

ÍNDICE DE THEIL

Referência Obrigatória: Hoffman cap 4 pags 99 a 116 e cap 3 pgs 42-44 (seção 3.4).

1. Conteúdo Informativo de uma mensagem

- Baseado na teoria da informação, que analisa o conteúdo informativo de uma mensagem
- Esse conteúdo depende da probabilidade de ocorrência de um evento

Ex: $p \sim 1 \Rightarrow$ “o evento ocorreu” tem baixo conteúdo informativo

$p \sim 0 \Rightarrow$ “o evento ocorreu” tem alto conteúdo informativo

- Fórmula

$$h(x) = \log \frac{1}{x} = -\log x$$

- Unidades

$\log_2 x \Rightarrow$ binário \Rightarrow Bits

$\log_e x \Rightarrow$ natural \Rightarrow Nits

- Exemplos

Pela série pluviométrica $x = 0,2$

$$h(x) = \log \frac{1}{0,2} = 1,6094 \text{ Nits}$$

Pela informação de chuva na véspera $y=0,6$

$$h(y) = \log \frac{1}{0,6} = 0,5108 \text{ Nits}$$

O conteúdo informativo da mensagem incerta em questão é

$$h(x) - h(y) = 1,0986 \text{ Nits}$$

2. Entropia de uma distribuição

$$H(x) = E[h(x_i)] = \sum_i x_i h(x_i) = \sum_i x_i \ln \frac{1}{x_i} = - \sum_i x_i \ln x_i$$

$$H(x) = - \sum_i x_i \ln x_i$$

$$\text{Max } H(x)$$

$$\text{s.a. } \sum x_i$$

$$\text{Max } \{- \sum_i x_i \ln x_i - \lambda (\sum_i x_i - 1)$$

$$\text{CPO : } \ln x_i = -(1 + \lambda)$$

O $H(y)$ máximo, ou seja, a entropia máxima, ocorre quando há um máximo de incerteza a respeito do que pode ocorrer, visto que entropia é a esperança do conteúdo informativo para uma distribuição. Este máximo ocorre quando todos os possíveis eventos são igualmente prováveis, e você não deriva nenhuma informação destes eventos equiprováveis: $0 \leq H(x) \leq \ln n$

3. Medidas de Desigualdade de Theil

Theil (1967) propôs então uma medida de desigualdade a partir da entropia da distribuição. Entretanto, a igualdade não equivale a desordem (imprevisibilidade) econômica. Portanto, faz-se a seguinte transformação, subtraindo esta entropia de seu valor máximo. Logo temos:

$$T = \log n - H(y) = \left(\sum_{i=1}^n y_i \right) \log n + \sum_{i=1}^n y_i \log y_i = \sum_{i=1}^n y_i [\log n + \log y_i] = \sum_{i=1}^n y_i \log ny_i$$

$$T = \sum_{i=1}^n y_i \log ny_i$$

$0 \leq T \leq \ln n$, isto é, verifica-se que $T = 0$ no caso de uma distribuição perfeitamente igualitária e que $T = \ln n$ no caso de máxima desigualdade.

E no caso em que $y_i = 0$ temos $y_i \log y_i = 0$, por convenção.

onde $y_i \Rightarrow$ parcela de i na renda total

De maneira mais intuitiva,

$$T = \ln n - H(x) = \sum_i y_i \ln \frac{y_i}{1/n}$$

isto é, o índice de Theil mede o quanto a distribuição de renda observada (cada indivíduo detendo uma fração igual a y_i da renda total) descola de uma distribuição perfeitamente uniforme (cada indivíduo detendo uma fração igual a $1/n$ da renda total), ou o grau de redundância em relação a esta última, ponderando-se cada observação pela parcela na renda.

