

Medindo Desigualdade e Pobreza com a PNAD Contínua

Guilherme Jacob

ENCE/IBGE

07/11/2022

- 1 Medindo Desigualdade
 - a) Definição e Axiomas
 - b) Curva de Lorenz
 - c) (Algumas) Medidas de Desigualdade
- 2 Medindo Pobreza
 - a) Definição e Axiomas
 - b) Principais Medidas de Pobreza

- 3 PNAD Contínua
 - a) O que é a PNAD Contínua?
 - b) Conceitos Básicos de Amostragem
 - c) Plano amostral da PNAD Contínua
- 4 Prática com o R
 - a) Pacotes survey e convey
 - b) Objeto de plano amostral com svrepdesign
 - c) Estimación de medidas de desigualdade e pobreza
 - d) Principais recomendações

Parte 1

Medindo Desigualdade

O que são medidas de desigualdade?

- Uma medida de desigualdade tenta expressar a desigualdade em uma distribuição como um número real.
 - Completude: $x, y \in \mathbb{R}$, $x \geq y$ ou $x \leq y$;
 - A distância entre medidas de desigualdade de duas distribuições tem significado.

O que são medidas de desigualdade?

- A exposição a seguir segue uma abordagem axiomática (Chakravarty, 1999).
 - Avaliamos medidas de desigualdade de acordo com sua aderência a um conjunto de axiomas considerados úteis.
- Esta não é a única abordagem possível:
 - Abordagem Subjetiva (Amiel, 1999): opiniões sobre desigualdade;
 - Abordagem Estatística (Giorgi, 1999): distribuição amostral das medidas, estatística descritiva;
 - Abordagem Normativa (Blackorby, Bossert e Donaldson, 1999): funções de bem-estar.

(Alguns) Axiomas

- 1 Simetria
- 2 Princípio de Pigou-Dalton
- 3 Invariância Populacional
- 4 Invariância de Escala
- 5 Invariância de Translação
- 6 Decomponibilidade por Grupo

Simetria

- Sob simetria, se permutarmos a renda de duas unidades da população, a medida de desigualdade deve permanecer inalterada.
 - Ou seja: não importa quem recebe, mas o quanto recebe.
- Este princípio também é chamado de princípio da anonímia (ou anonimato).

Princípio de Pigou-Dalton (Dalton, 1920; Pigou, 1912)

- Considere uma distribuição de renda A ;
- Gere uma distribuição de renda A' com uma transferência regressiva.
 - Transferência regressiva: transferir parte da renda de alguém mais pobre para alguém mais rico.
 - ~~“o de cima sobe e o de baixo desce”, com uma quantia fixa~~
- Sob o Princípio (Forte) de Pigou-Dalton (PPD), $I(A) < I(A')$.
 - Ou seja: uma transferência regressiva aumenta a medida de desigualdade.
- Sob o Princípio Fraco de Pigou-Dalton, $I(A) \leq I(A')$.
 - Ou seja: uma transferência regressiva não pode diminuir a medida de desigualdade.

Princípio de Pigou-Dalton (Dalton, 1920; Pigou, 1912)

- Não é o único “princípio” para entender desigualdade:
 - Princípio de Hammond (Hammond, 1976);
 - Chateauneuf e Moyes (2005) para alternativas ao PPD.

Invariância Populacional

- Sob invariância populacional, se repetirmos todas as observações m vezes, o índice de desigualdade é o mesmo;
- Isso significa que a medida é independente do tamanho da população;
 - Podemos comparar populações de tamanhos diferentes.

Invariância Populacional

Exemplo:

- Considere a população $A = [1, 2, 3]$
- Repetindo cada valor 3 vezes, temos $A' = [1, 1, 1, 2, 2, 2, 3, 3, 3]$
- Se uma medida de desigualdade $I(\cdot)$ é população-invariante,
 $I(A) = I(A')$.

Invariância Populacional

Apesar da sua simplicidade, ele ajuda a separar medidas de desigualdade e medidas de concentração industrial:

- Índice de Hirschman-Herfindahl (Hirschman, 1964):
 - $H = \sum_{i=1}^N q_i^2$, ($q_i = y_i / \sum_{j=1}^N y_j$)
 - Não atende Invariância Populacional;
 - Além disso: adicionar outros $y_j = 0$ não muda o valor.

Útil para concentração industrial, mas não para desigualdade.

Invariância de Escala

- Sob invariância de escala, se multiplicarmos a renda de cada indivíduo por uma constante $\lambda > 0$, a medida de desigualdade é a mesma.

Exemplo:

- Considere a população $A = [1, 2, 3]$
- Multiplicando cada valor por $\lambda = 3$, temos $A' = [3, 6, 9]$
- Se uma medida de desigualdade $I(\cdot)$ é escala-invariante, $I(A) = I(A')$.

Invariância de Escala

- Neste caso, o que importa são as distâncias relativas:
 - Ou seja: as razões entre as rendas do mais ricos e dos mais pobres.
- Medidas que atendem a invariância de escala são chamadas medidas de *desigualdade relativa*.

Invariância de Translação

- Alternativamente, podemos pensar uma medida de desigualdade que não mude se quantidades fixas forem adicionadas;
- Sob invariância de translação, se adicionarmos uma constante qualquer $\delta \neq 0$ a todas as rendas, a medida de desigualdade permanece a mesma.
 - Exemplo: variância.

Invariância de Translação

- Neste caso, o que importa é a distância absoluta entre as rendas:
 - Ou seja: a diferença entre as rendas dos mais ricos e dos mais pobres;
- Medidas que atendem a invariância de translação são chamadas medidas de *desigualdade absoluta*.

Desigualdade Relativa vs. Desigualdade Absoluta

Atkinson e Brandolini (2010) mostram que a opção entre desigualdade relativa vs. desigualdade absoluta pode mudar o modo como vemos a evolução da desigualdade.

