

**INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA
ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS**

Alexandre Pinto de Carvalho

Decomposição do Diferencial de Salários no Brasil em 2003:
*Uma aplicação dos procedimentos de Oaxaca e Heckman em Pesquisas
Amostrais Complexas.*

Dissertação de Mestrado em Estudos Populacionais e Pesquisas Sociais,
Área de Concentração em Pesquisas Sociais e Amostragem, apresentada
à Coordenação do Mestrado da ENCE. Orientador: Professor Doutor
Marcelo Cortes Néri e a Co-orientadora: Professora Doutora Denise Britz
do Nascimento Silva.

RIO DE JANEIRO

12 de setembro de 2005.

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Dedico essa dissertação a Ortélia, Alfredo Pinto, Eunice Carvalho, Alcides Pinto, Maria Lucia, Vânia Pinto, Maira e a Roberta Guedes. A vocês, todo o meu carinho.

Alexandre Pinto de Carvalho

AGRADECIMENTOS

A Deus, por seu infinito amor e por ter me proporcionado a realização deste trabalho;

A Maria Lucia e Alcides Pinto, por não medirem esforços para me educar;

A meu grande amor, Ortélia, que sempre me incentivou;

A Roberta Guedes, Aline Visconti, Wagner Tassinari, Dorian Borges, Hugo e Márcio Lopes, amigos na vida acadêmica;

Aos professores Eduardo Pontual e Pedro Silva, pela atenção e as valiosas sugestões.

Aos meus orientadores Marcelo Néri e Denise Britz por todo incentivo e a orientação.

“Eu sou brasileiro e nunca desisto”.

RESUMO

Esta dissertação calcula a decomposição do diferencial entre as médias dos logaritmos dos salários/hora, com ênfase na cor e no sexo dos trabalhadores brasileiros. Foram utilizados os microdados da Pesquisa Nacional por Amostra de Domicílio (PNAD), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2003. Participaram da análise as pessoas que declararam a cor branca, preta ou parda.

A metodologia consiste em estimar a equação de salários (Mincer, 1974) com a correção do viés de seleção das informações dos salários (Heckman, 1979). Em seguida, a decomposição do diferencial da média do logaritmo do salário/hora foi obtida pelo procedimento de Oaxaca (1973) apresentada em dois efeitos: características produtivas e discriminação. A análise empírica tem como foco o uso adequado de procedimentos de modelagem estatística em pesquisas, por amostragem complexa, conforme os trabalhos de Skinner e Smith (1989) e Pessoa e Silva (1998).

Os resultados desta Dissertação indicam a necessidade de se incorporar o plano amostral e a correção do viés de seleção da informação dos salários, visando melhorar a qualidade das estimativas das equações de salários e avaliar adequadamente as medidas de discriminação. Como exemplo, a estimativa do coeficiente de discriminação, D , entre homens e mulheres de cor branca é 0,37 sem a correção do viés e 0,30 com a correção do viés de seleção das informações dos salários.

ABSTRACT

This dissertation calculate the decomposition of the differential enter the averages of the logarithms of the hourly wages, with emphasis in the color and the sex of the Brazilian workers. The microdata of the National Research for Sample of Domicile had been used (PNAD), carried through for the Brazilian Institute of Geography and Estatística (IBGE) in 2003. The people had participated of the analysis who had declared the white color, black or medium brown. The methodology consists of estimate the equation of wages (Mincer, 1974) with the correction of the bias of election of the information of the wages (Heckman, 1979). After that, the decomposition of the differential of the average of the logarithms of the hourly wages was gotten by the procedure of Oaxaca (1973) presented in two effect: productive characteristics and discrimination. The empirical analysis has as focus the adequate use of modeling procedures statistics in research, for complex sampling, as the works of Skinner and Smith (1989) and Pessoa e Silva (1998). The results of this Dissertação indicate the necessity of if incorporating the amostral plan and the correction of the bias of election of the information of the wages, aiming at to improve the quality of the estimates of the equations of wages and to evaluate the measures of discrimination adequately. As example, the estimate of the discrimination coefficient, D , between men and women of white color is 0,37 without the correction of bias and 0,30 with the correction of the bias of selection of the information of the wages.

SUMÁRIO

Lista de Figuras	8
Lista de Tabelas	9
Lista de Gráficos	10
1 Introdução	11
1.1 Motivação e Objetivos	11
1.2 Estrutura da Dissertação	12
2 Análise Não Controlada do Diferencial de Salários	14
2.1 A Utilização de Estatísticas Descritivas	14
2.2 Inserção e Participação no Mercado de Trabalho	14
2.3 Análise das Características Produtivas dos Trabalhadores	16
3 Aspectos Metodológicos da Pesquisa Nacional por Amostra de Domicílios – 2003	21
3.1 A Utilização da PNAD 2003 em Pesquisas Sociais	21
3.2 O Esquema Amostral da PNAD 2003	21
4 Análise de Dados Amostrais Complexos	27
4.1 A Inferência em Dados Amostrais Complexos	27
4.2 Efeitos do Plano Amostral	28
4.3 Estatística teste	30
4.4 Método de Máxima Pseudo Verossimilhança	31
5 Análise Controlada do Diferencial de Salários e Decomposição Segundo o Procedimento de Oaxaca.	36
5.1 Análise de Regressão e a Teoria do Capital Humano	36
5.2 Estudos sobre Determinantes do Diferencial de Salários no Brasil	37
5.3 A forma funcional da Equação de Salários	38
5.4 Análise Controlada do Diferencial de Salários	41
5.5 Coeficiente de Discriminação	43
5.6 Procedimentos para o Cálculo da Decomposição de Salários	46
5.7 Ajuste das Equações de Salários para os grupos de cor e sexo	48
5.8 Análise da Equação de Salários com interação	52
5.9 Resultados da Decomposição do Diferencial de Salários	55
6 Decomposição do Diferencial de Salário com Correção do Viés de Seleção do Salário na Análise de Regressão	59
6.1 O viés de seleção da informação dos salários	59
6.2 A Correção do Viés de Seletividade da Informação dos Salários através do Procedimento de Heckman em Pesquisas Amostrais Complexas	60
6.3 Ajuste da Equação de Salários com a Correção de Heckman	64
6.3.1 Equação de Participação	64
6.3.2 Ajuste das Equações de Salário com Correção do Viés de Seletividade da Informação.	
6.4 Decomposição dos Diferenciais de Salários segundo Oaxaca	
7 Considerações Finais e Trabalhos Futuros	74
8 Referências Bibliográficas	76
Anexo 1 – Capítulo 1	78
Anexo 2 – Capítulo 2	84
Anexo 3 – Capítulo 5	86
Anexo 4 – Capítulo 6	88

Lista de Figuras

Figura 3.1 – Plano Amostral Estratificado e Conglomerado em dois estágios	23
Figura 3.2 – Esquema de Amostragem Conglomerada em três estágios	24
Figura 3.3 – Esquema de Amostragem para Novas Construções	25

Lista de Tabelas

Tabela 2.1 – População Economicamente Ativa, Ocupada e Desocupada no Brasil em 2003 segundo cor e sexo	15
Tabela 2.2 – Taxas de Participação e Desemprego no Brasil em 2003 segundo cor e sexo	16
Tabela 2.3 – Estatísticas descritivas das características produtivas dos trabalhadores	17
Tabela 2.4 – Salário/hora segundo grupos de cor e sexo	18
Tabela 2.5 – Diferença da média do salário/hora ¹ segundo cor e sexo	19
Tabela 2.6 – Razão entre as médias do salário/hora segundo cor e sex	19
Tabela 5.1 – Estimação das Equações de Salários	42
Tabela 5.2 - Equações de salários estimadas com o peso e sem a especificação dos estratos. Estimativas dos parâmetros e respectivas estatísticas de teste (negrito)	49
Tabela 5.3 - Equações de Salários para os grupos de cor e sexo Estimativas dos parâmetros e respectivas estatísticas de teste (negrito)	50
Tabela 5.4 – Ajuste do Modelo de Regressão da Equação dos Salários com a interação sexo*cor e as variáveis explicativas anos de estudo, experiência e local de residência.	54
Tabela 5.5 – Diferença do ln(salário/hora) entre os grupos de cor	55
Tabela 5.6 – Diferença entre as médias das características produtivas e coeficientes estimados para o Grupo Base (L) e o I-ésimo grupo de cor/sexo	56
Tabela 5.6 - Efeitos da Discriminação Estimados pelas Características Pessoais	57
Tabela 6.1 – Equação de Participação no Mercado de Trabalho Brasileiro em 2003	66
Tabela 6.2 – Equação de Salários com Correção de Heckman	67
Tabela 6.3 – Diferenças entre as médias dos regressores e coeficientes estimados com a correção de Heckman.	70
Tabela 6.4 – Efeitos da Discriminação Estimados pelas Características Pessoas – Correção de Heckman	71

Lista de Gráficos

Gráfico 5.1 – Retornos à escolaridade sobre o $\ln(\text{salário/hora})$	51
Gráfico 5.2 – Retornos à experiência sobre o $\ln(\text{salário/hora})$	52
Gráfico 6.1 – Comparação entre os coeficientes da Equação de Salários pelo Método Tradicional e com a Correção de Heckman.	68
Gráfico 6.2 - Efeito da Discriminação – Método Tradicional vs Heckman	72
Gráfico 6.3 – Estimativa do Coeficiente de Discriminação – Método Tradicional vs Procedimento de Heckman	73

1 Introdução

1.1 Motivação e Objetivos

Em 2001, a III Conferência Mundial Contra o Racismo e Xenofobia realizada em Durban, África do Sul, debateu medidas para o combate da discriminação no mundo. O Brasil, dentre outros países, comprometeu-se à implementação de políticas no combate a disparidades no acesso à educação e ao mercado de trabalho. Este fato motivou a elaboração desta dissertação com o intuito de contribuir no estudo da discriminação¹ dos salários no mercado de trabalho brasileiro segundo cor e sexo a partir de informações obtidas em pesquisas domiciliares amostrais complexas.

Nos últimos anos, a estimação da equação de salários² tem sido amplamente utilizada em estudos relativos à discriminação no mercado de trabalho brasileiro segundo cor e sexo. Os resultados revelam os efeitos dos determinantes do salário segundo características individuais e do posto de trabalho. O propósito destes trabalhos é avaliar a situação em que indivíduos com atributos produtivos semelhantes, exceto pela cor ou sexo, têm os salários diferentes. Este conhecimento auxilia na formulação de políticas públicas em nosso mercado de trabalho. No entanto, alguns dos estudos que utilizam dados provenientes de pesquisas amostrais complexas, como a Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), não incorporam o plano amostral em suas análises. Rodrigues (2003, cap. 5) revela em seu estudo que o fato de não incorporar

¹ Hélio Santos (2001) define o racismo e o preconceito como modos de ver certas pessoas ou grupos de cor/raça, enquanto que a discriminação seria a manifestação concreta de um ou outro. O racismo parte do pressuposto da “superioridade de um grupo racial sobre outro”. O preconceito racial restringe-se à construção realizada como padrão que é próprio aquele que julga. Desta forma, quando o racista ou preconceituoso externaliza sua atitude, agora transformada em manifestação, ocorre a discriminação”.

² Modelagem estatística através da Análise de Regressão.

os pesos e o plano amostral na análise gera estimativas incorretas tanto para os coeficientes como para as suas variâncias.

O objetivo principal desta dissertação é decompor o diferencial da média do logaritmo do salário/hora com ênfase na cor e sexo dos trabalhadores brasileiros (homens de cor branca, mulheres de cor branca, homens de cor preta ou parda e mulheres de cor preta ou parda). A metodologia, baseada no modelo de capital humano, consiste em estimar a equação de salários (Mincer, 1974) com a correção do *viés de seleção das informações* dos salários através do procedimento de Heckman (1979). O procedimento da decomposição do diferencial de salários é baseado no trabalho de Oaxaca (1973) composto por dois efeitos: características produtivas individuais e da discriminação. A análise empírica tem como foco o uso adequado de procedimentos de modelagem estatística em pesquisas por amostragem complexa, conforme os trabalhos de Skinner e Smith (1989) e Pessoa e Silva (1998).

1.2 Estrutura da Dissertação

A presente dissertação encontra-se organizada em mais seis capítulos, além desta introdução e as considerações finais. No capítulo 2, apresenta-se uma análise não-controlada do diferencial da média dos salários entre os homens de cor branca, homens de cor preta ou parda, mulheres de cor branca e mulheres de cor preta ou parda. Avalia-se a inserção e participação no mercado de trabalho e, adicionalmente, caracterizam-se os grupos de cor e sexo segundo estatísticas descritivas da escolaridade, idade, experiência de trabalho e local da residência.

O capítulo 3 destaca os aspectos metodológicos da principal fonte de dados utilizada nesta dissertação, a Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2003. Este

capítulo discute a importância do esquema amostral e o tratamento dos microdados para construção das variáveis de estratificação. Em seguida apresenta as expressões dos estimadores e respectivas variâncias dos indicadores utilizados nesta dissertação.

O capítulo 4 elucida a importância do uso adequado dos dados na inferência estatística com base em pesquisas amostrais complexas como a PNAD 2003. Basicamente, é um resumo dos trabalhos de Skinner, Holt e Smith (1989), Pessoa e Silva (1998), Phillippe (2001) e Rodrigues (2003).

No capítulo 5 é calculada a decomposição do diferencial da média do logaritmo do salário/hora através da modelagem estatística. O procedimento consiste em estimar a equação de salários para cada grupo e trabalhador com o uso da análise de regressão (*equação de salários* ou *equação minceriana*), através do método de Máxima Pseudo-Verossimilhança (MPV), incorporando o plano amostral. Em seguida, com base no trabalho de Oaxaca (1973), o diferencial é decomposto em dois efeitos: das características individuais e da discriminação.

O capítulo 6 destaca que para determinadas pessoas seria vantajoso trabalhar se o salário recebido (ou salário potencial) fosse maior que o custo de oportunidade de não trabalhar (ou salário reserva). Desta forma, existe um *viés de seleção das informações do salário*. Neste caso, as estimativas dos coeficientes da equação de salários obtidas a partir das informações dos indivíduos que trabalham na data de referência da pesquisa podem estar viesadas sob a ótica do *viés de seleção das informações* considerando-se que o modelo utilizado no capítulo 5 não incorpora a informação sobre a avaliação dos indivíduos que não trabalham no que se refere ao custo de oportunidade. Heckman (1979) apresentou este fenômeno e sua solução não apenas para pesquisas amostrais. De modo geral, o viés de seleção de informação pode ser decorrente de duas razões, como destaca Heckman: ou em virtude de seletividade das informações dos indivíduos ou devido ao desenho amostral da pesquisa. Cabe

ressaltar que, ao utilizar os dados da PNAD, o efeito que deseja-se incorporar na modelagem visa corrigir *o viés de seletividade da informação dos salários* para os indivíduos que, apesar de estarem devidamente representados na amostra da PNAD 2003, não trabalharam na data de referência da pesquisa em virtude da avaliação do salário potencial e do custo de oportunidade envolvido nesta escolha.

O capítulo 7, por sua vez, apresenta as considerações finais e as propostas de trabalhos futuros. Uma extensão natural desta dissertação seria o cálculo da decomposição para os percentis do logaritmo do salário/hora através do uso de regressões quantílicas com a correção de Heckman e a incorporação do plano amostral nas análises.

2 Análise Não Controlada do Diferencial de Salários

2.1 A Utilização de Estatísticas Descritivas

A análise não controlada do diferencial de salários é obtida através de estatísticas descritivas do rendimento do trabalho principal (salário) e indicadores do mercado de trabalho dos grupos de cor e sexo (cor/sexo). Estes resultados são obtidos sem o uso de modelagem estatística.

Considera-se quatro grupos no mercado de trabalho: mulheres de cor branca, mulheres de cor preta ou parda, homens de cor branca e homens de cor preta ou parda. A fonte de dados utilizada é a Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada pelo IBGE em 2003 (maiores detalhes no capítulo 3).

O capítulo está dividido em mais três seções. A seção 2.2 apresenta os indicadores de mercado de trabalho para avaliar a inserção e participação dos trabalhadores. Na seção 2.3, traça-se um perfil dos grupos através da escolaridade, experiência de trabalho, jornada de trabalho e local da residência (mora em área urbana ou rural). Em seguida, avalia-se o diferencial não controlado pela variável salário/hora (i.e., a razão entre o salário recebido no mês de referência e a jornada de trabalho no mesmo mês).

2.2 Inserção e Participação no Mercado de Trabalho

Oliveira, Porcaro e Costa (1981) no trabalho *“O lugar do negro na força de trabalho”* revelam que a análise da inserção segundo cor na força de trabalho a partir de categorias sócio-ocupacionais, rendimento e escolaridade são fundamentais para divisão social do trabalho no Brasil. De fato, ao longo dos últimos anos a participação

das mulheres no mercado cresceu. Como exemplo, 260% entre 1970 e 1990 contra 73% dos homens (Araújo e Ribeiro, 2003).

