7. Medidas de concentração e desigualdade

2012

Exemplo 1. Variável: renda do trabalho de pessoas.

Valores: $x_1,...,x_n$. Renda total: $T = x_1 + ... + x_n$.

(a) A renda total pode estar igualmente repartida entre as n pessoas, cada uma com renda = T / n $(= \bar{x})$.

(b) A renda total pode ser de uma única pessoa:
$x_1 = T$, $x_2 = x_3 = ... = x_n = 0$.

Estas duas situações são extremas. Em (b): concentração máxima.

É mais comum encontrarmos situações intermediárias.

Obs. Concentração está relacionada com variabilidade.

Exemplo 2. Variável: altura de pessoas.

Valores: $x_1,...,x_n$. Altura total: $T = x_1 + ... + x_n$.

$x_1 = T$, $x_2 = x_3 = ... = x_n = 0$ não faz sentido.

7.1. A curva de Lorenz

Valores ordenados: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. Total: $T = x_{(1)} + x_{(2)} + \ldots + x_{(n)}$.

Proporção acumulada de posições até a i-ésima posição ($p_0 = 0$):

$p_i = i / n$. $p_1 = 1 / n$, $p_2 = 2 / n$, ..., $p_{n-1} = (n-1) / n = 1 - 1 / n$, $p_n = n / n = 1$.

Proporção acumulada de valores até a i-ésima posição ($q_0 = 0$):

$q_i = (x_{(1)} + x_{(2)} + \ldots + x_{(i)}) / T$ $\quad$ ($q_n = T / T = 1$).

Obs. Se $x_i \geq 0$, então $p_i \geq q_i$, $i = 1, \ldots, n$.

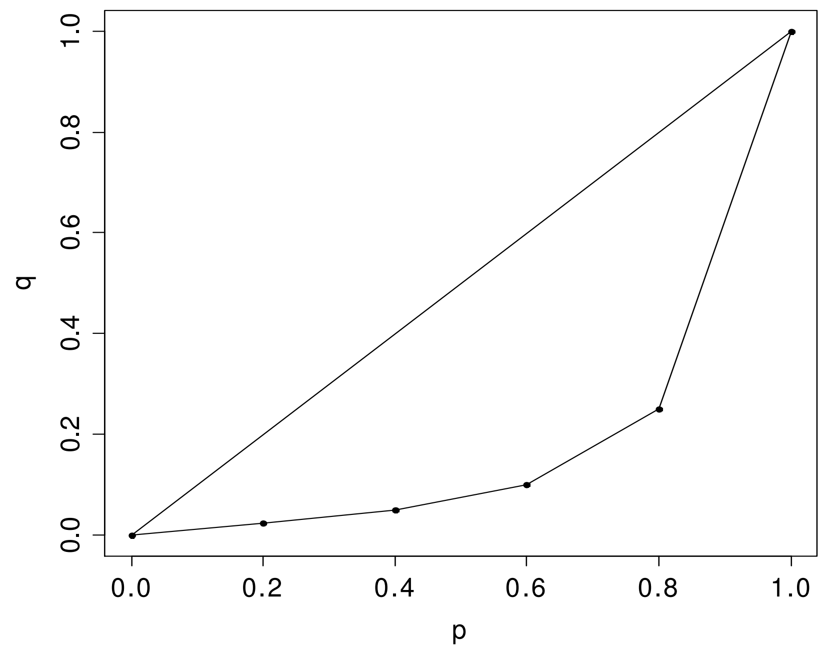
O gráfico formado pela união dos pontos $(0, 0)$, (p_1, q_1), (p_2, q_2), ..., (p_n, q_n) é chamado de curva de Lorenz ($p_n = q_n = 1$).

O segmento de reta unindo $(0, 0)$ e $(1, 1)$ também é incluído.