Portanto, o chamado índice de Theil-T é dado pela fórmula

$$T = \sum_{i=1}^n y_i \log ny_i$$

ou, alternativamente, por

$$T = \sum_{i=1}^n \frac{x_i}{N\mu} \log \frac{x_i}{\mu}$$

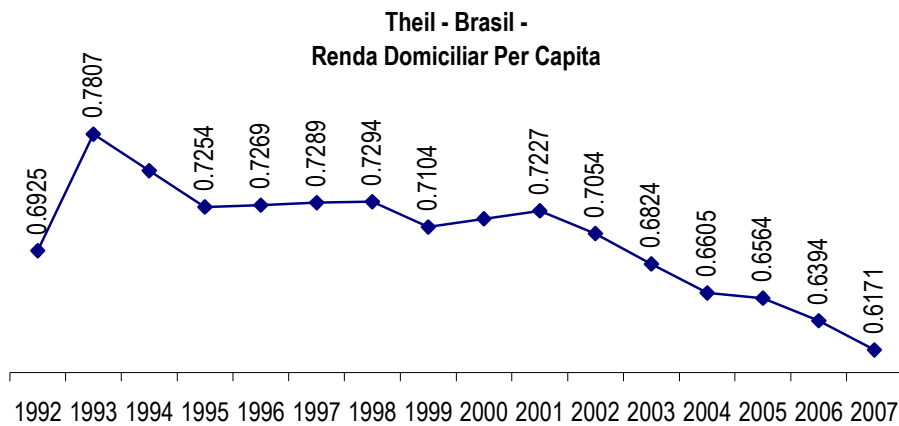
A segunda medida de desigualdade de Theil é chamada índice Theil L, e é dada pela fórmula:

$$L = \sum_{i=1}^n \frac{1}{n} \log \frac{1/n}{y_i} = \frac{1}{n} \sum_{i=1}^n \log \frac{y_i}{1/n}$$

ou, alternativamente, por

$$L = \sum_{i=1}^n \frac{1}{N} \log \frac{\mu}{x_i}$$

enquanto no Theil T os fatores de ponderação da desigualdade dentro dos grupos é a fração de renda apropriada, no Theil L os fatores de ponderação da desigualdade dentro dos grupos são as populações dos grupos.



OBS: 1994 e 2000 são médias interpoladas. Nesses anos a PNAD não foi a campo

4. Decomposição Intra e Entre Grupos

Suponha que eu tenha uma população de tamanho N , a qual seja dividida em K grupos tal que:

$N = \sum_{h=1}^K n_h$, em que n_h é o nº de pessoas no h -ésimo grupo. E a proporção da população correspondente ao h -ésimo grupo seria:

$\pi_h = \frac{n_h}{N}$. Seja, também x_{hi} a renda do i -ésimo indivíduo do h -ésimo grupo. Assim, a fração da renda total deste indivíduo seria:

$y_{hi} = \frac{x_{hi}}{N\mu}$, notando que o denominador nada mais é do que a renda total da população, sendo μ a renda média. Então a fração da renda total da população apropriada pelo h -ésimo grupo é:

$Y_h = \sum_{i=1}^{n_h} y_{hi}$, ou seja, eu estou somando a fração da renda total dos indivíduos que pertencem ao grupo h .

Temos o índice T de Theil:

$T = \sum_{i=1}^N y_i \log Ny_i = \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log Ny_{hi}$, estou apenas somando primeiro os indivíduos de um grupo, e depois de outro e outro e assim sucessivamente até somar toda a população.

Somando e subtraindo

(*) $\sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} = \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$ (do lado esquerdo para o direito, eu apenas abri o Y_h que

está fora do log, como definido acima ($Y_h = \sum_{i=1}^{n_h} y_{hi}$). Assim, a equação se torna fica:

$T = \sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^k \sum_{i=1}^{n_h} \frac{Y_h}{Y_h} y_{hi} \log Ny_{hi} - \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$, em que eu somei e subtraí (*) e

apenas dividi e multipliquei Y_h . Continuando:

$$T = \sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log Ny_{hi} - \sum_{h=1}^k \sum_{i=1}^{n_h} y_{hi} \log \frac{NY_h}{n_h}$$

$$T = \sum_{h=1}^k Y_h \log \frac{NY_h}{n_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log Ny_{hi} - y_{hi} \log \frac{NY_h}{n_h} \right]$$

$$T = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log \frac{Ny_{hi}}{\frac{NY_h}{n_h}} \right]$$

$$T = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h} + \sum_{h=1}^k Y_h \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \left[\log \frac{n_h y_{hi}}{Y_h} \right]$$

$$T = T_e + \sum_{h=1}^K Y_h T_h$$

Onde, $T_e = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h}$ é o Theil entre grupos e $T_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log n_h \frac{y_{hi}}{Y_h}$ é o Theil intra grupo,

Logo $\sum_{h=1}^K Y_h T_h$ é a média ponderada dos Theils intragrupos.