Desigualdade Relativa vs. Desigualdade Absoluta

FIGURE 2. Evolution of World Inequality, 1820–1992: Different Parameter Values

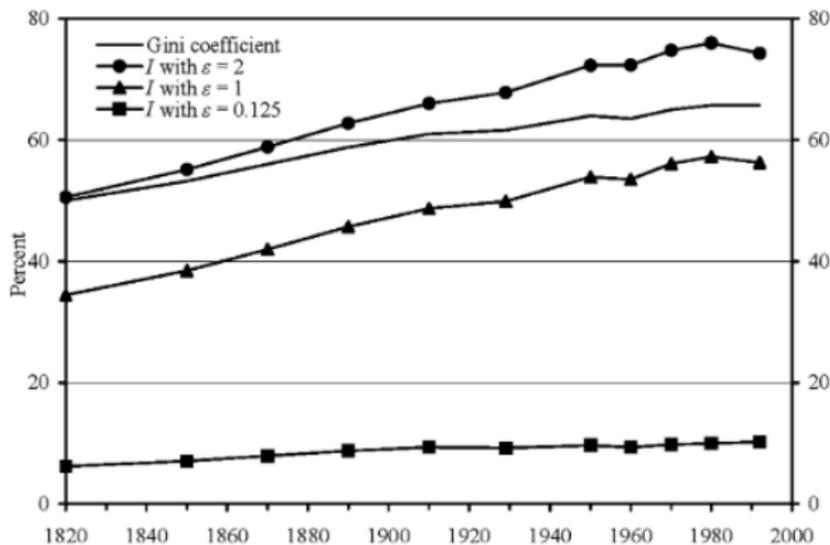


Figura 1: Série de medidas de desigualdade relativa

Desigualdade Relativa vs. Desigualdade Absoluta

FIGURE 3. Evolution of World Inequality, Absolute Measures, 1820–1992

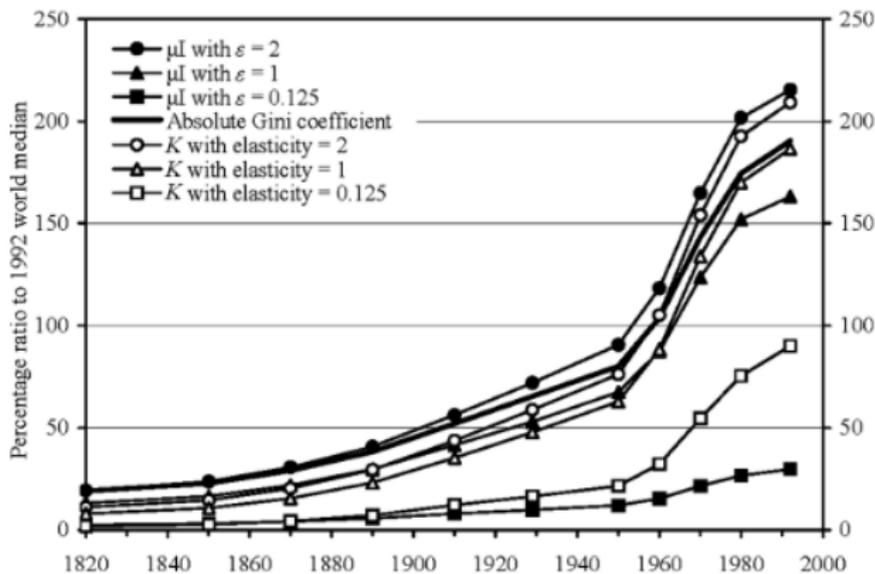


Figura 2: Série de medidas de desigualdade absoluta

Decomponibilidade por Grupo (Shorrocks, 1984)

- Sob este princípio, a medida de desigualdade pode ser decomposta em dois componentes:
 - Desigualdade Intra-grupos; e
 - Desigualdade Inter-Grupos.
- Isso ajuda a responder perguntas do tipo:
 - A desigualdade nesta região decorre da desigualdade dentro de cada município ou entre cada município?
 - Quanto da desigualdade total é atribuído à desigualdade dentro de cada setor ou entre setores?

Curva de Lorenz

- Curva de Lorenz: os $p\%$ menos ricos possuem $L(p)\%$ da renda total;
 - Por exemplo: a frase “os 20% mais pobres possuem 10% da renda total” é escrita como $L(20\%) = 10\%$.
- Por se basear em proporções, a curva de Lorenz:
 - Não depende do tamanho da população; e
 - Não depende da escala das rendas;
 - Podemos comparar distribuições com rendas totais diferentes.
- Quanto mais a curva observada se afasta da curva de igualdade perfeita, mais desigual é a distribuição.

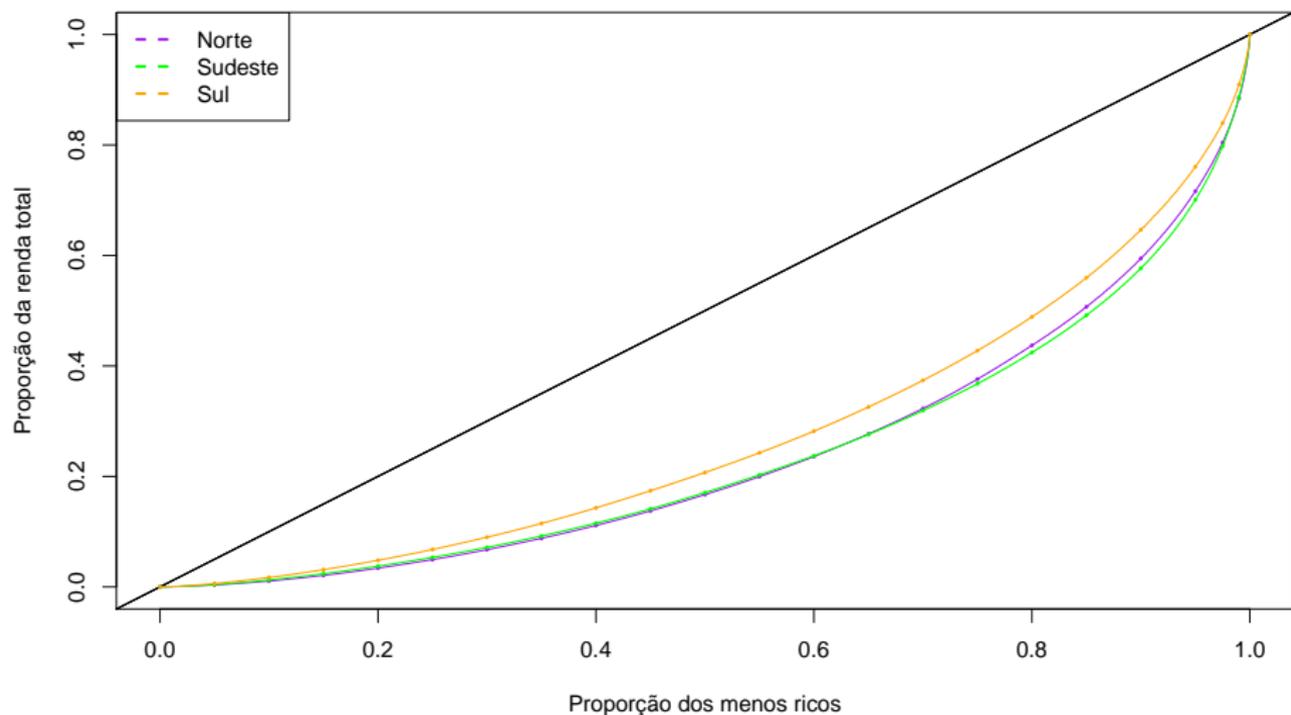
Lorenz-dominância (Atkinson, 1970; Dasgupta, Sen e Starrett, 1973)

- Suponha duas distribuições, A e B ;
 - Com curvas de Lorenz $L_A(p)$ e $L_B(p)$, respectivamente.
- Há Lorenz-dominância de A sobre B quando $L_A(p) \geq L_B(p), \forall p \in [0, 1]$
 - Ou seja: quando os $p\%$ menos ricos de A possuem uma proporção maior da renda total de A do que os mesmos $p\%$ em B em relação à renda total de B , para todo $p\%$ entre 0% e 100% .

Lorenz-dominância (Atkinson, 1970; Dasgupta, Sen e Starrett, 1973)

- Em notação, A Lorenz-domina B é escrito como $A \preceq B$;
- Quando $A \preceq B$, todas as medidas que atendam os quatro princípios (SIM, IP, IE e PPD) vão ordenar as distribuições da mesma maneira.

