Quartis, Decis e Percentis

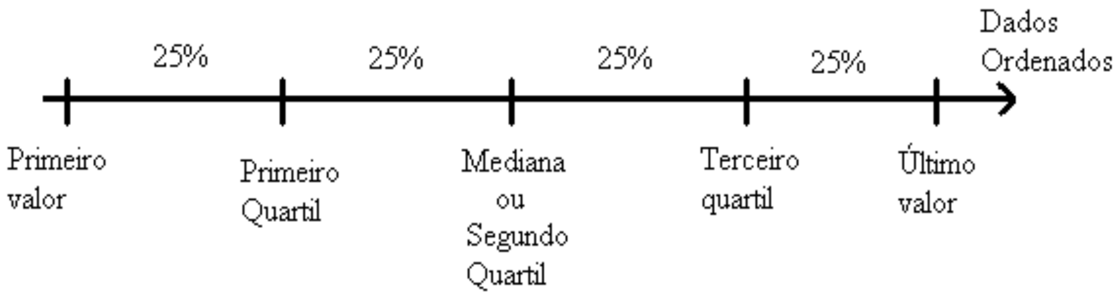
A mediana é o valor que separa a quantidade de dados em duas partes iguais: 50% dos dados abaixo dela e 50% acima. Assim como a mediana, existem outros valores que separam os dados em partes iguais. Eles são chamados genericamente de *quantis*. Os quantis mais importantes e usados são:

- Quartis: dividem os dados em quartas partes (cada parte tem 25% dos dados). São indicados por Q_1, Q_2 = Md e Q_3.

- Decis: dividem os dados em décimas partes (cada parte tem 10% dos dados). São indicados por D_1, D_2, ..., D_9.

- Percentis: dividem os dados em centésimas partes (cada parte tem 1% dos dados). São indicados por P_1, P_2, ..., P_{99}.

Um conjunto de dados pode ser dividido em 3 quartis, 9 decis e 99 percentis. Veja o exemplo a seguir para os quartis.



Divisão de um intervalo de dados em quartis

Para uma coleção de n dados discretos, os postos dos quartis, decis e percentis são calculados como:

Quartis: 1º quartil: $(n)/4$; 2º quartil (Md): $2n/4 = n/2$; 3º quartil: $3n/4$.

<u>Decis</u>: 1º decil (D_1): $n/10$; 2º decil: $2n/10$; ...; i-ésimo decil: $in/10$; ...;
9º decil: $9n/10$.

<u>Percentis</u>: 1º percentil (P_1): $n/100$; 2º percentil: $2n/100$; ...;
i-ésimo percentil: $in/100$; ... ; 99º percentil: $99n/100$.

A partir do posto, pode-se calcular o valor do quartil, do decil ou do percentil desejado. Como regra geral, se o posto coincide com um número inteiro i o valor a ser usado é o da média aritmética entre os dados que ocupam as posições i e i+1. Já se o posto não for um número inteiro a convenção que vamos usar é arredondar para a posição do número inteiro acima do posto e tomar o valor correspondente.

Por exemplo, sejam os 16 números ordenados:

0,5; 0,7; 0,7; 0,9; 1,0; 1,1; 1,1; 1,2; 1,3; 1,3; 1,5; 1,8; 2,1; 2,2; 2,5; 2,5.

Posto de $Q_1 = n/4 = 16/4 = 4$. Toma-se a média entre o 4º e o 5º valores: $Q_1 = (0{,}9 + 1{,}0)/2 = 0{,}95$.

Posto de $Q_2 = n/2 = 16/2 = 8$. $Q_2 = Md = (1{,}2 + 1{,}3)/2 = 1{,}25$.

Posto de $Q_3 = 3n/4 = 3.4 = 12$. $Q_3 = (1{,}8 + 2{,}1)/2 = 1{,}95$.

Posto de $D_1 = n/10 = 16/10 = 1{,}6$. Arredondando para 2, $D_1 = 0{,}7$.

Posto de $D_9 = 9.n/10 = 9.1{,}6 = 14{,}4$. Arredondando para 15, $D_9 = 2{,}2$.

Posto de $P_{95} = 95.n/100 = 95.0{,}16 = 15{,}2$. Arredondando para 16, $P_{95} = 2{,}5$.

Quando temos dados agrupados, os quartis, decis e percentis podem ser calculados por um raciocínio idêntico ao que foi usado para o cálculo da mediana.