Em 2003, a População Economicamente Ativa³ (PEA) no Brasil era em torno de 142 milhões de brasileiros⁴ dos quais 79 milhões tinham trabalho, 8 milhões procuravam trabalho e 55 milhões estavam fora da força de trabalho (tabela 2.1). A população ocupada era composta por aproximadamente 42 milhões de homens e mulheres que declararam sua cor branca e 37 milhões de homens e mulheres de cor preta ou parda.

Tabela 2.1 – População Economicamente Ativa, Ocupada e Desocupada no Brasil em 2003 segundo cor e sexo

População	Homens		Mulheres	
	Branca	Preta ou Parda	Branca	Preta ou Parda
Pessoas com 10 anos ou mais de idade	35.901.231	33.032.405	40.165.419	33.569.730
População Ocupada	24.283.013	22.034.281	18.264.154	14.515.279
População Desocupada	1.848.167	2.069.910	2.274.884	2.326.757

Fonte: Elaboração própria a partir dos microdados da PNAD/IBGE 2003

Na tabela 2.2, verifica-se que, embora a População Economicamente Ativa das mulheres seja superior a dos homens, a sua taxa de participação⁵ é inferior. A

³ Pessoas com 10 anos ou mais de idade.

⁴ Exceto as pessoas que declararam a cor ignorada ou indígena em 2003.

⁵ Taxa de participação = $\left(\frac{\text{Ocupados} + \text{Desocupados}}{\text{Pessoas com 10 anos ou mais}} \right) \times 100$

inserção no mercado de trabalho, representada pela taxa de desocupação⁶, revela que existe um efeito entre as variáveis cor e sexo. Destaca-se a situação das mulheres de cor preta ou parda que apresentam a menor taxa de participação (47,91%) e a maior taxa de desocupação (14,47%).

Tabela 2.2 – Taxas de Participação e Desocupação no Brasil em 2003 segundo cor e sexo

Indicadores	Homens		Mulheres	
	Branca	Preta ou Parda	Branca	Preta ou Parda
Taxa de Participação	68,87%	68,56%	49,41%	47,91%
Taxa de Desocupação	7,47%	9,14%	11,46%	14,47%

Fonte: Elaboração própria a partir dos microdados da PNAD/IBGE 2003

2.3 Análise das Características Produtivas dos Trabalhadores

As estatísticas descritivas para escolaridade, experiência de trabalho, idade, local da residência (área urbana ou rural), salário efetivamente recebido no mês, jornada de trabalho semanal são utilizadas para caracterizar os grupos de trabalhadores segundo cor e sexo. Em seguida, define-se a média da razão entre o salário e a jornada de trabalho mensal para o cálculo do diferencial não-controlado entre os trabalhadores.

Os resultados da tabela 2.3 revelam que existe uma disparidade entre os salários dos trabalhadores brasileiros segundo cor e sexo. Como exemplo, embora as mulheres de cor branca tenham um nível de escolaridade superior ao dos homens de cor branca

⁶ Taxa de desemprego = $\left(\frac{\text{Desocupados}}{\text{Ocupados} + \text{Desocupados}} \right) \times 100$

o seu salário é inferior, R\$657 versus R\$1.010. O mesmo resultado é encontrado entre homens e mulheres de cor preta ou parda. Além disso, os salários das pessoas de cor branca são superiores ao de pessoas de cor preta ou parda em nosso mercado de trabalho. Nota-se que a jornada de trabalho dos homens e das mulheres é diferente e, por isso, faz-se necessário calcular o diferencial através de uma medida padronizada denominada salário/hora. O salário/hora é a razão entre o salário recebido no mês multiplicado pela jornada de trabalho mensal (i.e., jornada de trabalho semanal multiplicada 4.2 semanas).

Tabela 2.3 – Estatísticas descritivas das características produtivas dos trabalhadores

Indicadores	Homens		Mulheres	
	Branca	Preta ou Parda	Branca	Preta ou Parda
Salário ¹	1.010,02	478,38	656,76	348,58
Jornada de trabalho semanal	45,40	44,55	38,10	37,31
Escolaridade ²	7,97	5,65	9,29	7,17
Experiência ³	23,54	24,25	20,71	22,23
Idade ⁴	37,51	35,90	36,00	35,40

Fonte: Elaboração própria a partir dos microdados da PNAD/IBGE 2003

Notas: 1 – rendimento do trabalho principal recebido no mês (R\$), 2 – escolaridade mensurada em anos de estudo, 3 – experiência dos anos de trabalho, 4 – idade em anos.

Na tabela 2.4, o indicador salário/anos de estudo destaca o quanto cada grupo de trabalhador ganha por seu investimento em escolaridade. Os homens de cor branca

são os que apresentam os maiores salários por horas trabalhadas (R\$6,09) e por anos de estudo (R\$122). No extremo, encontram-se as mulheres de cor preta ou parda que ganham R\$4,75 por hora trabalhada.

Tabela 2.4 – Salário/hora segundo grupos de cor e sexo

Indicadores	Homens		Mulheres	
	Branca	Preta ou Parda	Branca	Preta ou Parda
Salário/Hora ¹	6,09	2,83	4,75	2,55
ln(Salário/Hora) ²	1,17	0,58	1,02	0,48
Salário/Escolaridade ³	122,38	77,17	67,49	46,90

Fonte: Elaboração própria a partir dos microdados da PNAD/IBGE 2003.

Notas: 1 – razão entre o rendimento do trabalho principal (R\$) e a jornada de trabalho mensal (horas), 2 – logaritmo do salário/hora, 3 – razão entre o rendimento do trabalho principal e anos de estudo.

O diferencial não-controlado é apresentado na tabela 2.5 através da comparação entre homens de cor preta/parda, mulheres de cor branca e mulheres de cor preta/parda com o grupo dos homens de cor branca – definido como grupo base. O propósito desta análise é avaliar se a diferença entre o salário/hora é estatisticamente diferente de zero – o teste é realizado pela estatística t. Como exemplo, os homens de cor branca ganham R\$ 3,26 a mais que os homens de cor preta. O menor diferencial é registrado entre homens e mulheres de cor branca (R\$1,34). Vale ressaltar que todos os resultados são estatisticamente diferentes de zero.

Tabela 2.5 – Diferença da média do salário/hora¹ segundo cor e sexo

Diferença entre o salário/hora	Estimativa	Estatística t
Homens de cor branca - Homens de cor preta	3,26	26,50
Homens de cor branca - Mulheres de cor branca	1,34	8,52
Homens de cor branca - Mulheres de cor preta ou parda	3,54	28,70

Fonte: Elaboração própria a partir dos microdados da PNAD/IBGE 2003.

Notas: 1 – razão entre o rendimento do trabalho (R\$) e a jornada do trabalho (horas).

Os resultados da tabela 2.6 revelam, em termos percentuais, o salário do grupo base versus os demais grupos de cor/sexo. Os homens de cor branca ganham, em média, o dobro dos homens e mulheres de cor preta/parda e 28% a mais que as mulheres de cor branca.

Tabela 2.6 – Razão entre as médias do salário/hora segundo cor e sexo

Razão entre o salário/hora	Estimativa	Estatística t
Homens de cor branca / Homens de cor preta	2,15	44,25
Homens de cor branca / Mulheres de cor branca	1,28	32,42
Homens de cor branca / Mulheres de cor preta ou parda	2,38	42,78

Os resultados da análise não-controlada do diferencial de salários revelam as disparidades segundo cor e sexo. Em geral, as pessoas de cor preta/parda ganham menos que o grupo base. No entanto, tais comparações não são suficientes para avaliar o quanto deste diferencial é explicado pela discriminação ou pelas características individuais (escolaridade, experiência de trabalho, local de residência e idade). Os próximos capítulos têm por objetivo testar a existência da discriminação através da modelagem estatística. Além disso, decompor este diferencial na parcela referente à discriminação ou às características individuais.

3 Aspectos Metodológicos da Pesquisa Nacional por Amostra de Domicílios – 2003

3.1 A Utilização da PNAD 2003 em Pesquisas Sociais

O sistema de pesquisas domiciliares no Brasil, por amostragem, foi introduzido a partir de 1967 com a criação da PNAD (IBGE, 2003). As informações⁷ coletadas têm por finalidade subsidiar a apreciação do grau do desenvolvimento socioeconômico do nosso país, em suas próprias dimensões. A PNAD subsidia a produção de estudos que permitem balizar o desenho de políticas públicas, por exemplo, nortear a universalização da educação básica, o combate à discriminação no mercado de trabalho, ou ainda, a redução da pobreza e desigualdade de renda. A abrangência geográfica nacional da PNAD é um componente importante para sua utilização. Em 2003, foram coletadas informações de domicílios em todo país, com exceção das áreas rurais da federação da Região Norte⁸. Este capítulo tem por finalidade descrever o esquema amostral da PNAD e auxiliar na utilização dos microdados em pacotes estatísticos como SAS e STATA (2003).

3.2 O Esquema Amostral da PNAD 2003⁹.

O esquema de seleção da amostra da PNAD, conforme descreve Silva, Pessoa e Lilá (2002), é estratificado e conglomerado com um, dois ou três estágios de seleção, de acordo com o tipo de estrato. O “primeiro tipo de estrato” é composto pelas seguintes unidades da federação: Acre, Alagoas Amapá, Amazonas, Distrito Federal,

⁷ Escolaridade, cor, sexo, trabalho, rendimento e habitação, fazem parte do corpo fixo do questionário do PNAD. Outros temas como saúde, trabalho infantil ou mobilidade social são investigados de acordo com a necessidade percebida em suplementos especiais da pesquisa.

⁸ Rondônia, Acre, Amazonas, Roraima, Pará e Amapá.

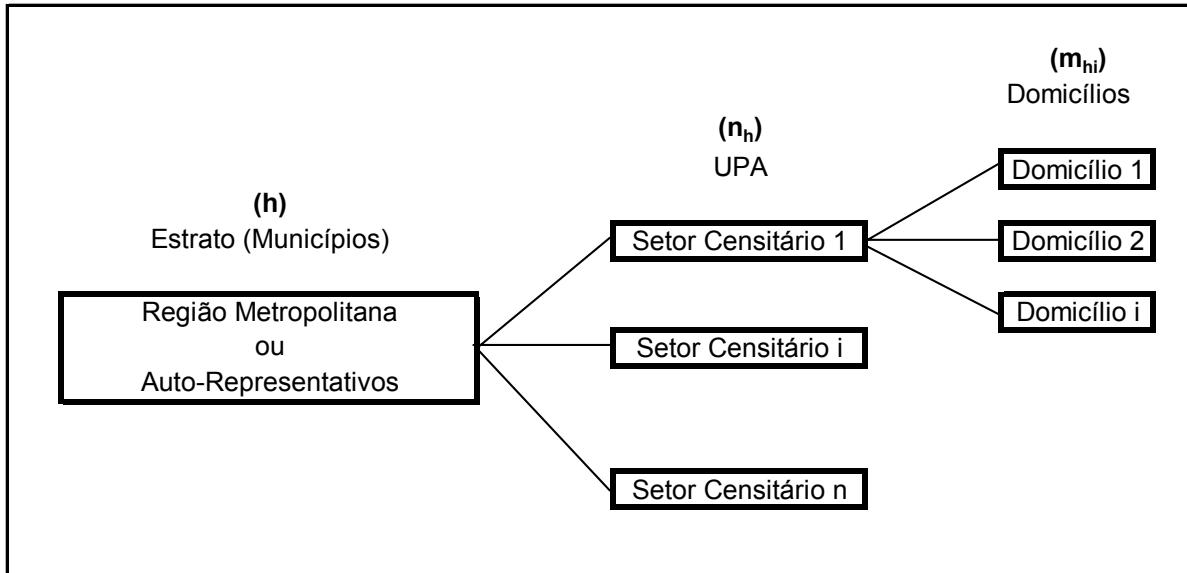
Espírito Santo, Goiás, Maranhão, Mato Grosso, Mato Grosso do Sul, Paraíba, Piauí, Rio Grande do Norte, Rondônia, Roraima, Santa Catarina, Sergipe e Tocantins. Enquanto que, o “segundo tipo de estrato” é formado pelas unidades da federação: Bahia, Ceará, Minas Gerais, Pará, Paraná, Pernambuco, Rio de Janeiro, Rio Grande do Sul e São Paulo.

As 18 unidades da federação, do primeiro tipo, formam 18 estratos independentes. No “segundo tipo de estrato”, cada unidade da federação forma dois tipos de estratos naturais: um com todos os municípios da Região Metropolitana na capital, e outro com os demais municípios. *Com isso, a PNAD é composta por 36 estratos “naturais”.*

A figura 3.1 ilustra o esquema de seleção da PNAD para os nove estratos naturais formados pelas regiões metropolitanas (segundo tipo de estrato). As Unidades Primárias de Amostragem (UPAs) são os setores censitários e as Unidades Secundárias de Amostragem (USAs) são os domicílios. A seleção das UPAs foi realizada através de uma amostragem sistemática com probabilidades proporcionais ao tamanho (PPT)¹⁰. Antes de efetuar a seleção dos setores em cada estrato (município), os setores foram ordenados segundo a situação (urbano ou rural) e o código geográfico, conferindo um efeito de estratificação implícita por situação devido ao uso do sorteio sistemático.

Figura 3.1 – Plano Amostral Estratificado e Conglomerado em dois estágios

¹⁰ Unidade Primária de Amostragem.

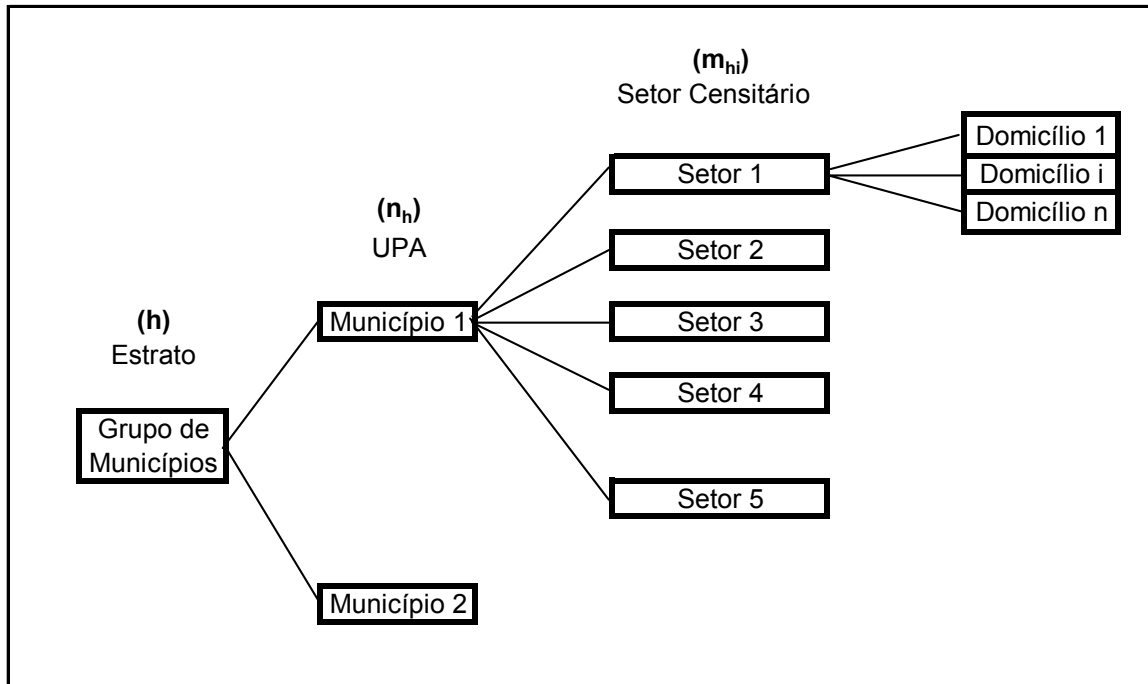


Para os 27 estratos naturais restantes, compostos pelos municípios que não são situados em regiões metropolitanas ou ficam nas unidades da federação sem região metropolitana, o plano amostral da PNAD é conglomerado em três estágios. As UPAs são os municípios, as USAs são os setores censitários e as unidades terciárias de amostragem são os domicílios. No entanto, alguns municípios considerados grandes em termos populacionais foram incluídos na amostra com certeza. Tais municípios são chamados de auto-representativos e seu plano amostral é igual ao utilizado nos municípios das regiões metropolitanas, isto é, conglomerado em dois estágios conforme a figura 3.1.

