

# ESTUDO TÉCNICO

N.º 15/2014

Análise Discriminante no tratamento  
dos grupos de domicílios Sem Rendimento (SR)  
e Sem Declaração (SD)

MDS

SAGI

MINISTÉRIO DO DESENVOLVIMENTO SOCIAL E COMBATE À FOME

SECRETARIA DE AVALIAÇÃO E GESTÃO DA INFORMAÇÃO

## **Estudo Técnico**

No. 15/2014 - Análise Discriminante no tratamento dos grupos de domicílios Sem Rendimento (SR) e Sem Declaração (SD)

## **Técnicos responsáveis**

Júlio César Gomes Fonseca

Márcia Valéria Sousa Barbosa

## **Revisão**

Paulo Jannuzzi

Estudos Técnicos SAGI é uma publicação da Secretaria de Avaliação e Gestão da Informação (SAGI) criada para sistematizar notas técnicas, estudos exploratórios, produtos e manuais técnicos, relatórios de consultoria e reflexões analíticas produzidas na secretaria, que tratam de temas de interesse específico do Ministério de Desenvolvimento Social e Combate à Fome (MDS) para subsidiar, direta ou indiretamente, o ciclo de diagnóstico, formulação, monitoramento e avaliação das suas políticas, programas e ações.

O principal público a que se destinam os Estudos são os técnicos e gestores das políticas e programas do MDS na esfera federal, estadual e municipal. Nesta perspectiva, são textos técnico-científicos aplicados com escopo e dimensão adequados à sua apropriação ao Ciclo de Políticas, caracterizando-se pela objetividade, foco específico e tempestividade de sua produção.

Futuramente, podem vir a se transformar em artigos para publicação: Cadernos de Estudos, Revista Brasileira de Monitoramento e Avaliação (RBMA) ou outra revista técnica-científica, para alcançar públicos mais abrangentes.

**Palavras-chave:** *Pobreza, Indicadores sociais, Discriminante, PNAD*

## **Unidade Responsável**

### **Secretaria de Avaliação e Gestão da Informação**

Esplanada dos Ministérios | Bloco A | Sala 307

CEP: 70.054-906 Brasília | DF

Fone: 61 2030-1501 | Fax: 2030-1529

[www.mds.gov.br/sagi](http://www.mds.gov.br/sagi)

### **Secretário de Avaliação e Gestão da Informação**

Paulo de Martino Jannuzzi

### **Secretária Adjunta**

Paula Montagner

## APRESENTAÇÃO

Este estudo técnico sistematiza a metodologia de Aplicação da Análise Discriminante para identificação de pessoas com perfil de extrema pobreza nos grupos de domicílios cuja declaração de rendimento domiciliar *per capita* sejam iguais a zero, doravante neste texto denominados Sem Rendimento (SR), e os grupos de domicílios informados como Sem Declaração (SD), ambos observados segundo os dados captados pela variável V4742 nos microdados das PNADs de 2004 a 2013. Desta forma, o estudo se divide em três seções principais: contextualização do problema, descrição da técnica e validação dos resultados.

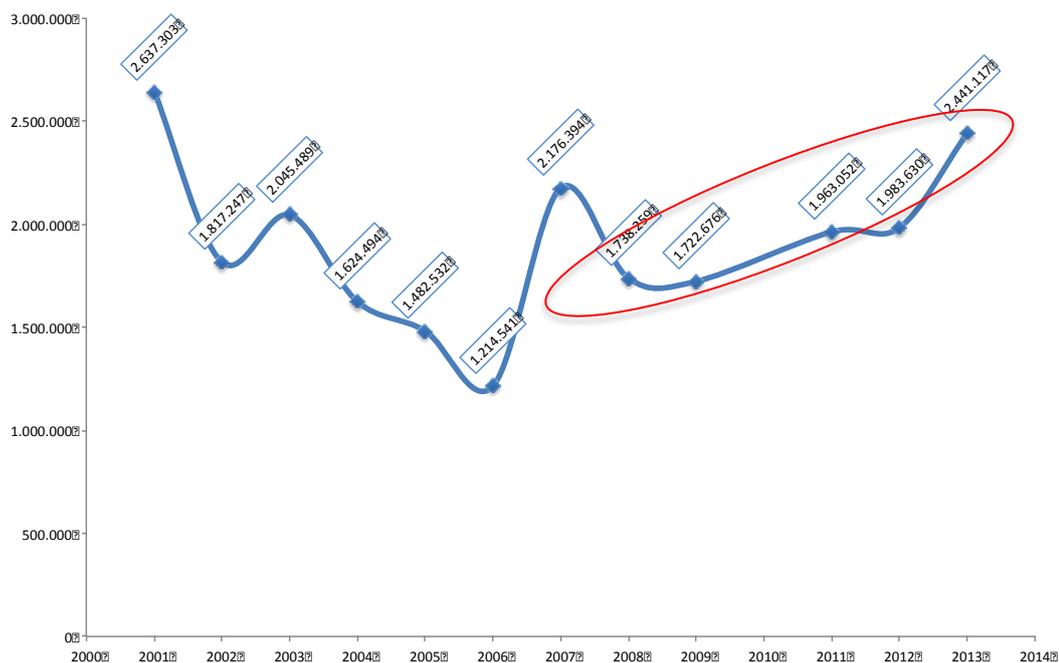
### 1. Contextualização do Problema

No Estudo Técnico n.º 08 de 2014 realizado pela SAGI/MDS, apontou-se as diversas questões metodológicas na computação de indicadores de pobreza e extrema pobreza monetária. Uma dessas questões estava relacionada com a possível superestimativa dos SR e com o tratamento do SD.

Como decisão metodológica aplicada à narrativa do estudo aqui apresentado, consideram-se excluídos das análises, os membros dos domicílios da PNAD cuja condição no domicílio de residência é a de pensionistas, empregados(as) domésticos(as) e parentes de empregados(as) domésticos(as). Partindo desta premissa, o presente estudo não pretende discutir as motivações conceituais, metodológicas ou até mesmo procedimentais para tal questão, valendo-se para tanto, de referências a outros estudos técnicos que tratam as possíveis metodologias passíveis de serem adotadas afim de se incorporar os demais indivíduos no cômputo das estimativas.

Estudos realizados pela SAGI/MDS sobre os segmentos tratados neste estudo, revelam comportamentos distintos na participação dos SR e dos SD, no estudo das estimativas de rendimento domiciliar *per capita* na PNAD nos últimos anos. Neste sentido, observa-se que a série dos microdados de 2001 à 2013 apresenta uma quebra de comportamento no quantitativo dos SR a partir de 2007 (Gráfico 1), onde se observa um aumento de 79,1% em relação a 2006, culminando na inversão do comportamento da série no período subsequente, a partir de 2008.

