ipea Instituto de Pesquisa Econômica Aplicada

Título do capítulo	**CAPÍTULO 4** **Avaliando a significância estatística da queda na desigualdade no Brasil**
Autores(as)	João Pedro Azevedo
DOI	

Título do livro	**DESIGUALDADE DE RENDA NO BRASIL: UMA ANÁLISE DA QUEDA RECENTE: UMA ANÁLISE DA QUEDA RECENTE**
Organizadores(as)	Ricardo Paes de Barros Miguel Nathan Foguel Gabriel Ulyssea
Volume	1
Série	-
Cidade	Brasília
Editora	Ipea
Ano	2006
Edição	-
ISBN	858617087-9
DOI	

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2006

As publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos). Acesse: http://www.ipea.gov.br/ portal/publicacoes

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério do Planejamento, Desenvolvimento e Gestão.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

CAPÍTULO 4

Avaliando a Significância Estatística da Queda na Desigualdade no Brasil

João Pedro Azevedo*

1 INTRODUÇÃO**

Embora em um patamar ainda bastante alto, a concentração de renda no Brasil vem apresentando uma trajetória de queda contínua no período mais recente. Como mostra Ipea (2006),[1] várias medidas de desigualdade de renda declinaram de forma expressiva, de fato, ao longo da primeira metade da década atual. A motivação para a elaboração do presente estudo foi, portanto, contribuir para a identificação do grau de significância estatística dessa queda.

A comparação de medidas de desigualdade ao longo do tempo exige que suas estimativas venham acompanhadas da análise de respectivas variâncias. Muitas vezes a falta de informação sobre os erros padrões de tais medidas acaba por relegá-las a um papel descritivo, sem nenhuma informação relativa ao grau de significância estatística sobre eventuais diferenças.

Um dos principais problemas do cálculo da variância dessas estimativas de desigualdade é a sua complexidade. Tais estimativas são funções não-lineares das observações, e muitas vezes dependem do ordenamento dos dados. Além disso, comumente as informações utilizadas para estimar essas medidas são geradas a partir de desenhos amostrais complexos, ou seja, de amostras probabilistísticas estratificadas – com múltiplos estágios de seleção – e conglomeradas. Com estratificação, a probabilidade de seleção das observações freqüentemente não é a mesma para todas as observações, violando, assim, a hipótese de i.i.d. Essa mesma hipótese é também violada quando há conglomeração das observações, uma vez que isso implica a dependência

*Pesquisador associado do Instituto de Pesquisa Econômica Aplicada (Ipea).

** Todos os resultados empíricos deste capítulo foram elaborados pelo autor com base nos dados da Pesquisa Nacional por Amostra de Domicílios (Pnad) do Instituto Brasileiro de Geografia e Estatística (IBGE).

[1] Nota técnica elaborada pelo Ipea, a qual é publicada neste livro com o mesmo título disponível na página do Ipea na internet.

de algumas observações. Por conta desses problemas, a variância das estimativas de desigualdade não pode ser expressa de maneira simples,[2] o que torna necessária a utilização de técnicas que aproximem a estimativa dessa variância.

Este artigo aplica a metodologia proposta por Binder e Kovačević (1995; 1997) para estimar o erro padrão de algumas medidas de desigualdade de amostras complexas, com base na linearização de Taylor via *Estimating Equations* (EE) – (Binder, 1991) e (Binder; Patak, 1994). Estudos realizados pelo *Statistics Canada* sugerem que o método proposto por Binder e Kovačević funcione tão bem quanto o *bootstrap*, tendo ainda a vantagem de não ser computacionalmente tão intensivo (Kovačević et al., 1995).

Além desta breve introdução, este capítulo contém outras três seções discursivas. A segunda dedica-se a relembrar a metodologia de cálculo da curva de Lorenz e do índice de Gini, bem como a apresentar o método de linearização de Taylor para a estimativa de erros padrões. A terceira expõe o resultado da aplicação dessa metodologia nos dados da Pesquisa Nacional por Amostragem de Domicílios (Pnad) de 2001, de 2004 e de 2005, e apresenta o grau de significância estatística da queda de desigualdade no Brasil. Já a quarta seção, essa resume as principais conclusões deste capítulo.