Por fim, os municípios que não pertencem à região metropolitana e não são auto-representativos são chamados de *não auto-representativos*. Tais municípios foram estratificados por tamanho e proximidade geográfica, buscando formar estratos com população total aproximadamente igual, conforme os dados do censo demográfico 2000.

No segundo estágio de seleção, o sorteio de setores foi feito dentro de cada município contido na amostra do primeiro estágio, usando o mesmo método já descrito para a seleção de setores nos estratos de regiões metropolitanas (figura 3.2).

Figura 3.2 – Esquema de Amostragem Conglomerada em três estágios

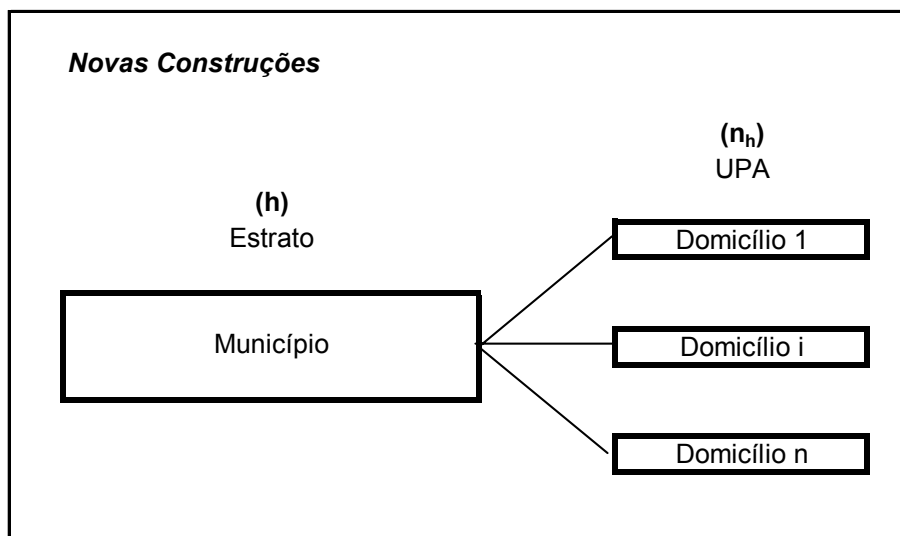


Antes da seleção da amostra (amostragem de domicílios), o IBGE realiza uma Operação de Listagem dentro de cada setor selecionado. Com isso, o cadastro é atualizado para seleção de domicílios em cada setor, permitindo assim localizar, identificar e quantificar as unidades domiciliares ali existentes no ano de realização da pesquisa.

Além disso, a amostra básica da PNAD é complementada com unidades domiciliares selecionadas do Cadastro de Projetos de Novas Construções. Este cadastro inclui projetos habitacionais com mais de 30 domicílios que surgiram após o censo realizado na década. O universo das Novas Construções é estratificado por municípios, e nesses estratos o plano amostral é conglomerado em apenas um estágio,

conforme a figura 3.3, as UPAs são os domicílios, cujo sorteio é efetuado mediante amostragem sistemática simples.

Figura 3.3 – Esquema de Amostragem para Novas Construções



Nos microdados da PNAD 2003, disponibilizados pelo IBGE, existem estratos (variável STRAT no dicionário de dados) que contêm apenas uma unidade primária de amostragem (variável PSU no dicionário de dados). Este fato impossibilita o cálculo da variância das estimativas para as análises exploratórias e modelos estatísticos.

A solução para este problema foi apresentada pela Coordenação de Trabalho e Rendimento (COREN/IBGE), responsável pela pesquisa, que identificou tais informações como Novas Construções (NCs). O procedimento (Anexo 1) utilizado para correção, que permite o cálculo da variância das estimativas, foi agregar os dados de duas formas: a primeira agrupa os estratos com apenas uma PSU em um único estrato na mesma unidade da federação; a segunda diz respeito aos casos de ocorrer somente um estrato de NC na UF. Neste caso, o procedimento foi agregá-lo a um outro estrato do mesmo município (identificado pela variável UPA nos microdados).

Por fim, através da descrição do plano amostral da PNAD e o conhecimento para o tratamento dos microdados, o próximo capítulo apresenta a importância da utilização de métodos adequados para realização de inferência estatística a partir de dados amostrais complexos, as estatísticas de teste, as variâncias e o Método de Máxima Pseudo-Verossimilhança.

4 Análise de Dados Amostrais Complexos

4.1 A Inferência em Dados Amostrais Complexos

O uso de dados de amostrais complexos, conforme destacado no capítulo anterior, envolve probabilidades distintas de seleção das unidades, conglomeração das unidades e estratificação. Para maiores detalhes ver os trabalhos de Skinner, Holt e Smith (1989) e Pessoa e Silva (1998).

A utilização de métodos adequados para realização de inferência em dados amostrais complexos permite estimar valores de uma variável de interesse e avaliar o grau de precisão das estimativas (através de suas variâncias). As estimativas das variâncias, por sua vez, são influenciadas pelo plano amostral utilizado. Com isso, é importante ressaltar a importância da incorporação do plano nos procedimentos de inferência com base em dados amostrais complexos como a PNAD. Já existem programas estatísticos que suportam tais análises, como exemplo, o SAS, STATA (2003).

A justificativa para a incorporação do desenho amostra nas inferências analíticas, partindo de dados amostrais complexos (Côrrea, 2001), é que os pesos podem ser usados para proteger *contra planos amostrais não-ignoráveis* (Pessoa e Silva, 1998), *que poderiam introduzir ou causar vícios, e má especificação do modelo*.

Este capítulo tem por objetivo apresentar o Efeito do Plano Amostral em pesquisas de dados amostrais complexos como a PNAD. Além disso, apresentar o procedimento para o cálculo dos intervalos de confiança, os testes de hipóteses e o Método de Máxima Pseudo-Verossimilhança (MPV) aplicado na inferência analítica (modelagem).

4.2 Efeito do Plano Amostral

O Efeito do Plano Amostral (EPA¹¹) tem por finalidade avaliar o impacto em desconsiderar o esquema de seleção da amostra no cálculo das estimativas. Esta medida foi proposta inicialmente por Kish (1965) e aperfeiçoada por Kish e Frankel (1974), para maiores detalhes ver Pessoa e Silva (1998). Na inferência estatística, para um parâmetro θ , o EPA é obtido pela razão entre a variância do plano amostral complexo (verdadeiro) e a variância da distribuição do estimador $\hat{\theta}$ de θ induzida pelo plano de amostragem aleatória simples (AAS¹²) - $V_{AAS}(\hat{\theta})$.

$$EPA(\hat{\theta}) = \frac{V_p(\hat{\theta})}{V_{AAS}(\hat{\theta})} \quad (4.1)$$

onde, $V_p(\hat{\theta})$ = variância da distribuição de $\hat{\theta}$ induzida pelo plano amostral complexo.

Skinner, Holt e Smith (1989, p.24) destacam que esta medida é importante para avaliar a eficiência quando comparamos desenhos alternativos na concepção das pesquisas. Além disso, o uso do EPA apresenta dificuldades no seu cálculo em inferências analíticas (modelagem) e, por isso, definiram o conceito do EPA ampliado (misspecification effect – meff). Esta medida mensura a tendência de um estimador usual (consistente), calculado sob hipótese de observações independentes e identicamente distribuídas (IID), subestimar ou superestimar a variância verdadeira do estimador pontual. O EPA ampliado (também denominado por meff - misspecification effect) é a razão entre a variância do estimador sob o plano amostral ou modelo correto

¹¹ É apresentado nos softwares estatísticos como design effect (deff).

¹² As informações são coletadas de forma independente e são identicamente distribuídas – IID

$V_{\text{VERD}}(\hat{\theta})$ sobre a esperança do estimador da variância de $\hat{\theta}$ sob a hipótese de observações IID da variância $E_{\text{VERD}}(v_0)$.

Dado que $v_0 = \hat{V}_{\text{IID}}(\hat{\theta})$ um estimador da variância de $\hat{\theta}$ para uma Amostra Aleatória Simples sem reposição. Então:

$$\text{meff} = \text{EPA}_{\text{ampliado}}(\hat{\theta}, v_0) = \frac{V_{\text{VERD}}(\hat{\theta})}{E_{\text{VERD}}(v_0)} = \frac{V_{\text{VERD}}(\hat{\theta})}{E_{\text{VERD}}(\hat{V}_{\text{IID}}(\hat{\theta}))} \quad (4.2)$$

O $\text{EPA}_{\text{ampliado}}(\hat{\theta}, v_0)$ mensura a tendência de v_0 subestimar ou superestimar $V_{\text{VERD}}(\hat{\theta})$, variância verdadeira sob o modelo e/ou plano amostral de $\hat{\theta}$. Quanto mais afastado de 1 for o valor de $\text{EPA}_{\text{ampliado}}(\hat{\theta}, v_0)$, mais incorreta será considerada a especificação do plano amostral ou do modelo nos procedimentos analíticos.

Desta forma, enquanto a medida proposta por Kish baseia-se nas distribuições induzidas pela aleatorização dos planos amostrais comparados, o $\text{EPA}_{\text{ampliado}}(\hat{\theta}, v_0)$ pode ser calculado com respeito a distribuições de aleatorização ou do modelo envolvido.

Em geral, são esperadas as seguintes consequências sobre o Efeito do Plano Amostral ao ignorar o plano amostral efetivamente adotado e admitir que o desenho da amostra foi AAS:

1. Ignorar os pesos em v_0 pode inflacionar o meff (ou EPA ampliado);
2. Ignorar conglomerações em v_0 pode inflacionar o meff;
3. Ignorar estratificação em v_0 pode reduzir o meff.

4.3 Estatística teste

Para uma população finita com um parâmetro de interesse θ e sua estimativa pontual $\hat{\theta}$ o intervalo de confiança com nível de confiança $(1 - \alpha)$ a partir da distribuição

assintótica de $t_0 = \frac{\hat{\theta} - \theta}{\sqrt{v_0^{0,5}}}$, sob a hipótese de que as observações são Independente e

Identicamente distribuídas (IID) com distribuição $N(0;1)$, é dado por:

$\left[\hat{\theta} - z_{\alpha/2} \sqrt{v_0^{0,5}}; \hat{\theta} + z_{\alpha/2} \sqrt{v_0^{0,5}} \right]$ onde $z_{\alpha/2} = \int_{-\infty}^{+\infty} \varphi(t) dt$ e φ é uma função de densidade da

distribuição normal padrão.

Para um plano amostral complexo a estatística de teste é dada por:

$$t_0 = \frac{\hat{\theta} - \theta}{[\hat{V}_{\text{VERD}}(\hat{\theta})]^{1/2}} \text{ tal que } t_0 \sim N[0, \text{EPA}(\hat{\theta}, v_0)] \text{ e } \text{EPA} = \frac{\hat{V}_{\text{verd}}(\hat{\theta})}{v_0}$$

Desta forma, ao ignorar os pesos e o efeito de conglomeração do desenho amostral pode-se inflacionar o EPA, ampliando-se os intervalos de confiança para os parâmetros de interesse.

4.4 Método de Máxima Pseudo Verossimilhança

Esta seção é um resumo dos trabalhos apresentados por Phillippe (2001) e Rodrigues (2003). A incorporação do plano amostral na inferência analítica consolidada por Skinner, Holt e Smith (1989), propõem um método para a estimação dos

parâmetros do modelo denominado Método de Pseudo Máxima Verossimilhança e será detalhado adiante.

Seja i o índice do elemento de uma população, o qual contém as informações dos estratos, das Unidades Primárias de Amostragem (UPA) no estrato e dos elementos dentro das UPA's. Dada as variáveis (Y_i, X_i) provenientes da população U , onde Y_i é a variável de interesse (ou resposta) e X_i é um vetor de características (variáveis explicativas) associadas a cada i , tal que $i \in U$. Assumindo que Y_1, \dots, Y_N são IID com função de densidade $f(y_i, \underline{\beta})$, onde $\underline{\beta}$ é um vetor de parâmetros desconhecidos de interesse. Com isso, a verossimilhança e o logaritmo da verossimilhança populacionais são descritos por:

$$L(\underline{\beta}; y_i, X_i) = \prod_{i \in U} f(y_i, \underline{\beta}) \quad (4.6)$$

$$l(\underline{\beta}; y_i, X_i) = \sum_{i \in U} \log[f(y_i, \underline{\beta})] \quad (4.7)$$

Sendo $l(\underline{\beta}; y_i, X_i)$ o logaritmo da função de verossimilhança associado ao modelo, onde $\underline{\beta}$ é o vetor de parâmetros com dimensão $p \times 1$, X_i é um vetor de dimensão $1 \times p$, definido para todo $i \in U$. Então para uma população finita, os parâmetros $\underline{\beta}$ são gerados através da solução de um sistemas de equações definido por:

$$G(\underline{\beta}) = \sum_{i \in U} u_i(\underline{\beta}; y_i, X_i) \quad (4.8)$$

onde $u_i = \frac{\partial l(\underline{\beta}; y_i, X_i)}{\partial \underline{\beta}}$ é o vetor $p \times 1$ dos escores do elemento i , $i \in U$

A solução deste sistema, quando $G(\underline{\beta}) = 0$, para todo $i \in U$ é o estimador de máxima verossimilhança para $\underline{\beta}$ no caso de um censo.

Como $G(\underline{\beta}) = \sum_{i \in U} u_i(\underline{\beta}; y_i, \underline{x}_i)$ é a soma dos escores, que por sua vez é um vetor de totais, e para estimá-lo ($i \in s$) - amostra - pode-se utilizar um estimador linear ponderado da forma $\hat{G}(\underline{\beta}) = \sum_{i \in s} w_i u_i(\underline{\beta}; y_i, \underline{x}_i)$, onde w_i são os pesos¹³ propriamente definidos. O estimador de Máxima Pseudo-Verossimilhança de $\underline{\hat{\beta}}$ é a solução da equação:

$$\hat{G}(\underline{\beta}) = \sum_{i \in s} w_i u_i(\underline{\beta}; y_i, \underline{x}_i) = 0 \quad (4.9)$$

Binder (1983), Pessoa e Silva (1998) apresentam a matriz de primeira ordem da expansão da série de Taylor para o estimador de pseudo máxima-verossimilhança. Com isto, encontra-se um estimador para a variância assintótica, sob o plano amostral, de $\underline{\hat{\beta}}$, que é dado por:

$$\hat{V}(\underline{\hat{\beta}}) = \left[\left. \frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \right|_{\underline{\beta} = \underline{\hat{\beta}}} \right]^{-1} \hat{V} \left[\sum_{i \in s} w_i u_i(\underline{\hat{\beta}}; y_i, \underline{x}_i) \right] \left[\left. \frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \right|_{\underline{\beta} = \underline{\hat{\beta}}} \right]^{-1} \quad (4.10)$$

onde $\hat{V} \left[\sum_{i \in s} w_i u_i(\underline{\hat{\beta}}; y_i, \underline{x}_i) \right]$ é um estimador consistente para a variância do estimador do total populacional dos escores.

Binder (1983) mostrou que a distribuição assintótica de $\hat{\underline{\beta}}$ é normal multivariada, fornecendo uma base para inferência sobre $\underline{\beta}$ sob amostras grandes, tal que:

$$\hat{V}(\hat{\underline{\beta}})^{1/2}(\hat{\underline{\beta}} - \underline{\beta}) \sim NM(0;I) \quad (4.11)$$

Os pesos w_i devem ser tais que satisfaçam as condições de que os estimadores sejam assintoticamente normais e não viciados, além de possuírem estimadores de variância consistentes. Estas condições são satisfeitas quando a probabilidade de inclusão na amostra da *i-ésima* unidade amostral da população, $i=1,2,..,N$, é maior do que zero ($\pi_i > 0$) e, simultaneamente, que a probabilidade de inclusão conjunta da *i-ésima* e *j-ésima* unidades amostrais da população, $i \neq j$, seja maior do que zero ($\pi_{ij} > 0$). Os estimadores de máxima pseudo verossimilhança não são únicos, pois existem diversas maneiras de definir os pesos w_i .