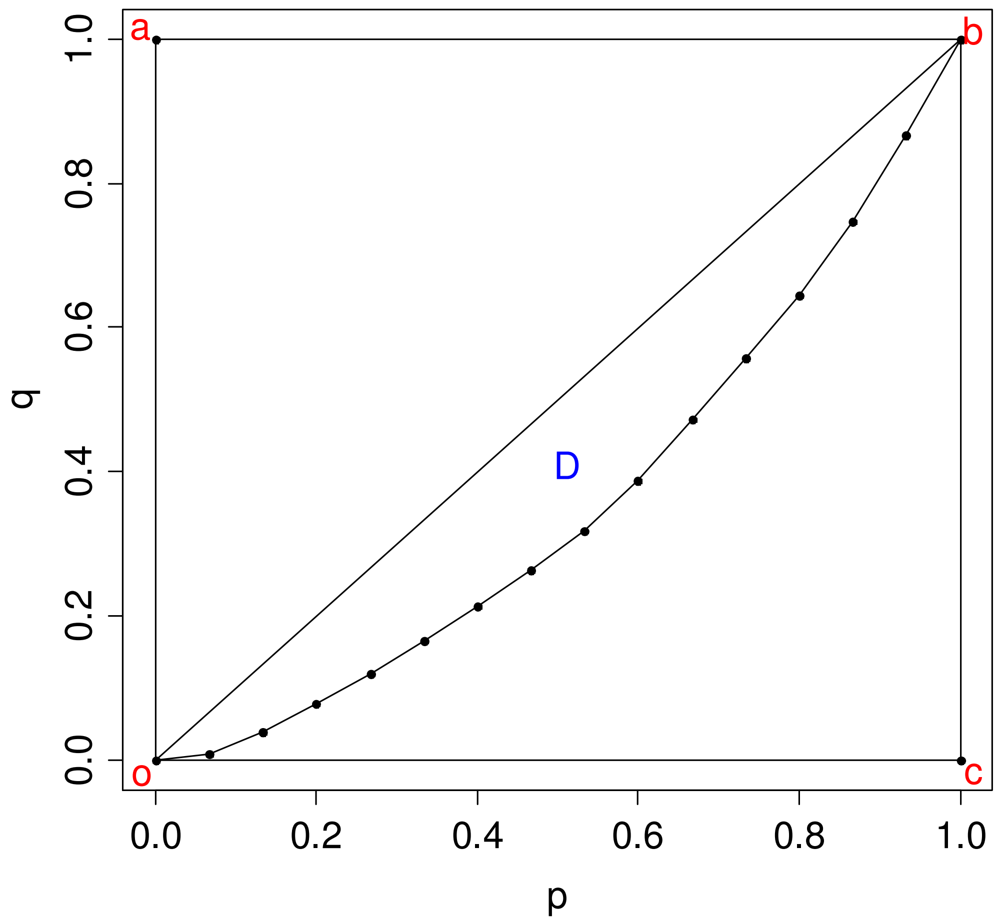
Exemplo

Dados ordenados: 1, 1, 2, 6, 30 ($n = 5$, $T = 40$ e média = $T / n = 8$).

i	$X_{(i)}$	p_i	q_i
1	1	1 / 5 = 0,2	1 / 40 = 0,025
2	1	2 / 5 = 0,4	(1 + 1) / 40 = 0,05
3	2	3 / 5 = 0,6	(1 + 1 + 2) / 40 = 0,1
4	6	4 / 5 = 0,8	(1 + 1 + 2 + 6) / 40 = 0,25
5	30	5 / 5 = 1	(1 + 1 + 2 + 6 + 30) / 40 = 1



Área de desigualdade



Área compreendida entre ob e a curva de Lorenz: área de desigualdade (D).

(a) $x_{(1)} = x_{(2)} = ... = x_{(n)} = T / n$: proporções de posições = proporções acumuladas de valores ($q_i = p_i$, i = 1,...,n).

$\Rightarrow$ curva de Lorenz = segmento ob (linha da igualdade perfeita).

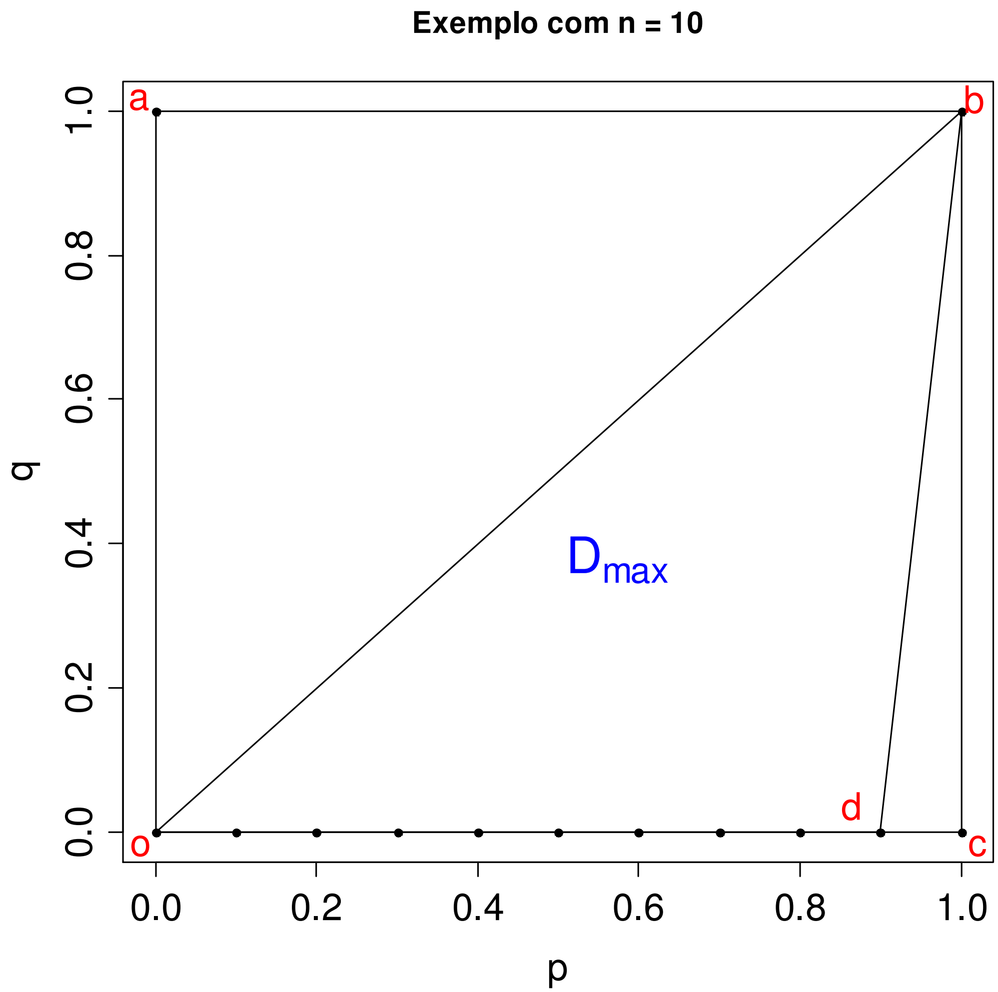
(b) $x_{(1)} = x_{(2)} = ... = x_{(n-1)} = 0$ e $x_{(n)} = T$:

curva de Lorenz é formada pelos pontos (0, 0), (1 – 1 / n, 0) e (1, 1): curva da desigualdade perfeita.

Quando $n \to \infty$: curva da desigualdade perfeita coincide com ocb.

Quanto mais a curva de Lorenz estiver afastada de ob, maior o grau de desigualdade.

7.2. Índice de Gini



Curva da desigualdade perfeita: odb.

Como a área do triângulo obc = ½, temos que $0 \leq D < ½$.

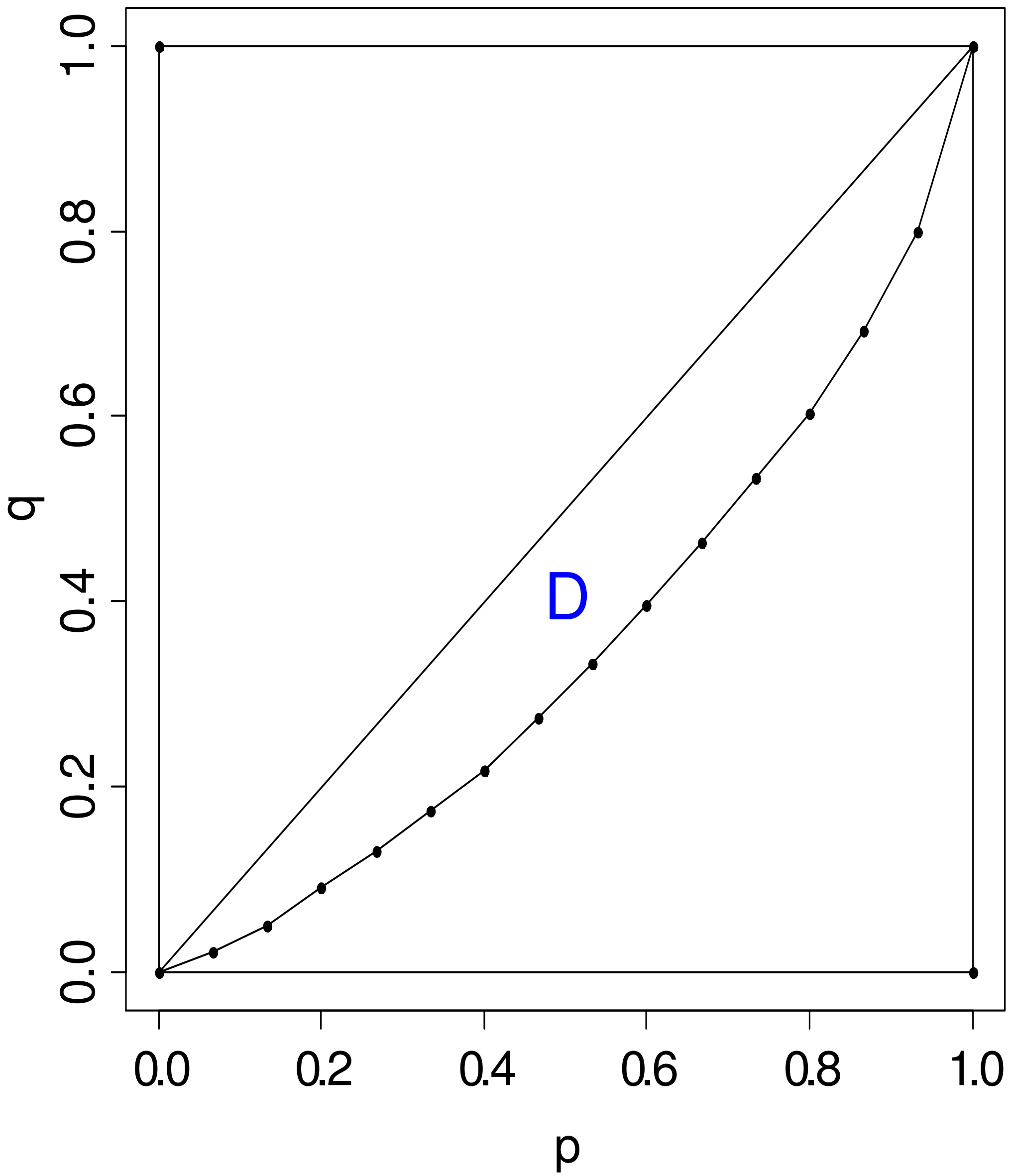
Valor máximo de D (desigualdade perfeita):