Lorenz-dominância: Exemplo



Lorenz-dominância: Exemplo

- Neste caso:
 - Sul \succeq Sudeste;
 - Sul \succeq Norte;
- Mas: Norte ? Sudeste.

Lorenz-dominância: Exemplo

- Quando curvas de Lorenz se cruzam:
 - Não há Lorenz-dominância;
 - Não há ordenamento unânime de medidas de desigualdade das duas distribuições.
- Ou seja:
 - Para medidas de desigualdade $I_1(\cdot)$ e $I_2(\cdot)$; e
 - Populações A e B ;
 - Podemos ter $I_1(A) > I_1(B)$ e $I_2(A) < I_2(B)$ ao mesmo tempo.
- Medidas de desigualdade podem discordar!

Lorenz-dominância: Exemplo

Região	Medidas			Ranking de Desigualdade		
	Gini	Theil-L	Theil-T	Gini	Theil-L	Theil-T
Sul	0.454	0.364	0.398	3	3	3
Sudeste	0.523	0.496	0.553	1	2	1
Norte	0.520	0.504	0.540	2	1	2

Fonte: PNAD Contínua 2021, 5ª Visita. Microdados.

Algumas medidas de desigualdade

- 1 Razão de Palma;
- 2 Índice de Gini;
- 3 Índices de Entropia Generalizada.

Razão de Palma (Cobham, Schlögl e Sumner, 2016; Palma, 2011)

- Analisando as distribuições de renda dos países ao longo do tempo, Palma (2011) encontrou um padrão:
 - O grupo entre os decis 5 e 9 costumam possuir aproximadamente 50% da renda total
- A variabilidade das distribuições é explicada por duas frações:
 - Total da renda apropriado pelos 40% mais pobres; e
 - Total da renda apropriado pelos 10% mais ricos.

Razão de Palma (Cobham, Schlögl e Sumner, 2016; Palma, 2011)

- Em termos da curva de Lorenz:
 - A fração da renda dos 40% mais pobres: $L(40\%)$;
 - A fração da renda dos 10% mais ricos: $1 - L(90\%)$.
- Razão de Palma:

$$P_R = \frac{1 - L(90\%)}{L(40\%)}$$

- Vantagem: fácil de explicar.

Razão de Palma (Cobham, Schlögl e Sumner, 2016; Palma, 2011)

- Esta ideia pode ser generalizada de várias maneiras:
 - *Quantile share ratios*;
 - Índice de Zenga (2007);
 - Índice de Gastwirth-Davydov-Greselin (Davydov e Greselin, 2020; Gastwirth, 2016)

Índice de Gini

- Sem dúvida, é a medida de desigualdade mais utilizada.
- Existe uma dúzia de maneiras de ver o índice de Gini, mas duas são mais interessantes:
 - Como áreas abaixo da curva de Lorenz sob igualdade perfeita;
 - Como média de distância entre rendas.

Índice de Gini

- Lembrando: quanto mais afastada da diagonal, mais desigual é a sociedade.
- O índice de Gini é proporcional à área entre a diagonal e a curva observada.
 - No caso de igualdade perfeita, a área é 0.
 - No caso de desigualdade perfeita, a área entre as curvas é de 0.5.
- O índice de Gini é o dobro destas áreas.

Índices de Entropia Generalizada

- Shorrocks (1980) provou a seguinte afirmação:
 - Uma medida de desigualdade que atende simultaneamente:
 - Princípio de Pigou-Dalton;
 - Invariância populacional;
 - Invariância de escala;
 - Decomponibilidade por Grupo;
 - ... só pode ser um Índice de Entropia Generalizada.
 - Ou uma transformação dele.

Índices de Entropia Generalizada

Os índices de Entropia Generalizada têm a seguinte fórmula:

$$IEG_{\alpha} = \begin{cases} \frac{1}{\alpha^2 - \alpha} \frac{1}{n} \sum_{i=1}^n \left[\left(\frac{y_i}{\mu} \right)^{\alpha} - 1 \right], & \alpha \in \mathbb{R} \setminus \{0, 1\} \\ -\frac{1}{n} \sum_{i=1}^n \ln \frac{y_i}{\mu}, & \alpha \rightarrow 0 \\ \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\mu} \ln \frac{y_i}{\mu}, & \alpha \rightarrow 1. \end{cases}$$

onde α é um parâmetro de aversão/sensibilidade à desigualdade.

Índices de Entropia Generalizada

- Quanto menor o valor de α , mais importância é dada para as transferências entre os mais pobres.
 - Ou seja: transferências regressivas entre os mais pobres afetam mais a desigualdade do que entre os mais ricos.
- Quanto maior o valor de α , mais importância é dada para as transferências entre os mais ricos.
- Os casos especiais IEG_0 e IEG_1 são chamados de Theil-L e Theil-T.
 - Também podemos mostrar que $IEG_2 = CV^2$.

Último comentário sobre medir desigualdade

*Medir de **desigualdade** não é apenas medir concentração. Por trás da matemática, existe um **juízo ético** sobre quais distribuições são mais ou menos justas.*

Parte 2

Medindo Pobreza

Identificação e Agregação

- De acordo com Sen (1976), a mensuração de pobreza consiste em duas etapas:
 - Identificação: quem são os pobres e o quão pobres eles são?
 - Agregação: como combinar as informações dos pobres em um único número?
- Se identificação é importante do ponto de vista individual, a agregação é importante do ponto de vista da formulação de políticas públicas (Medeiros, 2015, p. 16)

Identificação e Agregação

- Na prática, isso significa:
 - Definir uma “pontuação” para o nível de pobreza individual em relação a uma linha de pobreza;
 - Definir uma função que combine níveis individuais de pobreza.
- Matematicamente, uma medida de pobreza é uma função $P(\mathbf{y}; z)$, onde:
 - $\mathbf{y} = (y_1, \dots, y_n)$, uma “lista” das rendas dos n indivíduos na população;
 - z é a linha de pobreza.

Identificação e Agregação

- Linha de pobreza é um limite que define alguém como pobre ou não-pobre.
- Linhas de pobreza podem ser absolutas ou relativas:
 - Absoluta: um valor fixo no espaço onde avaliamos o bem-estar.
 - Ex.: 1/2 salário mínimo, US\$ 1.00/dia.
 - Relativa: um valor estabelecido a partir da distribuição do bem-estar.
 - Ex.: 60% da média, 60% da mediana.

Principais Axiomas

- 1 Foco
- 2 Independência de escala
- 3 Monotonicidade
- 4 Princípio da Transferência Mínima
- 5 Princípio da Sensibilidade à Transferência

Foco

- O princípio do foco estabelece que a medida de pobreza só depende da renda dos pobres.
- Exemplo:
 - Linha de pobreza: \$ 100;
 - Qualquer mudança não-empobrecedora da renda das pessoas não-pobres deve manter a medida de pobreza inalterada.

Independência de escala

- Se as rendas e a linha de pobreza são multiplicadas por uma constante $\lambda > 0$, a medida de pobreza permanece inalterada.
 - Matematicamente, $P(\mathbf{y}; z) = P(\mathbf{y} \cdot \lambda; z \cdot \lambda), \lambda > 0$.
- Exemplo: considere uma taxa de pobreza calculada com linha de pobreza e rendas em R\$.
 - Se as rendas e linha de pobreza forem convertidas para US\$, a medida deve permanecer inalterada.