GROSS RATES OF CONTRIBUTION THEIL-T

Universe : Per Capita - All Income Sources

	GROSS							
	1976	1985	1990	1993	1997	2002	2003	2004
Groups:								
Gender	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Race	---	---	11.2%	10.8%	12.1%	10.7%	11.6%	10.2%
Age	0.2%	0.1%	0.2%	0.4%	0.9%	1.7%	2.0%	1.8%
Schooling	36.6%	42.4%	40.3%	36.8%	41.3%	38.2%	36.7%	35.2%
Working Class	12.0%	15.1%	13.4%	11.9%	14.2%	13.2%	14.7%	13.9%
Sector of Activity	13.7%	11.3%	10.3%	7.8%	10.2%	---	---	---
Population Density	17.6%	13.6%	13.5%	9.1%	11.1%	8.2%	6.7%	6.4%
Region	10.2%	8.4%	8.0%	6.9%	8.3%	7.2%	7.8%	7.0%

Source: PNAD

MARGINAL RATES OF CONTRIBUTION

THEIL-T

Universe : Per Capita - All Income

Sources

	MARGINAL							
	1976	1985	1990	1993	1997	2002	2003	2004
Groups:								
Age	1.6%	1.9%	2.4%	2.6%	2.8%	3.7%	5.8%	3.9%
Schooling	30.6%	34.7%	34.4%	32.3%	34.9%	33.1%	31.2%	30.4%
Working Class	4.9%	5.2%	4.9%	4.8%	5.3%	5.0%	5.9%	5.6%

Source: PNAD

5. Decomposição de Theil-T: Conceitos de Renda e Unidades de Análise

Trabalharemos com duas medidas básicas de desigualdade: o coeficiente Gini e o Theil-T. A popularidade do coeficiente Gini, e o fato de que permite a incorporação de rendas nulas à análise justifica a sua utilização. O Theil-T é a medida central usada aqui, considerando a sua propriedade exata de decomposição. A PNAD será a nossa principal fonte de dados neste estudo, sendo que a análise estará cobrindo os anos de 1976, 1985, 1990, 1993 e 1997.

Trabalharemos com cinco pares de conceitos de população-renda usando a PNAD:

Conceito de Renda	Conceito de População			
	Economicamente			
	Ocupados	Ativos	Idade Ativa	Total
Trabalho NH*				
Trabalho				
Indivíduos Todas as Fontes				
Per Capita Todas as Fontes				

*NH = Normalizado por Horas de Trabalho

Como valor de referência central, usaremos o Theil-T baseado nos economicamente ativos e todas as fontes de renda.

RATES OF CONTRIBUTION THEIL-T - 1997

GROSS RATES

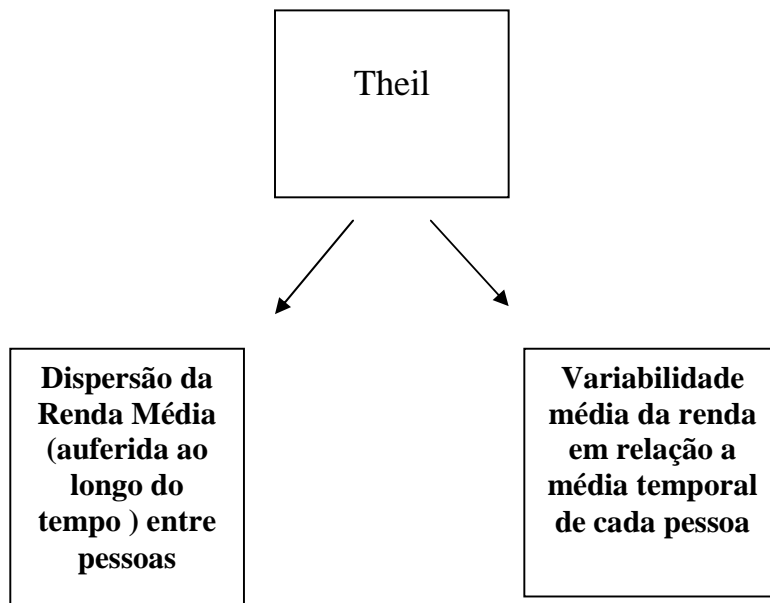
Population Concept	Occupied	Occupied	Economically A	Active Age	Total - Per Capita
Income Concept	Labor NH1	Labor	All Sources	All Sources	All Sources
Groups:					
Gender	0,6%	2,7%	2,7%	3,3%	0,0%
Race	8,3%	9,4%	9,4%	8,5%	12,1%
Age	6,6%	7,8%	8,2%	7,3%	0,9%
Schooling	35,0%	34,6%	34,7%	36,0%	41,3%
Working Class	16,8%	21,0%	21,4%	19,8%	14,2%
Sector	5,9%	5,1%	5,6%	6,0%	10,2%
Population Density	6,9%	7,5%	7,8%	7,5%	11,1%
Region	4,0%	5,4%	5,4%	4,9%	8,3%

