5.1, para as variáveis resposta.

O resultado dessa análise é apresentado nos 4 gráficos a seguir, que demonstram que a inspeção visual não permite constatar a presença de saltos das covariadas no ponto de corte da diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito – o que é favorável para o emprego de RDD aos dados. Apesar das covariadas apresentarem alta variabilidade em função da *running variable*, observa-se que, no entorno do ponto de corte, a variação dos dados é pequena.

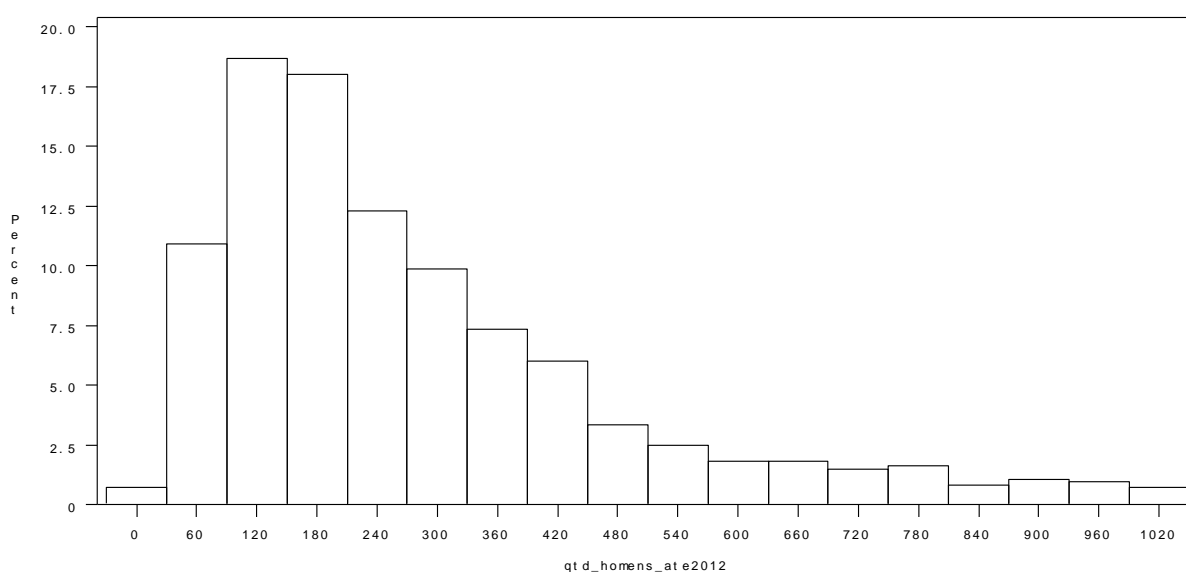


Gráfico 11 - Histograma da quantidade de homens filiados aos partidos políticos antes de 2012.

Fonte: Elaboração própria.

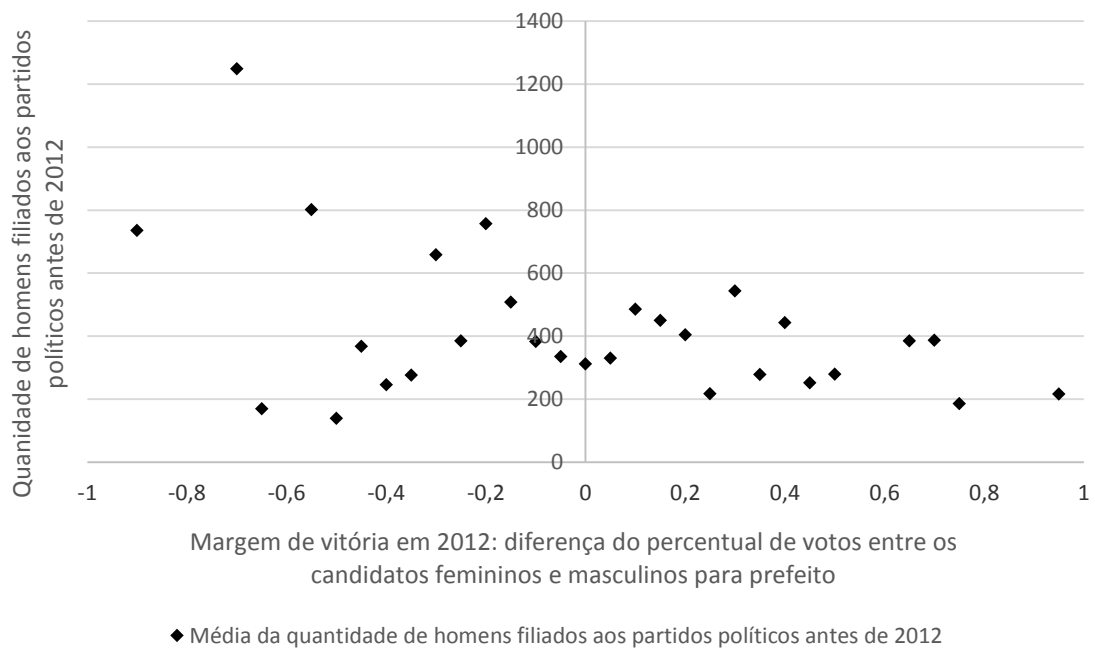


Gráfico 12 - Média da quantidade de homens filiados aos partidos políticos antes de 2012 por faixa da margem de vitória (running variable).