2 DESIGUALDADE DE RENDA E SEUS ERROS PADRÕES

2.1 Curva de Lorenz

A curva de Lorenz é um dos principais métodos para descrição e análise de distribuições de renda. Mostra, simplesmente, como a proporção da renda total aumenta em razão da proporção da população, considerando rendas crescentes *per capita*. Se todas as pessoas tivessem exatamente a mesma renda, a proporção acumulada da renda seria sempre igual à proporção acumulada da população, o que configuraria uma situação de "perfeita igualdade". Em contrapartida, se toda a renda fosse apropriada por uma única pessoa, e as demais não recebessem absolutamente nada, configurar-se-ia uma situação de "prefeita desigualdade". A formalização da curva de Lorenz pode ser dada por

$$L(p) = \frac{1}{N\mu} \sum_{U} y_i I\{y_i \leq \xi_p\}, \quad 0 \leq p \leq 1, \tag{1}$$

[2] Vale notar que a negligência do desenho amostral tende a levar à superestimativa da variância de medidas de desigualdade como a curva de Lorenz (Zheng, 2002, p. 1235-1243).

em que μ é a renda média da população, $I(.)$ denota uma função indicador, e ξ_p é o quantil de renda da população p.

Para estimar a variância dessa medida de desigualdade, Kovačević e Binder (1997) propõem um método de linearização de Taylor por meio do EE, método esse para cuja utilização as ordenadas da curva de Lorenz devem ser expressas como a solução de um sistema de equações do tipo

$$\begin{cases} \sum_U [I\{y_i \le \xi_p\} - L(p)]y_i & = 0 \\ \sum_U [I\{y_i \le \xi_p\} - p] & = 0 \quad \text{para } 0 \le p \le 1. \end{cases} \tag{2}$$

A variância da curva de Lorenz é estimada pela expressão

$$var(\hat{\theta}) = var(\hat{\theta} - \theta_N) \approx var\left(\sum_s w_i u^*(y_i, \theta_N) \right), \tag{3}$$

em que u^* é dada por

$$u_i^* = \frac{1}{\hat{N}\hat{\mu}}[(y_i - \xi_p)I\{y_i \le \xi_p\} + p\xi_p - y_i\hat{L}(p)]. \tag{4}$$

Mais detalhes sobre a derivação da expressão anterior, podem ser encontradas em Binder e Kovačević (1995).

2.2 Índice de Gini

Uma das medidas mais populares de desigualdade de renda, o índice de Gini pode ser definido como a área de uma curva de Lorenz padronizada; isto é, a razão entre a área obtida e a maior área possível da curva de Lorenz (que é ½). Assim, o índice de Gini assume o valor [0,1]. Segundo Glasser (1962), a forma de cálculo para populações finitas pode ser dada por

$$G = \frac{1}{N}\sum_U (2F_i - 1)y_i/\mu, \tag{5}$$

em que $F_i = F(y_i) = (1/N)\sum_{j \in U} I\{y_j \le y_i\}$ é o valor função de distribuição para população finita em y_i. Logo, o índice de Gini pode ser definido como solução da equação:

$$U_1(G,\{F_i\}_{i\in U},\mu)=\sum_U[(2F_i-1)y_i/\mu-G]=0, \quad (6)$$

na qual o parâmetro $\lambda=\{\{F_i\}_{i\in U},\mu\}$ é a solução do sistema de equações

$$\begin{cases}\left\{\sum_{j\in U}[I\{y_j\le y_i\}-F_i\right\}_{i\in U}=0\\ \sum_U(y_i-\mu)=0,\end{cases} \quad (7)$$

em que o número de parâmetros desconhecidos é N. Como os valores de y são para uma população de tamanho N, é possível solucionar esse sistema de equações.