Monotonicidade

- Considere a lista de rendas \mathbf{y} . Seja \mathbf{y}' uma distribuição obtida retirando uma quantia da renda de uma pessoa pobre em \mathbf{y} .
- Assim, $y_j = y'_j$ para todo $j \neq i$, mas $y_i > y'_i$ para algum i ;
- Segundo este princípio, $P(\mathbf{y}'; z) > P(\mathbf{y}; z)$.
- Em outras palavras, *coeteris paribus*, diminuir a renda de alguém pobre aumenta a pobreza.

Princípio da Transferência Mínima

- Sen (1976) propôs o seguinte princípio:
 - Se transferirmos a renda de alguém mais pobre para alguém menos pobre (sem tirar essa pessoa da pobreza), a medida de pobreza deve aumentar.
- Este é um princípio parecido com o princípio de Pigou-Dalton.

Princípio da Sensibilidade à Transferência

- Na distribuição \mathbf{y} , considere 4 pessoas com rendas $y_i < y_j < y_k < y_l$;
- Fazendo transferências, obtemos duas novas distribuições:
 - \mathbf{y}' : a partir de \mathbf{y} , uma transferência *regressiva* de $\$ \delta$ de y_i para y_j ;
 - \mathbf{y}'' : a partir de \mathbf{y} , uma transferência *regressiva* de $\$ \delta$ de y_k para y_l ;
 - Sem permitir que alguém cruze a linha de pobreza.
- Pelo Princípio da Sensibilidade à Transferência: $P(\mathbf{y}'; z) > P(\mathbf{y}''; z)$;
 - Mais peso na desigualdade entre as pessoas mais pobres. . .
 - Do que na desigualdade entre pessoas pobres mais próximas da linha de pobreza.

Princípio da Sensibilidade à Transferência

Indivíduos	Distribuições		
	y	y'	y''
l	9	9	11
k	7	7	5
j	6	8	6
i	4	2	4

Desta forma:

- Transferência Mínima: $P(y'; z) > P(y; z)$ e $P(y''; z) > P(y; z)$;
- Sensibilidade à Transferência: $P(y'; z) > P(y''; z)$;

Algumas medidas

- 1 Taxa de Pobreza
- 2 Hiato de Pobreza
- 3 Hiato Quadrático de Pobreza
- 4 Classe FGT
- 5 Índice de Watts

Taxa de Pobreza

- Sem dúvida, é a medida de pobreza mais utilizada;
- É o percentual de pessoas abaixo da linha de pobreza;
- É uma medida que ressalta a incidência/extensão da pobreza;
- Vamos denotar a taxa de pobreza como P_{HC} .

Taxa de Pobreza

- Vantagem: fácil de comunicar;
- Problema: Monotonicidade.
 - Reduzir a renda de alguém pobre não altera a medida;
 - Trata igualmente quem está R\$ 1 ou R\$ 100 abaixo da linha de pobreza.

Taxa de Pobreza

Região	Medidas	Ranking de Pobreza
	P_{HC}	P_{HC}
Norte	45.4	2
Nordeste	48.3	1
Sudeste	20.0	4
Sul	13.4	5
Centro-Oeste	20.2	3

Fonte: PNAD Contínua 2021, 5ª Visita. Microdados.

Nota: ¹ Linha de Pobreza: R\$ 499 a preços médios de 2021.

Hiato de Pobreza

- A insensibilidade da taxa de pobreza é um problema;
- Além da *incidência*, é preciso capturar a *intensidade* da pobreza;
- O Hiato de Pobreza é a média das distâncias padronizadas em relação à linha de pobreza:

$$P_G = \frac{1}{n} \sum_{i=1}^n \frac{z - y_i^*}{z} = 1 - \frac{1}{n} \sum_{i=1}^n \frac{y_i^*}{z}$$

onde $y_i^* = \min(y_i, z)$.

Hiato de Pobreza

- $y_i^* = \min(y_i, z)$ implica em indiferença às rendas maiores que z ;
 - Princípio do Foco, como a taxa de pobreza.
- Mudança na agregação:
 - Taxa de pobreza: “peso” binário;
 - Hiato: “peso” linear e contínuo.
- Ao contrário da taxa de pobreza, este atende ao princípio da monotonicidade.
 - Diminuir a renda de uma pessoa pobre aumenta o Hiato de Pobreza.

Hiato de Pobreza

Região	Medidas		Ranking de Pobreza	
	P_{HC}	P_G	P_{HC}	P_G
Norte	45.4	20.5	2	2
Nordeste	48.3	23.2	1	1
Sudeste	20.0	7.4	4	3
Sul	13.4	4.5	5	5
Centro-Oeste	20.2	7.1	3	4

Fonte: PNAD Contínua 2021, 5ª Visita. Microdados.

Nota: ¹ Linha de Pobreza: R\$ 499 a preços médios de 2021.

Hiato de Pobreza

- Considere uma população A ;
- Faça uma transferência regressiva entre os pobres;
 - Ou seja: transfira a renda de alguém muito pobre para alguém menos pobre;
 - Mas sem tirar ninguém da pobreza.

Neste caso:

- A Taxa de Pobreza é a mesma;
- O Hiato de Pobreza é o mesmo;
- A média da renda entre os pobres é a mesma;
- Mas os mais pobres ficaram ainda mais pobres!

Hiato Quadrático de Pobreza

- O Hiato Quadrático de Pobreza tenta resolver esse problema;
 - Princípio da Transferência Mínima.

$$P_{SG} = \frac{1}{n} \sum_{i=1}^n \left(\frac{z - y_i^*}{z} \right)^2$$

onde $y_i^* = \min(y_i, z)$.

Hiato Quadrático de Pobreza

- Peso da distância em relação à linha de pobreza:
 - Hiato: $1 - y_i^*/z$
 - Hiato Quadrático: $(1 - y_i^*/z)^2$
- Considere $y_i = 80$ e $z = 100$.
 - Peso do Hiato: 0.20;
 - Peso do Hiato Quadrático: 0.04.
- Agora, considere $y_i = 40$ e $z = 100$.
 - Peso do Hiato: 0.60;
 - Peso do Hiato Quadrático: 0.36.

Hiato Quadrático de Pobreza

Região	Medidas			Ranking de Pobreza		
	P_{HC}	P_G	P_{SG}	P_{HC}	P_G	P_{SG}
Norte	45.4	20.5	12.3	2	2	2
Nordeste	48.3	23.2	14.5	1	1	1
Sudeste	20.0	7.4	4.0	4	3	3
Sul	13.4	4.5	2.2	5	5	5
Centro-Oeste	20.2	7.1	3.5	3	4	4

Fonte: PNAD Contínua 2021, 5ª Visita. Microdados.

Nota: ¹ Linha de Pobreza: R\$ 499 a preços médios de 2021.