Fonte: Elaboração própria.

Nota: Os pontos representam a média da variável quantidade de homens filiados aos partidos políticos antes de 2012, mensurada no eixo y contra o ponto médio em cada faixa da margem de vitória.

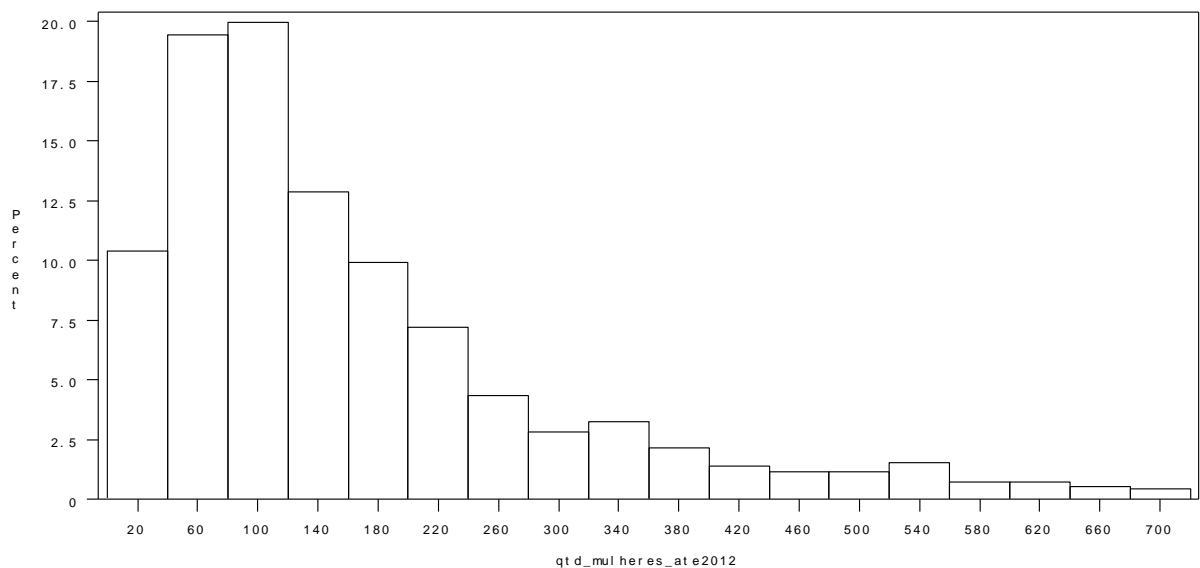


Gráfico 13 - Histograma da quantidade de mulheres filiadas aos partidos políticos antes de 2012.

Fonte: Elaboração própria.

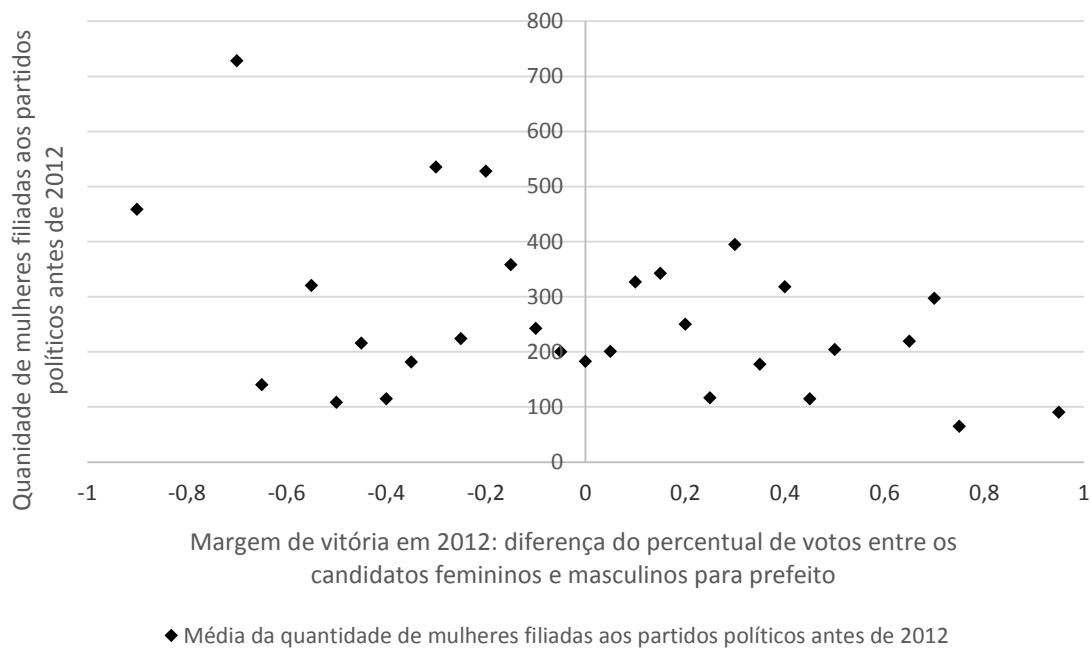


Gráfico 14 - Média da quantidade de mulheres filiadas aos partidos políticos antes de 2012 por faixa da margem de vitória (running variable).