Usualmente utilizamos os pesos do estimador de Horwitz-Thompson para totais, dado por:

$$w_i = \frac{1}{\pi_i}, \forall i \in s \quad (4.12)$$

Substituindo (4.12) em (4.9) encontramos o estimador pontual $\hat{\underline{\beta}}_{\pi}$ do parâmetro $\underline{\beta}$ no modelo, e substituindo (4) e $\hat{\underline{\beta}}_{\pi}$ em (4.10) temos que:

¹³ Peso é igual ao inverso da probabilidade de seleção do indivíduo em uma amostra.

$$\hat{V}(\hat{\underline{\beta}}_{\pi}) = \left[\frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\underline{\beta} = \hat{\underline{\beta}}_{\pi}} \right]^{-1} \hat{V} \left[\sum_{i \in S} \frac{1}{\pi_i} u_i(\hat{\underline{\beta}}_{\pi}; y_i; \underline{x}_i) \right] \left[\frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\underline{\beta} = \hat{\underline{\beta}}_{\pi}} \right]^{-1} \quad (4.13)$$

$$\hat{V}(\hat{\underline{\beta}}) = \left\{ \left[\frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\underline{\beta} = \hat{\underline{\beta}}_{\pi}} \right]^{-1} \sum_{i \in S} \sum_{j \in S} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} [u_i(\hat{\underline{\beta}}_{\pi}, y_i, \underline{x}_i)] [u_j(\hat{\underline{\beta}}_{\pi}, y_j, \underline{x}_j)]' \left[\frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\underline{\beta} = \hat{\underline{\beta}}_{\pi}} \right]^{-1} \right\} \quad (4.14)$$

A equação (4.14) fornece o estimador da variância do estimador de pseudo máxima verossimilhança $\hat{\underline{\beta}}_{\pi}$ de $\underline{\beta}_{\pi}$.

5 Análise Controlada do Diferencial de Salários e Decomposição Segundo o Procedimento de Oaxaca.

5.1 Análise de Regressão e a Teoria do Capital Humano

O capital humano é o conjunto das habilidades do indivíduo ligadas à capacidade produtiva, e incorporadas no conhecimento e qualificação para determinadas atividades (Becker (1993)). Até 1950 os economistas geralmente assumiam que o poder do salário como dados e não adquiridos. As análises sofisticadas do investimento em educação e treinamento por Adam Smith, Alfred Marshall e Milton Friedman não incluíam em suas discussões a produtividade. Então Theodore W. Schultz, entre outros, iniciaram uma exploração pioneira nas implicações do investimento do capital humano nas questões econômicas.

Este capítulo tem por finalidade testar a existência de discriminação dos trabalhadores segundo cor e sexo através da análise controlada do diferencial de salários. Os resultados da análise controlada são obtidos pelo uso de modelagem estatística considerando o plano amostral da PNAD 2003. Adicionalmente, decompõe-se o diferencial de salários em efeitos provenientes das características individuais (escolaridade, experiência ou local de residência) e da discriminação.

A metodologia consiste em estimar a equação de salários (Mincer, 1974) e decompor o diferencial de salários através da metodologia apresentada por Oaxaca (1973). Para isso, incorpora-se o esquema amostral complexo da PNAD 2003 apresentado no capítulo 3. Desta forma, a sinergia entre as técnicas econométricas e as ferramentas estatísticas, fundamentadas na teoria econômica, resulta em medidas concretas que propiciam um maior conhecimento do nosso mercado de trabalho. A

principal ferramenta utilizada é a Análise de Regressão¹⁴ com modelos estimados pelo Método de Máxima Pseudo-Verssimilhança descrito no capítulo 4.

A análise de regressão é utilizada no estudo da relação de uma variável resposta com um conjunto de variáveis denominadas explicativas. O seu propósito é estimar ou prever o valor médio populacional da variável dependente em termos dos valores conhecidos¹⁵ das explicativas.

A *teoria do capital humano* aplicada à análise de regressão fornece um arcabouço para avaliar como decisões individuais influenciam nos retornos dos rendimentos. Os trabalhos de Becker (1962), Mincer (1964) e Ramos (1996) descrevem a ligação entre o ciclo da vida de um indivíduo e os investimentos em capital humano. Ben-Porath (1973) formaliza a equação de produção do capital humano baseado nos trabalhos de Becker e Mincer.

5.2 Estudos sobre Determinantes do Diferencial de Salários no Brasil

Diversos estudos procuram obter informações sobre os determinantes do salário dos trabalhadores, a partir de suas características individuais (escolaridade, experiência, cor/raça ou local de moradia) e informações sobre o mercado de trabalho (setor de atividade ou ocupação). O entendimento do processo pelo qual as pessoas desenvolvem suas habilidades na escola e no trabalho são fundamentais para investigar não apenas o porquê das diferenças de salário, mas para fomentar idéias no desenvolvimento econômico e social do país. A literatura existente sobre o tema revela que outros fatores também influenciam o salário como, por exemplo, habilidade inata ou discriminação.

¹⁴ O termo Regressão, por sua vez, amplamente utilizado em estudos do mercado de trabalho foi introduzido Galton (1886) – Gujarati (2000).

¹⁵ Conhecido ou fixo no sentido de não estocástico.

Coelho e Corseuil (2002) apresentam um resumo dos estudos sobre Diferencial de Salários nos últimos trinta anos no Brasil. Os autores encontraram diferentes abordagens no uso da equação de salário. Em alguns artigos o foco da análise é a avaliação das características dos indivíduos na determinação do salário, enquanto outros têm como objetivo mensurar diferenciais entre grupos de trabalhadores de acordo com suas características sócio-demográficas.

Rodrigues (2003) revela a estrutura salarial do Brasil a partir da Pesquisa de Padrão de Vida (PPV), realizada em 1996 pelo IBGE, levando em consideração os efeitos do esquema amostral da pesquisa. A contribuição do trabalho é a avaliação do diferencial de salário não-controlado (razão entre os salários do grupo) e o diferencial de salário controlado quando estimado considerando-se ou não o esquema amostral da pesquisa.

A equação de salários (Mincer, 1974), estimada pela Análise de Regressão, é um importante instrumento para decompor o diferencial de salários entre os grupos de cor. A próxima seção discute a forma funcional da variável resposta e as variáveis explicativas utilizadas.

5.3 A forma funcional da Equação de Salários

Mincer (1974) integrou a teoria do investimento em capital humano dentro de um contexto empírico, compatível com a teoria econômica. Desde então, o seu trabalho passou a ser amplamente utilizada em estudos do mercado de trabalho que foi denominada como “função salário do capital humano”, ou popularmente conhecida como “equação minceriana”.

O logaritmo da razão entre o salário e a jornada de trabalho (hora) é a variável de interesse para o estudo do diferencial de salários que também é denominada por variável resposta.

Seja:

Z_i - salário/hora do i-ésimo indivíduo,

$Y_i = \ln(z_i)$ - logaritmo de Z_i ,

\underline{X}_i - vetor de variável explicativa do i-ésimo indivíduo,

Considerando um modelo com apenas uma variável explicativa:

$$Y_i = \beta_0 + \beta_k \underline{X}_i + \varepsilon_i, \quad \varepsilon_i \sim N(0,1) \quad (5.1)$$

A forma funcional da expressão (5.1) é chamada de modelo semilog, pois a variável resposta aparece na forma logarítmica. Desta forma, o coeficiente de inclinação da variável explicativa \underline{X}_i mede a variação proporcional constante em Y para uma dada variação absoluta no valor da variável explicativa, como exemplo:

$$\beta_k = \frac{\text{Variação relativa na variável resposta}}{\text{Variação absoluta no regressor}} \quad (5.2)$$

Se multiplicarmos a variação relativa em Y por 100, a expressão (5.2) fornecerá a variação percentual, ou taxa de crescimento/decrécimo em Y para uma variação absoluta em X, o regressor.

Oaxaca (1973) destaca a experiência, escolaridade, posição na ocupação, setor de atividade, grupos de ocupação, migração e estado civil como controles importantes

da fonte da discriminação segundo cor e sexo. O controle definido pela ocupação elimina alguns dos efeitos das barreiras ocupacionais como fonte de discriminação. Desta forma, estaríamos sub-estimando os efeitos da discriminação quando incluímos tais efeitos na análise. Ele apresentou um outro conjunto de equações de salários que não considera a ocupação e setor de atividade dos trabalhadores denominado por “equação de salários de características individuais” que é adotado na dissertação.

A partir das evidências apresentadas por Oaxaca e o princípio da parcimônia, as variáveis explicativas utilizadas nesta dissertação são: educação, experiência e local de moradia (área urbana ou rural).

A escolaridade dos trabalhadores é definida pelos anos de estudo completos. Tal informação encontra-se definida nos microdados da PNAD 2003. A experiência do trabalhador é obtida pela expressão: experiência = idade em anos completos - anos de estudo completos – 6¹⁶.

5.4 Análise Controlada do Diferencial de Salários

Apresenta-se a seguir, os resultados da estimação da equação de salários ajustada para os trabalhadores entrevistados na PNAD 2003. A distinção entre os modelos 1 e 2 (tabela 5.1) está na utilização das variáveis cor e sexo, além das demais características como escolaridade, experiência e local de moradia. Esta análise compara indivíduos com atributos semelhantes, por exemplo, escolaridade, experiência e local de moradia, exceto a cor e sexo. Os resultados revelam que mulheres de cor branca e homens e mulheres de cor preta e parda ganham menos do que o grupo base (homens brancos).

Para ambos os modelos os sinais dos coeficientes da escolaridade e experiência estão de acordo com os resultados da Teoria do Capital Humano. A interpretação dos

coeficientes de escolaridade (linear e quadrático) revela que o salário aumenta conforme a escolaridade a taxas crescentes. No entanto, os sinais dos coeficientes da experiência indicam que o salário aumenta conforme a experiência de trabalho a taxas decrescentes.

Tabela 5.1 – Estimação das Equações de Salários

Características produtivas	Estimativas dos coeficientes	
	Modelo (1)	Modelo (2)
Escolaridade	0.05335	0.04619
	15.07	13.67
Escolaridade ²	0.00622	0.00640
	30.21	32.46
Experiência	0.04951	0.04905
	85.31	86.65
Experiência ²	-0.00055	-0.00056
	-49.45	-51.65
Área Urbana	0.1898968	0.2007406
	27.37	30.07
Sexo (1=homens; 0=mulheres)	-	0.29914
	-	56.03
Cor (1=cor branca; 0=caso contrário)	-	0.26297
	-	43.59
Constante	-0.79963	-1.06887
	-45.32	-60.37
R ²	0.40	0.44
P>F	0.00	0.00

Nota: As informações em negrito correspondem a estatística t

¹⁶ Idade média em que os indivíduos iniciam os seus estudos.

Além do uso de modelagem é possível descrever um indicador para avaliar o grau de discriminação entre os grupos de cor e sexo. A próxima seção descreve o Coeficiente de Discriminação proposto por Becker (1962) e sua generalização descrita por Oaxaca (1973).

5.5 Coeficiente de Discriminação

O Coeficiente de Discriminação apresentado por Becker é definido como a porcentagem do diferencial salarial entre dois tipos de mercados perfeitamente substitutos. Para os casos nos quais os dois fatores não são necessariamente substitutos perfeitos, Becker definiu o coeficiente de discriminação como uma simples diferença entre os salários e a razão dos salários na ausência de discriminação.

A formalização do coeficiente de discriminação entre os grupos de cor e sexo segue o mesmo procedimento da seção 2.3. Atribui-se como grupo base (L) os homens de cor branca para comparações e L-1 *grupos* definidos pelos homens de cor preta/parda, mulheres de cor branca e mulheres de cor preta/parda.

$$D = \frac{\bar{Z}_L / \bar{Z}_I - (\bar{Z}_L / \bar{Z}_I)^0}{(\bar{Z}_L / \bar{Z}_I)^0}, \quad I = 1, 2, 3 \text{ e } L \quad (5.3)$$

onde:

- \bar{Z}_I = média do salário/hora no I-ésimo grupo;
- $I = 1$ (homens de cor preta/parda), 2 (mulheres de cor branca), 3 (mulheres de cor preta/parda) e L (homens de cor branca ou grupo base);
- $(\bar{Z}_L / \bar{Z}_I)^0$ = razão entre as médias do salário/hora do grupo base (L) e o I-ésimo grupo;

- $(\bar{Z}_L / \bar{Z}_I)^0$ = razão entre as médias do salário/hora do grupo base(L) e o I-ésimo grupo na ausência da discriminação;

Oaxaca (1973) apresenta uma generalização da medida de Becker admitindo substitutos perfeitos como caso especial (sendo mais flexível para trabalhos empíricos).

Para isso, aplica-se o logaritmo na expressão 5.3.

$$\ln(D) = \ln \left[\frac{\bar{Z}_L / \bar{Z}_I - (\bar{Z}_L / \bar{Z}_I)^0}{(\bar{Z}_L / \bar{Z}_I)^0} \right] \quad (5.4)$$

$$\ln(D) = \ln \left[\frac{\bar{Z}_L / \bar{Z}_I}{(\bar{Z}_L / \bar{Z}_I)^0} - \frac{(\bar{Z}_L / \bar{Z}_I)^0}{(\bar{Z}_L / \bar{Z}_I)^0} \right]$$

$$\ln(D) = \ln \left[\frac{\bar{Z}_L / \bar{Z}_I}{(\bar{Z}_L / \bar{Z}_I)^0} - 1 \right]$$

$$\ln(D) = \frac{\ln \left[\frac{\bar{Z}_L / \bar{Z}_I}{(\bar{Z}_L / \bar{Z}_I)^0} \right]}{\ln[1]}$$

$$\ln(D) * \ln[1] = \ln \left[\frac{\bar{Z}_L / \bar{Z}_I}{(\bar{Z}_L / \bar{Z}_I)^0} \right]$$

$$\ln(D + 1) = \ln(\bar{Z}_L / \bar{Z}_I) - \ln(\bar{Z}_L / \bar{Z}_I)^0 \quad (5.4)$$

Como não conhecemos $(\bar{Z}_L / \bar{Z}_I)^0$, para estimar D faz-se necessário obter valores para $(\bar{Z}_L / \bar{Z}_I)^0$. A estimativa é realizada considerando ou não a ausência de discriminação. Se não existisse discriminação, a estrutura salarial de um determinado grupo padrão poderia ser aplicada aos demais grupos, ou a estrutura destes grupos poderia ser aplicada ao grupo base.

A estimação pelo Método de Máxima Pseudo-Verossimilhança da equação de salários para determinados grupos de trabalhadores, fornece uma estimativa da estrutura salarial que pode ser aplicada a outros grupos. Este procedimento consiste em estimar separadamente para cada grupo de cor/sexo a equação de Mincer. A definição do modelo é dada pela expressão 5.5.

$$\ln(Z_{ij}) = \underline{X}_{ij}' \underline{\beta}^{(l)} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0,1) \quad (5.5)$$

Onde:

- $\ln(Z_{ij})$ = logaritmo do salário/hora para o i-ésimo indivíduo pertencente ao l-ésimo grupo de cor e sexo;
- \underline{X}_{ij}' = matriz das variáveis explicativas (escolaridade, experiência) do i-ésimo indivíduo do l-ésimo grupo de cor e sexo;
- $\underline{\beta}^{(l)}$ = parâmetros a serem estimados do l-ésimo grupo de cor e sexo;
- ε_{ij} = erro aleatório

5.6 Procedimentos para o Cálculo da Decomposição de Salários

Através da Análise de Regressão, equação de salários, é possível inferir para cada grupo de cor/sexo quais os retornos para cada característica produtiva. Isto é, através dos coeficientes estimados (betas). Com isso seja:

$$G = \frac{\bar{Z}_L - \bar{Z}_I}{\bar{Z}_I} \quad (5.6)$$

Aplicando o logaritmo na expressão (5.6):

$$\ln(G + 1) = \ln(\bar{Z}_L) - \ln(\bar{Z}_I) \quad (5.7)$$

Onde \bar{Z}_L e \bar{Z}_I são as médias dos logaritmos dos salários/hora para o grupo-padrão e do I-ésimo grupo de cor/sexo. A partir das propriedades dos estimadores de máximo pseudo-versossimilhança, podemos escrever:

$$\ln(\bar{Z}_L) = \bar{X}'_L \hat{\beta}_L \quad (5.8)$$

$$\ln(\bar{Z}_I) = \bar{X}'_I \hat{\beta}_I \quad (5.9)$$

Seja:

$$\bar{X}'_L \hat{\beta}_L - \bar{X}'_I \hat{\beta}_I$$

Substituindo (5.8) e (5.9) na expressão (5.7), obtem-se:

$$\ln(G + 1) = \bar{X}'_L \hat{\beta}_L - \bar{X}'_I \hat{\beta}_I \quad (5.10)$$

Considere:

Efeito devido às diferenças das características produtivas

$$\ln\left(\frac{Z_L}{Z_I}\right)^0 = \Delta\bar{X}'\hat{\beta}_I \quad (5.11)$$

Efeitos devido às diferenças da discriminação/retornos

$$\ln(G + 1) = -\bar{X}'_L \Delta\hat{\beta} \quad (5.12)$$

As expressões (5.11) e (5.12) representam a decomposição do diferencial devido às diferenças das características produtivas e ao efeito da discriminação, respectivamente.