Exemplo: Um teste de raciocínio abstrato foi aplicado a 816 alunos de uma escola de 1º grau, dando os seguintes resultados:

Pontos Alcançados no Teste	Nº de Alunos	Freqüência Acumulada
4 ├ 8	10	10
8 ├ 12	89	99
12 ├ 16	206	305
16 ├ 20	219	524
20 ├ 24	155	679
24 ├ 28	78	757
28 ├ 32	30	787
32 ├ 36	18	805
36 ├ 40	11	816
Total	816	816

a) Qual é o máximo de pontos que classifica um aluno entre os 25% mais fracos?
   O valor pedido é o do primeiro quartil, Q_1. O Posto do Q_1 é 816/4 = 204. Portanto, o Q_1 é média aritmética entre o 204º e o 205º elementos, ou seja é o 204,5º elemento, que cai entre 12 e 16.

$$Q_1 = L_i + (P - f_{ai})\frac{h}{f_q} = 12 + (204{,}5 - 99)\frac{4}{206} = 12 + 2{,}05 = 14{,}05.$$

b) Qual é o mínimo de pontos necessários para um aluno se classificar entre os 25% mais fortes?
   O valor pedido é o do terceiro quartil, Q_3. O posto do Q_3 é P = 3.816/4 = 612. Portanto, o Q_3 é o 612,5º elemento, que cai na quinta classe, entre 20 e 24.

$$Q_3 = L_i + (P - f_{ai})\frac{h}{f_q} = 20 + (612{,}5 - 524)\frac{4}{155} = 20 + 2{,}28 = 22{,}28.$$

c) Qual é o máximo de pontos que ainda classifica um aluno entre os 10% mais fracos?

O valor pedido é o do primeiro decil, D_1. O posto do D_1 é P = 816/10 = 81,6, que será arredondado para 82. Portanto, o D_1 cai na segunda classe, entre 8 e 12.

$$D_1 = L_i + (P - f_{ai})\frac{h}{f_d} = 8 + (82 - 10)\frac{4}{89} = 8 + 3{,}24 = 11{,}24.$$

d) Qual é o mínimo de pontos para que um aluno esteja entre os 10% mais fortes?

O valor pedido é o do nono decil, D_9. O posto do D_9 é P = 9.(816/10) = 734,4, arredondado para 735. Portanto, o D_9 cai na sexta classe, entre 24 e 28.

$$D_9 = L_i + (P - f_{ai})\frac{h}{f_d} = 24 + (735 - 679)\frac{4}{78} = 24 + 2{,}87 = 26{,}87.$$

e) Qual é o máximo de pontos que ainda classifica o aluno entre os 1% mais fracos?

O valor pedido é o do primeiro percentil, C_1. O posto do C_1 é P = 816/100 = 8,16, arredondado para 9. Portanto, o C_1 cai na primeira classe, entre 4 e 8.

$$C_1 = L_i + (P - f_{ai})\frac{h}{f_c} = 4 + (9 - 0)\frac{4}{10} = 4 + 3{,}6 = 7{,}6.$$

f) Qual é o mínimo de pontos para que um aluno esteja entre os 5% mais fortes?

O valor pedido é o do 95º percentil, C_{95}. O posto do C_{95} é P = 95.(816/100) = 775,2, arredondado para 776. Portanto, o C_{95} cai na sétima classe, entre 28 e 32.

$$C_{95} = L_i + (P - f_{ai})\frac{h}{f_c} = 28 + (776 - 757)\frac{4}{30} = 28 + 2{,}5 = 30{,}5.$$

Exemplo. Distribuição de renda de uma população: Construção da curva de Lorenz e o índice de Gini

Uma aplicação importante do conceito quantis ocorre na caracterização da distribuição de renda de uma população através da construção da chamada curva de Lorenz.

Considere a população economicamente ativa de um país, por exemplo, o Brasil. Seja N o tamanho dessa população. Suponha que as rendas de todas essas N pessoas sejam conhecidas; vamos denominar a renda da *i*-ésima pessoa de r_i. Podemos então ordenar as rendas em ordem crescente, desde a mais baixa até a mais alta:

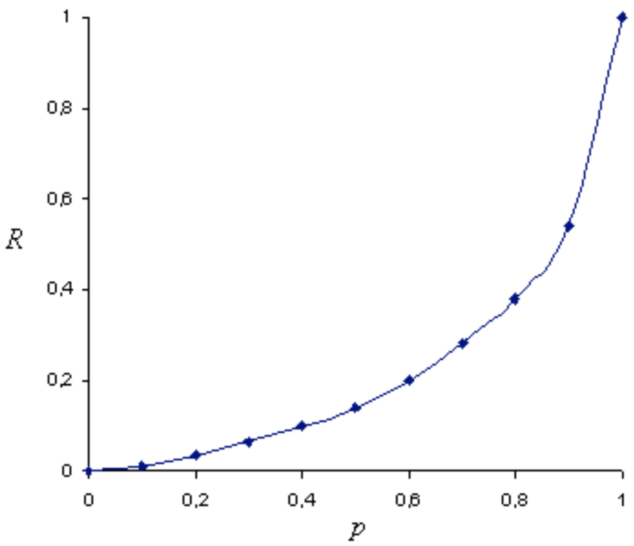
$$r_1, r_2, r_3, r_4, \ldots, r_{N-1}, r_N.$$