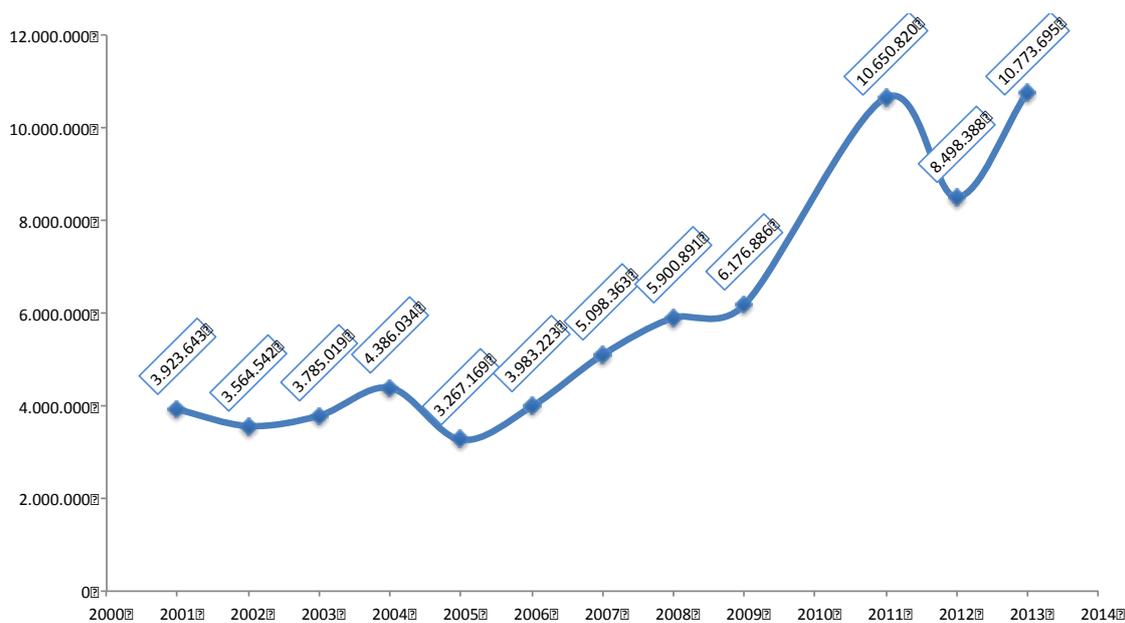
**Gráfico 1: Evolução do quantitativo de indivíduos sem rendimentos no rendimento domiciliar *per capita* – Brasil, 2001-2013**



Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

Já o grupo dos SD apresenta um padrão de comportamento crescente desde o início da série (Gráfico 2), com expressivo aumento em 2011, onde se observava um quantitativo de 6,1 milhões de indivíduos em 2009, se elevar a 10,6 milhões em 2011, representando um aumento de 72,4% entre estes anos.

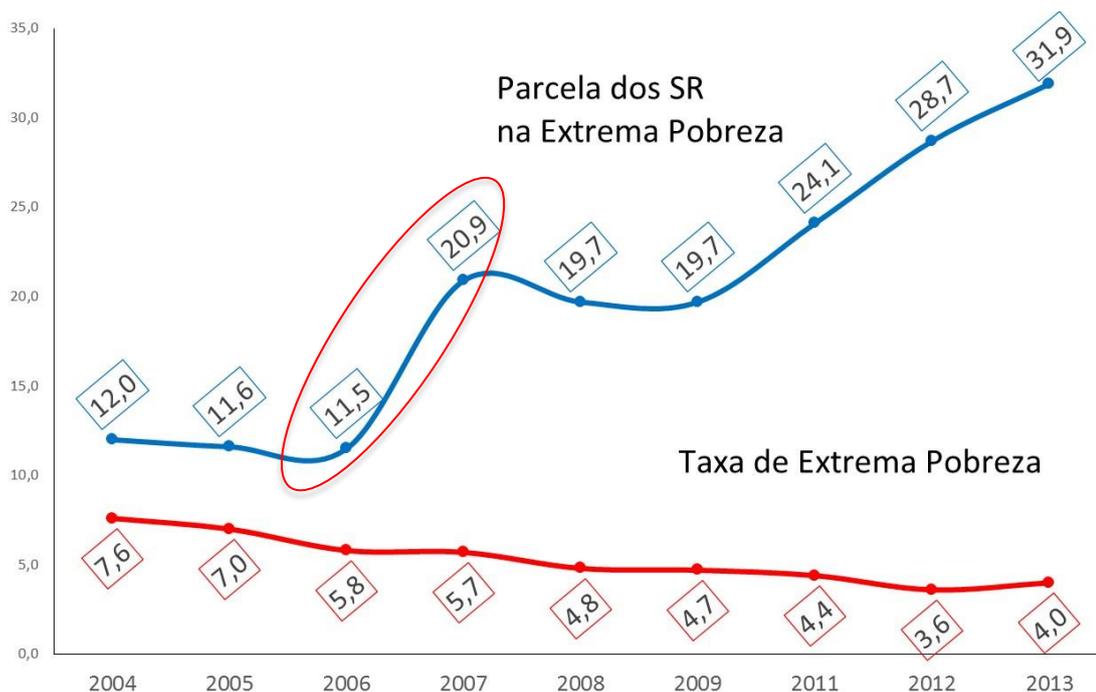
**Gráfico 2: Evolução do quantitativo de indivíduos sem declaração no rendimento domiciliar *per capita* – Brasil, 2001-2013**



Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

O cenário apresentado acima pode ser uma das razões pelas quais no cômputo da taxa de extrema pobreza, o impacto da parcela dos rendimentos declarados como zero chega a 31,9% no ano de 2013 (Gráfico 3). Estes dados revelam o alto grau de superestimação das taxas calculadas utilizando-se pesquisas como a PNAD, na avaliação de vulnerabilidades no país, elevando assim os índices de pobreza e principalmente os de extrema pobreza, sempre que incorporados a quaisquer procedimentos metodológicos a serem adotados. Uma alternativa a ser considerada como estudo comparativo neste sentido, seria o uso de registros administrativos como balizadores para pesquisas com desenhos amostrais complexos. O departamento de monitoramento da SAGI/MDS tem feito esforços nesta direção, apesar de ainda se tratar de uma questão pouco explorada no âmbito do Cadastro Único, quanto a estudos comparativos da renda domiciliar *per capita* em relação a pesquisas de desenhos complexos como a PNAD.

Gráfico 3: **Impacto do quantitativo de indivíduos sem rendimento no rendimento domiciliar *per capita* na Taxa de Extrema Pobreza – Brasil, 2004-2013**