$$D_{\max} = \frac{1}{2}\left(1 - \frac{1}{n}\right).$$

$D_{max} \to ½$ quando $n \to \infty$ ($d \to c$).

max D_{max} = ½.

7.2. Índice de Gini



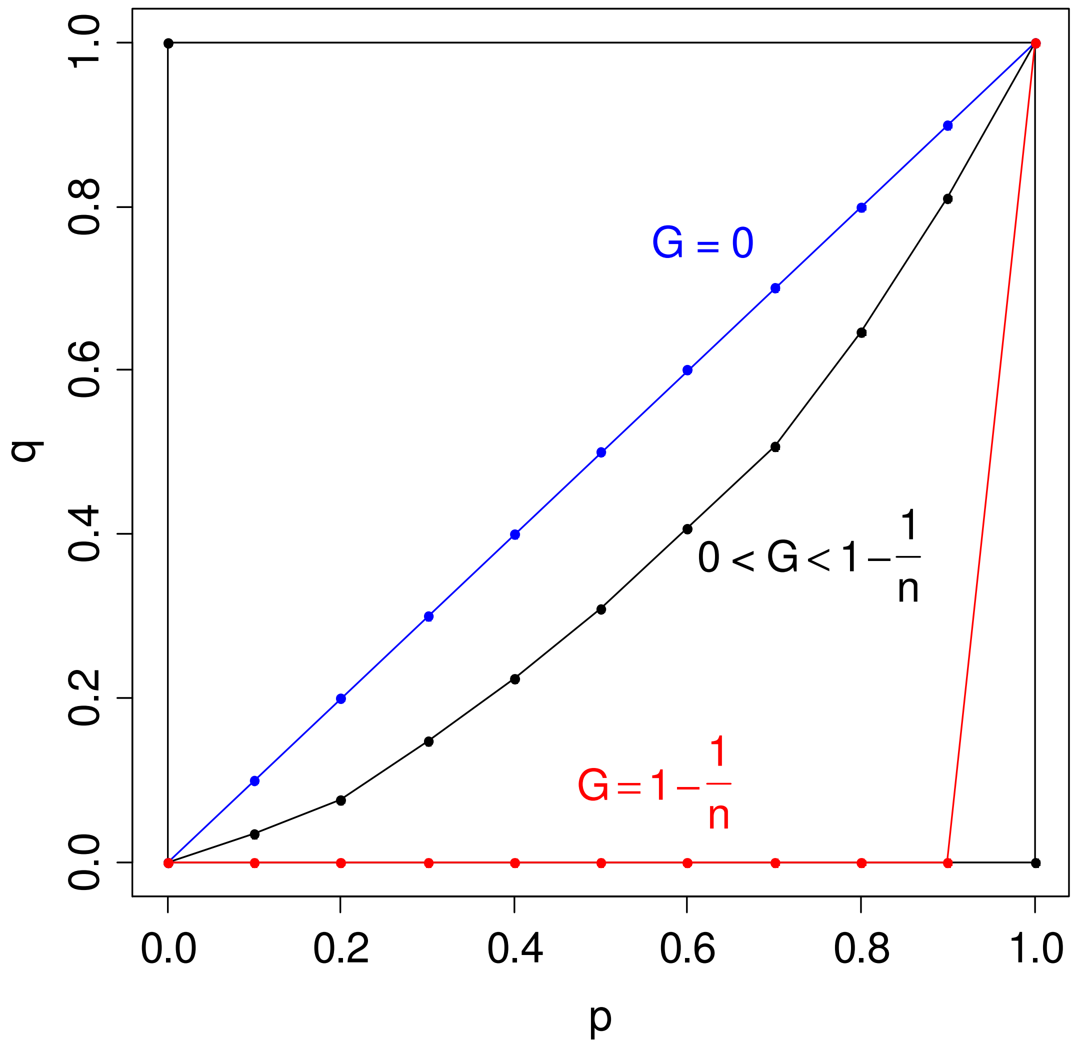
Proposto por C. Gini em 1914.

$G = D / \max D_{max} = D / ½ = 2D$.

Propriedades. (a) $0 \leq G < 1$ e (b) $0 \leq G \leq 1 - 1/n$.