Classe FGT (Foster, Greer e Thorbecke, 1984)

Foster, Greer e Thorbecke (1984) propuseram uma classe geral que inclui a Taxa de Pobreza, Hiato e Hiato Quadrático de Pobreza:

$$P_{FGT}^{\gamma} = \frac{1}{n} \sum_{i=1}^n \left(\frac{z - y_i^*}{z} \right)^{\gamma}$$

onde:

- $y_i^* = \min(y_i, z)$;
- γ é um número natural não-negativo que denota um nível de sensibilidade à pobreza.

Classe FGT (Foster, Greer e Thorbecke, 1984)

Região	Medidas			Ranking de Pobreza		
	P_{FGT}^0	P_{FGT}^1	P_{FGT}^2	P_{FGT}^0	P_{FGT}^1	P_{FGT}^2
Norte	45.4	20.5	12.3	2	2	2
Nordeste	48.3	23.2	14.5	1	1	1
Sudeste	20.0	7.4	4.0	4	3	3
Sul	13.4	4.5	2.2	5	5	5
Centro-Oeste	20.2	7.1	3.5	3	4	4

Fonte: PNAD Contínua 2021, 5ª Visita. Microdados.

Nota: ¹ Linha de Pobreza: R\$ 499 a preços médios de 2021.

Índice de Watts (Watts, 1968)

- Atende diversos axiomas importantes para a mensuração de pobreza (Zheng, 1993, 1997);
- Tem interpretações interessantes.

$$P_{Watts} = \frac{1}{n} \sum_{i=1}^n \ln \frac{z}{y_i^*}$$

onde $y_i^* = \min(y_i, z)$.

Índice de Watts (Watts, 1968)

Índice de Watts como composição de outros índices (Blackburn, 1989):

$$P_{Watts} = P_{HC}(G_{Watts} + L^*)$$

onde:

- Hiato de Pobreza de Watts: $G_{Watts} = \ln \frac{z}{\mu_p}$
- Índice de Theil-L das rendas dos pobres: $L^* = \frac{1}{n_p} \sum_{i=1}^{n_p} \ln \frac{\mu_p}{y_i}$

Índice de Watts (Watts, 1968)

- O índice de Watts também pode produzir uma estimativa do *Tempo de Saída da Pobreza* (T_r) (Morduch, 1998);
- Tomando r como uma estimativa do crescimento da renda entre os pobres, tem-se:

$$T_r = \frac{P_{Watts}}{r}$$

Índice de Watts (Watts, 1968)

Região	P_{Watts}	P_{HC}	G_{Watts}	Theil-L
Norte	34.8	45.4	60.3	16.5
Nordeste	40.4	48.3	65.2	18.3
Sudeste	11.7	20.0	46.1	12.5
Sul	6.7	13.4	41.2	9.1
Centro-Oeste	10.6	20.2	42.9	9.3

Fonte: PNAD Contínua 2021, 5ª Visita. Microdados.

Nota: ¹ Linha de Pobreza: R\$ 499 a preços médios de 2021.

Impacto de distância para a linha de pobreza

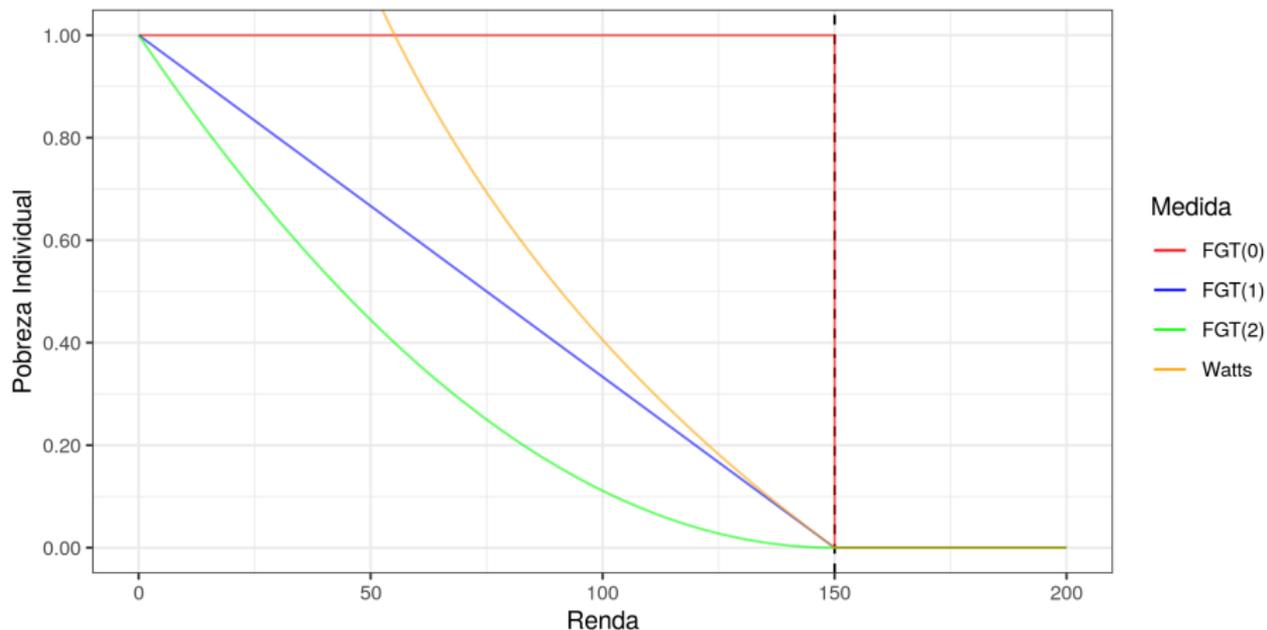


Figura 3: Pontuações individuais em medidas de pobreza

Relações entre Medidas de Pobreza e Axiomas

Axioma	P_{HC}	P_G	P_{SG}	$P_{FGT}^{\gamma>2}$	P_{Watts}
Foco	S	S	S	S	S
Simetria	S	S	S	S	S
Invariância Populacional	S	S	S	S	S
Invariância de Escala	S	S	S	S	S
Monotonicidade	N	S	S	S	S
Transferência Mínima	N	N	S	S	S
Sens. à Transferência	N	N	N	S	S

Fonte: Adaptado de Zheng (1997).

Comentários Finais

- Decomposições:
 - Incidência + Intensidade + Desigualdade: p. ex., Aristondo, De La Vega e Urrutia (2010);
 - *Poverty Profiles*: Kanbur (1987), Bidani e Ravallion (1993).
- Dominância
- Pobreza Multidimensional
- Nunca é uma escolha “puramente técnica”
 - Justo vs. injusto

Leituras sugeridas

- Textos introdutórios:
 - Medeiros (2015)
 - Foster *et al.* (2013): Capítulo 2
- Abordagens formais:
 - Villar (2017): Capítulo 7
 - Chakravarty (2009): Capítulo 2
 - Zheng (1997)

Parte 3

PNAD Contínua

Amostragem

- Uma população finita \mathcal{U} de N indivíduos, identificados pelos rótulos 1, 2, ..., N .
 - $\mathcal{U} = \{1, 2, \dots, N\} = \{i\}_{i=1}^N$
- Se dispomos de recursos ilimitados, podemos calcular diversas quantidades nesta população;
- Denotando o valor da variável y para o indivíduo $i \in \mathcal{U}$ por y_i , podemos calcular, por exemplo:
 - O total de y na população: $Y = \sum_{i=1}^N y_i$;
 - A média populacional: $\bar{Y} = \sum_{i=1}^N y_i / N$;
 - O índice de Gini de y , etc.