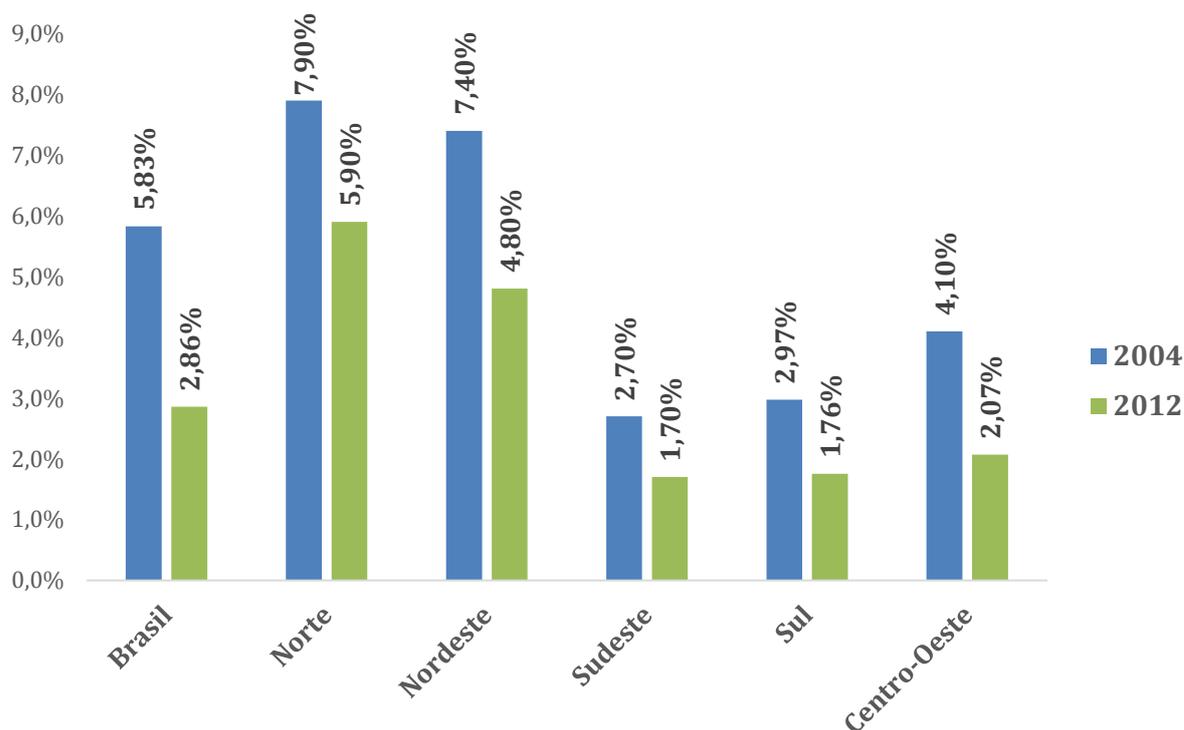


Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

Desconsiderados os SD do cômputo das taxas de extrema pobreza, de forma a não serem contabilizados no denominador do cálculo, a avaliação do impacto dos SR apresentado no gráfico acima, revela uma mudança de patamar a partir de 2007, ano este onde o quantitativo de indivíduos que se auto declararam com rendimento domiciliar *per capita* iguais a zero, passaram de 1.214.541 declarados em 2006 para 2.176.394 em 2007, representando um aumento de 79,19%. Dado este aumento, o impacto na taxa de extrema pobreza causado

por esta parcela, salta de 11,5% para 20,9%, e a partir deste ponto, apesar de todos os esforços revertidos pelas mais diferentes linhas metodológicas<sup>1</sup> para o cômputo das taxas de pobreza e extrema pobreza, os dados da pesquisa parecem de fato apontar para algum tipo de procedimento que remeta à imputação de renda para subsequente avaliação de taxas que apresentam a vulnerabilidade sob diversos aspectos e/ou eixos temáticos, elevando assim o tratamento da pobreza a dimensões multidimensionais.

Gráfico 4: Evolução da pobreza multidimensional, segundo grandes regiões – Brasil, 2004 e 2012



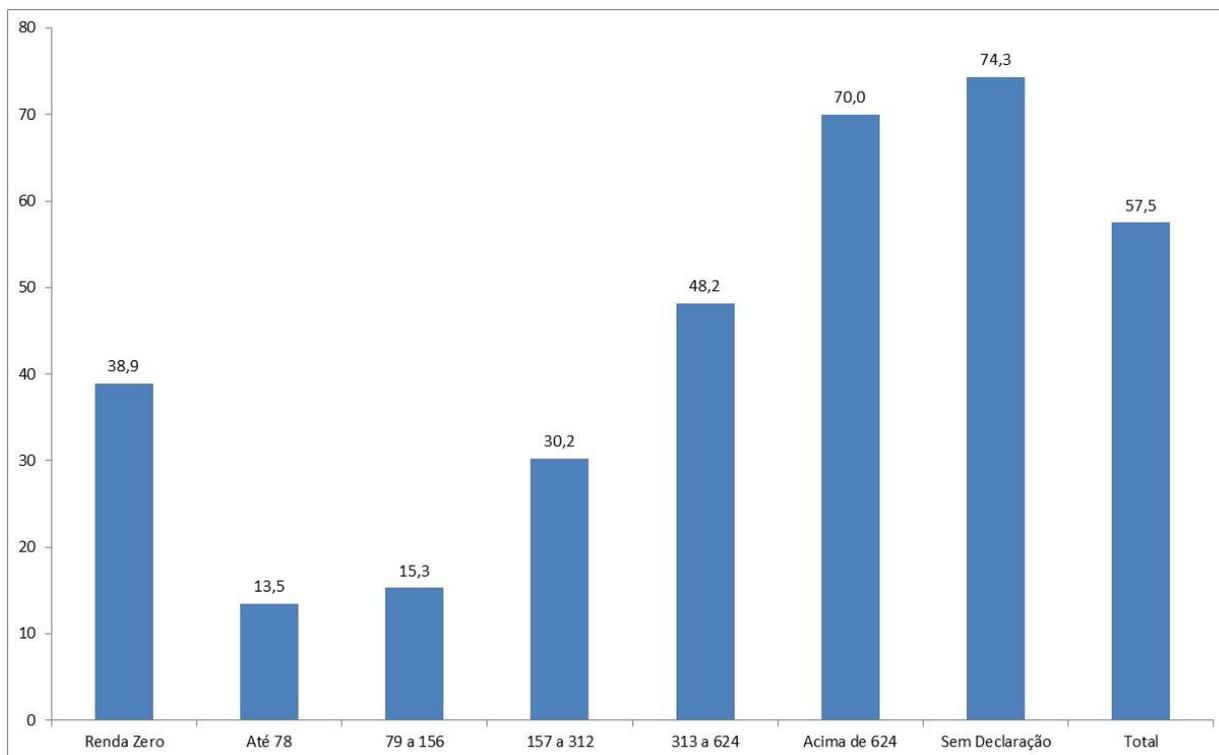
Fonte: Relatório de Desenvolvimento Humano/PNUD, com processamento da PNAD/IBGE

Demais evidências do problema exposto até então, podiam ser checadas avaliando-se a distribuição do rendimento domiciliar *per capita* por algumas variáveis captadas pela pesquisa, de modo que melhor representassem a exposição da população de baixa renda à situação de vulnerabilidade (Gráfico5). A maior parte destas variáveis apresentava um alto percentual de sub-declaração de renda, superior aos níveis apurados nas faixas de renda que definem as linhas de pobreza discutidas pelos principais pesquisadores e instituições com publicações no âmbito da questão da vulnerabilidade social<sup>2</sup>.

<sup>1</sup> Citar publicações que tratem de linhas metodológicas adotadas no cálculo das taxas de EP e PO

<sup>2</sup> Vide publicações do CEPAL, IPEA, Banco Mundial, PNUD, entre outros.

**Gráfico 5: Distribuição percentual do rendimento domiciliar *per capita* entre os indivíduos que possuem máquina de lavar – Brasil, 2013**



**Fonte:** IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

Considerando o período de estudo, foram avaliados alguns dos itens quanto a suscetibilidade à vulnerabilidade social, através uma análise do perfil dos indivíduos que se encontravam abaixo da linha pobreza, e destes, de um conjunto de cerca de 18 (dezoito) variáveis, elencou-se 10 (dez) deles através de estudos técnicos anteriores<sup>3</sup> sobre monitoramento das taxas de pobreza e extrema pobreza. Dentre todos os resultados obtidos, a distribuição apresentada pelo questionamento quanto à posse de *máquina de lavar* foi o mais expressivo e portanto um balizador para se mitigar a utilização de outras técnicas que pudessem apontar possíveis soluções para a grande questão a ser respondida : “*dado o perfil semelhante apresentado por indivíduos com declaração de rendimento domiciliar per capita iguais a zero nas (10) dez variáveis que melhor classificam a vulnerabilidade social, em particular aquelas que tratam da pobreza monetária quanto à ausência de itens no domicílio, considerados como necessidades básicas de uma sobrevivência digna, estariam de fato estes indivíduos, à época da pesquisa, sem rendimento? Ou poderiam ser reclassificados como não extremamente pobres dado seus padrões sociais quanto à presença de itens caracterizados como de luxo em seus domicílios?*”.

<sup>3</sup> Estudo Técnico n.º 10 de 2011.