Igualdade perfeita: $G = 0$.
Desigualdade perfeita:
$G = 1 - 1 / n$
($\rightarrow 1$ quando $n \rightarrow \infty$).

7.2. Índice de Gini



Valores ordenados:

$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

Como calcular G?

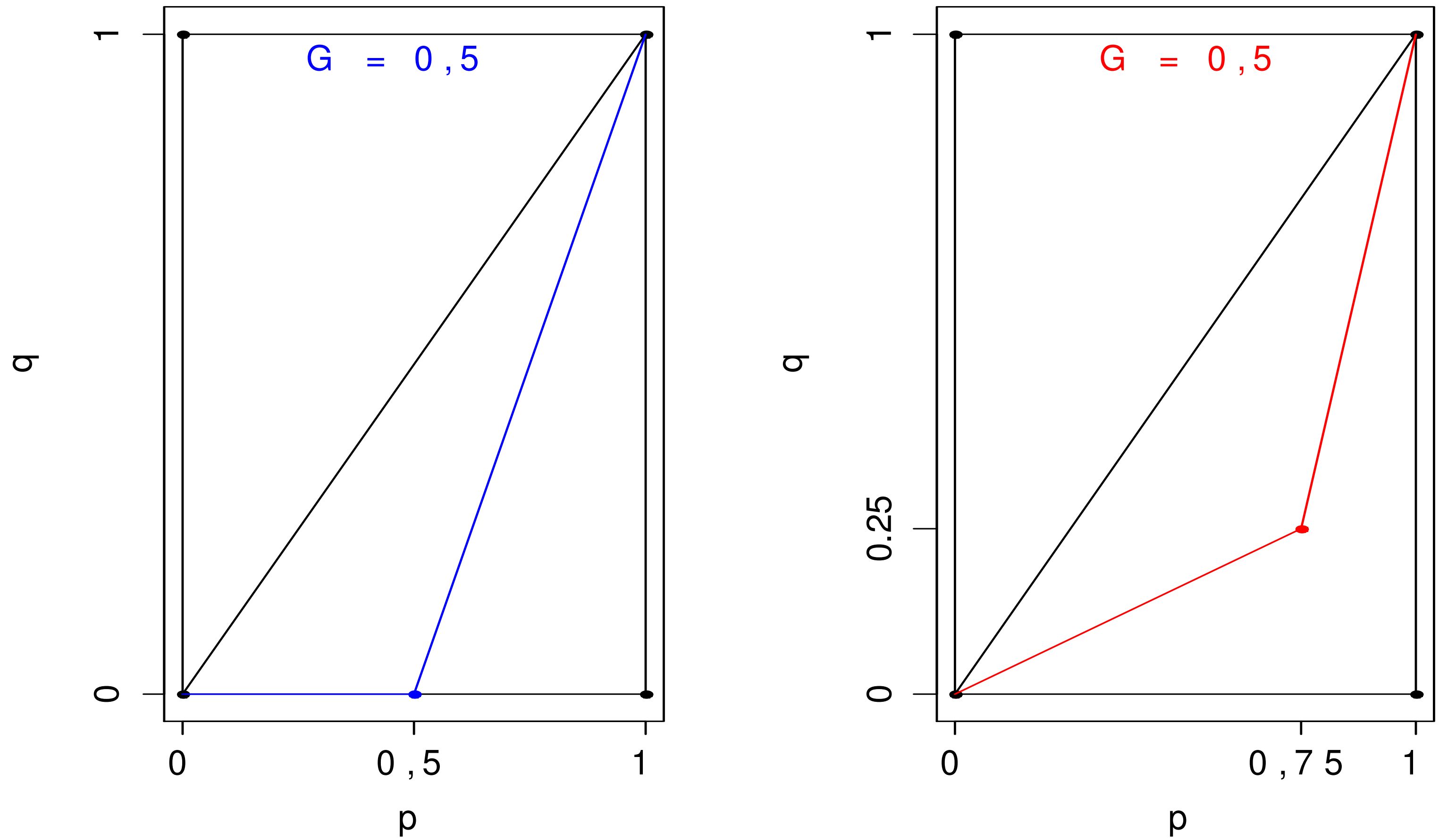
$$G = 1 - \frac{1}{n} \sum_{i=1}^{n} (q_i + q_{i-1}),$$