Amostragem

- Porém, não podemos coletar informação sobre N indivíduos.
- Portanto, extraímos uma amostra S de n indivíduos de acordo com um plano amostral:
 - Conhecemos as probabilidades de seleção $\pi_i > 0, \forall i \in \mathcal{U}$;
 - Idealmente, também conhecemos as probabilidades conjuntas de seleção $\pi_{ij} > 0, \forall i, j \in \mathcal{U}$.
 - Sorteamos os indivíduos de acordo com estas probabilidades.
- Planos amostrais probabilísticos: AAS, Estratificada, Conglomerados, Multi-estágios, etc.

Amostragem

- Para inferir sobre o parâmetro θ da população finita a partir da amostra probabilística \mathcal{S} , utilizamos um estimador $\hat{\theta}$;
 - Estratégia de estimação: plano amostral + estimador.
- Por não conhecermos os valores para todos os N indivíduos, nossas estimativas têm erros;
 - $\text{EQM}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{B}(\hat{\theta})^2$
- Se o nosso estimador $\hat{\theta}$ é não-viesado, $\text{B}(\hat{\theta}) = 0$.
 - Logo, $\text{EQM}(\hat{\theta}) = \text{Var}(\hat{\theta})$
- Estimador de Total de Horvitz-Thompson:
 - n unidades distintas;
 - Não-viesado se $\pi_i > 0, \forall i \in \mathcal{U}$;
 - Variância estimável sem viés se $\pi_{ij} > 0, \forall i, j \in \mathcal{U}$.

Amostragem

Pontos principais:

- Importância da estratégia: plano amostral + estimador;
- Sob condições de regularidade, funções de estimadores de HT são:
 - *Assintoticamente* não-viesadas;
 - Variância assintótica pode ser estimada.
- A palavra “modelo” não foi mencionada.
 - Inferência baseada no plano amostral (*design-based inference*)

O que é a PNAD Contínua? (IBGE, 2014)

- Pesquisa Nacional por Amostra de Domicílios Contínua
- População-Alvo:
 - Todas as pessoas moradoras em domicílios particulares permanentes da área de abrangência da pesquisa.
- Abrangência geográfica:
 - Todo o território nacional, excluídas áreas com características especiais.

Plano Amostral da PNAD Contínua

- Conglomerado em Dois Estágios com Estratificação das Unidades Primárias de Amostragem.

O que isso significa?

Plano Amostral da PNAD Contínua

Estratificação

- Estratificação: sorteio em cada estrato.
- Estratos: agrupamentos baseados em
 - Dependências administrativas;
 - Características sociodemográficas;
 - Razões Estatísticas.
- Finalidade:
 - Operacional: Garantir amostras para determinados domínios (UFs, capitais, regiões metropolitanas, etc.)
 - Estatística: melhorar a precisão dos estimadores

Plano Amostral da PNAD Contínua

Conglomerado em Dois estágios

- Dois estágios de seleção:
 - Sorteio de unidades primárias (UPAs)
 - Sorteio de unidades secundárias (USAs)
- Na PNADC:
 - UPAs são setores ou grupos de setores censitários;
 - Setores com menos de 60 domicílios particulares permanentes foram combinados até atingirem o tamanho mínimo.
 - USAs são domicílios particulares permanentes ocupados.

Plano Amostral da PNAD Contínua

- UPAs selecionadas com probabilidade proporcional ao tamanho;
 - Tamanho: número de domicílios particulares permanentes.
- Em cada UPA selecionada, são sorteados 14 USAs por amostragem aleatória simples;
 - Usando o CNEFE.

Plano Amostral da PNAD Contínua

Por que isso é importante?

- A inferência não se baseia em modelos, mas no plano amostral;
- A distribuição amostral depende da probabilidade de seleção de cada unidade na população;
- As hipóteses usuais (independência, por exemplo) não são satisfeitas;
- É assim que (praticamente) todos os institutos de estatística operam.

Leitura sugerida

- IBGE (2014): documentação da PNAD Contínua;
- Silva, Pessoa e Lila (2002): como funcionava na antiga PNAD anual;
- Hoffmann, Botassio e Jesus (2019) , Capítulo 10: dados de renda na PNAD Contínua e suas limitações;
- Deaton (2019), Capítulo 1: planejamento e estimação com pesquisas domiciliares;
- West, Sakshaug e Kim (2017): impacto da especificação incorreta do plano amostral sobre as estimativas.

Parte 4

Prática usando o R

Pacote survey (Lumley, 2004, 2021)

- Estratégia de estimação: plano amostral + estimador
- Problema: os estimadores mudam de acordo com o plano amostral.
 - Principalmente os estimadores de variância.
- O pacote survey cria um ambiente para aplicar estratégias de estimação.

Pacote survey (Lumley, 2004, 2021)

- Neste pacote, destacamos algumas funções:
 - `svydesign`, que cria objetos que descrevem o plano amostral;
 - `svrepdesign`, que cria objetos com pesos de replicação (baseados no plano amostral);
 - Funções de estimação: `svytotal`, `svymean`, `svyquantile`, `svycdf`, `svyglm`, etc.
- Usando funções de estimação com objetos de plano amostral, é possível implementar estratégias de estimação adequadas.

Objeto de plano amostral com svrepdesign

```
# carrega pacote
library( survey )

# cria objeto com pesos de replicação
pnadc.design <-
  svrepdesign( repweights = "v1032[0-9]{3}" ,
              weights = ~v1032 ,
              type = "bootstrap" ,
              mse = TRUE ,
              data = pnadc.df )
```

Objeto de plano amostral com svrepdesign

- Cada argumento da função descreve um aspecto do plano amostral;
- Esta função cria um objeto de plano amostral;
 - No nosso caso, o objeto `pnadc.design`.

Esse é o “print” do objeto:

```
## Call: svrepdesign.default(repweights = "v1032[0-9]{3}", weights = ~v1032,  
##      type = "bootstrap", mse = TRUE, data = pnadc.df)  
## Survey bootstrap with 200 replicates and MSE variances.
```

- Note o tipo e número de réplicas e o tipo de estimador de variância.

Estimando médias com a função svymean

- svymean: função que estima médias;
- Variável: rendimento domiciliar per capita deflacionado def.rdpc;
- na.rm = TRUE: tratamento de valores ausentes.

```
svymean( ~def.rdpc , pnadc.design , na.rm = TRUE )
```

```
##           mean      SE  
## def.rdpc 1353.9 14.441
```

Estimando médias com a função `svymean`

- Por que `na.rm = TRUE`?
 - NA: valor ausente ou não-aplicável;
 - Alguns moradores têm valor ausente para o rendimento domiciliar per capita.
 - Por exemplo: empregados que moram no domicílio do empregador.
 - Dupla contagem do rendimento.