Para dar respostas a estas perguntas, iniciou-se um trabalho investigativo de métodos que se aplicassem ao problema em questão, e que pudessem apresentar soluções concomitantemente razoáveis aos resultados já investigados em outras publicações que trataram do mesmo tema. Neste contexto discutiu-se sobre a utilização de técnicas conhecidas e utilizadas como alternativas para a aplicação no problema proposto.

## 2. Descrição da técnica

Após trabalho investigativo sobre a aplicabilidade de um modelo estatístico preditivo, por se tratar de estimação da probabilidade de um indivíduo “ser extremamente pobre” ou não, optou-se pela utilização de um método multivariado com caráter classificatório e não exatamente preditivo, em função da facilidade de uso e aplicabilidade no problema proposto.

Assim, a metodologia utilizada para esse estudo balizou-se em técnicas provenientes do ramo da estatística que trata de análises multivariadas. Tais técnicas têm como premissa, em sua grande maioria, a predição, classificação, agrupamento e/ou correspondência de um ou mais grupos em função da presença de outros fatores que juntos, formam padrões multivariados, dado seu conjunto estrutural de ocorrências, contendo determinado perfil característico.

A técnica adotada foi a *Análise de Discriminante* por se tratar de método estritamente de investigação da classificação de grupos, segundo características determinísticas. Como a principal questão a ser investigada tratava da classificação dos indivíduos cujo rendimento domiciliar *per capita* foi declarado como “zero”, enquanto possibilidade de pertencerem, em verdade, ao grupo dos “*extremamente pobres*” em função das características intrínsecas deste grupo, definidas pelas variáveis elencadas como determinísticas do grupo, definiu-se como variável de interesse (*variável resposta*) a presença ao grupo cuja renda domiciliar *per capita* pertencesse ao intervalo de R\$ 1,00 a R\$ 70,00<sup>4</sup>.

Dessa forma, uma variável indicadora do tipo “*dummy*” foi definida como “1” para todos os indivíduos que pertencessem à faixa de renda do grupo acima citada, e “0” para os demais. Por ser a variável de interesse uma variável dicotômica, sabe-se pela literatura estatística, haver duas técnicas amplamente utilizadas para abordar esse tipo de problemática: a Análise de Discriminante (AD) e Análise de Regressão Logística (RL).

Ambas as técnicas são amplamente utilizadas em métodos estatísticos multivariados para dados categóricos, sendo portanto apropriadas para o desenvolvimento de modelos de

---

<sup>4</sup> Limite ajustado ano a ano segundo inflação calculada pelo INPC. O valor em questão remete ao ano de 2011.

classificação. No entanto, as duas técnicas diferem em sua ideia básica. Enquanto a RL não precisa de suposições sobre a distribuição dos dados explicativos, a AD pressupõe que as variáveis explicativas tenham uma distribuição normal. Dessa forma, é razoável esperar que a AD apresente melhores resultados quando os pressupostos de normalidade são cumpridos. (HARRELL & LEE, 1985)

Escolhida a técnica, utilizou-se o conjunto das 10 (dez) variáveis abordadas nas análises de perfis, recodificando-se as mesmas afim de facilitar sua utilização e interpretação (Quadro 1).

Quadro 1: Variáveis utilizadas para AD

Variável na PNAD	Descrição	Construção metodológica	Dummy utilizada
V8005	Idade do Trabalhador	0 - tem de 1 a 3 anos 1 - tem mais de 3 filhos	V1_CRIANCA
V0230	Tem máquina de lavar roupa?	0 - não tem 1 - tem	V2_MAQLAVAR
V2032	Tem carro ou motocicleta de uso pessoal?	0 - não tem 1 - tem	V3_CARRO
V0228	Tem geladeira?	0 - não tem 1 - tem	V4_GELADEIRA
V0221 / V0222	Tem fogão de duas ou mais bocas? Tem fogão de uma boca?	0 - não tem 1 - tem (tem fogão com gás ou elétrico)	V5_FOGAOMBOM
V4624	Forma de abastecimento de água	0 - não tem 1 - tem	V6_AGUA
V0232	Microcomputador é utilizado para acessar a Internet?	0 - não tem 1 - tem	V7_INTERNET
V4745	Nível de instrução mais elevado alcançado (todas as pessoas)	0 - não tem (se 1 e 2) 1 - tem (demais)	V8_ESCOLARIDADE
V2020	Tem telefone fixo convencional?	0 - não tem 1 - tem	V9_TELFIXO
V0215	Tem banheiro ou sanitário no domicílio ou na propriedade?	0 - não tem 1 - tem	V10_BANHEIRO

Tomou-se como referência o ano de 2011 para início da modelagem, ano para o qual o limite da linha de extrema pobreza considerou o valor nominal de R\$ 70,00. Inicialmente foram isolados todos os integrantes do grupo dos SR, dos SD e os missings das bases de dados, criando-se assim sub populações doravante neste texto denominadas por “*universo*” de

indivíduos dos quais se separaria uma parte para o processo de “*criação*” e outro para o processo de “*aplicação*” do modelo<sup>5</sup>.

O universo separado para “*criação*” do modelo levou em consideração os seguintes pontos:

- 1) O grupo dos SR seria o ponto focal de aplicação do modelo, pertencendo portanto apenas ao universo “*aplicação*”.
- 2) O perfil a ser “*classificado*” deveria ser balizado pelos indivíduos cujo rendimento domiciliar *per capita* fosse descritivo do grupo dos “*extremamente pobres*”, doravante denominado por grupo “*sucesso*”, composto pelos indivíduos cuja renda domiciliar *per capita* estivesse no intervalo de R\$ 1,00 a R\$ 70,00.
- 3) Os demais indivíduos comporiam o grupo “*fracasso*”, e neste caso, considerados todos os indivíduos cuja renda domiciliar *per capita* fosse superior a R\$ 70,00 até o maior valor de renda domiciliar *per capita* captada pela amostra.

A partir do universo “*criação*”, separou-se dois grupos. De acordo com Hair (2005), a base foi dividida na proporção 75% e 25% aplicando-se procedimento de amostra aleatória simples, selecionando-se registros ao atribuir “0” ou “1” a uma variável que definiria a separação dos grupos conforme distribuição uniforme<sup>6</sup>, de forma a permitir uma posterior validação cruzada. Assim, uma parte da amostra foi utilizada para se desenvolver as funções discriminantes (grupo contendo 75% do universo “*criação*”) e a outra parte utilizada como teste para validação (grupo contendo 25% do universo “*criação*”).

Definido o modelo AD de 2011, calculou-se as estimativas de sensibilidade, especificidade, valores preditos positivos e valores preditivos negativos, afim de se avaliar a acurácia do modelo. Realizados estes cálculos, verificou-se que o modelo em criado não apresentava “uma boa performance”, pois apresentava baixíssima sensibilidade e alta especificidade, conduzindo a um grande número de falsos positivos.

Dessa forma, após o cálculo dos EP e nEP foi possível avaliar a qualidade do score do modelo. Para essa etapa, utilizou-se a estatística de *Kolmogorov-Smirnov* (KS), que é definida como a diferença absoluta máxima das funções de distribuição acumulada (fda) entre os EP e nEP.

---

<sup>5</sup> Denota-se por “*modelo*” o resultado final das funções discriminantes ajustadas pelo software estatístico utilizado (SPSS v22). Neste caso, como a variável de interesse possui dois grupos (compostos pelo “*sucesso*” aqui representado pelo fato de pertencer ao grupo alvo (indivíduos extremamente pobres), e o “*fracasso*” representado pelos demais, ao final do procedimento, apenas as probabilidades estimadas para o grupo “*sucesso*” e os coeficientes relacionados a este grupo, foram utilizados na atribuição ao universo “*aplicação*”.