sendo que $q_0 = 0$ e

$$q_i = \frac{1}{T} \sum_{j=1}^{i} x_{(j)}.$$

7.2. Índice de Gini

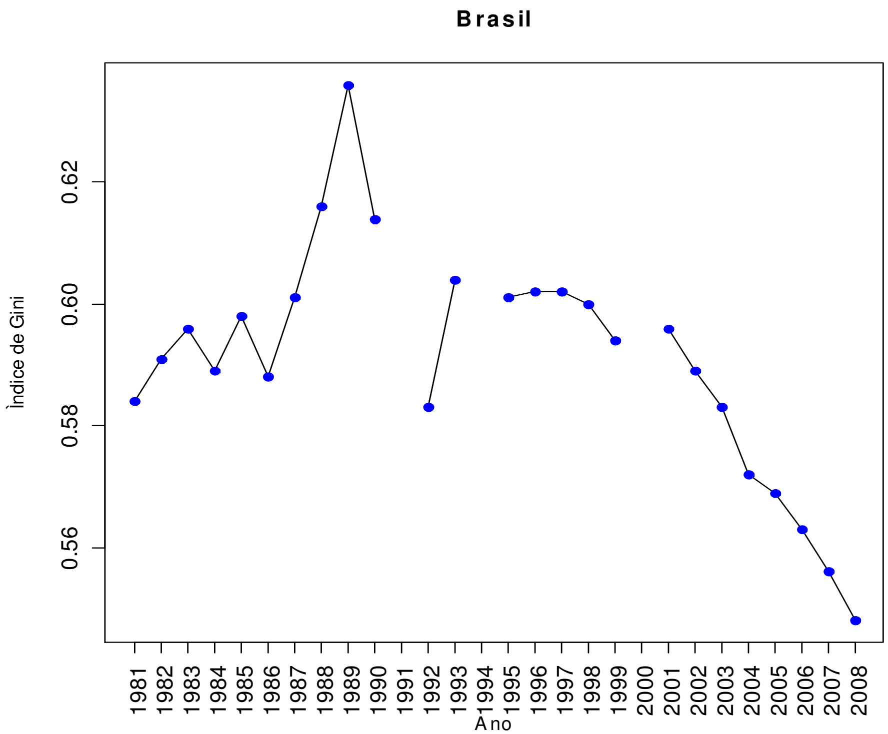
Obs. (a) Diferentes curvas de Lorenz podem gerar o mesmo valor de G.

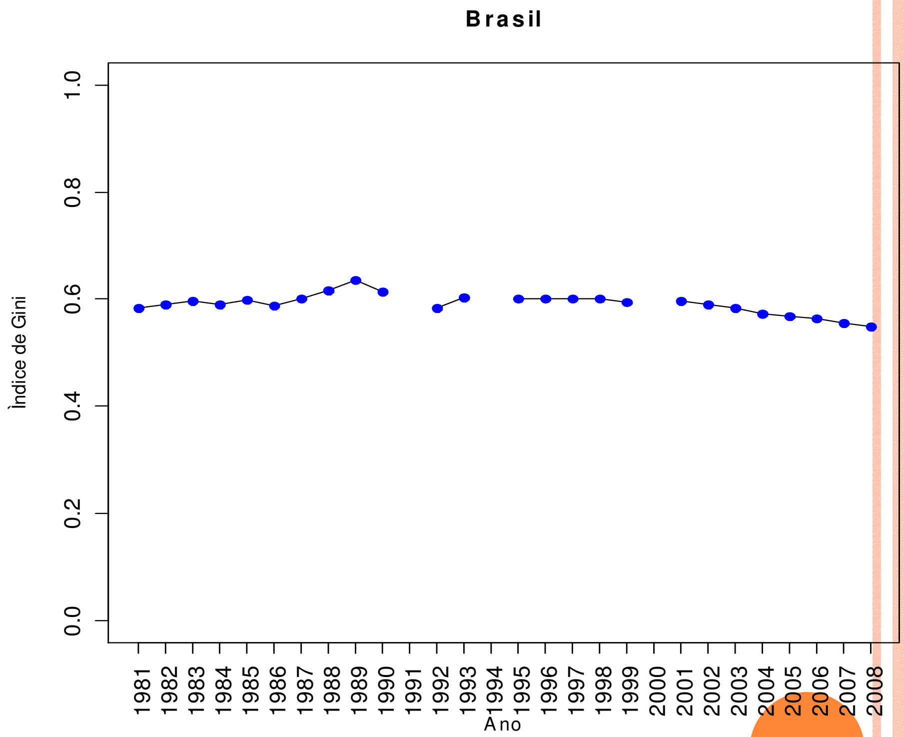


(b) G mede apenas desigualdade. Por exemplo, diferentes países podem ter valores de G semelhantes e diferentes níveis de riqueza.

7.2. Índice de Gini

Mede o grau de desigualdade existente na distribuição de indivíduos segundo a renda domiciliar *per capita*. Seu valor varia de 0, quando não há desigualdade (a renda de todos os indivíduos tem o mesmo valor), a 1, quando a desigualdade é máxima (apenas um um indivíduo detém toda a renda da sociedade e a renda de todos os outros indivíduos é nula). Fonte: http://www.pnud.org.br/popup/pop.php?id_pop=97.