Fonte: Elaboração própria.

Nota: Os pontos representam a média da variável quantidade de mulheres filiadas aos partidos políticos antes de 2012, mensurada no eixo y contra o ponto médio em cada faixa da margem de vitória.

Em seguida à análise gráfica, regressões de especificação similar àquelas das equações (11) e (12) foram realizadas substituindo-se as variáveis resposta pelas covariadas. Para cada regressão testou-se a hipótese nula de que os coeficientes associados à *dummy* de tratamento (indicador se a mulher venceu a eleição) para cada covariada são nulos. Os modelos testados são descritos abaixo e os resultados são apresentados nas tabelas subsequentes.

Modelo1: $qtd_mulheres = \alpha + \tau mulher_venceu + \varepsilon$

Modelo2: $qtd_mulheres = \alpha + \tau mulher_venceu + \beta margin + \varepsilon$

Modelo3: $qtd_mulheres = \alpha + \tau mulher_venceu + \beta margin + \gamma inter + \varepsilon$

Modelo4: $qtd_mulheres = \alpha + \tau mulher_venceu + \beta margin + \gamma margin^2 + \varepsilon$

Modelo5: $qtd_mulheres = \alpha + \tau mulher_venceu + \beta margin + \gamma margin^2 + \delta inter + \theta inter^2 + \varepsilon$

Modelo6: $qtd_homens = \alpha + \tau mulher_venceu + \varepsilon$

Modelo7: $qtd_homens = \alpha + \tau mulher_venceu + \beta margin + \varepsilon$

Modelo8: $qtd_homens = \alpha + \tau mulher_venceu + \beta margin + \gamma inter + \varepsilon$

Modelo9: $qtd_homens = \alpha + \tau mulher_venceu + \beta margin + \gamma margin^2 + \varepsilon$

Modelo10: $qtd_homens = \alpha + \tau mulher_venceu + \beta margin + \gamma margin^2 + \delta inter + \theta inter^2 + \varepsilon$

mulher_venceu (variável de tratamento): Indicador binário que recebe o valor 1 quando a prefeita eleita na cidade no ano de 2012 é do gênero feminino, 0 caso contrário;

margin (*running variable*): margem de vitória em 2012 que é a diferença do percentual de votos entre os candidatos femininos e masculinos para prefeito;

margin²: $margin * margin$;

inter: $mulher_venceu * margem \text{ de vitória}$;

inter²: $mulher_venceu * margem \text{ de vitória ao quadrado}$;

qtd_mulheres: quantidade de mulheres filiadas aos partidos políticos 4 anos antes à eleição de 2012;

qtd_homens: quantidade de homens filiados aos partidos políticos 4 anos antes à eleição de 2012.

Da mesma forma que fizemos nas análises das estimações utilizando as variáveis resposta na seção 5.2, vamos direcionar as análises para o parâmetro τ , com o intuito de examinar o efeito do tratamento sobre as covariadas. As tabelas a seguir demonstram os resultados das estimações para:

- I. O efeito da vitória de uma mulher em 2012 sobre a quantidade de mulheres filiadas aos partidos políticos nos 4 anos antes à eleição de 2012 (Modelo 1 ao Modelo5);
- II. O efeito da vitória de uma mulher em 2012 sobre a quantidade de homens filiados aos partidos políticos nos 4 anos antes à eleição de 2012 (Modelo 6 ao Modelo10);

Tabela 39 - Resultados dos coeficientes da variável *mulher_venceu* em 2012 sobre as covariadas observadas nos 4 anos antes à eleição de 2012

Covariada1: quantidade de mulheres filiadas aos partidos políticos 4 anos antes à eleição de 2012										
	Modelo1		Modelo2		Modelo3		Modelo4		Modelo5	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	23.03	0.2336	13.11	0.6586	14.58	0.6242	13.02	0.6622	23.76	0.5497

Covariada2: quantidade de homens filiados aos partidos políticos 4 anos antes à eleição de 2012										
	Modelo6		Modelo7		Modelo8		Modelo9		Modelo10	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	27.26	0.2795	23.91	0.5343	27.12	0.4816	25.39	0.5111	30.93	0.5480

Conforme se observa, ao nível de significância de 10%, não se rejeita a hipótese nula de que os coeficientes da *dummy* de tratamento (indicador se a mulher venceu a eleição em 2012) são nulos para as covariadas.