5.7 Ajuste das Equações de Salários para os grupos de cor e sexo

Para a realização do procedimento de Oaxaca e obtenção dos efeitos descritos na seção 5.6 é necessário ajustar as equações de salários para os diferentes grupos de cor e sexo. Utilizou-se o programa estatístico STATA 8.0, procedimento `svyregress` (anexo 3), para o ajuste do modelo de regressão com duas finalidades: ajustar o modelo com o peso e sem a especificação dos estratos (tabela 5.2) e a forma adequada com a especificação dos pesos, estratos e unidades primárias de amostragem (tabela 5.3).

A comparação entre as tabelas 5.2 e 5.3 revela que o fato de não incorporar a estratificação no ajuste dos modelos não altera os valores dos coeficientes, mas as estatísticas de teste (t) são afetadas pela não especificação no ajuste do modelo. Como exemplo, enquanto que a estatística de teste para a variável escolaridade na tabela 5.2 é **13.18** na tabela 5.3 é **10.41** e isso significa que o desvio padrão no primeiro caso é menor quando estimamos os coeficientes considerando que tenham sido dados obtidos através de uma amostra aleatória simples. Na tabela 5.3, a interpretação dos coeficientes de escolaridade e experiência de trabalho é realizada em conjunto com os termos quadráticos. Como exemplo, o efeito marginal dos anos de estudo das mulheres de cor branca (denominado por *Ef marg MB*) sobre a média do $\ln(\text{salário/hora})$ das mulheres é obtido pela expressão:

$$Ef\ marg\ MB = -0.0191 * \text{anos de estudo} + 0.0098 * (\text{anos de estudo}^2) \quad (5.13)$$

Tabela 5.2 - Equações de salários estimadas com o peso e sem a especificação dos estratos. Estimativas dos parâmetros e respectivas estatísticas de teste (negrito)

Variáveis	Cor branca		Cor Preta ou Parda	
	Homens	Mulheres	Homens	Mulheres
Anos de estudo	0,0519 13,18	-0,0191 -3,65	0,0643 19,26	0,0261 5,38
Anos de estudo2	0,0065 28,10	0,0098 34,96	0,0046 18,74	0,0069 22,87
Experiência	0,0548 60,78	0,0382 36,00	0,0524 58,18	0,0460 37,25
Experiência2	-0,0006 -35,28	-0,0004 -18,84	-0,0006 -40,71	-0,0005 -22,48
Área Urbana	0,2719 21,02	0,1972 9,77	0,3408 32,72	0,3494 18,31
Constante	-0,8393 -39,59	-0,5315 -16,63	-1,0332 -60,84	-1,1365 -40,20
R ²	0,44	0,42	0,34	0,32
Prob>F	0,0001	0,0001	0,0001	0,0001
Amostra	44.311	31.115	46.257	28.129
População	22.080.841	15.297.084	19.311.338	11.455.643

Tabela 5.3 - Equações de Salários para os grupos de cor e sexo

Estimativas dos parâmetros e respectivas estatísticas de teste (negrito)

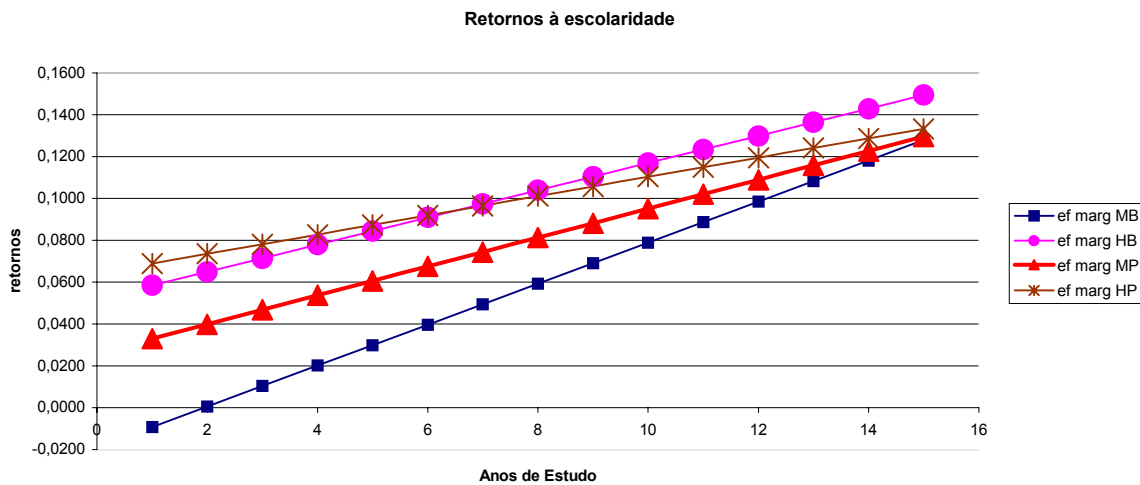
Variáveis	Cor branca		Cor Preta ou Parda	
	Homens	Mulheres	Homens	Mulheres
Anos de estudo	0,0519 10,41	-0,0191 -3,29	0,0643 15,39	0,0261 4,18
Anos de estudo ²	0,0065 22,26	0,0098 30,98	0,0046 15,61	0,0069 18,95
Experiência	0,0548 56,97	0,0382 35,64	0,0524 47,73	0,0460 35,07
Experiência ²	-0,0006 -33,58	-0,0004 -18,44	-0,0006 -32,06	-0,0005 -21,65
Área Urbana	0,2719 12,57	0,1972 7,47	0,3408 17,36	0,3494 10,43
Constante	-0,8393 -27,50	-0,5315 -14,10	-1,0332 -35,88	-1,1365 -24,73
R ²	0,44	0,42	0,34	0,32
Prob>F	0,0001	0,0001	0,0001	0,0001
Amostra	44.311	31.115	46.257	28.129
População	22.080.841	15.297.084	19.311.338	11.455.643

Os gráficos 5.1 e 5.2 apresentam os efeitos marginais sobre o $\ln(\text{salário/hora})$ para a escolaridade e experiência, respectivamente. Vale ressaltar que, como o efeito marginal é a derivada da equação de salários ele é uma função linear. No gráfico 5.1, as mulheres de cor branca apresentam os menores retornos em todos os anos de estudo ao comparar com outros de grupos de sexo/cor. Além disso, exceto para um ano de estudo, o valor dos retornos é positivo conforme a teoria do capital humano (quanto maior a escolaridade maior o rendimento).

Os homens apresentam os maiores retornos quando comparados às mulheres. Os retornos dos homens de cor branca superam o dos homens de cor preta/parda a partir dos seis anos de estudo.

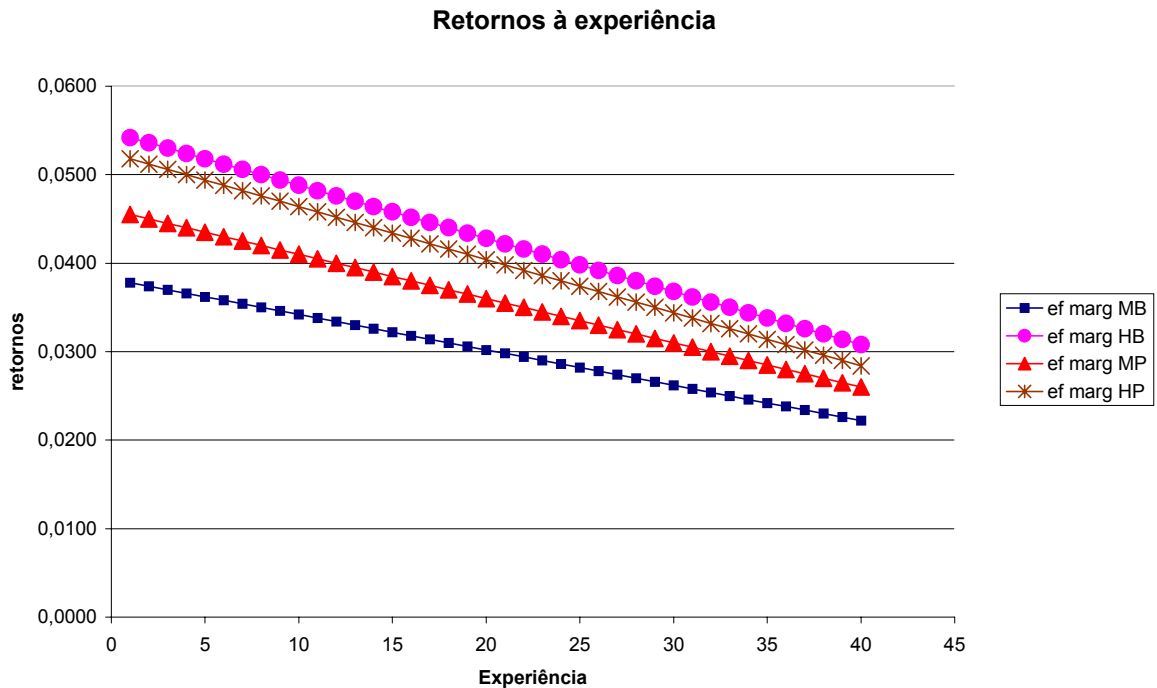
Verifica-se no gráfico 5.2 que, para todos os grupos, o retorno da experiência sobre o $\ln(\text{salário/hora})$ diminui conforme o aumento da experiência. A interpretação do efeito marginal da experiência reflete que o retorno é positivo (coeficiente da experiência linear) e cresce a taxas decrescentes (sinal negativo do termo quadrático da experiência).

Gráfico 5.1 – Retornos à escolaridade sobre o $\ln(\text{salário/hora})$



Notas: *ef marg MB* = efeito marginal das mulheres cor branca, *ef marg HB* = efeito marginal dos homens de cor branca, *ef marg MP* = efeito marginal das mulheres de cor preta/parda e *ef marg HP* = efeito marginal dos homens de cor preta/parda.

Gráfico 5.2 – Retornos à experiência sobre o $\ln(\text{salário/hora})$



5.8 Análise da Equação de Salários com interação

Na seção anterior os resultados foram obtidos de forma independente para grupo de cor e sexo. No entanto, para o cálculo do efeito devido às diferenças dos retornos (discriminação – expressão 5.12) torna-se necessário o cálculo do desvio-padrão incorporando o plano amostral.

A solução para obter a diferença entre os coeficientes dos grupos e o respectivo desvio-padrão com a incorporação do plano amostral é a estimação da equação de salários com interação das variáveis explicativas: escolaridade, experiência, idade e local de residência com o sexo e a cor dos trabalhadores.

Os resultados da tabela 5.4 em negrito representam as estimativas dos coeficientes para o grupo base L (homens de cor branca). As linhas em itálico representam a diferença entre o grupo base (L) e o l-ésimo grupo (homens de cor preta/parda, mulheres de cor branca e mulheres de cor preta/parda).

Como exemplo, o coeficiente para os anos de estudo dos homens de cor branca é **0,0519** (tabela 5.2 e tabela 5.3) enquanto que o para os homens de cor preta/parda é obtido por: **0,0519** + 0,0124 (tabela 5.3) = **0,0643** (tabela 5.2). A amostra considera todos os grupos, aproximadamente 149 mil trabalhadores.

Tabela 5.4 – Ajuste do Modelo de Regressão da Equação dos Salários com a interação sexo*cor e as variáveis explicativas anos de estudo, experiência e local de residência.

Variáveis	Estimativas	Estatística-t
Anos de estudo (Homens de cor branca)	0,0519	10,41 *
Homens de cor preta/parda	0,0124	2,13 *
Mulheres de cor branca	-0,0711	-10,24 *
Mulheres de cor preta/parda	-0,0258	-3,34 *
Anos de estudo² (Homens de cor branca)	0,0065	22,26 *
Homens de cor preta/parda	-0,0020	-5,21 *
Mulheres de cor branca	0,0033	8,7 *
Mulheres de cor preta/parda	0,0004	0,78
Experiência (Homens de cor branca)	0,0548	56,97 *
Homens de cor preta/parda	-0,0024	-1,67
Mulheres de cor branca	-0,0165	-11,75 *
Mulheres de cor preta/parda	-0,0088	-5,47 *
Experiência² (Homens de cor branca)	-0,00059	-33,58 *
Homens de cor preta/parda	-0,00004	-1,53
Mulheres de cor branca	0,00016	5,82 *
Mulheres de cor preta/parda	0,00005	1,57
Área Urbana (Homens de cor branca)	0,2719	12,57 *
Homens de cor preta/parda	0,0689	2,71 *
Mulheres de cor branca	-0,0747	-2,75 *
Mulheres de cor preta/parda	0,0775	2,09 *
Constante (Homens de cor branca)	-0,8393	-27,5 *
Homens de cor preta/parda	-0,1939	-5,2 *
Mulheres de cor branca	0,3078	7,28 *
Mulheres de cor preta/parda	-0,2972	-5,7 *
R ²	0,4457	
Prob>F	0,0000	
Amostra	149.812	
População	68.144.906	

Notas: * significativo ao nível de 5% de confiança.

5.9 Resultados da Decomposição do Diferencial de Salários

O objetivo desta seção é detalhar os procedimentos para decomposição da média do logaritmo do salário/hora através da análise controlada (modelagem estatística). Em seguida, investigar o quanto desta diferença é explicada pelas características individuais dos trabalhadores e quanto é proveniente dos retornos dos coeficientes.

Os resultados da tabela 5.5 avaliam se a diferença entre a média do logaritmo do salário/hora do grupo base (homens de cor branca) e o l-ésimo grupo de cor e sexo é estatisticamente significativa.

Tabela 5.5 – Diferença do $\ln(\text{salário/hora})$ entre os grupos de cor

Diferenças	Estimativa	Estatística t
Homens de cor branca - Homens de cor preta	0.59	50.72
$\ln(\text{salário/hora})$ Homens de cor branca - Mulheres de cor branca	0.15	18.83
Homens de cor branca - Mulheres de cor preta ou parda	0.69	56.74

Fonte: Processado a partir dos microdados da PNAD 2003.

O cálculo da diferença entre as médias das características produtivas (expressão 5.11) foi obtido com a incorporação do plano amostral da PNAD (anexo 3). Esta análise tem por finalidade testar se a diferença entre as características dos grupos de cor e sexo são estatisticamente diferentes de zero.

Verifica-se na tabela 5.5 que os homens têm, em média, uma escolaridade inferior as mulheres de cor branca.

Tabela 5.6 – Diferença entre as médias das características produtivas e coeficientes estimados para o Grupo Base (L) e o l-ésimo grupo de cor/sexo

Características	(Grupo padrão) - (l-ésimo grupo de cor/sexo)					
	Entre as médias			Entre os coeficientes estimados		
	Mulher de cor branca	Mulher de cor preta ou parda	Homem de cor preta ou parda	Mulher de cor branca	Mulher de cor preta ou parda	Homem de cor preta ou parda
Anos de estudo	-1,32 -36,17	0,80 15,32	2,32 44,55	-0,0711 -10,2	-0,02578 -3,3	0,0124 2,1
Anos de estudo ²	-21,83 -36,92	12,77 15,87	33,41 44,21	0,0033 8,7	0,00035 0,8	-0,0020 -5,2
Experiência	2,84 23,89	1,31 9,92	-0,71 -5,35	-0,0165 -11,8	-0,00883 -5,5	-0,0024 -1,7
Experiência ²	159,82 23,66	85,74 11,11	-40,09 -4,91	0,0002 5,8	0,00005 1,57	-0,000039 -1,5
Área Urbana	-0,07 -21,54	-0,03 -5,46	0,06 9,77	-0,0747 -2,8	0,07745 2,1	0,0689 2,7
Constante				0,3078 7,28	-0,29722 -5,70	-0,1939 -5,20

Com base nos resultados das tabelas 5.5 e 5.6 é possível a construção da decomposição do diferencial da média do $\ln(\text{salário/hora})$ de forma adequada. A finalidade da tabela 5.7 é avaliar o quanto das características produtivas explicam o diferencial e o que é proveniente da discriminação. Entre as pessoas de cor branca, homens e mulheres, as características produtivas explicam -107,1% e o efeito da discriminação explica 207% do diferença entre a média do $\ln(\text{salário/hora})$. Para esta análise vale ressaltar que o efeito da discriminação (0,31) é maior que a diferença dos logaritmos dos salários (0,15), pois as mulheres têm mulheres atributos pessoais (escolaridade media superior) quando comparados aos homens, e se fossem remuneradas igualmente seus rendimentos seriam maiores.

A diferença da média do ln(salário/hora) entre os trabalhadores do grupo e as pessoas de cor preta e parda revela que as mulheres sofrem uma maior discriminação. Enquanto que o efeito da discriminação explica 84% do diferencial entre as mulheres de cor branca e o grupo base, entre os homens é responsável por 47%.