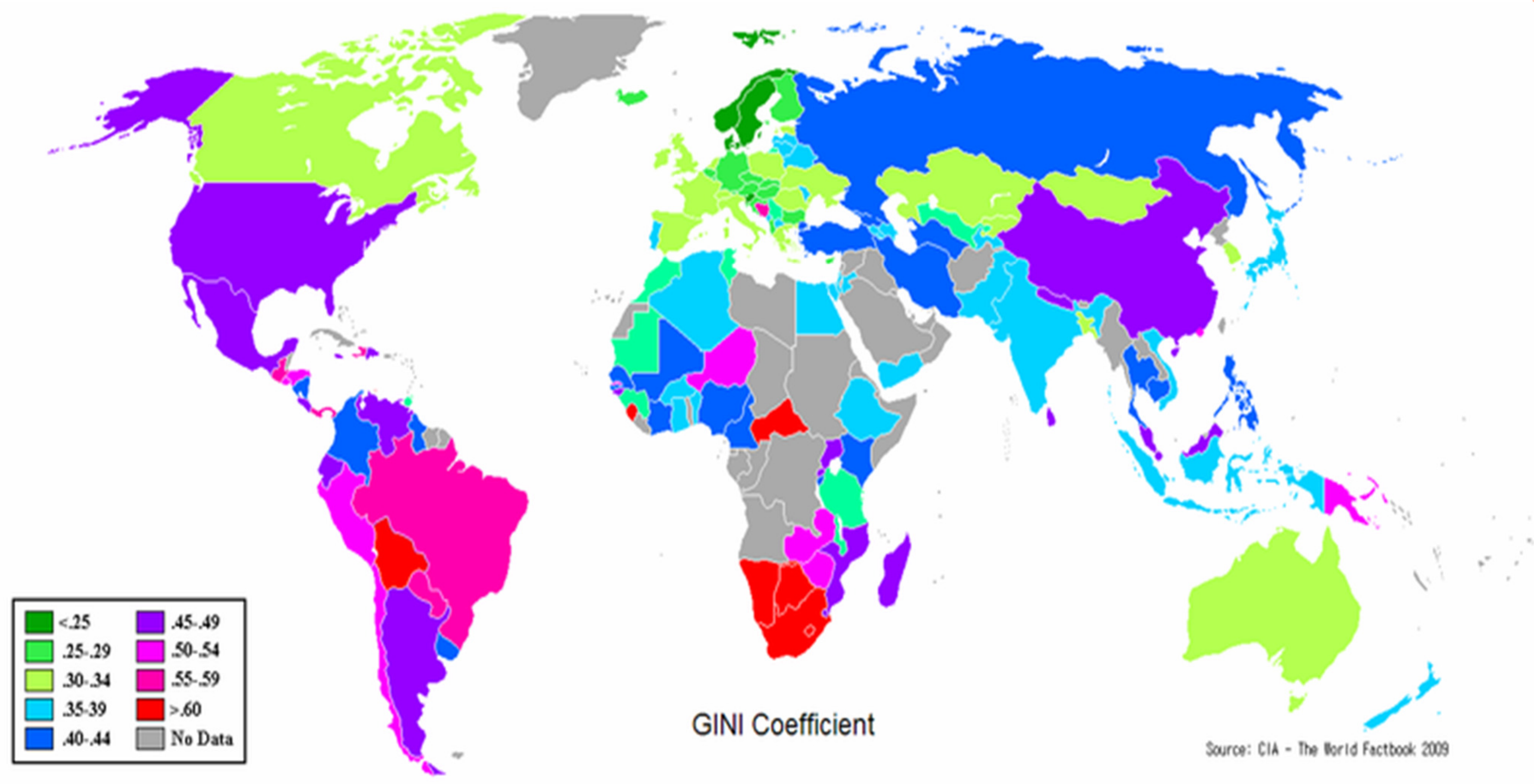




Redução importante nos últimos anos.

Pouca variação.

7.2. Índice de Gini



Gini Coefficient World CIA Report 2009.

Obs. Exemplo de um cartograma.

7.2. Índice de Gini

Diferença média. Medida de dispersão dada por

$$\bar{d} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} | x_i - x_j |.$$

Diferenças ($n = 5$):