Da mesma forma que foi feito na seção 5.2, além dos resultados apresentados nas tabelas acima, também foram testados diferentes cortes para a *running variable margin*. Para isso, calculou-se a média e o desvio padrão da variável margem de vitória em 2012 conforme apresentado a seguir:

Tabela 40 - Média e desvio padrão para a variável margem de vitória em 2012

Variável: <i>margin_2012</i>	
Média	-0.0334
Desvio Padrão	0.2044

A partir do desvio padrão, construiu-se um novo intervalo que restringe as observações num raio de $0,5 \times DP$ do zero ou seja, $0,5 \times DP$ para cima e $0,5 \times DP$ para baixo do zero, conforme calculado abaixo:

$$Intervalo = 0,5 \times DP = 0,5 \times 0,2044 = 0,1022 \therefore Intervalo = [-0,1022; 0,1022]$$

Utilizando o intervalo acima, geramos os modelos apresentados anteriormente e obtivemos os seguintes resultados:

Tabela 41 - Resultados dos coeficientes da variável *mulher_venceu* em 2012 sobre as covariadas observadas nos 4 anos antes à eleição de 2012 considerando o intervalo: $-0,1022 \leq \text{margin_2012} \leq 0,1022$ como corte

Covariada1: quantidade de mulheres filiadas aos partidos políticos 4 anos antes à eleição de 2012

	Modelo1		Modelo2		Modelo3		Modelo4		Modelo5	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	24.77	0.2356	24.54	0.5517	27.20	0.5114	25.60	0.5358	-12.22	0.8550

Covariada2: quantidade de homens filiados aos partidos políticos 4 anos antes à eleição de 2012

	Modelo6		Modelo7		Modelo8		Modelo9		Modelo10	
Variável	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor	τ	p-valor
mulher_venceu	33.42	0.2668	28.54	0.6309	32.96	0.5806	30.20	0.6117	-25.73	0.7894

Considerando os resultados obtidos na tabela 41, em que restringimos as observações para a *running variable* *margin_2012*, o efeito do tratamento sobre as covariadas nos traz resultados também indistinguíveis de zero. Novamente, tem-se um resultado que respalda a especificação de regressão descontínua utilizada no presente trabalho, ao demonstrar que não há efeito do tratamento sobre variáveis que estavam completamente determinadas anteriormente à sua ocorrência.

5.4 Discussão dos resultados

Como mencionado nos tópicos anteriores, a regressão descontínua permite que estimemos e testemos uma relação causal quando há uma forte descontinuidade em determinado corte de uma variável explicativa. Como vimos nas especificações dos modelos apresentados neste trabalho, analisamos a participação política feminina nos partidos e legendas em função de algumas variáveis selecionadas, tais quais uma *dummy* referente à vitória de uma mulher e, principalmente, a margem de vitória de uma candidata em determinada eleição. Esta última é importante pois serve de base para expurgarmos eventuais elementos não contemplados no modelo e que também afetam a participação política das mulheres (aspectos socioculturais ou políticos não abordados neste trabalho).

No trabalho de Brookman (2013), o autor realiza alguns cortes na variável referente à margem percentual de vitória das mulheres. O objetivo é utilizar, mediante a regressão

descontínua, um cenário quase experimental, vale dizer, selecionar casos de eleições extremamente competitivas (com margem de vitória muito apertada) para tentar isolar demais influências que afetam a filiação partidária de mulheres. Em outras palavras, eleições com margem de vitória muito apertada podem ser encaradas como quase randômicas: testar o efeito que uma vitória apertada de uma candidata tem sobre as filiações seguintes é o mais próximo que se pode chegar de um teste experimental em um ambiente genérico e pouco influenciado por outros fatores não contemplados pelo modelo. São os elementos mais próximos do corte que devem se sobressair na estimação do efeito de tratamento de acordo com a metodologia das regressões descontínuas.

Brookman (*ibidem*) realiza cortes em suas regressões com o objetivo de analisar o limiar dos 50% de votos recebidos, ou ainda, o limite em que a margem de vitória é muito próxima de zero. E o resultado obtido, na análise do caso americano, é que não há argumentos estatísticos robustos o suficiente para se inferir que a eleição de mulheres afeta o número de filiações de mulheres em pleitos subsequentes.

Como visto anteriormente, utilizamos em nosso trabalho o mesmo conceito: realizamos cortes na variável de margem de vitória de 0,5, 0,25 e 0,10 desvios-padrão com o intuito de afunilarmos a análise e comparar como a significância estatística varia com este afunilamento da margem de vitória. Em seguida, avaliamos como estes cortes afetam a significância estatística do estimador da variável binária que diz se a mulher venceu ou não a eleição. Os resultados são dúbios.