A estimativa do índice de Gini é a solução da equação (5), ou seja:

$$\hat{G}=\frac{1}{\hat{N}\hat{\mu}}\sum_s(2\hat{F}_i-1)y_i\,, \quad (8)$$

em que $\hat{\mu}$ e $\hat{F}_i$ são a solução do sistema de equações estimadas (7).

Estima-se a variância do índice de Gini pela expressão (3) , caso em que u^*

$$u_i^*=\frac{2}{\hat{N}\hat{\mu}}[\hat{A}(y_i)y_i+\hat{B}(y_i)-\frac{\hat{\mu}}{2}(\hat{G}+1)]\,, \quad (9)$$

em que $\hat{A}(y)=\hat{F}(y)-(\hat{G}+1)/2$ e $\hat{B}(y)=\sum_s w_i y_i I\{y_i\ge y\}/\hat{N}$. Kovačević e Binder (1997) contém maiores detalhes.

3 A QUEDA DA DESIGUALDADE DE RENDA NO BRASIL

Aplicou-se o método descrito na seção anterior para estimar os erros padrões das estimativas da curva de Lorenz e do índice de Gini, utilizando, para isso, dados das Pnads de 2001, de 2004 e de 2005. Trata-se a Pnad de uma pesquisa anual por amostragem probabilística de domicílios, realizada em todo o território nacional, exclusive a área rural da Região Norte (a partir de 2004 o Norte rural foi incluído na pesquisa), cuja população-alvo é composta por domicílios de sua área de abrangência, bem como das pessoas neles

residentes. A Pnad adota um plano amostral estratificado e conglomerado com um, dois ou três estágios de seleção, em conformidade com o estrato (Silva et al., 2002).

Anualmente são entrevistados cerca de 105 mil domicílios, obtendo-se, dessa forma, informação detalhada sobre diversas fontes de renda de todos os seus membros. As medidas de desigualdade apresentadas neste capítulo foram calculadas a partir da renda domiciliar *per capita* de tais domicílios, os quais se situam em cerca de 600 estratos e 5 mil unidades primárias de amostragem. Para o número exato de domicílios, de estratos e de conglomerados entrevistados nesses três anos, observe-se a tabela 1. Vale destacar que neste exercício os domicílios localizados no Norte rural de 2004 e 2005 foram excluídos das estimativas.

Associada a esses domicílios, criou-se também uma variável que apresenta a soma dos pesos de todos os indivíduos de cada um deles. Esses dados são de fundamental importância para a estimativa pontual das medidas de desigualdade e respectivos erros padrões.

TABELA 1
Número de observações, de estratos e de Unidades Primárias de Amostragem (UPA) da Pnad

	2001	2004	2005
Domicílios	101,316	107,952	112,048
UPA	5,116	5,428	5,476
Estrato	545	793	838
Estrato[1]	545	615	586

Nota:[1] É importante notar que, entre 2002 e 2005, alguns estratos foram criados, em particular para comportar o surgimento de novas construções. Grande parte desses estratos conta apenas com um conglomerado, impossibilitando, com isso, a estimativa de sua variância. Assim, adotou-se o procedimento de agregar os estratos mais próximos com apenas um conglomerado. Maiores detalhes a esse respeito podem ser obtidos diretamente com o autor.

A seguir são apresentadas as estimativas pontuais, com respectivos erros padrões, para a curva de Lorenz e do índice de Gini.

3.1 Curva de Lorenz e índice de Gini

A tabela 2 apresenta as estimativas da curva de Lorenz e do índice de Gini, bem como seu intervalo de confiança, de 95%, para 2001, 2004 e 2005. Os gráficos 1 e 2 apresentam a comparação da curva de Lorenz para, respectivamente, os anos de 2001 e 2005, e de 2001 e 2004.