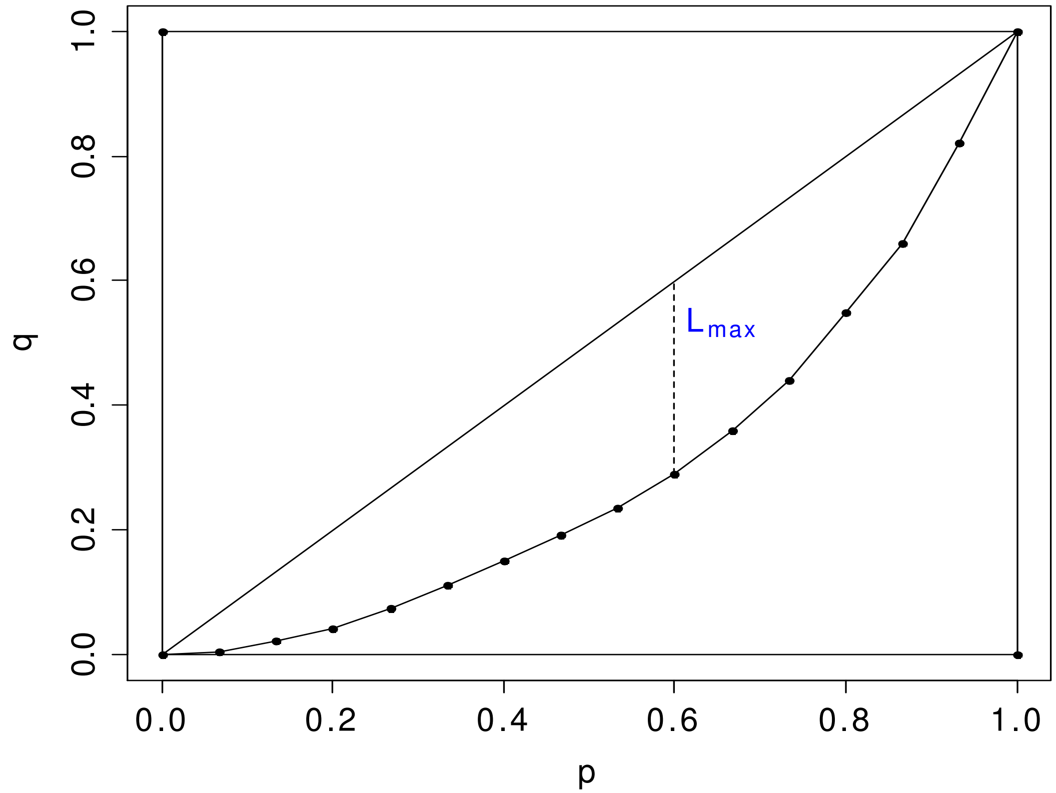
	x_1	x_2	x_3	x_4	x_5
x_1	$x_1 - x_1 = 0$	$x_1 - x_2$	$x_1 - x_3$	$x_1 - x_4$	$x_1 - x_5$
x_2	$x_2 - x_1$	$x_2 - x_2 = 0$	$x_2 - x_3$	$x_2 - x_4$	$x_2 - x_5$
x_3	$x_3 - x_1$	$x_3 - x_2$	$x_3 - x_3 = 0$	$x_3 - x_4$	$x_3 - x_5$
x_4	$x_4 - x_1$	$x_4 - x_2$	$x_4 - x_3$	$x_4 - x_4 = 0$	$x_4 - x_5$
x_5	$x_5 - x_1$	$x_5 - x_2$	$x_5 - x_3$	$x_5 - x_4$	$x_5 - x_5 = 0$

Pode ser provado que $G = \dfrac{\bar{d}}{2\,\bar{x}}$.

G é uma medida de dispersão relativa.

7.3. Discrepância máxima

Medida associada à curva de Lorenz. Valor máximo da diferença entre a proporção acumulada de posições e a proporção acumulada de valores: $L_{max} = \max(p_i - q_i)$, $i = 1,...,n$.



Declividade da curva:

$$B_i = \frac{q_i - q_{i-1}}{p_i - p_{i-1}} = \frac{x_{(i)}}{\bar{x}}, \; i = 1,...,n.$$

$$x_{(i)} \le \bar{x} \Rightarrow B_i \le 1.$$

$$x_{(i)} > \bar{x} \Rightarrow B_i > 1.$$

Encontrar j tal que $x_{(1)} \le x_{(2)} \le ... \le x_{(j)} \le \bar{x} < x_{(j+1)} \le ... \le x_{(n)}$. $L_{max} = p_j - q_j$.

Pode ser provado que $L_{max} = \dfrac{\text{dm}}{2\bar{x}}$. L_{max} é uma medida de dispersão relativa.

Medidas de desigualdade em R

Pacote `ineq`

```
> library(ineq)
```

15 observações

```
> x = c(2.8, 13.7, 6.8, 12.1, 1.1, 5.9, 4.5, 9.6, 2.3, 28.9, 6.7, 0.4, 5.6, 8.0, 10.3)
> summary(x)
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  0.400    3.650    6.700    7.913    9.950   28.900
```

Curva de Lorenz: função `Lc`.

```
> clorenz = Lc(x)
```

Índice de Gini:

```
> Gini(x)
[1] 0.4213423
> names(clorenz)
[1] "p"        "L"        "L.general"
```

= p   = q

```
> (jmax = which.max(clorenz$p - clorenz$L))
 [1] 10
> (Lmax = clorenz$p[jmax] - clorenz$L[jmax])
[1] 0.2958719
> c(clorenz$p[jmax], clorenz$L[jmax])
[1] 0.6000000 0.3041281
```

Medidas de desigualdade em R

Curva de Lorenz e discrepância máxima (L_{max}):

```
> plot(clorenz, main = "", ylab = "q")
> segments(clorenz$p[jmax], clorenz$L[jmax],  clorenz$p[jmax],
clorenz$p[jmax],lty = 2)
```