As regressões gerais, ou seja, aquelas em que utilizamos todo o espectro da *running variable margin*, possuem significância estatística no que tange ao estimador supracitado. Ou seja, num plano estatístico mais abrangente, quando uma mulher vence uma eleição, pode-se dizer que o efeito deste evento sobre as filiações de mulheres é significativo ou, no máximo, há correlação entre as duas pontas. Entretanto, essa relação deixa de ser relevante quando nos aproximamos de casos em que a margem de vitória das mulheres é muito pequena. Neste corte quase experimental em que buscamos a relação causal no bojo das regressões descontínuas, a hipótese que testamos não é corroborada.

6 CONCLUSÃO

Este trabalho teve como objetivo utilizar a metodologia de regressões descontínuas para testar em que medida existe causalidade entre a vitória de mulheres em eleições municipais e a filiação de novas candidatas em pleitos subsequentes no Brasil. Após examinar a literatura sobre a desigualdade de gênero no mundo político e verificar em que países a metodologia proposta foi aplicada, definiu-se passos conceituais aplicados ao caso brasileiro de modo a testar a hipótese supracitada.

Conforme mencionado acima, restringimos a análise às eleições municipais e utilizamos dados do TSE para montar uma base condizente com as necessidades de nossa metodologia. A causalidade é inferida se identificamos descontinuidades – estatisticamente significantes – entre a vitória feminina e novas filiações. Tal descontinuidade é verificada quando analisamos uma variável binária que indica se a mulher venceu ou não uma eleição: se o estimador for robusto, a hipótese é corroborada.

Entretanto, objetivamos aplicar intervalos à nossa análise estatística, pois as regressões descontínuas refletem melhor a existência de causalidade em intervalos específicos. No nosso caso, trabalhamos uma variável que indica a margem de vitória de uma candidata em uma dada eleição. Quanto mais aproximarmos de zero essa margem, mais expurgamos demais fatores exógenos ao modelo e testamos no que chamamos de ambiente quase experimental a hipótese inicial.

A aplicação desta metodologia encontrou aderência em alguns casos, como o indiano (Bhalotra, 2013), mas também encontrou insucessos, como no caso americano (Brookman, 2013). As conclusões deste trabalho se assemelham às deste último, vale dizer, não encontramos significância estatística para inferir que a vitória de candidatas em eleições municipais tem um efeito de causalidade na filiação de novas mulheres em eleições subsequentes.

A afirmação acima pode ser verificada na análise gráfica e estatística dos estimadores. A descontinuidade supracitada não é sensível quando a margem de

vitória é apertada, apesar de, num âmbito geral, existir correlação entre a eleição de mulheres e a filiação política feminina.

Como agenda de pesquisa, sugere-se analisar em que medida outras barreiras implícitas perpetuam a dificuldade de inclusão feminina na política. Sabendo que igualdade de gênero não é uma máxima no caso brasileiro – vide a análise bibliográfica deste trabalho – provamos que o contraexemplo, ou seja, a eleição de novas mulheres, não é suficiente para reduzir a barreira entre homens e mulheres no cenário político brasileiro.

REFERÊNCIAS

- ANGRIST, J.D.; LAVY, V. *Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement*. Quarterly Journal of Economics 114, p. 533-575, 1999.
- AVELINO, G.; BIDERMAN, C.; BARONE, L. S. *Articulações Intrapartidárias e Desempenho Eleitoral no Brasil*. DADOS – Revista de Ciências Sociais, Rio de Janeiro, v. 55, n. 4, p. 987-1013, 2012.
- BHALOTRA, S.; FIGUEIRAS, C.; IYER, L. *Path-Breakers: How does women's political participation respond to electoral success?* IZA Discussion Paper, n. 7771, 2013.
- BLACK, S. *Do Better Schools Matter? Parental Valuation of Elementary Education*, Quarterly. Journal of Economics 114, p. 577-599, 1999.
- BRAMBOR, T.; CENEVIVA, R. *Reeleição e continuísmo nos municípios brasileiros*. Novos Estudos - CEBRAP, n. 93, p. 9-21, 2012.
- BROOCKMAN, D.E. *Do female politicians empower women to vote or run for office?* A regression discontinuity approach. Elsevier – Electoral Studies, 2013.
- CARD, D.; MAS, A.; ROTHSTEIN, J. *Tipping and the Dynamics of Segregation in Neighborhoods and Schools*, Unpublished Manuscript, Department of Economics. Princeton University, 2006.
- CATTANEO, M. D.; FRANDSEN, B. R.; TITIUNIK, R. *Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate*. Journal of Causal Inference, 2014. No prelo. Disponível em < http://www-personal.umich.edu/~cattaneo/papers/Cattaneo-Frandsen-Titiunik_2015_JCI.pdf>. Acesso em 30 nov. 2015.
- CAVENAGHI, S.; ALVES, J.E.D. *Desigualdades de gênero na política na América Latina e as eleições presidenciais do Brasil em 2010*. Dezembro, 2012.
- CHAY, K.; GREENSTONE, M. *Does Air Quality Matter: Evidence from the Housing Market*. Journal of Political Economy 113, p. 376-424, 2005.
- CHAY, K.; MCEWAN, P.; URQUIOLA, M. *The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools*. American Economic Review 95, p.1237-1258, 2005.