MARGINAL RATES

Population Concept	Occupied	Occupied	Economically A	Active Age	Total - Per Capita
Income Concept	Labor NH1	Labor	All Sources	All Sources	All Sources
Groups:					
Age	3,9%	4,7%	5,9%	5,7%	2,8%
Schooling	26,6%	25,7%	26,4%	28,0%	34,9%
Working Class	5,6%	8,7%	8,7%	8,5%	5,3%

1/ Normalized by Hours

6. Aplicação de Técnicas de Decomposição: Variabilidade Temporal



Ex: Decomposição do intragrupos



Cada pessoa é como se fosse um grupo de várias observações temporais.

$$T = \sum_{i=1}^N \sum_{t=1}^T Y_{it} \log \frac{Y_{it}}{\frac{1}{NT}}$$

$$\bar{y}_i = \sum_{t=1}^T y_{it}$$

Resumo $T = T_{SS} + \sum_{t=1}^T Y_t T_t$, onde $T_{SS} = \sum_{i=1}^n \bar{Y}_i \log \frac{\bar{Y}_i}{\frac{1}{n}}$ e $T_t = \sum_{i=1}^n \frac{y_{it}}{\bar{Y}_i} \log \frac{\frac{y_{it}}{\bar{Y}_i}}{\frac{1}{T}}$

A

		THEIL-T INDEX						
Population Concept - Income Concept		1985	1990	1993	1994	1996	1997	1998
Theil total	Always Occupied - Month by Month	0.504	0.651	0.709	0.787	0.533	0.545	0.547
Theil media 4 meses	Always Occupied - Mean Earnings	0.448	0.580	0.551	0.646	0.497	0.508	0.512
Theil dispersão de renda média	resíduo inst temporal	0.056	0.071	0.158	0.142	0.037	0.037	0.035

Participação na desigualdade total %

		THEIL-T INDEX						
Population Concept - Income Concept		1985	1990	1993	1994	1996	1997	1998
Theil total	Always Occupied - Month by Month	100	100	100	100	100	100	100
Theil media 4 meses	Always Occupied - Mean Earnings	88.806	89.069	77.704	82.019	93.086	93.220	93.563
Theil dispersão de renda média		11.194	10.931	22.296	17.981	6.914	6.780	6.437

B

		GINI COEFFICIENT						
Population Concept - Income Concept		1985	1990	1993	1994	1996	1997	1998
Gini total	Always Occupied - Month by Month	0.520	0.566	0.592	0.618	0.527	0.530	0.527
Gini media 4 meses	Always Occupied - Mean Earnings	0.496	0.541	0.529	0.566	0.510	0.514	0.512
Gini dispersão de renda média	resíduo não exato	0.025	0.025	0.064	0.052	0.017	0.016	0.015

B Participação na desigualdade total %

		GINI COEFFICIENT						
Population Concept - Income Concept		1985	1990	1993	1994	1996	1997	1998
Gini total	Always Occupied - Month by Month	100	100	100	100	100	100	100
Gini media 4 meses	Always Occupied - Mean Earnings	95.256	95.555	89.273	91.609	96.865	96.942	97.144
Gini dispersão de renda média	resíduo não exato	4.744	4.445	10.727	8.391	3.135	3.058	2.856

PME's Longitudinal Aspect and Inequality Comparisons

We have also decided to incorporate PME data because its longitudinal aspect provide relevant insights of what happened to inequality in Brazil in recent years, specially allowing for pre and post stabilization inequality comparisons. We used the micro-longitudinal aspect of PME in two alternative ways: first, the 4 consecutive observations of the same individuals were treated independently before the inequality measures were assessed; second, we considered earnings average over four months before the inequality measures were calculated. The Theil-T is decomposed as follows: Month by Month Theil-T equals Mean Earnings Theil-T plus Individual Earnings Over Time Theil-T. In other words, the difference in the levels of inequality measures between month by month and average over four months is explained by the variability component of individual earnings over the four month period.