Estimação para domínios

- Às vezes, estamos interessados em domínios (i.e., subpopulações) específicas;
 - Por exemplo, moradores de domicílios na área rural.
- Podemos filtrar essas observações usando a função `subset` sobre o objeto de plano amostral:

```
svymean( ~def.rdpc ,
         subset( pnadcdesign , v1022 == "Rural" ) ,
         na.rm = TRUE )
```

```
##           mean      SE
## def.rdpc 694.85 8.6319
```

Estimação para domínios

Posso deletar observações na base de dados em vez de usar `subset` no objeto de plano?

- Em regra: **não**.
 - Motivo: risco de obter estimativas de variância incorretas.
- Em planos amostrais complexos, sempre use a função `subset`:
 - Em regra, ela atribui peso zero para as observações que não nos interessam;
 - Quando possível, ela deleta observações da base para criar um objeto “mais leve”.

Estimativas para vários domínios

- O método de `subset` é excelente quando temos um domínio específico.
- Mas isso é um problema quando temos vários domínios:
 - Por exemplo, um comando para cada UF é pouco prático.
- Solução: função `svyby`

Estimativas para vários domínios

Média do rendimento domiciliar per capita por Grande Região:

```
svyby( ~def.rdpc , ~regiao , pnadc.design ,
       svymean , na.rm = TRUE , deff = TRUE )
```

##	regiao	def.rdpc	se	DEff.def.rdpc
## 1	Norte	871.0693	20.45840	10.415021
## 2	Nordeste	843.7826	17.89427	15.605197
## 3	Sudeste	1645.3248	31.69645	13.175618
## 4	Sul	1656.7517	24.79909	8.316918
## 5	Centro-Oeste	1534.6421	33.30193	7.863063

Estimativas para vários domínios

Por que $deff = TRUE$?

- $DEff$ é uma abreviação para *Design Effect* (Kish, 1965);
 - EPA: Efeito do Plano Amostral.
- Não é algo exclusivo da *svyby*, mas de qualquer estimador.
- Ele apresenta o impacto do plano amostral sobre a estimativa da variância do estimador:
 - Por exemplo: um $DEff = 13$ indica que a estimativa correta da variância é 13 vezes maior do que sob AAS.
- Ele pode ser usado para indicar o tamanho do erro em ignorar o plano amostral complexo.

Estimando medidas de desigualdade e pobreza com convey (Pessoa, Damico e Jacob, 2021)

Para estimar o índice de Gini e medidas com linhas de pobreza relativa, a library convey pode ser útil.

O primeiro passo é usar a função convey_prep para preparar o objeto de plano amostral.

\footnotesize

```
pnadc.design <- convey_prep( pnadc.design )
```

Essa função cria uma cópia interna do objeto de plano amostral no objeto principal. Por quê?

Estimando medidas de desigualdade e pobreza com convey (Pessoa, Damico e Jacob, 2021)

- A razão principal pode ser entendida com o uso de linhas de pobreza relativas;
- Em geral, a linha de pobreza relativa é estimada no nível do país;
- ... mas a taxa de pobreza de interesse pode ser sobre o Rio de Janeiro!
- Ou seja: a estimação da taxa de pobreza (relativa!) no Rio depende de observações fora do Rio!
 - Depois da `convey_prep`, a função `subset` entende isso e trata o domínio corretamente.
- Essa função também é necessária no caso de métodos de replicação!

Exemplo: Índice de Gini

A lógica é a mesma da survey. Por exemplo, para estimar o índice de Gini por Região, fazemos:

```
svyby( ~def.rdpc , ~regiao , pnadc.design ,
       svygini , na.rm = TRUE )
```

... que retorna:

```
##           regiao  def.rdpc se.def.rdpc
## 1           Norte 0.5285300 0.007629401
## 2      Nordeste 0.5556985 0.006702216
## 3           Sudeste 0.5329958 0.005259823
## 4             Sul 0.4616618 0.004601518
## 5 Centro-Oeste 0.5142519 0.006797145
```

Exemplo: Classe FGT

Hiato Quadrático de Pobreza ($\gamma = 2$) por situação do domicílios, com linha de pobreza absoluta de R\$ 499:

```
svyby( ~def.rdpc , ~v1022 , pnadc.design ,
       svyfgt , abs_thresh = 499 , g = 2 ,
       na.rm = TRUE )
```

... que retorna:

```
##      v1022  def.rdpc se.def.rdpc
## 1 Urbano 0.07650826 0.001205142
## 2 Rural  0.17715193 0.003309522
```

Exemplo: Classe FGT

E com linha de pobreza baseada em 60% da renda mediana?

```
svyby( ~def.rdpc , ~v1022 , pnadc.design ,  
       svyfgt , type_thresh = "relq" , percent = .60 , g=2 ,  
       na.rm = TRUE )
```

Principais recomendações

- O plano amostral importa!
- Sempre use a base de dados completa na `svydesign`;
 - E, em geral, na `svrepdesign`;
 - Filtre domínios com a função `subset` no objeto de plano amostral.
- Consulte a documentação do pacote `survey`;
- Em geral, os estimadores dependem de amostras suficientemente grandes;
 - O viés pode ser importante em amostras pequenas.
- Na dúvida, converse com especialistas.

Principais recomendações

- Silva (2004) para uma introdução sobre calibração, cobrindo aspectos teóricos e práticos;
- IBGE (2021) para detalhes sobre calibração na PNADC.
- Escrevi um pouco sobre isso:
guilhermejacob.github.io/2021/12/pnadc-raking-bootstrap.

Parte 5

Contato e Referências

Contato

- O link para os slides ficarão disponíveis neste post: guilhermejacob.github.io/2022/11/minicurso-abep-2022/.
- Meu e-mail é guilhermejacob91@gmail.com.
- Três possibilidades:
 - Respondo com o que sei;
 - Respondo dizendo que não sei;
 - Esqueço de responder (> 1 semana).
 - Neste caso, mande outro e-mail!

Referências

- AMIEL, Y. The Measurement of Income Inequality: The Subjective Approach. *Em*: SILBER, J. (Ed.). **Handbook of Income Inequality Measurement**. Dordrecht: Springer Netherlands, 1999. p. 227–243.
- ARISTONDO, O.; DE LA VEGA, C. L.; URRUTIA, A. A new multiplicative decomposition for the Foster–Greer–Thorbecke poverty indices. **Bulletin of Economic Research**, v. 62, n. 3, p. 259–267, 2010.
- ATKINSON, A. B. On the Measurement of Inequality. **Journal of Economic Theory**, v. 2, n. 3, 3, p. 244–263, set. 1970.
- ATKINSON, A. B.; BRANDOLINI, A. On Analyzing the World Distribution of Income. **The World Bank Economic Review**, v. 24, n. 1, p. 1–37, jan. 2010.
- BIDANI, B.; RAVALLION, M. A Regional Poverty Profile for Indonesia. **Bulletin of Indonesian Economic Studies**, v. 29, n. 3, p. 37–68, 1993.

Referências

BLACKBURN, M. L. Poverty measurement: an index related to a Theil measure of inequality. **Journal of Business & Economic Statistics**, v. 7, n. 4, p. 475–481, 1989.

BLACKORBY, C.; BOSSERT, W.; DONALDSON, D. Income Inequality Measurement: The Normative Approach. *Em*: SILBER, J. (Ed.).