TABELA 2

Estimativas da renda acumulada e respectivo erro padrão

p	Estimativa	Erro Padrão	Z	P-valor	Intervalo de confiança 95%	
2001						
1	0.007	0.006	1.19	0.236	-0.005	0.019
2	0.024	0.006	4.02	0.000	0.012	0.035
3	0.049	0.006	8.27	0.000	0.037	0.060
4	0.083	0.006	14.26	0.000	0.071	0.094
5	0.128	0.006	22.42	0.000	0.116	0.139
6	0.186	0.006	33.60	0.000	0.175	0.196
7	0.262	0.005	49.49	0.000	0.251	0.272
8	0.369	0.005	75.60	0.000	0.359	0.378
9	0.528	0.004	126.82	0.000	0.520	0.536
10	1.000					
Gini	0.593	0.003	195.44	0.000	0.588	0.599
2004						
1	0.009	0.006	1.48	0.140	-0.003	0.020
2	0.028	0.006	4.78	0.000	0.017	0.039
3	0.056	0.006	9.61	0.000	0.044	0.067
4	0.094	0.006	16.52	0.000	0.083	0.105
5	0.141	0.006	25.26	0.000	0.130	0.152
6	0.202	0.005	37.45	0.000	0.192	0.213
7	0.281	0.005	54.74	0.000	0.271	0.291
8	0.392	0.005	82.36	0.000	0.383	0.401
9	0.551	0.004	133.59	0.000	0.543	0.559
10	1.000					
Gini	0.569	0.003	195.49	0.000	0.563	0.574
2005						
1	0.009	0.006	1.43	0.153	-0.003	0.022
2	0.029	0.006	4.58	0.000	0.016	0.041
3	0.057	0.006	9.28	0.000	0.045	0.069
4	0.094	0.006	15.85	0.000	0.083	0.106
5	0.143	0.006	24.99	0.000	0.132	0.154
6	0.209	0.005	38.12	0.000	0.198	0.220
7	0.290	0.005	55.86	0.000	0.280	0.300
8	0.390	0.005	79.33	0.000	0.381	0.400
9	0.551	0.005	119.03	0.000	0.542	0.560
10	1.000					
Gini	0.566	0.003	183.21	0.000	0.560	0.572

Como pode ser observado, a curva de Lorenz apresenta estimativas estatisticamente significativas, exceto no primeiro decil da população (p=1). Em relação ao índice de Gini, as estimativas pontuais foram 0,593, 0,569 e 0,566 para 2001, 2004 e 2005, respectivamente. Já o erro padrão, esse foi de, aproximadamente, 0,003, em todos os anos.

Um outro exercício analisou o erro padrão da variação das curvas de Lorenz de 2001 a 2005, bem como de 2001 a 2004, para cada centil da população acumulada. Além de apresentarem tal comparação, os gráficos mostram também que, tanto no período 2001-2005 (gráfico 3) quanto no período 2001-2004 (gráfico 4), essas curvas foram estatisticamente diferentes. Vale destacar que a variação da porcentagem da renda acumulada se deu a partir do 40º centil da população.