The main result here is that the fall of month to month inequality measures observed after the fall of inflation in 94 drastically overestimates the fall of inequality when one compares it with mean earnings over four months. The greater fall of traditional inequality measures on a monthly basis in comparison to measures on a four months basis is explained by the fall of the individual volatility measures following the sharp fall of inflation rates observed in this period. In sum, stabilization produced more stable earnings trajectories (i.e., lower temporal inequality (in fact, volatility) of individual earnings). On the other hand, the observed fall of

inequality *stricto sensu* was much smaller than what inequality measures based on monthly measures would have suggested.

Another way of looking at the effects of inflation and stabilization is to note that most of the fall in inequality measures is attributed to the within groups component, specially in the month by month inequality measures.

Table 13 below summarizes this information in terms of the gross and the marginal contribution of different groups characteristics. For example, in the case of month by month income concept presented in part B of table, during 1993 the sum of the marginal contributions of between groups component relative to schooling, working class and age (i.e., the three main characteristics) explains only 31.5% of total inequality. This statistic rises to 42.3% in 1997 which correspond to a 34.3% increase of relative contributive power to total inequality. In the case of the corresponding measures based on mean earnings across 4 months presented in table 13 part A, the relative rise of explanatory power is 12%. These results seems to confirm the idea that the explained share of total inequality tends to increase as we approach the permanent income concept.

Overall, the main point of this section is that most of the monthly earnings inequality fall observed after stabilization may be credited to a reduction of earnings volatility and not to a fall in the permanent income inequality (or *strictu senso* inequality).

Table 13

A - GROSS AND MARGINAL RATES OF CONTRIBUTION THEIL-T														
Universe : Longitudinal Data - 4 Observations - Always Occupied														
Mean Earnings Across 4 Months														
	GROSS							MARGINAL						
	1985	1990	1993	1994	1996	1997	1998	1985	1990	1993	1994	1996	1997	1998
Groups:														
Age	9,7%	8,7%	7,1%	6,7%	9,1%	9,2%	9,0%	10,4%	7,0%	6,3%	5,7%	6,9%	7,1%	7,6%
Schooling	34,5%	35,8%	32,2%	30,7%	37,5%	38,7%	37,8%	31,5%	30,7%	28,8%	26,8%	32,5%	33,2%	33,1%

B - GROSS AND MARGINAL RATES OF CONTRIBUTION THEIL-T														
Universe : Longitudinal Data - 4 Observations - Always Occupied														
Month by Month Labor Earnings														
	GROSS							MARGINAL						
	1985	1990	1993	1994	1996	1997	1998	1985	1990	1993	1994	1996	1997	1998
Groups:														
Age	8,6%	7,8%	5,5%	5,5%	8,4%	8,6%	8,5%	9,3%	6,2%	4,9%	4,7%	6,4%	6,6%	7,1%
Schooling	30,6%	31,9%	25,0%	25,2%	34,9%	36,1%	35,4%	27,9%	27,4%	22,4%	22,0%	30,2%	30,9%	31,0%

Source: PME

Table

VARIATION OF THEIL-T INDEX - Between 93 and 97							
Universe : Longitudinal Data - 4 Observations - Always Occupied							
		Mean Earnings			Month by Month		
	Gender	Total	Between	Within	Total	Between	Within
	Age						
Age	Up to 24 years	-0,006	0,003	-0,009	-0,019	0,003	-0,023
	25 to 34 years	-0,049	-0,019	-0,030	-0,085	-0,019	-0,066
	35 to 59 years	0,011	0,021	-0,010	-0,057	0,021	-0,078
	More than 60 years	0,001	0,002	-0,001	-0,002	0,002	-0,005
Total		-0,043	0,007	-0,050	-0,164	0,007	-0,171
	Schooling						
Schooling	0 Years	0,004	0,006	-0,002	0,001	0,006	-0,005
	1 to 4 years	-0,014	0,010	-0,024	-0,034	0,010	-0,044
	5 to 8 years	-0,017	-0,009	-0,008	-0,041	-0,009	-0,033
	9 to 12 years	-0,053	-0,038	-0,015	-0,087	-0,038	-0,049
	13 to 16 years	0,015	0,028	-0,013	-0,021	0,028	-0,049
	More than 16 years	0,022	0,021	0,000	0,019	0,021	-0,003
Total		-0,043	0,019	-0,062	-0,164	0,019	-0,183

Source: PME