Tabela 5.6 - Efeitos da Discriminação Estimados pelas Características Pessoais

	Efeitos	Mulher branca		Homem preto		Mulher preta	
		(1)a	(2)b	(1)a	(2)b	(1)a	(2)b
	Diferencial de salários = (3)	0,152	100,0%	0,593	100,0%	0,688	100,0%
Efeitos da características produtivas	Anos de estudo	0,0254	16,7%	0,1489	25,1%	0,0209	3,04%
	Anos de estudo ²	-0,2143	-140,9%	0,1526	25,7%	0,0878	12,76%
	Experiência	0,1085	71,4%	-0,0371	-6,3%	0,0602	8,74%
	Experiência ²	-0,0686	-45,1%	0,0253	4,3%	-0,0467	-6,79%
	Área Urbana	-0,0138	-9,1%	0,0212	3,6%	-0,0108	-1,58%
	Somatório do efeitos = (4)	-0,1629	-107,1%	0,3110	52,4%	0,1113	16,2%
Efeito Discriminação	Estimativa de ln(D+1) = (5)	0,3150	207,1%	0,2825	47,6%	0,577	83,8%
Coefficiente de Discriminação	Estimativa de D = (6)	0,37		0,33		0,78	

Notas: (a) é igual ao produto entre o coeficiente da variável explicativa e a diferença das características médias da tabela 5.5

(b) é igual a coluna (a) expressa como percentual da diferencial de salários.

(3) Representa o diferencial de salários entre o grupo padrão e o l-ésimo grupo de cor/sexo

(4) Somatório dos efeitos sobre o diferencial de salários

(5) É igual a (3) - (4)

(6) É igual a exponencial do item (5)

A estimativa do coeficiente de discriminação (D), por sua vez, sintetiza os resultados da análise controlada da discriminação presente no mercado de trabalho brasileiro em 2003. A conclusão é que a discriminação no mercado de trabalho é mais evidente contra as mulheres de cor preta. Desta forma, considerando como grupo base os homens de cor branca, o maior valor registrado de D foi 0,78 entre as mulheres de cor preta e parda e o grupo base. Em seguida, 0,37 para as mulheres de cor branca e 0,33 para os homens de cor preta e parda.

Este capítulo desenvolveu e implementou os procedimentos teóricos e operacionais para o cálculo da decomposição da média do $\ln(\text{salário/hora})$ em pesquisas como a PNAD 2003. Foi identificado que a melhor maneira para estimar as equações de salários é através de um modelo de interação que já fornece as estimativas das diferenças dos coeficientes com a incorporação do plano amostral e facilita na construção do quadro de decomposição. O próximo capítulo tem por finalidade avaliar corrigir os possíveis vieses de seleção da informação do salário em pesquisas como a PNAD 2003 que registra os salários daqueles que trabalham.

6 Decomposição do Diferencial de Salário com Correção do Viés de Seleção do Salário na Análise de Regressão

6.1 O viés de seleção da informação dos salários

Este capítulo tem por finalidade apresentar os procedimentos necessários para corrigir o viés de seleção das informações dos salários reportados pelos indivíduos. Com isso, avaliar os impactos nas equações de salários e nos resultados da decomposição de salários. Com relação à variável resposta, $\ln(\text{salário/hora})$, vale ressaltar que na PNAD, como em outras pesquisas, as informações coletadas são fornecidas pelas pessoas que tinham trabalho na época da pesquisa. Isto é, os salários observados na PNAD 2003 estão relacionados com a decisão de um indivíduo trabalhar ou não.

De fato, para determinadas pessoas seria vantajoso trabalhar se o salário recebido (ou salário potencial) fosse maior que o custo de oportunidade (ou salário reserva). Desta forma, existe um viés de seleção das informações do salário. Neste caso, as estimativas dos coeficientes da equação de salários obtidas a partir das informações dos indivíduos que trabalham na data de referência da pesquisa podem estar viesadas sob a ótica do **viés de seleção das informações**, considerando-se que o modelo utilizado no capítulo 5 não incorpora a informação sobre a avaliação dos indivíduos que não trabalham no que se refere ao custo de oportunidade.

Heckman (1979) apresentou o fenômeno e sua solução não apenas para pesquisas amostrais. De modo geral, o viés de seleção de informação pode ser decorrente de duas razões, como destaca Heckman: ou em virtude de seletividade das informações dos indivíduos ou devido ao desenho amostral da pesquisa. Cabe ressaltar que, ao utilizar os dados da PNAD, o efeito que deseja-se incorporar na

modelagem visa corrigir o viés de seletividade da informação dos salários para os indivíduos que, apesar de estarem devidamente representados na amostra da PNAD 2003, não trabalhavam na data de referência da pesquisa supostamente devido a uma avaliação do salário potencial e do custo de oportunidade envolvido nesta escolha.

6.2 A Correção do Viés de Seletividade da Informação dos Salários através do Procedimento de Heckman em Pesquisas Amostrais Complexas

Conforme apresentado na seção anterior, o problema de estimar a equação de salários é que não observamos o salário para toda a amostra, mas apenas para aqueles que trabalham. A metodologia apresentada é um resumo dos trabalhos de Heckman (1979) e Kassouf (1994) e sua motivação tem como base a incorporar a complexidade do desenho amostral (capítulo 3) na estimação dos modelos. Os únicos programas estatísticos que incorporam o plano amostral no procedimento de Heckman são o STATA 8.0 e a versão 9 do SAS. Nesta dissertação adotou-se o programa STATA 8.0.

Considere que para o indivíduo que trabalha o salário pago ($\ln(\text{salário/hora})$), ou salário potencial (Y_{ji}^{pot}), no mercado de trabalho tem que ser maior que o custo de oportunidade, ou salário reserva ($\ln(\text{salário/hora})$) (Y_{ji}^{res}). Desta forma, para a equação (6.1) Y_{ji}^{pot} é igual ao $\ln(\text{salário/hora})$ do i -ésimo indivíduo do l -ésimo grupo de cor e sexo, X_{ji}^{pot} é o vetor com a especificação da variável explicativa do i -ésimo indivíduo do l -ésimo grupo de cor e sexo, $\beta^{(l)pot}$ é vetor do parâmetro a ser estimado do l -ésimo grupo de cor e sexo e τ_{ji} é o erro aleatório da equação de salários do i -ésimo indivíduo do l -ésimo grupo de cor e sexo.

$$Y_{li}^{pot} = \underline{X}_{li}^{pot} \underline{\beta}^{(l)pot} + \tau_{li}, \tau_{li} \sim N(0, \sigma_{\tau}^2) \quad (6.1)$$

Na equação (6.2) temos a equação do salário reserva, dado que Y_{li}^{res} é igual ao ln(salário/hora) do i-ésimo indivíduo do l-ésimo grupo de cor e sexo, \underline{X}_{li}^{res} é o vetor com a especificação da variável explicativa do i-ésimo indivíduo do l-ésimo grupo de cor e sexo, $\underline{\beta}^{(l)res}$ vetor do parâmetro a ser estimado e $\underline{\psi}_{li}$ o erro aleatório.

$$Y_{li}^{res} = \underline{X}_{li}^{res} \underline{\beta}^{(l)res} + \psi_{li}, \psi_{li} \sim N(0, \sigma_{\psi}^2) \quad (6.2)$$

Entre as equações (6.1) e (6.2) a correlação é dada por:

$$cor(\tau_{li}, \psi_{li}) = \rho \quad (6.3)$$

A decisão de trabalhar ou não é dada pela diferença entre o salário potencial e reserva.

$$(Y_{li}^{pot} - Y_{li}^{res}) \quad (6.4)$$

Substituindo as expressões 6.1 e 6.2 na expressão 6.4 temos:

$$(\underline{X}'_{li}{}^{pot} \underline{\beta}^{(l)pot} + \underline{\tau}_{li}) - (\underline{X}'_{li}{}^{res} \underline{\beta}^{(l)res} + \underline{\psi}_{li}) = \quad (6.5)$$

$$(\underline{X}'_{li}{}^{pot} \underline{\beta}^{(l)pot}) \cdot (-\underline{X}'_{li}{}^{res} \underline{\beta}^{(l)res}) + (\underline{\tau}_{li} - \underline{\psi}_{li}) = \quad (6.6)$$

$$-(\underline{X}'_{li}{}^{pot} \underline{X}'_{li}{}^{res}) \cdot (\underline{\beta}^{(l)res} \underline{\beta}^{(l)pot}) + (\underline{\tau}_{li} - \underline{\psi}_{li}) = \quad (6.7)$$

Seja:

$$(\underline{X}'_{li}{}^{pot} \underline{X}'_{li}{}^{res}) = \underline{R}'_{li} \quad (6.8)$$

$$(\underline{\beta}^{res} \underline{\beta}^{pot}) = \underline{\gamma} \quad (6.9)$$

$$(\underline{\tau}_{li} - \underline{\psi}_{li}) = \underline{v}_{li} \quad (6.10)$$

Com isso, a equação de participação é dada por:

$$(Y_{li}^{pot} - Y_{li}^{res}) = \underline{R}'_{li} \underline{\gamma} + v_{li} \quad (6.11)$$

$$Q_{li}^* = \underline{R}'_{li} \underline{\gamma} + v_{li} \quad (6.12)$$

A esperança da informação do salário ser registrada dado que o salário potencial é observado é:

$$E(Y_{li}^{pot} | Y_{li}^{pot} \text{ observado}) = E(Y_{li}^{pot} | (Y_{li}^{pot} - Y_{li}^{res}) > 0) \quad (6.13)$$

$$= E[Y_{li}^{pot} | (v_{li} > -\underline{R}'_{li} \underline{\gamma})] \quad (6.14)$$

$$= E[\underline{X}'_{li}{}^{pot} \underline{\beta}^{(l)pot} + \underline{\tau}_{li} | v_{li} > -\underline{R}'_{li} \underline{\gamma}] \quad (6.15)$$

$$= \underline{X}'_{li}{}^{pot} \underline{\beta}^{(l)pot} + E[\underline{\tau}_{li} | v_{li} > -\underline{R}'_{li} \underline{\gamma}] \quad (6.16)$$

$$= \underline{X}_{ii}^{pot} \underline{\beta}^{(l)pot} + \rho\sigma_{\tau} + \lambda_{ii}(\alpha_v) \quad (6.17)$$

onde:

$$\alpha_v = \frac{-R_{ii}'\gamma}{\sigma_v} \quad (6.18)$$

$$\lambda_{ii}(\alpha_v) = \frac{\phi\left(\frac{R_{ii}'\gamma}{\sigma}\right)}{\Phi\left(\frac{R_{ii}'\gamma}{\sigma}\right)} \text{ ou } \frac{\phi\left(\frac{-R_{ii}'\gamma}{\sigma}\right)}{1-\Phi\left(\frac{-R_{ii}'\gamma}{\sigma}\right)}, \quad (6.19)$$

A variável lambda também é denominada como a razão inversa de Mill. Seja ϕ e Φ são respectivamente, a função de densidade e função de distribuição acumulada para uma distribuição normal.

$$Y_{ii} = E\left[Y_{ii}^{pot} \mid (Y_{ii}^{pot} - Y_{ii}^{res}) > 0\right] + \eta_{ii} \quad (6.20)$$

$$= X_{ii}'\beta + \beta_{\lambda} \lambda_{ii}(\alpha_v) + \eta_{ii} \quad (6.21)$$

Então, pelo método tradicional de estimação da Equação de Salário, mesmo incorporando o plano amostral, se omitir o termo $\beta_{\lambda} \lambda_{ii}(\alpha_v)$ implica em estimadores inconsistentes.

Desta forma, baseado nos trabalhos de Heckman (1979) o procedimento proposto consiste em dois passos:

- ⇒ Primeira etapa - cálculo de λ_{ii} através de um modelo Probit considerando o desenho amostral pelo Método de Máxima Pseudo Verossimilhança.

Inicialmente define-se uma equação de participação, através do modelo probit, para estimar a probabilidade de uma pessoa participar do mercado de trabalho. Em seguida, a partir dos coeficientes estimados, calcula-se uma nova variável, denominada *função inversa de Mills*;

⇒ Segunda etapa: estimar o modelo de regressão pelo método de máxima pseudo-verossimilhança incluindo a função inversa de Mills como variável explicativa do modelo.

6.3 Ajuste da Equação de Salários com a Correção de Heckman

6.3.1 Equação de Participação

A equação de participação avalia a probabilidade do indivíduo trabalhar segundo algumas variáveis explicativas¹⁷. O modelo utilizado é o probit com a incorporação do plano amostral. A variável dependente assume o valor “1” se o indivíduo tem rendimento (ocupado) e “0” caso contrário (descocupados ou inativos).

Com base nos trabalhos de Heckman (1979) e Kassouf (1994) as variáveis explicativas selecionadas foram:

- Escolaridade em anos de estudo;
- Escolaridade ao quadrado;
- Experiência
- Experiência ao quadrado
- Chefe - condição no domicílio chefe;
- Filho – condição no domicílio filho;
- Crianças no domicílio entre 0 e 5 anos – assume o valor 1 se há crianças e o caso contrário.

¹⁷ Ressalta-se que para identificação é necessário que algumas destas variáveis não estejam incluídas na equação de salários.

Os resultados da equação de salários – modelo probit – estão descritos na tabela 6.1. Verifica-se que conforme aumenta a experiência de trabalho maior é a probabilidade de um indivíduo participar no mercado de trabalho. O sinal negativo do termo quadrático da experiência indica que a probabilidade de participar no mercado de trabalho cresce a taxas decrescentes. O efeito do termo linear da escolaridade em todos os grupos de cor reflete que a probabilidade aumenta conforme os anos de estudo. No entanto, o termo quadrático indica que para os homens a probabilidade cresce a taxas decrescentes e as mulheres a taxas crescentes. Ser chefe do domicílio é uma variável importante para determinar a probabilidade do indivíduo participar do mercado de trabalho. Com isso, independente do sexo e da cor a probabilidade de participar no mercado de trabalho de chefes de família é positiva em todos os grupos de cor e sexo. Para os indivíduos que declararam ter uma relação de filho com o chefe do domicílio o comportamento é o contrário. Isto é, o sinal negativo expressa que estes indivíduos estão menos propensos a participar do mercado de trabalho.

Além das características pessoais e responsabilidades no domicílio foi incluída uma informação que tem impacto sobre todas as pessoas que vivem no mesmo domicílio: *crianças de 0 a 5 anos de idade*. Enquanto os homens estão mais propensos a trabalhar, as mulheres têm um comportamento contrário. Heckman (1979) e Kassouf (1994) justificam este comportamento pela regra de decisão entre o salário potencial e o salário reserva descrito na seção anterior.

Tabela 6.1 – Equação de Participação no Mercado de Trabalho Brasileiro em 2003

Estimadores	Cor branca		Cor Preta ou Parda	
	Homens	Mulheres	Homens	Mulheres
Anos de estudo	0,1484 30,05	0,0752 16,49	0,1027 19,38	0,0560 11,36
Anos de estudo ²	-0,0043 -12,14	0,0017 5,51	-0,0001 -0,31	0,0044 11,26
Experiência	0,0944 63,52	0,0745 50,16	0,1087 74,94	0,0911 65,62
Experiência ²	-0,0016 -61,00	-0,0014 -48,44	-0,0017 -61,37	-0,0015 -56,98
Área Urbana	-0,2875 -13,75	0,2482 9,57	-0,1858 -8,47	0,2398 10,28
Chefe do domicílio	0,5883973 25,89	0,4216303 23,2	0,5646465 25,7	0,4643732 27,02
Filho	-0,26 -14,37	-0,12 -6,56	-0,23 -13,83	-0,05 -2,81
Criança de 0 a 5 anos	0,02 1,53	-0,18 -12,79	0,04 3,01	-0,21 -14,27
Constante	-1,1674 -43,90	-1,7858 -54,57	-1,3662 -47,77	-2,0279 -62,18
Amostra	87.601	96.528	98.593	99.112

A partir dos coeficientes da equação de participação (modelo probit) – tabela 6.1 – é calculado a variável lambda (ou razão inversa de Mills) e utilizado como variável explicativa para estimação da equação de salários que é apresentado na próxima seção.