GRÁFICO 1
Curva de Lorenz – 2001 e 2005

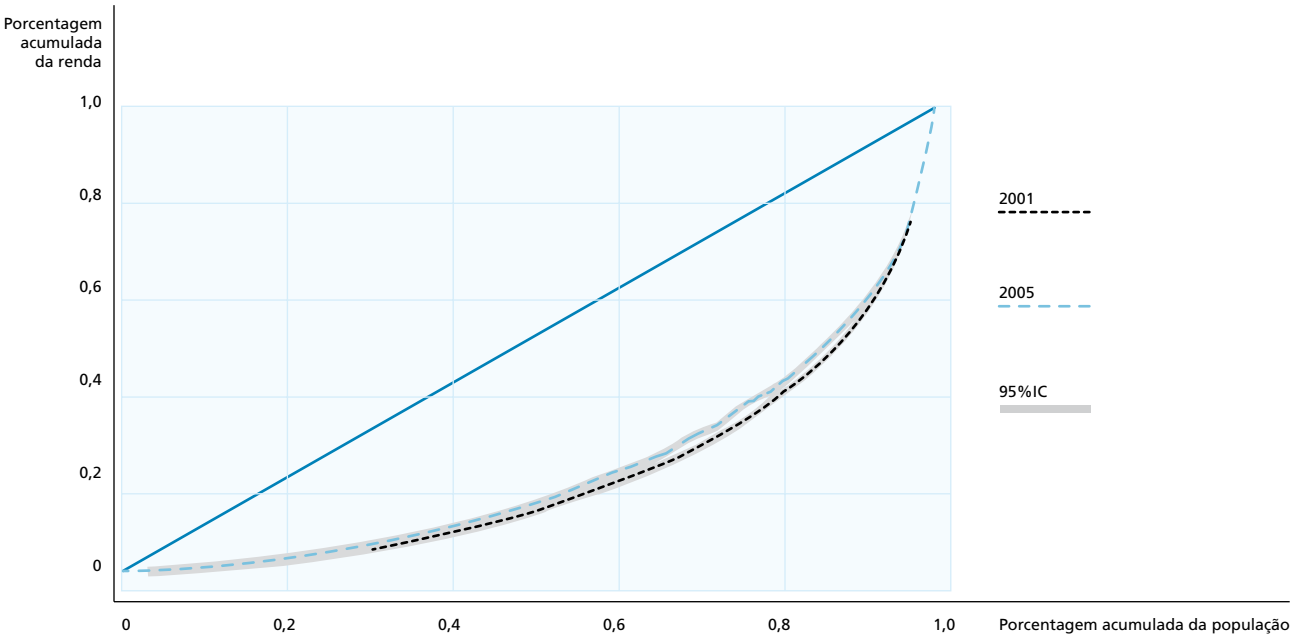


GRÁFICO 2
Curva de Lorenz – 2001 e 2004

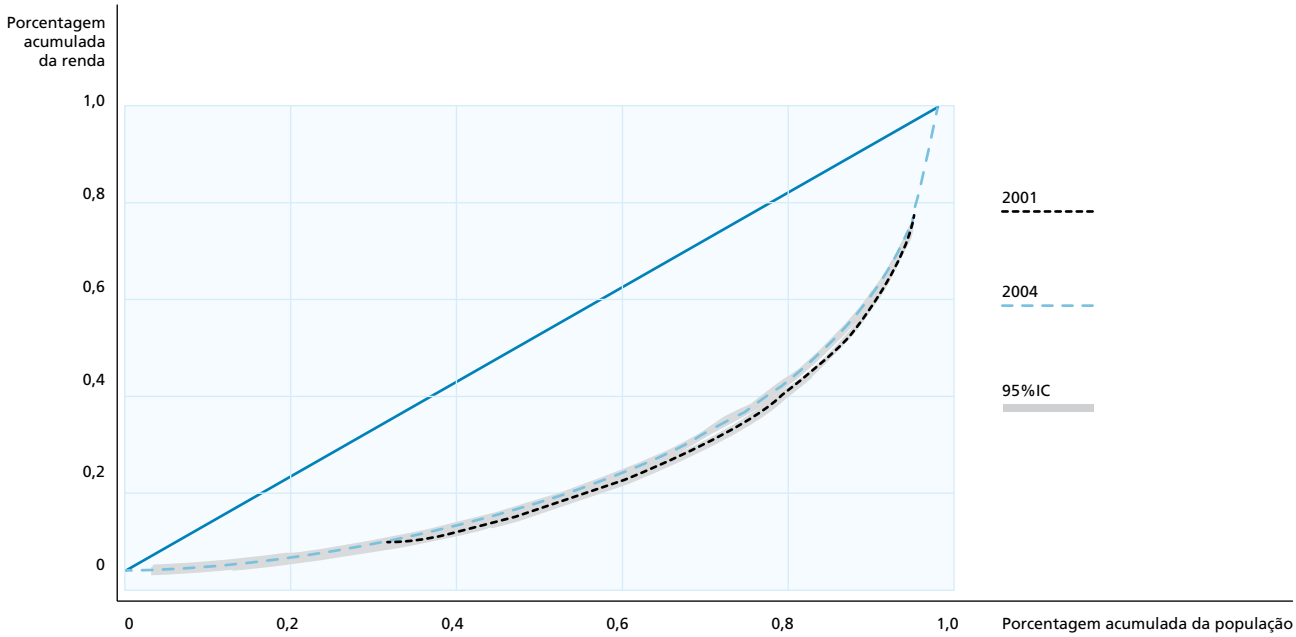