<sup>6</sup> Permite-se assim dar chances iguais aos elementos amostrados de pertencerem a ambos os grupos.

$$KS = \max |F_{m.nEP}(a) - F_{n.nEP}(a)|$$

Verificou-se, portanto, que a estimativa KS apresentava uma boa separação entre os grupos dicotômicos, ao alterar o ponto de corte onde as probabilidades criadas pela da função discriminante, associada ao “sucesso” (EP), deveriam de fato classificar um SR como um EP. O problema persistia, todavia, porque o software estatístico utilizado ainda não havia escolhido o ponto ótimo para a separação dos grupos. Para solucionar este problema, utilizou-se o indicador de qualidade do modelo, denominado Lift.

O Lift acumulado indica quantas vezes, em um determinado nível de rejeição, a seleção de pontuação do modelo é melhor que a seleção aleatória. Na prática, o Lift é computado correspondendo aos decis das famílias de melhor score, ou seja, a base é dividida em 10 partes (decil), calculando-se o indicador.

$$\text{Lift}(a) = F_{m.nEP}(a) / F_{n.nEP}(A)$$

Calculado este indicador, considerou-se como medida final para avaliação da qualidade do modelo estimado, o valor Lift superior a de 1,5, valor este que segundo especialistas do uso da técnica, combinado com o valor KS superior a 40%, produzem bons resultados em aplicações que utilizam-se de técnicas de data mining. Dessa foram estabelecido um patamar mínimo para se aceitar um ponto de corte aceitável para o modelo criado.

Aplicadas as etapas descritas acima, em todos os modelos da série, verificou-se que os grupos que foram fortemente considerados como EP no universo “*criação*”, validaram-se como EPs no grupo separado para validação. Assim, após a utilização do novo ponto de corte, que reclassificava os indivíduos entre EP e nEP, avaliou-se, novamente, as estimativas de performance, melhorando consideravelmente as estimativas de sensibilidade, e embora as estimativas de especificidade tem reduzido um pouco a acurácia final do modelo não sofreu grandes alterações, permitindo assim a aplicação do mesmo no universo “*aplicação*”, composto unicamente pelos SR.

Deste ponto em diante, consideradas realizadas as validações do modelo e reclassificação dos SR como possíveis EP, de acordo com os resultados do modelo, utilizou-se os mesmos procedimentos na construção do restante da série.

Consideradas as estimativas produzidas através de aplicação da técnica no universo que não continha nem SR, nem SD, replicou-se a aplicação do modelo gerado no grupo SD tal como se fez no grupo SR, ou seja, a mesma função discriminante utilizada para classificar um

SR com base na aplicação da combinação linear dos coeficientes gerados pela função discriminante nas variáveis preditoras deste grupo, foi aplicada no grupo dos SD.

### **3. Validação dos resultados**

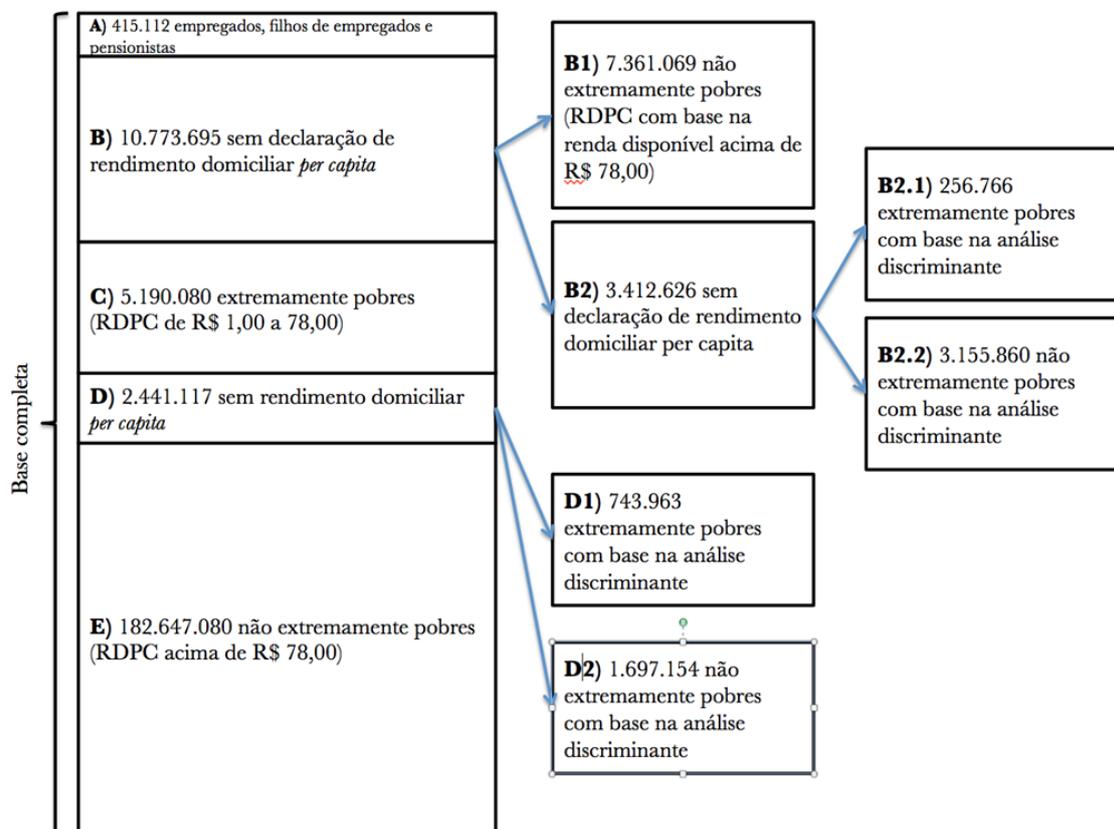
Concomitante ao estudo aqui apresentado, outros esforços – determinísticos e probabilísticos – vinham sendo realizados paralelamente ao trabalho investigativo das técnicas aqui apresentadas, para identificação de pessoas em Extrema Pobreza entre aqueles com Renda Zero e aqueles Sem Declaração, a saber:

1. Cálculo de renda computável entre os SD
2. Cálculo de renda computável e imputação de rendimentos em casos específicos
3. Uso do Filtro aplicado ao Censo Demográfico 2010
4. Desenvolvimento e uso de filtro determinístico na PNAD
5. Aplicação da metodologia de Tratamento de Volatilidade na Extrema Pobreza

Dentre os diversos procedimentos investigados pela SAGI/MDS, aplicou-se como soma de esforços na busca pela solução do problema apresentado, a Análise de Discriminante em conjunto com o cálculo de renda computável entre os SD, os quais produziram resultados bastante promissores no cômputo das taxas de extrema pobreza e pobreza. Dessa forma, o modelo estimado foi aplicado aos SD após o cálculo de renda computável entre os SD, procedimento este que leva em consideração o recálculo da renda domiciliar *per capita* para todos os indivíduos que apresentaram pelo menos um rendimento recebido nas variáveis que captam as transferências de renda obtidas do governo federal.

Após a aplicação destes procedimentos, um levantamento minucioso de alguns resultados foram feitos de forma a expandir os resultados em frentes diferentes de análises, produzindo impactos bastante satisfatórios no contexto investigativo a que se propõem a SAGI/MDS, conforme mostra a figura a seguir.