6.3.2 Ajuste das Equações de Salário com Correção do Viés de Seletividade da Informação.

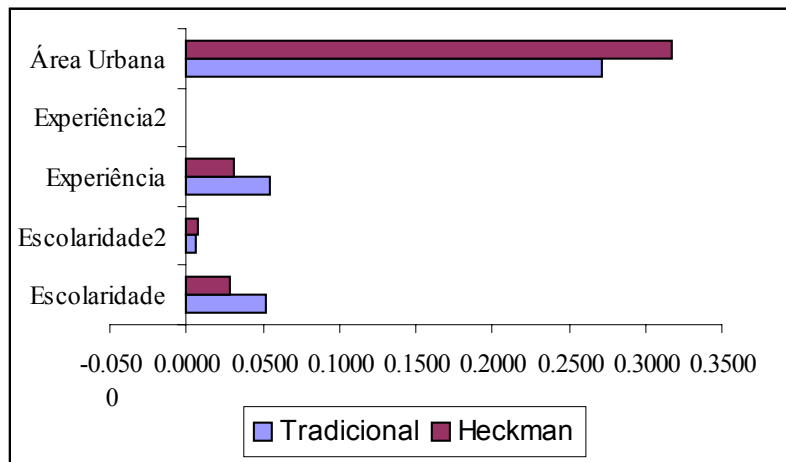
Os resultados da tabela 6.2 apresentam as equações de salários com a correção de Heckman. Vale ressaltar que o coeficiente da variável lambda é estatisticamente significativo, indicando que a inclusão é necessária para correção do viés de seleção da informação do salário.

Tabela 6.2 – Equação de Salários com Correção de Heckman

Estimadores	Cor branca		Cor Preta ou Parda	
	Homens	Mulheres	Homens	Mulheres
Anos de estudo	0,0280 5,42	-0,0257 -3,98	0,0467 10,88	0,0217 3,44
	0,0051697	0,0064646	0,0042961	0,0063076
Anos de estudo2	0,0072 24,06	0,0098 30,82	0,0046 15,66	0,0067 18,10
	0,0003012	0,000318	0,0002966	0,0003717
Experiência	0,0311 24,51	0,0324 11,71	0,0255 14,54	0,0401 17,94
	0,0012673	0,0027657	0,0017507	0,0022357
Experiência2	-0,0002 -9,96	-0,0003 -6,28	-0,0002 -8,62	-0,0005 -11,94
	0,0000214	0,0000517	0,0000275	0,000038
Área Urbana	0,3169 14,42	0,1750 6,1	0,3711 18,89	0,3312 9,93
	0,0219826	0,0286843	0,0196445	0,0333521
Lambda	-0,3456 -24,38	-0,1068 -2,32	-0,3569 -19,39	-0,0908 -3,10
	0,0141757	0,0460311	0,0184066	0,0293129
Constante	-0,3256 -9,12	-0,3040 -2,87	-0,4560 -11,78	-0,9317 -12,92

O sinal negativo da variável *lambda* indica que os fatores não mensurados, na equação de salários, por um aumentam a probabilidade de participação diminuem os retornos do salário. Como exemplo, nos gráficos 6.1 e 6.2, para os homens de cor branca, o coeficiente da escolaridade pelo método tradicional era 0.052 (capítulo 5) e com a correção de Heckman passa a ser 0.028 (tabela 6.2). Quando realizamos a mesma análise para a variável experiência de trabalho, para todos os grupos de cor e sexo, os coeficientes também sofrem reduções. A comparação com o método tradicional (tabela 5.4) revela que tanto os coeficientes, quanto às variâncias, se alteram com a utilização deste procedimento. Mesmo assim todas os coeficientes continuam significativos a 95% de confiança.

Gráfico 6.1 – Comparação entre os coeficientes da Equação de Salários pelo Método Tradicional e com a Correção de Heckman.



De posse das estimativas da equação de salários corrigidas do viés de seleção da informação do salário o próximo passo é decompor o diferencial entre os grupos de trabalhadores na próxima seção.

6.4 Decomposição dos Diferenciais de Salários segundo Oaxaca

Os resultados da tabela 6.3 são referentes às diferenças entre as médias das características produtivas e a diferença entre os coeficientes. Vale destacar que, por limitações operacionais o STATA não incorpora o procedimento de interação no Heckman incluindo a variável lambda. Desta forma, a hipótese adotada é que os dados são “paired data” entre os regressores, ou seja, não há um correspondente homem negro (ou l-ésimo grupo de cor e sexo) para cada homem branco (grupo base). Seja:

$$Var(\underline{\beta}^{(L)} - \underline{\beta}^{(l)}) = Var(\underline{\beta}^{(L)}) - Var(\underline{\beta}^{(l)}) - 2Cov(\underline{\beta}^{(L)} - \underline{\beta}^{(l)}) \quad (6.22)$$

Onde:

$\underline{\beta}^{(L)}$ = coeficiente do grupo base

$\underline{\beta}^{(l)}$ = coeficiente do l-ésimo grupo de cor e sexo.

Assumindo que:

$$Cov(\underline{\beta}^{(L)} - \underline{\beta}^{(l)})$$

Tabela 6.3 – Diferenças entre as médias dos regressores e coeficientes estimados com a correção de Heckman.

	(Grupo padrão) - (I-ésimo grupo de cor/sexo)					
	Entre as médias			Entre os coeficientes estimados		
	Mulher de cor branca	Mulher de cor preta ou parda	Homem de cor preta ou parda	Mulher de cor branca	Mulher de cor preta ou parda	Homem de cor preta ou parda
Anos de estudo	-1,32	0,80	2,32	-0,0537	0,0187	-0,0064
	-36,17	15,32	44,55	-4,6193	1,9769	-0,5544
Anos de estudo ²	-21,83	33,41	33,41	0,0116	0,0095	0,0115
	-36,92	15,87	44,21	0,0026	-0,0026	-0,0005
Experiência	2,84	-0,71	-0,71	0,0006	0,0006	0,0007
	23,89	9,92	-5,35	0,0013	-0,0056	0,0090
Experiência ²	159,82	-40,09	-40,09	0,0040	0,0030	0,0035
	23,66	11,11	-4,91	-0,0001	0,0000	-0,0002
Área Urbana	-0,07	0,06	0,06	0,0001	0,0000	0,0001
	-21,54	-5,46	9,77	-0,1420	0,0542	0,0143
Lambda	-0,36	-0,42	-0,42	0,0507	0,0416	0,0553
				0,2388	-0,0114	0,2547
Constante				3,9667	-0,3489	5,8578
		-	-	0,0602	0,0326	0,0435
				0,0216	-0,60614	-0,1305

Na tabela 6.4, o diferencial observado entre homens de cor branca e mulheres de cor preta/parda o efeito da discriminação explica 97% contra apenas 3% das características pessoais. Entre homens e mulheres de cor branca o efeito da discriminação explica 175% do diferencial de salários.

Tabela 6.4 – Efeitos da Discriminação Estimados pelas Características Pessoas –
Correção de Heckman

	Efeitos	Mulher branca		Homem preto		Mulher preta	
		(1)a	(2)b	(1)a	(2)b	(1)a	(2)b
	Diferencial de salários = (3)	0,152	100,0%	0,593	100,0%	0,688	100,0%
Efeitos da características produtivas	Anos de estudo	0,0341	22,4%	0,1083	18,3%	0,00040551	0,05891%
	Anos de estudo2	-0,2139	-140,7%	0,1551	26,1%	-0,00001751	-0,00254%
	Experiência	0,0919	60,4%	-0,0180	-3,0%	-0,00022505	-0,03269%
	Experiência2	-0,0519	-34,1%	0,0095	1,6%	0,00000001	0,00000%
	Área Urbana	-0,0123	-8,1%	0,0231	3,9%	0,01795235	2,60798%
	Lambda	0,0382	25,1%	0,1488	25,1%	0,00103260	0,15001%
	Somatório do efeitos = (4)	-0,1139	-74,9%	0,4268	71,9%	0,0191	2,8%
Efeito Discriminação	Estimativa de $\ln(D+1)$ = (5)	0,27	174,9%	0,17	28,1%	0,67	97,2%
Coefficiente de Discriminação	Estimativa de D = (6)	0,30		0,18		0,95	

Notas: (a) é igual ao produto entre o coeficiente da variável explicativa e a diferença das características médias da tabela 5.5

(b) é igual o ajuste de (a) expresso como percentual da diferencial de salários.

(3) Representa o diferencial de salários entre o grupo padrão e o l-ésimo grupo de cor/sexo

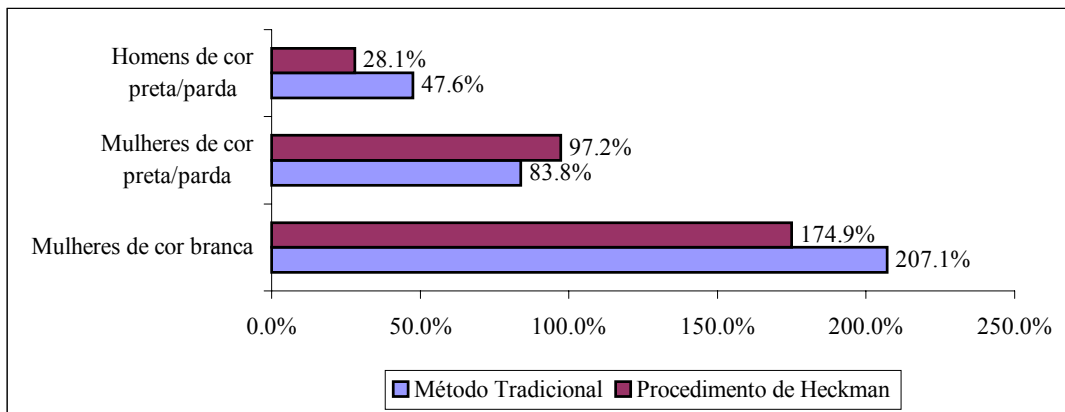
(4) Somatório dos efeitos sobre o diferencial de salários

(5) É igual a (3) - (4)

(6) É igual a exponencial do item (5)

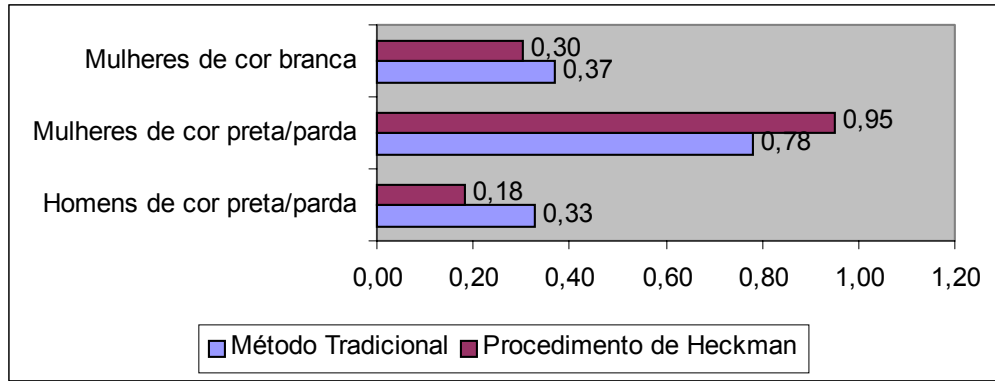
O gráfico 6.2 revela que se a decomposição fosse realizada pelo método tradicional a conclusão seria que 84% do diferencial entre homens de cor branca e mulheres de cor preta é provenientes da discriminação. No entanto, com a correção de Heckman este indicador é de 97%, ou seja, sem ela estaríamos subestimando o efeito da discriminação. Destaca-se o efeito da discriminação entre homens e mulheres de cor branca que caiu 7 pontos percentuais com a correção de Heckman.

Gráfico 6.2 - Efeito da Discriminação – Método Tradicional vs Heckman



A estimativa do coeficiente de discriminação, pelo método tradicional, estaria subestimada pelo método tradicional (gráfico 6.3), exceto na comparação entre os homens de cor branca e as mulheres de cor preta cujo coeficiente aumentou de 0.78 (método tradicional) para 0.95 (correção de Heckman).

Gráfico 6.3 – Estimativa do Coeficiente de Discriminação – Método Tradicional vs Procedimento de Heckman



7 Considerações Finais e Trabalhos Futuros

Em geral, as análises da equação de salários que utilizam os microdados da PNAD não incorporam a complexidade da pesquisa. Além disso, os resultados são utilizados para o estudo da decomposição do diferencial de salários. Contudo, esse procedimento pode não ser o mais adequado na avaliação da discriminação no mercado de trabalho brasileiro.

Esta dissertação avaliou que o procedimento de Heckman (detalhes no capítulo 6) corrige o viés de seleção das informações dos salários e é mais adequado para obter estimativas consistentes dos parâmetros da equação. Como exemplo, a estimativa do coeficiente de discriminação a partir da equação de salários sem correção, (0.37) para 0.30 (com correção) entre homens e mulheres de cor branca. A análise controlada da equação de participação no mercado de trabalho revela que quanto maior a escolaridade mais propensos os indivíduos estão para trabalhar. A discriminação é responsável por 97% do diferencial de salários entre homens de cor branca e as mulheres de cor preta ou parda.

Registra-se como uma extensão natural deste trabalho a análise da decomposição do salário/hora para os diferentes percentis da distribuição do rendimento de trabalho. Esta idéia já foi realizada por Soares (2000) como uma extensão da metodologia de Oaxaca, porém sem incorporar o plano amostral na modelagem. O procedimento consiste em estimar equações de salários por percentis (regressões quantílicas) da distribuição do salário/hora com a correção do viés de seleção da informação do salário pelo procedimento de Heckman. O que se pretende é avaliar os efeitos da discriminação entre os 10% mais pobres, os 10% mais ricos e comparar os resultados com a média da população.

Referências Bibliográficas

COCHRAN, W.G. Sampling techniques. New York: John Wiley & Sons, 1977 3ª Edição. 428p.

DOBSON, A. An Introduction to Statistical Modelling. London: Chapman and Hall, 1996.

HECKMAN, J.J. & LOCHNER, L.J. Fifty Years of Mincer Earnings Regressions

HECKMAN, J.J. Sample Selections Bias as a Specification Error
Econometrica, vol. 47, nº 1, 1979.

HOFFMANN, R. e KASSOUF, A.L. Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure.

JACCOUND, L. e BEGHIN N. Desigualdades raciais no Brasil: um balanço
Intervenção governamental. Brasília: IPEA, 002.

JOHNSTON, J. and DINARDO, J. Econometrica Methods
The McGraW-Hill Companies – New York

KASSOUF, A. L. The Wage Rate Estimation Using The Heckman Procedure
Revista de Econometria Rio de Janeiro, v.14, nº1, p.89-107, 1994.

MINCER, J. Schooling, Experience, and Earnings. New York 1974

MAGGIE, Y. e REZENDE, C.B. Raça como retórica a construção da diferença
Rio de Janeiro: Civilização Brasileira, 2001.

MENEZES, W.F. e FILHO, L.M.B. Diferenciais de Rendimentos na
Ocupação Não Registrada de Salvador.

MENEZES, W.F., F., J.C. e DEDECCA, C. Diferenciais de Rendimentos do
Trabalho nas Regiões Metropolitanas de São Paulo e de Salvador.

OAXACA, R. Male-Female Wage Differentials in Urban Labor Market.
In International Economic Review, v. 14, n.3, p. 693-709. 1973

OLIVEIRA, L.E.G.; PORCARO, R.M. e COSTA, T.C.N.A O lugar do Negro
na força de trabalho. Rio de Janeiro: IBGFE, 193, 86P.

PESSOA, D.G.C.; SILVA, P.L.N. Análise de dados amostrais complexos. São Paulo: ABE, 1998. (13º Simpósio Nacional de Probabilidade e Estatística – Caxambu) 187p.

RAMOS, L.R.A. A Distribuição de Rendimentos no Brasil 1976/85. Rio de Janeiro, IPEA, 1993. 135 p.

RODRIGUES, S.C. Análise da Estrutura Salarial Revelada pela PPV Incorporando peso e plano amostral. Rio de Janeiro, 2003. (Dissertação de Mestrado em Estudos Populacionais e Pesquisas Sociais).

SKINNER, C.J.; HOLT, D e SMITH, T.M.F. Analysis of Complex Surveys. Chichester: John Wiley & Sons, 1989.