GRÁFICO 3
Diferença das Curvas de Lorenz – 2001 e 2005

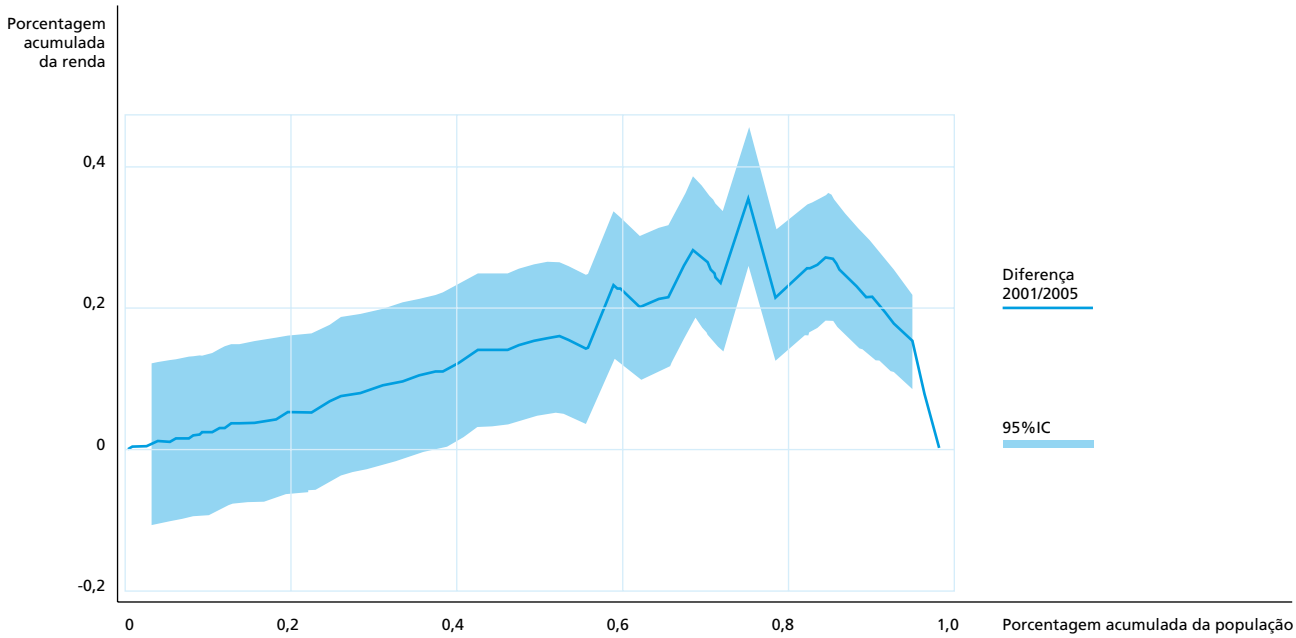
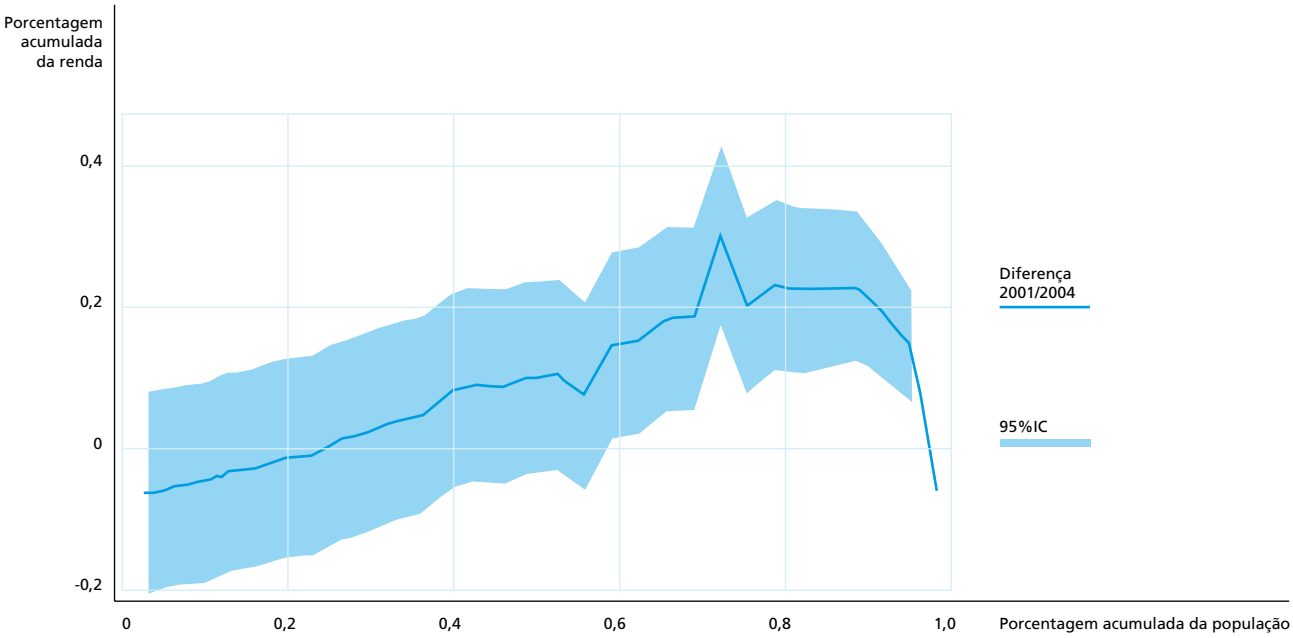