Figura 1: Fluxograma dos procedimentos adotados no contingente populacional dos SR e SD no cômputo da estimativa da taxa de extrema pobreza, Brasil - 2013



A aplicação do primeiro procedimento (item B) entre os 10,8 milhões de indivíduos classificados como SD na PNAD 2013, permitiu a identificação de 7,4 milhões de indivíduos com rendimento domiciliar *per capita* acima do parâmetro de referência da extrema pobreza (R\$ 78 em 2013)<sup>7</sup>. Haveria, pois, entre os SD cerca de 3,4 milhões de indivíduos que, pela renda disponível dos declarantes, não ultrapassavam a linha de extrema pobreza. Com a aplicação do modelo, desses 3,4 milhões de pessoas, somente 257 mil indivíduos seriam efetivamente classificáveis como extremamente pobres. Tal procedimento aplicado no conjunto dos 2,4 milhões de indivíduos declarados como SR, apontou que somente 745 mil seriam, de fato, extremamente pobres. Com os procedimentos adotados, em 2013, a taxa de extrema pobreza seria de 3,1% e de pobreza 8,1%<sup>8</sup> (Figura 2). A aplicação de procedimentos

<sup>7</sup> Equivale ao parâmetro de R\$ 70 atualizado pelo INPC de junho de 2011 a setembro de 2013.

<sup>8</sup> Com base no fluxo apresentado a taxa de extrema pobreza deveria ser calculada pela fórmula: Tx Ext Pobreza = (C+B2.1+D1) / (B+C+D+E), diferente pois da fórmula tradicional Tx Ext Pobreza = (C+D) / (C+D+E). Assim, a taxa de 3,1% é obtida pela razão entre 6.190.809 milhões de extremamente pobres sobre um total populacional de 201.051.972 de indivíduos. Sem os aprimoramentos propostos no tratamento dos SD e SR, a taxa seria de 4,0% de extremamente pobres, obtido pela razão entre 7.631.197 milhões de extremamente pobres sobre um total de 190.278.277 milhões de indivíduos. Vale observar que a diferença entre as taxas originais e as aprimoradas aumenta ao longo dos anos, refletindo as dificuldades da captação da variável rendimento ao longo do período.

análogos sobre os remanescentes não extremamente pobres dos grupos SD e SR, estão sendo estudadas e os primeiros resultados revelam que a larga maioria deles também não seria pobre. Assim, adotou-se como estimativa adicional de população pobre aquela resultante do primeiro procedimento, com renda computável entre R\$ 78 e R\$ 156 (570 mil pessoas em 2013)<sup>9</sup>.

No intuito de validar da técnica aplicada, através de seus resultados intrínsecos, observa-se pelo quadro de correlações (Quadro 2) entre os variáveis preditoras e a função discriminante canônica padronizada, que com exceção das variáveis, V3\_CARRO para a qual só se tem informação na PNAD após o ano de 2008; V7\_INTERNET para a qual apesar de o software apresentar estimativas de correlação nos anos de 2007 a 2009, a variável não foi ajustada para o modelo final, ou seja, no processo de estimação dos parâmetros do modelo, em alguma etapa da seleção de variáveis, a variável foi considerada, tendo portanto sido geradas as estimativas, porém, não aparece no modelo final; e V1\_CRIANCA cujo os coeficientes apontam uma correlação negativa; todas as demais variáveis apresentam correlações superiores a 0,4 e em sua grande maioria apresentando uma redução ao longo do tempo, indicando o impacto do ajuste.

Quadro 2: Correlações do ajuste do modelo de Análise de Discriminante - Brasil, 2004-2013

Correlações entre as Variáveis Discriminantes e as Funções Discriminantes									
VARIÁVEIS	2004	2005	2006	2007	2008	2009	2011	2012	2013
V1_CRIANCA	-0,153	-0,197	-0,204	-0,222	-0,251	-0,259	-0,311	-0,146	-0,163
V2_MAQLAVAR	0,433	0,428	0,405	0,452	0,458	0,496	0,534	0,531	0,601
V3_CARRO	-	-	-	-	0,414	0,425	0,411	0,418	0,449
V4_GELADEIRA	0,749	0,73	0,747	0,726	0,717	0,693	0,593	0,538	0,453
V5_FOGAOBOM	0,664	0,668	0,654	0,674	0,63	0,581	0,581	0,543	0,477
V6_AGUA	0,727	0,68	0,7	0,647	0,665	0,645	0,633	0,673	0,674
V7_INTERNET	0,246	0,251	0,263	0,326	0,362	0,39	0,444	0,481	0,545
V8_ESCOLARIDAC	0,416	0,405	0,396	0,421	0,393	0,413	0,379	0,377	0,414
V9_TELFIXO	0,55	0,543	0,495	0,497	0,493	0,479	0,435	0,414	0,465
V10_BANHEIRO	0,679	0,692	0,675	0,656	0,616	0,598	0,604	0,691	0,653

Em valores absolutos os principais quantitativos apresentados pelo procedimento conjunto do cálculo da renda computável e a aplicação do modelo no ano de 2013 são:

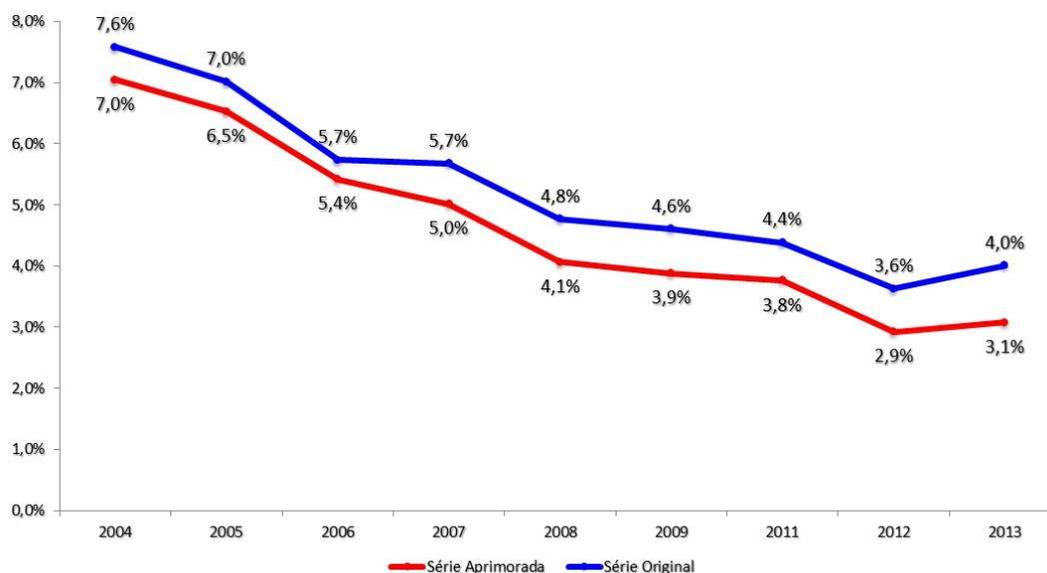
1. Cálculo da Renda Computável: Dos 10,7 milhões de pessoas em domicílios Sem Declaração de Renda, 7,3 milhões não seriam EP pela renda declarada por pelo menos uma pessoa no domicílio.

<sup>9</sup> Os resultados obtidos com a aplicação da Análise Discriminante apontam um caminho promissor para continuidade do refinamento das estatísticas de pobreza, seja no tratamento dos SR e SD, como também na subdeclaração de recebimento do Programa Bolsa Família. Estudos preliminares apontam que o impacto na estimativa poderia ser até 1 ponto percentual menor.