SOARES, S.S.D. O perfil da discriminação no mercado de trabalho – homens negros, mulheres brancas e mulheres negras. Brasília: IPEA, 2000. (Texto para discussão nº 769)

SOARES, S.S.D. Simulando – O Mundo Maravilhoso das Distribuições Contrafactuais. Brasília: IPEA, 2001. (Texto para discussão nº 780)

Anexo 1 – Capítulo 1

O algoritmo a seguir foi processado pelo SAS e tem por objetivo identificar estratos com apenas uma unidade primário de amostragem e novas construções.

```
DATA DOMICS;
  INFILE 'C:\TEMP_MDADOS\PNAD2003\DOM2003.TXT' LRECL=182 MISSEVER;

  INPUT

    @00005 NDOM    11. /* IDENTIFICAÇÃO DO DOMICILIO    */

          @00005 UF      2. /* UNIDADE DA FEDERAÇÃO      */

          @00166 UPA     3. /* DELIMITAÇÃO DO MUNICÍPIO      */
    @00169 STRAT   7. /* IDENTIFICAÇÃO DE AUTO E NÃO AUTO */
    @00176 PSU     7. /* UNIDADE PRIMÁRIA DE AMOSTRAGEM */

    @00005 V0102   8. /* NÚMERO DE CONTROLE          */
    @00013 V0103   3. /* NÚMERO DE SÉRIE            */
    @00016 V0104   2. /* TIPO DE ENTREVISTA         */
    @00018 V0105   2. /* TOTAL DE MORADORES         */
    @00020 V0106   2. /* TOTAL MORADORES 10 ANOS OU + */

;

*FILTRO PARA AS ENTREVISTAS REALIZADAS;
IF V0104=1;

Select;
when (strat=110006) nstrat=119999;
when (strat=110007) nstrat=119999;
when (strat=110009) nstrat=119999;
when (strat=130005) nstrat=130004;
when (strat=130007) nstrat=130006;
when (strat=140004) nstrat=140003;
when (strat=150011) nstrat=159999;
when (strat=150014) nstrat=159999;
when (strat=150018) nstrat=159999;
when (strat=160003) nstrat=169999;
when (strat=160005) nstrat=169999;
```

when (strat=170004) nstrat=179999;

when (strat=170005) nstrat=179999;
when (strat=170007) nstrat=179999;
when (strat=170012) nstrat=179999;
when (strat=170014) nstrat=179999;
when (strat=210003) nstrat=219999;
when (strat=210005) nstrat=219999;
when (strat=220002) nstrat=220001;
when (strat=230015) nstrat=239999;
when (strat=230016) nstrat=239999;
when (strat=230017) nstrat=239999;
when (strat=230019) nstrat=239999;
when (strat=230020) nstrat=239999;
when (strat=230022) nstrat=239999;
when (strat=230029) nstrat=239999;
when (strat=230032) nstrat=239999;
when (strat=230035) nstrat=239999;
when (strat=230037) nstrat=239999;
when (strat=230039) nstrat=239999;
when (strat=230041) nstrat=239999;
when (strat=240004) nstrat=249999;
when (strat=240005) nstrat=249999;
when (strat=240006) nstrat=249999;
when (strat=240012) nstrat=249999;
when (strat=240014) nstrat=249999;
when (strat=240016) nstrat=249999;
when (strat=250003) nstrat=259999;
when (strat=250004) nstrat=259999;
when (strat=250007) nstrat=259999;
when (strat=250010) nstrat=259999;
when (strat=250014) nstrat=259999;
when (strat=260016) nstrat=269999;
when (strat=260017) nstrat=269999;
when (strat=260018) nstrat=269999;
when (strat=260019) nstrat=269999;
when (strat=260020) nstrat=269999;
when (strat=260021) nstrat=269999;
when (strat=260031) nstrat=269999;
when (strat=260040) nstrat=269999;
when (strat=270003) nstrat=279999;

when (strat=270008) nstrat=279999;
when (strat=280004) nstrat=289999;
when (strat=280005) nstrat=289999;
when (strat=280006) nstrat=289999;
when (strat=280008) nstrat=289999;
when (strat=280010) nstrat=289999;
when (strat=280012) nstrat=289999;
when (strat=290019) nstrat=299999;
when (strat=290020) nstrat=299999;
when (strat=290021) nstrat=299999;
when (strat=290023) nstrat=299999;
when (strat=290025) nstrat=299999;
when (strat=290040) nstrat=299999;
when (strat=310026) nstrat=319999;
when (strat=310027) nstrat=319999;
when (strat=310028) nstrat=319999;
when (strat=310029) nstrat=319999;
when (strat=310031) nstrat=319999;
when (strat=310035) nstrat=319999;
when (strat=310036) nstrat=319999;
when (strat=310037) nstrat=319999;
when (strat=310038) nstrat=319999;
when (strat=310039) nstrat=319999;
when (strat=310047) nstrat=319999;
when (strat=310050) nstrat=319999;
when (strat=310054) nstrat=319999;
when (strat=310057) nstrat=319999;
when (strat=310059) nstrat=319999;
when (strat=310073) nstrat=319999;
when (strat=310076) nstrat=319999;
when (strat=310082) nstrat=319999;
when (strat=310090) nstrat=319999;
when (strat=310095) nstrat=319999;
when (strat=320008) nstrat=329999;
when (strat=320010) nstrat=329999;
when (strat=320015) nstrat=329999;
when (strat=330030) nstrat=339999;
when (strat=330031) nstrat=339999;
when (strat=330032) nstrat=339999;
when (strat=330033) nstrat=339999;
when (strat=330034) nstrat=339999;

when (strat=330035) nstrat=339999;
when (strat=330036) nstrat=339999;
when (strat=330037) nstrat=339999;
when (strat=330038) nstrat=339999;
when (strat=330039) nstrat=339999;
when (strat=330041) nstrat=339999;
when (strat=330042) nstrat=339999;
when (strat=330043) nstrat=339999;
when (strat=330044) nstrat=339999;
when (strat=330048) nstrat=339999;
when (strat=330053) nstrat=339999;
when (strat=350052) nstrat=359999;
when (strat=350054) nstrat=359999;
when (strat=350055) nstrat=359999;
when (strat=350056) nstrat=359999;
when (strat=350058) nstrat=359999;
when (strat=350060) nstrat=359999;
when (strat=350061) nstrat=359999;
when (strat=350062) nstrat=359999;
when (strat=350063) nstrat=359999;
when (strat=350064) nstrat=359999;
when (strat=350067) nstrat=359999;
when (strat=350069) nstrat=359999;
when (strat=350070) nstrat=359999;
when (strat=350071) nstrat=359999;
when (strat=350072) nstrat=359999;
when (strat=350073) nstrat=359999;
when (strat=350075) nstrat=359999;
when (strat=350076) nstrat=359999;
when (strat=350077) nstrat=359999;
when (strat=350084) nstrat=359999;
when (strat=350086) nstrat=359999;
when (strat=350097) nstrat=359999;
when (strat=350101) nstrat=359999;
when (strat=350105) nstrat=359999;
when (strat=350112) nstrat=359999;
when (strat=350118) nstrat=359999;
when (strat=350123) nstrat=359999;
when (strat=350126) nstrat=359999;
when (strat=410021) nstrat=419999;
when (strat=410022) nstrat=419999;

when (strat=410024) nstrat=419999;
when (strat=410025) nstrat=419999;
when (strat=410027) nstrat=419999;
when (strat=410029) nstrat=419999;
when (strat=410031) nstrat=419999;
when (strat=410033) nstrat=419999;
when (strat=410035) nstrat=419999;
when (strat=410040) nstrat=419999;
when (strat=410052) nstrat=419999;
when (strat=410057) nstrat=419999;
when (strat=410059) nstrat=419999;
when (strat=420009) nstrat=429999;
when (strat=420010) nstrat=429999;
when (strat=420012) nstrat=429999;
when (strat=420013) nstrat=429999;
when (strat=420014) nstrat=429999;
when (strat=420015) nstrat=429999;
when (strat=420019) nstrat=429999;
when (strat=420026) nstrat=429999;
when (strat=420028) nstrat=429999;
when (strat=430034) nstrat=439999;
when (strat=430035) nstrat=439999;
when (strat=430037) nstrat=439999;
when (strat=430038) nstrat=439999;
when (strat=430039) nstrat=439999;
when (strat=430042) nstrat=439999;
when (strat=430043) nstrat=439999;
when (strat=430044) nstrat=439999;
when (strat=430045) nstrat=439999;
when (strat=430049) nstrat=439999;
when (strat=430050) nstrat=439999;
when (strat=430051) nstrat=439999;
when (strat=430052) nstrat=439999;
when (strat=430056) nstrat=439999;
when (strat=430066) nstrat=439999;
when (strat=430076) nstrat=439999;
when (strat=500007) nstrat=509999;
when (strat=500008) nstrat=509999;
when (strat=500012) nstrat=509999;
when (strat=500016) nstrat=509999;
when (strat=500018) nstrat=509999;


```
when (strat=510007) nstrat=519999;  
when (strat=510011) nstrat=519999;  
when (strat=520015) nstrat=529999;  
when (strat=520016) nstrat=529999;  
when (strat=520017) nstrat=529999;  
when (strat=520020) nstrat=529999;  
when (strat=520021) nstrat=529999;  
when (strat=520026) nstrat=529999;  
when (strat=520035) nstrat=529999;
```

```
when (strat=170008) nstrat=179999;  
when (strat=170009) nstrat=179999;
```

```
when (strat=430031) nstrat=439999;  
when (strat=430054) nstrat=439999;
```

```
when (strat=500019) nstrat=509999;  
when (strat=500020) nstrat=509999;
```

```
when (strat=520022) nstrat=529999;  
when (strat=520023) nstrat=529999;  
otherwise      nstrat=strat;  
end;
```

```
RUN;
```

Anexo 2 – Capítulo 2

log using C:\Mestrado\Tese_Alex\STATA\log\PNAD2003_LOG_STATA_TESE_ALEXP_CAP1.log,
replace

```
set virtual on
set more off
set memory 200m
set matsize 4000
set maxvar 30000
#delimit ;
```

```
use "C:\Temp_Mdados\PNAD2003\br2003.dta", clear;
```

```
compress;
```

```
svyset [pweight=peso], strat(nstrat) psu(psu);
```

```
/***** TABELA 2.1 - Total por cor e sexo *****/
```

```
/*Total - Por cor e sexo*/
```

```
svytotal    tot_pes
            tot_pes_10
            pia
            pop_ocup
            pop_desemp,
            available meff by(cor_sexo);
```

```
/***** TABELA 2.2 - Estimador de razão *****/
```

```
svyratio pea/tot_pes_10
          pop_desemp/pea
          fpub/pop_ocup
          cc/pop_ocup
          sc/pop_ocup
          domes/pop_ocup
          cp/pop_ocup
          ep/pop_ocup,
          available meff by(cor_sexo);
```

```
/*Estimador de média por cor e sexo*/
```

```
svymean  salario jornada anos_estudo experiencia idade
          sal_hor ln_sal_hor
          sal_educ,
          complete meff by(cor_sexo);
```

```
/*Estimador do Diferencial Bruto */
```

```
lincom [sal_hor]1-[sal_hor]2;
lincom [sal_hor]1-[sal_hor]3;
lincom [sal_hor]1-[sal_hor]4;
```

```
lincom [ln_sal_hor]1-[ln_sal_hor]2;
lincom [ln_sal_hor]1-[ln_sal_hor]3;
lincom [ln_sal_hor]1-[ln_sal_hor]4;
```

```
/*Estimador do ln_sal_hor - Coeficiente de Discriminação*/
```

```
nlcom ([sal_hor]1/[sal_hor]2);
```

```
nlcom ([sal_hor]1/[sal_hor]3);
```

```
nlcom ([sal_hor]1/[sal_hor]4);
```

```
nlcom ([ln_sal_hor]1/[ln_sal_hor]2);
```

```
nlcom ([ln_sal_hor]1/[ln_sal_hor]3);
```

```
nlcom ([ln_sal_hor]1/[ln_sal_hor]4);
```

```
log close;
```

Anexo 3 – Capítulo 5

```
log using C:\MESTRADO\TESE_ALEX\STATA\log\PNAD2003_LOG_STATA_CAP2_mincer.log, replace
```

```
set virtual on
set more off
set memory 200m
set matsize 4000
set maxvar 30000
#delimit ;
```

```
use "C:\Temp_Mdados\PNAD2003\br2003.dta", clear;
```

```
compress;
```

```
svyset [pweight=peso], strat(nstrat) psu(psu);
```

```
rename ln_sal_hor      y;
rename cor_sexo        cs;
    rename anos_estudo  es;
    rename anos_estudo2 es2;
rename idade           id;
rename idade2          id2;
    rename metrop       me;
rename experiencia     ex;
rename experiencia2    ex2;
```

```
/*Modelo clássico - Mincer(1)
svyregress y es es2 ex ex2 situ, prob meff; */
```

```
/*Modelo clássico - Mincer(2) */
xi: svyregress y es es2 ex ex2 situ i.cs, prob meff;
```

```
/*Regressão - subpopulação=Homens + Cor branca
svyregress y es es2 ex ex2 situ, subpop(ho_br) prob meff;*/
```

```
/*Regressão - subpopulação=Homens + Cor preta
svyregress y es es2 ex ex2 situ, subpop(ho_pr) prob meff;*/
```

```
/*Regressão - subpopulação=Mulheres + cor branca
svyregress y es es2 ex ex2 situ, subpop(mu_br) prob meff;*/
```

```
/*Regressão - subpopulação=Mulheres + cor branca - Interação na escolaridade
xi: svyregress y i.es es2 ex ex2 situ, subpop(mu_br) prob meff;*/
```

```
/*Regressão - subpopulação=Mulheres + cor preta
svyregress y es es2 ex ex2 situ, subpop(mu_pr) prob meff;*/
```

```
/*Regressão - Interação entre cor e sexo
xi: svyregress y
    i.cs*es i.cs*es2
    i.cs*ex i.cs*ex2
    i.cs*situ;*/
```

```
/*Médias - Cor e Sexo  
svymean es es2 ex ex2 situ salario sal_hor y , complete meff by(cs);
```

```
lincom [es]1-[es]2, meff;  
lincom [es]1-[es]3, meff;  
lincom [es]1-[es]4, meff;
```

```
lincom [es2]1-[es2]2, meff;  
lincom [es2]1-[es2]3, meff;  
lincom [es2]1-[es2]4, meff;
```

```
lincom [ex]1-[ex]2, meff;  
lincom [ex]1-[ex]3, meff;  
lincom [ex]1-[ex]4, meff;
```

```
lincom [ex2]1-[ex2]2, meff;  
lincom [ex2]1-[ex2]3, meff;  
lincom [ex2]1-[ex2]4, meff;
```

```
lincom [situ]1-[situ]2, meff;  
lincom [situ]1-[situ]3, meff;  
lincom [situ]1-[situ]4, meff;*/
```

```
log close;
```

Anexo 4 – Capítulo 6

```
log using C:\MESTRADO\TESE_ALEX\STATA\log\PNAD2003_LOG_STATA_CAP3_lixo.log, replace
set virtual on
set more off
set memory 200m
set matsize 4000
set maxvar 30000
#delimit ;
```

```
use "C:\Temp_Mdados\PNAD2003\br2003.dta", clear;
compress;
```

```
/*Procedimento de Heckman - Incorporando os pesos e plano amostral*/
svyset [pweight=peso], strat(nstrat) psu(psu);
```

```
rename ln_sal_hor          y;
rename cor_sexo           cs;
    rename anos_estudo     es;
    rename anos_estudo2   es2;
rename idade              id;
rename idade2            id2;
    rename metrop          me;
rename experiencia        ex;
rename experiencia2      ex2;
```

```
/*Heckman com desenho - Equação de seleção dos homens de cor branca
```

```
svyheckman y es es2 ex ex2 situ,
select(      es es2 ex ex2 situ
            chefe filho dchild_0a5)
subpop(ho_br) prob meff;*/
```

```
svyheckman y es es2 ex ex2 situ,
select(      es es2 ex ex2 situ chefe filho dchild_0a5)
mills(lambda_hb)
subpop(ho_br) prob meff;
```

```
svymean lambda_hb, subpop(ho_br);
svymean lambda_hb;
```

```
mean lambda_hd;
```

```
/*Heckman com desenho - Equação de seleção dos homens de cor preta ou parda
```

```
svyheckman y es es2 ex ex2 situ,
select(      es es2 ex ex2 situ chefe filho dchild_0a5)
mills(lambda_hp)
subpop(ho_pr) prob meff;
```

```
svymean lambda_hp, subpop(ho_pr);*/
```

```
/*Heckman com desenho - Equação de seleção das mulheres de cor branca
```

```
svyheckman y es es2 ex ex2 situ,
select(      es es2 ex ex2 situ chefe filho dchild_0a5)
mills(lambda_mb)
subpop(mu_br) prob meff;
```

```
svymean lambda_mb, subpop(mu_br);*/
```

```

/*Heckman com desenho - Equação de seleção das mulheres de cor preta ou parda
svyheckman y es es2 ex ex2 situ,
select(      es es2 ex ex2 situ  chefe filho dchild_0a5)
mills(lambda_mp)
subpop(mu_pr) prob meff;

svymean lambda_mp, subpop(mu_pr);

rename chefe      ch;
rename dchild_0a5 chi;*/

```

```

/*Heckman com desenho - Equação de seleção Interação cor e sexo
xi: svyheckman y
      i.cs*es
      i.cs*es2
      i.cs*ex
      i.cs*ex2
      i.cs*situ,
select(
      i.cs*es          i.cs*es2
      i.cs*ex
      i.cs*ex2
      i.cs*situ
      i.cs*ch
      i.cs*filho
      i.cs*chi
      )
mills(lambda_t)
i.cs*lambda_t
prob meff;*/

```

```
log close;
```

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)