GRÁFICO 4
Diferença das Curvas de Lorenz –2001 e 2004



3.2 Variação da queda da desigualdade

Esta subseção apresenta o grau de significância estatística da queda da desigualdade entre 2001 e 2005 (tabela 3), bem como entre 2001 e 2004 (tabela 4). Observa-se, de 2001 a 2005, uma queda de 0,027 do índice de Gini. Assumindo-se a inexistência de co-variância entre as observações entre anos da Pnad,[3] pela simples propagação dos erros encontra-se uma estimativa do erro padrão da diferença de 0,0043.[4] Tal estimativa gera uma estatística-T de 6.3, e permite rejeitar a hipótese de que a diferença das estimativas do índice de Gini seja igual a zero, com 99% de confiança (tabela 3). O mesmo exercício foi feito entre 2001 e 2004, e encontrou-se um resultado similar (tabela 4). Vale ressaltar que a estimativa do erro padrão de todos os índices de Gini foi obtida também por *bootstrap* (300 replicações), e os resultados foram qualitativamente equivalentes.

[3] Essa não é uma hipótese leve, tendo em vista que os setores censitários visitados pela Pnad em uma mesma década são os mesmos. Não obstante, a inclusão de uma co-variância entre os anos da Pnad faria que o erro padrão dessas diferenças fosse ainda menor. Assim, pode-se interpretar o erro padrão obtido como o limite superior dessa estimativa.