Handbook of Income Inequality Measurement. Dordrecht: Springer Netherlands, 1999. p. 133–161.

CHAKRAVARTY, S. R. Measuring Inequality: The Axiomatic Approach. *Em*: SILBER, J. (Ed.). **Handbook of Income Inequality Measurement**. Dordrecht: Springer Netherlands, 1999. p. 163–186.

_____. **Inequality, Polarization and Poverty: Advances in Distributional Analysis**. 1. ed. Nova York: Springer-Verlag, 2009.

Referências

- CHATEAUNEUF, A.; MOYES, P. **Measuring Inequality Without the Pigou-Dalton Condition**. [s.l.] World Institute for Development Economic Research (UNU-WIDER), 2005. Disponível em: <<https://ideas.repec.org/p/unu/wpaper/rp2005-02.html>>.
- COBHAM, A.; SCHLÖGL, L.; SUMNER, A. Inequality and the Tails: the Palma Proposition and Ratio. **Global Policy**, v. 7, n. 1, p. 25–36, 2016.
- DALTON, H. The Measurement of the Inequality of Incomes. **The Economic Journal**, v. 30, set. 1920.
- DASGUPTA, P.; SEN, A.; STARRETT, D. Notes on the measurement of inequality. **Journal of Economic Theory**, v. 6, n. 2, p. 180–187, 1973.
- DAVYDOV, Y.; GRESELIN, F. Comparisons Between Poorest and Richest to Measure Inequality. **Sociological Methods & Research**, v. 49, n. 2, p. 526–561, 2020.

Referências

- DEATON, A. **The Analysis of Household Surveys (Reissue Edition with a New Preface): A Microeconometric Approach to Development Policy**. Washington, D.C.: The World Bank, 2019.
- FOSTER, J. *et al.* **A Unified Approach to Measuring Poverty and Inequality**. Washington, D.C.: The World Bank, 2013.
- FOSTER, J. E.; GREER, J.; THORBECKE, E. A Class of Decomposable Poverty Measures. **Econometrica**, v. 52, n. 3, p. 761–766, 1984.
- GASTWIRTH, J. L. Measures of Economic Inequality Focusing on the Status of the Lower and Middle Income Groups. **Statistics and Public Policy**, v. 3, n. 1, p. 1–9, 2016.
- GIORGI, G. M. Income Inequality Measurement: The Statistical Approach. *Em*: SILBER, J. (Ed.). **Handbook of Income Inequality Measurement**. Dordrecht: Springer Netherlands, 1999. p. 245–267.
- HAMMOND, P. J. Equity, Arrow's Conditions, and Rawls' Difference Principle. **Econometrica**, v. 44, n. 4, p. 793–804, 1976.

Referências

HIRSCHMAN, A. O. The Paternity of an Index. **The American Economic Review**, v. 54, n. 5, p. 761–762, 1964.

HOFFMANN, R.; BOTASSIO, D. C.; JESUS, J. G. DE. **Distribuição de Renda: Medidas de Desigualdade, Pobreza, Concentração, Segregação e Polarização**. 2. ed. São Paulo: Editora da Universidade de São Paulo, 2019.

IBGE. **Pesquisa Nacional por Amostra de Domicílios Contínua: Notas metodológicas**. Rio de Janeiro: IBGE; Instituto Brasileiro de Geografia e Estatística, 2014. Disponível em: <https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Notas_metodologicas/notas_metodologicas.pdf>.

_____. **Sobre a alteração do método de calibração dos fatores de expansão da PNAD Contínua**. Rio de Janeiro: IBGE; Instituto Brasileiro de Geografia e Estatística; IBGE, nov. 2021. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101882.pdf>>.

Referências

- KANBUR, S. M. R. Measurement and Alleviation of Poverty: With an Application to the Effects of Macroeconomic Adjustment (Evaluation quantitative de la pauvreté et remèdes possibles: analyse des effets d'un ajustement macroéconomique) (Medición y alivio de la pobreza, con una aplicación a los efectos del ajuste macroeconómico). **Staff Papers - International Monetary Fund**, v. 34, n. 1, p. 60, mar. 1987.
- KISH, L. **Survey Sampling**. Nova York: John Wiley & Sons, 1965.
- LUMLEY, T. Analysis of Complex Survey Samples. **Journal of Statistical Software**, v. 9, n. 1, p. 1–19, 2004.
- _____. **survey: analysis of complex survey samples**, 2021.
- MEDEIROS, M. Como medir a pobreza na sociedade? *Em*: JOHANSSON, K.; MAHUMANA, N.; MEDEIROS, M. (Eds.). **O que são Pobreza e Pobres?** Cadernos de Ciências Sociais. Lisboa: Escolar, 2015. p. 91–121.
- MORDUCH, J. Poverty, economic growth, and average exit time. **Economics Letters**, v. 59, n. 3, p. 385–390, 1998.

Referências

PALMA, J. G. Homogeneous Middles vs. Heterogeneous Tails, and the End of the “Inverted-U”: It’s All About the Share of the Rich. **Development and Change**, v. 42, n. 1, p. 87–153, 2011.

PESSOA, D.; DAMICO, A.; JACOB, G. **convey: Income Concentration Analysis with Complex Survey Samples**. [s.l.: s.n.].

PIGOU, A. C. **Wealth and Welfare**. London: Macmillan, 1912.

SEN, A. K. Poverty: An Ordinal Approach to Measurement. **Econometrica**, v. 44, mar. 1976.

SHORROCKS, A. F. The Class of Additively Decomposable Inequality Measures. **Econometrica**, v. 48, n. 3, p. 613–625, 1980.

SHORROCKS, A. F. Inequality Decomposition by Population Subgroups. **Econometrica**, v. 52, n. 6, p. 1369–1385, 1984.

Referências

- SILVA, P. L. N. **Calibration Estimation: When and Why, How Much and How**: Textos para discussão. Rio de Janeiro: IBGE, Diretoria de Pesquisas, 2004. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv66414.pdf>>.
- SILVA, P. L. N.; PESSOA, D. G. C.; LILA, M. F. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. **Ciência & Saúde Coletiva**, v. 7, n. 4, p. 659–670, 2002.
- VILLAR, A. **Lectures on Inequality, Poverty and Welfare**. Cham, Suíça: Springer International Publishing, 2017.
- WATTS, H. W. **An economic definition of poverty**. Madison, Wisconsin: University of Wisconsin, Institute For Research on Poverty, 1968. Disponível em: <<https://www.irp.wisc.edu/publications/dps/pdfs/dp568.pdf>>.

Referências

- WEST, B. T.; SAKSHAUG, J. W.; KIM, Y. Analytic Error as an Important Component of Total Survey Error. *Em: BIEMER, P. P. et al. (Eds.). Total Survey Error in Practice*. Hoboken, Nova Jersey: John Wiley & Sons, 2017. p. 487–510.
- ZENGA, M. Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. **Statistica e Applicazioni**, v. 1, n. 4, p. 3–27, 2007.
- ZHENG, B. An axiomatic characterization of the Watts poverty index. **Economics Letters**, v. 42, 1993.
- _____. Aggregate Poverty Measures. **Journal of Economic Surveys**, v. 11, 1997.