2. A aplicação Análise Discriminante: nos remanescentes 3,4 milhões de SD apontou que 257 mil teriam perfil EP, ou seja, dos 10,7 milhões de SD somente 2,4% teriam perfil de EP ( 257 mil).
3. A aplicação Análise Discriminante: nos 2,4 milhões com Renda Zero apontou a existência de 1,7 milhão sem perfil de EP (69,5%), ou seja, dos 2,4 milhões com Renda Zero, somente 744 mil (30,5%) teriam perfil EP.
4. A taxa de extrema pobreza: passa a ser, de fato, bem mais representativa da população brasileira, a população sob análise - denominador da taxa - passa de 190 milhões para 201 milhões, 99,8% da população brasileira estimada para 2013

Comparativamente a série das taxas de extrema pobreza visualizadas lado a lado, demonstram uma clara evidência do ajuste do modelo, evidenciando o peso da subdeclaração sobre o impacto realmente decisivo na avaliação das políticas públicas, visto que estas se baseiam também nestes resultados como metas de resultados.

Gráfico 6: Comparação das séries da taxa de extrema pobreza antes e pós aplicação da análise discriminante – Brasil, 2013

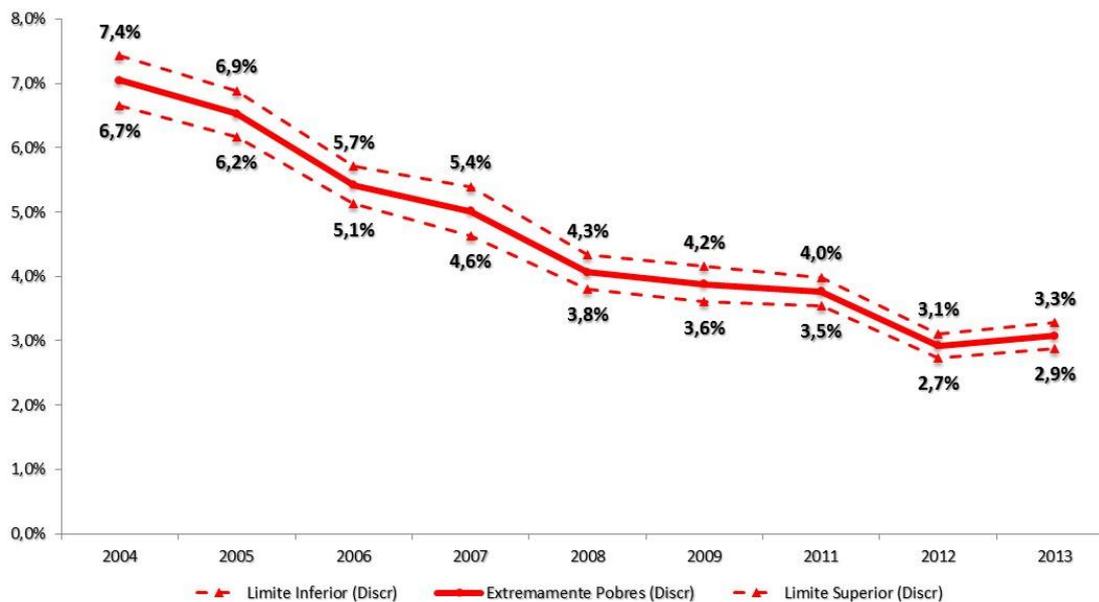


Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

Exaurindo os esforços de análise, calculou-se também os intervalos de confiança de 95% (Gráfico 7) considerando-se o desenho complexo da amostra da PNAD (SILVA; PESSOA 2002, vol. 7, n.4, pp. 659-670), segundo o qual observou-se que em 2004, a taxa de extrema pobreza estava entre 6,7% e 7,4% culminando em 2013 entre 2,9% e 3,3%. Estes resultados corroboram a hipótese de que a questão da subdeclaração avaliada conjuntamente sob todos os aspectos apresentados neste e em outros estudos, demonstra evidência da

necessidade de ajustes no desenho da pesquisa, seja no âmbito da formulação das perguntas ou mesmo na captação das informações.

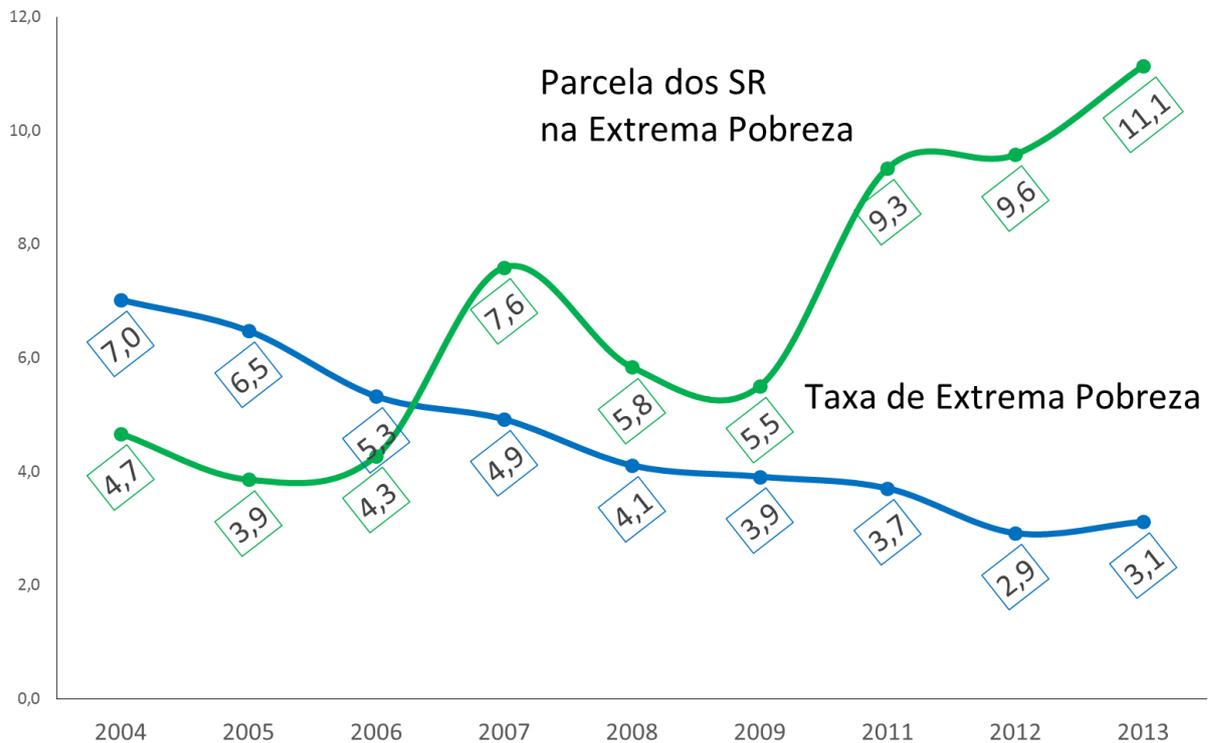
**Gráfico 7: Evolução da taxa de extrema pobreza e seus intervalos de confiança a 95% – Brasil, 2004-2013**



**Fonte:** IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

Recalculado o impacto dos SR (Gráfico 8) na taxa de extrema pobreza após a análise de discriminante, observa-se que comparativamente ao apresentado no gráfico 3, que o quantitativo de indivíduos que se auto declararam com rendimento domiciliar *per capita* iguais a zero, passaram de 783.276 declarados em 2006 para 1.438.949 em 2007, representando um aumento de 83,7%. Embora este aumento seja em termos relativos maior que o apresentado antes da aplicação da técnica, em número absolutos observa-se que a redução entre os quantitativos de mesmo ano, antes e depois da técnica, é realmente muito espantosa e da ordem de 55,1% no ano de 2006 e 51,2% para o ano de 2007, mostrando que a questão está além apenas da estimação da subdeclaração, apontando também para algum problema na estrutura dos dados, visto que o comportamento da série não se altera apesar de haver deslocamento em seus níveis.