[4] $erro = \sqrt{erro_1^2 + erro_2^2}$

TABELA 3
Significância estatística da queda recente na desigualdade de renda do Brasil – 2001-2005

	Estimativa		Erro Padrão		Redução na desigualdade			
	2001	2005	2001	2005	Estimativa	Erro padrão	Estatística T	Pval
Estimating Equations								
Gini	0.593	0.566	0.0030	0.0031	0.027	0.0043	6.315	0.0
Bootstrap (300 rep.)								
Gini	0.587	0.565	0.0030	0.0030	0.027	0.0042	6.477	0.0

TABELA 4
Significância estatística da queda recente na desigualdade de renda do Brasil – 2001-2004

	Estimativa		Erro Padrão		Redução na desigualdade			
	2001	2004	2001	2004	Estimativa	Erro padrão	Estatística T	Pval
Estimating Equations								
Gini	0.593	0.569	0.0030	0.0029	0.025	0.0042	5.877	0.0
Bootstrap (300 rep.)								
Gini	0.587	0.565	0.0030	0.0024	0.025	0.0038	6.494	0.0

4 CONCLUSÃO

Aplicou-se, aqui, a metodologia proposta por Binder e Kovačević (1995; 1997) para estimar o erro padrão da curva de Lorenz e do índice de Gini, considerando-se, para isso, o desenho amostral da Pnad (pesquisa probabilística estratificada e conglomerada com um, dois ou três estágios de seleção). Foram estimadas, ao todo, as curvas de Lorenz e os índices de Gini da renda domiciliar *per capita* brasileira para 2001, 2004 e 2005.

Com base nisso, conclui-se que a queda da desigualdade brasileira é estatisticamente significativa, tanto quando analisada no período 2001-2005 quanto no período 2001-2004, sendo possível rejeitar, portanto, a hipótese de a queda da desigualdade ser igual a zero, com 99% de confiança estatística. Vale destacar, ainda, que tais resultados foram corroborados com a estimativa, por *bootstrasp*, da variância das mesmas medidas de desigualdade.

Por último, a análise da diferença da desigualdade por meio da curva de Lorenz indica que a diferença da renda acumulada só foi estatisticamente diferente de zero a partir do 40º centil da população. Esse resultado sugere ter sido a diminuição da renda apropriada pelos 60% mais ricos da população aquilo que levou à redução da desigualdade no Brasil no período analisado.

5 REFERÊNCIAS

BINDER, D. A. Use of estimating functions for interval estimation from complex surveys. In: **Proceedings of the survey research methods section**, American Statistical Association, p. 34-42, 1991.

BINDER, D. A.; KOVAČEVIĆ, E. M. S. Estimating some measures of income inequality from survey data: an application of the estimating equations approach. **Survey methodology**, n. 21, p. 137-145, 1995.

BINDER, D. A.; PATAK, E. Z. Use of estimating functions for interval estimation from complex surveys. **Journal of the American Statistical Association**, v. 89, p. 1.035-1.043, 1994.

GLASSER, G.J. Variance formulas for the mean difference and coefficient of concentration. **Journal of the American Statistical Association**, v. 57, p. 648-654, 1962.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA – IPEA. **Sobre a recente queda da desigualdade de renda no Brasil.** Brasília: Ipea, ago. 2006. (Nota Técnica). Disponível também em versão eletrônica em: <http://www.ipea.gov.br>. Acesso em nov. 2006

KOVAČEVIĆ, E. M. S; BINDER, D. A. Variance estimation for measures of income inequality and polarization: the estimating equations approach. **Journal of Official Statistics**, v. 13, n. 1, p. 41-58, 1997.

KOVAČEVIĆ, E. M. S. et al. **Estimating the sampling variances of measures of income inequality and polarization:** an empirical study. Statistics Canada, Methodology Branch, 1995. (Working Paper, HSMD-95-007E).

SILVA, P.; PESSOA, D.; LILA, M. Análise estatística de dados da Pnad: incorporando a estrutura do plano amostral. **Ciência & Saúde Coletiva**, v. 7, n. 4, p. 659-670, 2002.

ZHENG, B. Testing Lorenz curves with non-simple random samples. **Econometrica**, v. 70, n. 3, p. 1.235-1.243, 2002.