DINARDO, J.; LEE, D.S. *Economic Impacts of New Unionization on Private Sector Employers: 1984-2001*. Quarterly Journal of Economics 119, p. 1383-1441, 2004.

IMBENS, G.; KALYANARAMAN, K. *Optimal Bandwidth Choice for the Regression Discontinuity Estimator*. Review of Economic Studies, v. 79, n. 3, p. 933-959, 2012.

IMBENS, G. W.; LEMIEUX, T. *Regression discontinuity designs: A guide to practice*. Journal of Econometrics, v. 142, n. 2, p. 615-635, 2008.

INGLEHART, R.; NORRIS, P. *The Developmental theory of the gender gap: women's and men's voting behavior in global perspective*. International Political Science Review, v.21, n.4, oct. 2000.

INTER-PARLIAMENTARY UNION. *The world's women MPs call for the full participation of the women in countries in conflict and post-conflict situations*, n. 2, April 2004.

LEE, D. S. *Randomized experiments from non-random selection in U.S. House elections*. Journal of Econometrics, v. 142, n. 2, p. 675-697, 2008.

LEE, D. S.; LEMIEUX, T. *Regression Discontinuity Designs in Economics*. Journal of Economic Literature, American Economic Association, v. 48, n. 2, p. 281-355, 2010.

MCEWAN, P.; SHAPIRO, J. *The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates*. Wellesley College and LSE working paper, 2007.

PANDE, R.; FORD, D. *Gender quotas and female leadership: a review*. Background paper for the world development report on gender. Harvard University, April 7, 2011.

PIERI, G. R. *Ensaio sobre economia política aplicada*. 2015. 95 f. Dissertação (Doutorado em Economia) – Escola de Economia de São Paulo da Fundação Getúlio Vargas, São Paulo, 2015.

SCOTT FILHO, R, A. *Efeito da eleição do candidato da oposição sobre a apresentação e aprovação de projetos do prefeito na câmara*. 2015. 106 f. Dissertação (Mestrado em Economia) - Escola de Economia de São Paulo da Fundação Getúlio Vargas, São Paulo, 2015.

THISTLETHWAITE, D. L.; CAMPBELL, D. T. *Regression-discontinuity analysis: an alternative to the ex post facto experiment*. The Journal of Educational Psychology, v. 51, n. 6, p. 309-317, 1960.

TRIBUNAL SUPERIOR ELEITORAL. Repositório de dados eleitorais. Disponível em: <<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais>>. Acesso em 20 ago. 2015.

UNITED NATIONS. *Women & Elections*: Guide to promoting the participation of women in election, v. 04-67240, March 2005.

Van Der KLAAUW, W. *Estimating the Effect of Financial Aid Offers on College Enrollment*: A Regression-discontinuity Approach, *International Economic*, v 43, p. 1249-1287, 2002.

Van Der KLAAUW, W. *Regression-Discontinuity Analysis*: A Survey of Recent developments in Economics. *LABOUR*, v. 22, n. 2, p. 219-245, 2008.

APÊNDICES

Apêndice A – LAYOUT DAS TABELAS EXISTENTES NO REPOSITÓRIO DE DADOS ELEITORAIS E SUAS RESPECTIVAS CARACTERÍSTICAS

- ✓ A codificação de caracteres dos arquivos é “Latin 1”;
- ✓ Os campos estão entre aspas e separados por ponto e vírgula, inclusive os campos numéricos;
- ✓ Campos preenchidos com “#NULO” significam que a informação está em branco no banco de dados. O correspondente para #NULO nos campos numéricos é “-1”;
- ✓ Campos preenchidos com “#NE” significam que naquele ano a informação não era registrada em banco de dados pelos sistemas eleitorais. O correspondente para #NE nos campos numéricos é “- 3”;
- ✓ O campo UF, além das unidades da federação pode conter alguma das seguintes situações:
 - BR – Quando se tratar de informação a nível nacional;
 - VT – Quando se tratar de voto em trânsito;
 - ZZ – Quando se tratar de Exterior.
- ✓ Todas as informações referentes a candidaturas são de responsabilidade do candidato;
- ✓ As variáveis em negrito seguidas de (*) são variáveis que podem ser utilizadas para relacionar os arquivos uns com os outros;
- ✓ Os arquivos estão em constante processo de atualização e aperfeiçoamento. Alguns arquivos podem estar em branco ou com mensagem de erro devido a indisponibilidade temporária na base de algum estado ou à inexistência daquele arquivo para a época pretendida.