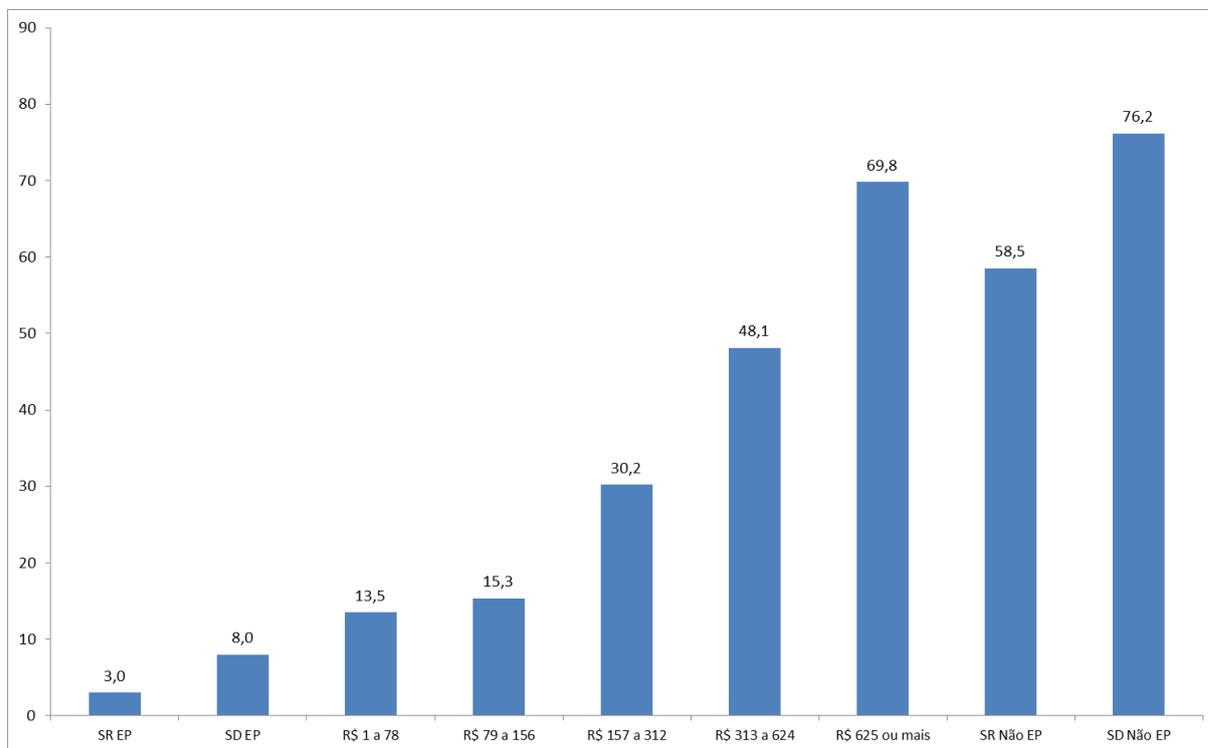
**Gráfico 8: Impacto do quantitativo de indivíduos sem rendimento no rendimento domiciliar *per capita* na Taxa de Extrema Pobreza, pós aplicação da Análise Discriminante – Brasil, 2004-2013**



**Fonte:** IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

Por fim, observando o comportamento da distribuição da renda domiciliar *per capita* frente à variável “*máquina de lavar*” (Gráfico 9), tem-se mais uma clara evidência da importância da questão, visto que o impacto os SR antes avaliado em 38,9% dentre os indivíduos que apesar da declaração do rendimento domiciliar *per capita* zerado, possui bens de consumo de padrão elevado, cai para 3,0% após a aplicação do modelo, mostrando-se abaixo dos níveis do percentual apresentado pelos extremamente pobres. O fato ainda mais relevante neste gráfico observa-se quanto ao percentual dos SD após o procedimento comportam-se também abaixo do mesmo patamar.

**Gráfico 9: Distribuição percentual do rendimento domiciliar *per capita* entre os indivíduos que possuem máquina de lavar, pós aplicação da Análise Discriminante – Brasil, 2013**



**Fonte:** IBGE, Pesquisa Nacional por Amostra de Domicílios (PNAD). Elaboração SAGI/MDS.

#### **4. Considerações finais**

Abordado o tema da modelagem como possível solução frente as demais técnicas atualmente utilizadas no tratamento de casos que consideram a problemática da subdeclaração e imputação de renda para os indivíduos captados pela PNAD, o presente estudo almeja ter apresentado uma nova direção na qual as técnicas estatísticas multivariadas possam trazer à luz da questão, uma proposta para se considerar os registros válidos de uma pesquisa complexa, sem se valer da utilização de procedimentos que utilizem de quaisquer metodologias de seleção determinística ou mesmo probabilística, na imputação de renda dos registros que aparentemente possam ser considerados como “erros sistêmicos”, fazendo assim um cômputo mais justo para a pobreza monetária avaliada sob a ótica unilateral declaração de rendimentos.

Assim, como os patamares da extrema pobreza monetária estão se diluindo cada vez mais ao longo dos anos, atingindo limites onde diferenças tendem mais e mais a serem observadas apenas no contexto do erro amostral das estimativas, o uso de novas técnicas deve ser estimulado e difundido entre os pesquisadores do tema, fomento novas maneiras de se

abordar um assunto que certamente toma proporções globais dia após dia, não se limitando ao escopo de abordagens apenas das pesquisas domiciliares por amostras, mas que tenhamos neste contexto a motivação para a avaliação também dos cálculos destas estimativas nos registros administrativos.

## 5. Bibliografia

HAIR, Jr., J.F.; ANDERSON, R.E.; TATHAM, R.L. BLACK, W.C. **Análise Multivariada de dados**. Tradução: Adonai Schlup Sant'Anna e Anselmo Chaves Neto. 5.ed. Porto Alegre: Bookman, 2005.

HARRELL, F.E. and LEE, K.L. A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality. In P. K. Sen (Ed.): **Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences**. North-Holland: Elsevier Science Publishers, 1985.

JOHNSON, R A. and WICHERN, D.W. **Applied Multivariate Statistical Analysis**. 4ed . Upper Saddle River: Prentice Hall, 1984.

MEDEIROS, M.; OSÓRIO, R. **Arranjos Domiciliares e Arranjos Nucleares no Brasil: Classificação e Evolução de 1977 a 1998**. Brasília, IPEA, 2001 (TEXTO PARA DISCUSSÃO N.788)

SILVA, Pedro Luis do Nascimento; PESSOA, Djalma Galvão Carneiro and LILA, Maurício Franca. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. **Ciênc. saúde coletiva [online]**. 2002, vol.7, n.4, pp. 659-670 .

SOARES, S.S.D. Metodologias para estabelecer a linha de pobreza: objetivas, subjetivas, relativas, multidimensionais. Brasília, 2009 (**Texto para Discussão** 1381).

VAZ, A.C.N.. **Metodologia de estimação de população em extrema pobreza: um estudo dos "Sem Declaração" e dos "Sem Rendimento" na PNAD**. Estudo Técnico SAGI n. 24/2012.

SOUZA, M.; JANNUZZI, P.M. **Questões metodológicas acerca do dimensionamento da extrema pobreza no Brasil nos anos 2000**. Estudo Técnico SAGI n. 08/2014.

SOUZA, M.; JANNUZZI, P.M. **Perfil estrutural da Extrema Pobreza no Brasil: análise exploratória da Pesquisa Nacional de Amostra de Domicílios (IBGE) sobre a parcela dos 5% mas pobres**. Estudo Técnico SAGI n. 10/2011.