Apêndice B – VERSÕES ANTERIORES DO LAYOUT DA BASE DE CANDIDATOS

Tabela 42 - Descrição das variáveis contidas na base de dados dos candidatos nas eleições até 2010

VARIÁVEL	DESCRIÇÃO
DATA_GERACAO	Data de geração do arquivo (data da extração)
HORA_GERACAO	Hora de geração do arquivo (hora da extração) - Horário de Brasília
ANO_ELEICAO	Ano da eleição
NUM_TURNO (*)	Número do turno
DESCRICAO_ELEICAO (*)	Descrição da eleição
SIGLA_UF	Sigla da Unidade da Federação em que ocorreu a eleição
SIGLA_UE (*)	Sigla da Unidade Eleitoral (Em caso de eleição majoritária é a sigla da UF que o candidato concorre (texto) e em caso de eleição municipal é o código TSE do município (número)). Assume os valores especiais BR, ZZ e VT para designar, respectivamente, o Brasil, Exterior e Voto em Trânsito
DESCRICAO_UE	Descrição da Unidade Eleitoral
CODIGO_CARGO (*)	Código do cargo a que o candidato concorre
DESCRICAO_CARGO	Descrição do cargo a que o candidato concorre
NOME_CANDIDATO	Nome completo do candidato
SEQUENCIAL_CANDIDATO (*)	Número sequencial do candidato gerado internamente pelos sistemas eleitorais. Não é o número de campanha do candidato.
NUMERO_CANDIDATO	Número do candidato na urna
NOME_URNA_CANDIDATO	Nome de urna do candidato
COD_SITUACAO_CANDIDATURA	Código da situação de candidatura
DES_SITUACAO_CANDIDATURA	Descrição da situação de candidatura
NUMERO_PARTIDO	Número do partido
SIGLA_PARTIDO	Sigla do partido
NOME_PARTIDO	Nome do partido
CODIGO_LEGENDA	Código sequencial da legenda gerado pela Justiça Eleitoral
SIGLA_LEGENDA	Sigla da legenda
COMPOSICAO_LEGENDA	Composição da legenda
NOME_LEGENDA	Nome da legenda
CODIGO_OCUPACAO	Código da ocupação do candidato
DESCRICAO_OCUPACAO	Descrição da ocupação do candidato
DATA_NASCIMENTO	Data de nascimento do candidato
NUM_TITULO_ELEITORAL_CANDIDATO	Número do título eleitoral do candidato
IDADE_DATA_ELEICAO	Idade do candidato da data da eleição
CODIGO_SEXO	Código do sexo do candidato
DESCRICAO_SEXO	Descrição do sexo do candidato
COD_GRAU_INSTRUCAO	Código do grau de instrução do candidato. Gerado internamente pelos sistemas eleitorais
DESCRICAO_GRAU_INSTRUCAO	Descrição do grau de instrução do candidato
CODIGO_ESTADO_CIVIL	Código do estado civil do candidato
DESCRICAO_ESTADO_CIVIL	Descrição do estado civil do candidato
CODIGO_NACIONALIDADE	Código da nacionalidade do candidato
DESCRICAO_NACIONALIDADE	Descrição da nacionalidade do candidato
SIGLA_UF_NASCIMENTO	Sigla da UF de nascimento do candidato
CODIGO_MUNICIPIO_NASCIMENTO	Código TSE do município de nascimento do candidato
NOME_MUNICIPIO_NASCIMENTO	Nome do município de nascimento do candidato
DESPESA_MAX_CAMPANHA	Despesa máxima de campanha declarada pelo partido para aquele cargo. Valores em Reais.
COD_SIT_TOT_TURNO	Código da situação de totalização do candidato naquele turno
DESC_SIT_TOT_TURNO	Descrição da situação de totalização do candidato naquele turno

Tabela 43 - Descrição das variáveis contidas na base de dados dos candidatos nas eleições em 2012

VARIÁVEL	DESCRIÇÃO
DATA_GERACAO	Data de geração do arquivo (data da extração)
HORA_GERACAO	Hora de geração do arquivo (hora da extração) - Horário de Brasília
ANO_ELEICAO	Ano da eleição
NUM_TURNO (*)	Número do turno
DESCRICAO_ELEICAO (*)	Descrição da eleição
SIGLA_UF	Sigla da Unidade da Federação em que ocorreu a eleição
SIGLA_UE (*)	Sigla da Unidade Eleitoral (Em caso de eleição majoritária é a sigla da UF que o candidato concorre (texto) e em caso de eleição municipal é o código TSE do município (número)). Assume os valores especiais BR, ZZ e VT para designar, respectivamente, o Brasil, Exterior e Voto em Trânsito
DESCRICAO_UE	Descrição da Unidade Eleitoral
CODIGO_CARGO (*)	Código do cargo a que o candidato concorre
DESCRICAO_CARGO	Descrição do cargo a que o candidato concorre
NOME_CANDIDATO	Nome completo do candidato
SEQUENCIAL_CANDIDATO (*)	Número sequencial do candidato gerado internamente pelos sistemas eleitorais. Não é o número de campanha do candidato.
NUMERO_CANDIDATO	Número do candidato na urna
CPF_CANDIDATO	CPF do candidato
NOME_URNA_CANDIDATO	Nome de urna do candidato
COD_SITUACAO_CANDIDATURA	Código da situação de candidatura
DES_SITUACAO_CANDIDATURA	Descrição da situação de candidatura
NUMERO_PARTIDO	Número do partido
SIGLA_PARTIDO	Sigla do partido
NOME_PARTIDO	Nome do partido
CODIGO_LEGENDA	Código sequencial da legenda gerado pela Justiça Eleitoral
SIGLA_LEGENDA	Sigla da legenda
COMPOSICAO_LEGENDA	Composição da legenda
NOME_LEGENDA	Nome da legenda
CODIGO_OCUPACAO	Código da ocupação do candidato
DESCRICAO_OCUPACAO	Descrição da ocupação do candidato
DATA_NASCIMENTO	Data de nascimento do candidato
NUM_TITULO_ELEITORAL_CANDIDATO	Número do título eleitoral do candidato
IDADE_DATA_ELEICAO	Idade do candidato da data da eleição
CODIGO_SEXO	Código do sexo do candidato
DESCRICAO_SEXO	Descrição do sexo do candidato
COD_GRAU_INSTRUCAO	Código do grau de instrução do candidato. Gerado internamente pelos sistemas eleitorais
DESCRICAO_GRAU_INSTRUCAO	Descrição do grau de instrução do candidato
CODIGO_ESTADO_CIVIL	Código do estado civil do candidato
DESCRICAO_ESTADO_CIVIL	Descrição do estado civil do candidato
CODIGO_NACIONALIDADE	Código da nacionalidade do candidato
DESCRICAO_NACIONALIDADE	Descrição da nacionalidade do candidato
SIGLA_UF_NASCIMENTO	Sigla da UF de nascimento do candidato
CODIGO_MUNICIPIO_NASCIMENTO	Código TSE do município de nascimento do candidato
NOME_MUNICIPIO_NASCIMENTO	Nome do município de nascimento do candidato
DESPESA_MAX_CAMPANHA	Despesa máxima de campanha declarada pelo partido para aquele cargo. Valores em Reais.
COD_SIT_TOT_TURNO	Código da situação de totalização do candidato naquele turno
DESC_SIT_TOT_TURNO	Descrição da situação de totalização do candidato naquele turno
NM_EMAIL	E-mail para comunicação com o candidato

Apêndice C – MACRO UTILIZADA PARA EXTRAÇÃO DOS DADOS

Figura 2 - Exemplo da macro utilizada para extrair as bases dos filiados ao PDT utilizando um arquivo executável

[illegible]

Apêndice D – PARTIDOS POLÍTICOS CONSIDERADOS NAS ANÁLISES

Tabela 44 - Siglas e nomes dos partidos políticos levados em consideração para a extração dos dados e análises presentes neste estudo

Sigla	Nome
DEM	DEMOCRATAS
PC do B	PARTIDO COMUNISTA DO BRASIL
PCB	PARTIDO COMUNISTA BRASILEIRO
PCO	PARTIDO DA CAUSA OPERÁRIA
PDT	PARTIDO DEMOCRÁTICO TRABALHISTA
PEN	PARTIDO ECOLÓGICO NACIONAL
PHS	PARTIDO HUMANISTA DA SOLIDARIEDADE
PMDB	PARTIDO DO MOVIMENTO DEMOCRÁTICO BRASILEIRO
PMN	PARTIDO DA MOBILIZAÇÃO NACIONAL
PP	PARTIDO PROGRESSISTA
PPL	PARTIDO PÁTRIA LIVRE
PPS	PARTIDO POPULAR SOCIALISTA
PR	PARTIDO DA REPÚBLICA
PRB	PARTIDO REPUBLICANO BRASILEIRO
PROS	PARTIDO REPUBLICANO DA ORDEM SOCIAL
PRP	PARTIDO REPUBLICANO PROGRESSISTA
PRTB	PARTIDO RENOVADOR TRABALHISTA BRASILEIRO
PSB	PARTIDO SOCIALISTA BRASILEIRO
PSC	PARTIDO SOCIAL CRISTÃO
PSD	PARTIDO SOCIAL DEMOCRÁTICO
PSDB	PARTIDO DA SOCIAL DEMOCRACIA BRASILEIRA
PSDC	PARTIDO SOCIAL DEMOCRATA CRISTÃO
PSL	PARTIDO SOCIAL LIBERAL
PSOL	PARTIDO SOCIALISMO E LIBERDADE
PSTU	PARTIDO SOCIALISTA DOS TRABALHADORES UNIFICADO
PT	PARTIDO DOS TRABALHADORES
PT do B	PARTIDO TRABALHISTA DO BRASIL
PTB	PARTIDO TRABALHISTA BRASILEIRO
PTC	PARTIDO TRABALHISTA CRISTÃO
PTN	PARTIDO TRABALHISTA NACIONAL
PV	PARTIDO VERDE
SD	SOLIDARIEDADE