Com as rendas ordenadas, podemos dividi-las em quantis. A título de exemplo, vamos considerar aqui que elas foram divididas em decis: o primeiro decil separa os 10% mais pobres do resto da população, o segundo decil separa os 20% mais pobres e assim por diante. Para cada decil, podemos calcular qual a proporção da renda total da população que corresponde aos indivíduos delimitados por ele. Levantando essa informação para todos os decis, podemos montar uma tabela como a dada abaixo (os dados mostrados são fictícios).

Estrato da população	Proporção da população	Proporção da renda	Proporção acumulada da população (*p*)	Proporção acumulada da renda (*R*)
Até o 1º decil	0,10	0,01	0,10	0,01
Até o 2º decil	0,10	0,025	0,20	0,035
Até o 3º decil	0,10	0,03	0,30	0,065
Até o 4º decil	0,10	0,035	0,40	0,1
Até o 5º decil	0,10	0,04	0,50	0,14
Até o 6º decil	0,10	0,06	0,60	0,2
Até o 7º decil	0,10	0,08	0,70	0,28
Até o 8º decil	0,10	0,1	0,80	0,38
Até o 9º decil	0,10	0,16	0,90	0,54
Acima do 9º decil	0,10	0,46	1,00	1,00

A tabela mostra a porcentagem da renda total da população que corresponde a cada estrato contendo 10% da população. Ela também mostra, para cada decil, os valores acumulados da proporção da população (chamados de p) e da proporção da renda (chamados de R).

Com os dados da tabela pode-se montar o gráfico de R *versus* p, mostrado abaixo.



A curva formada pela união dos pontos (p, R) foi proposta inicialmente por Max Lorenz em 1905 para representar a distribuição de renda de uma população. Por isso ela é denominada de *curva de Lorenz* para a distribuição de renda da população. Ela ilustra como a proporção da renda total aumenta em função da proporção da população.

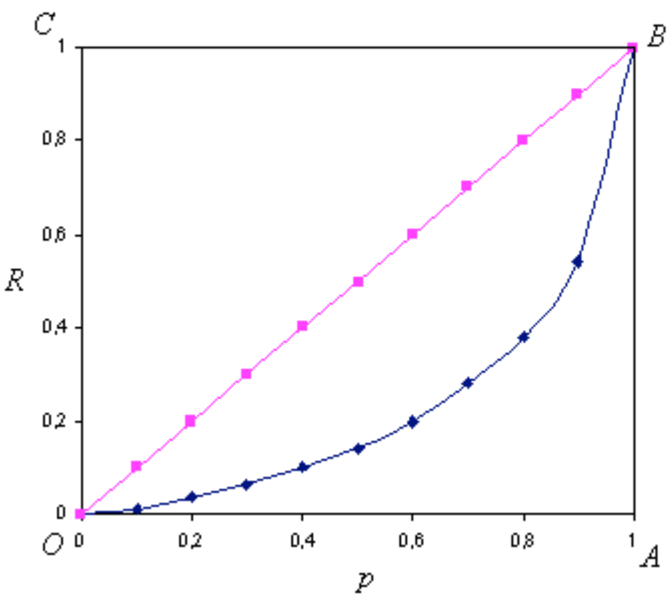
A partir da curva de Lorenz pode-se definir um índice que caracterize a distribuição de renda da população.

Do ponto de vista teórico, há dois casos extremos com relação à renda de uma população: aquele em que todos os indivíduos têm exatamente a mesma renda (distribuição perfeitamente equilibrada) e aquele em que apenas um indivíduo tem toda a renda da população e os demais nada possuem (distribuição de máxima desigualdade).

A tabela dando a distribuição de renda para a população com distribuição perfeitamente equilibrada é a seguinte:

Estrato da população	Proporção da população	Proporção da renda	Proporção acumulada da população (*p*)	Proporção acumulada da renda (*R*)
Até o 1º decil	0,10	0,10	0,10	0,10
Até o 2º decil	0,10	0,10	0,20	0,20
Até o 3º decil	0,10	0,10	0,30	0,30
Até o 4º decil	0,10	0,10	0,40	0,40
Até o 5º decil	0,10	0,10	0,50	0,50
Até o 6º decil	0,10	0,10	0,60	0,60
Até o 7º decil	0,10	0,10	0,70	0,70
Até o 8º decil	0,10	0,10	0,80	0,80
Até o 9º decil	0,10	0,10	0,90	0,90
Acima do 9º decil	0,10	0,10	1,00	1,00

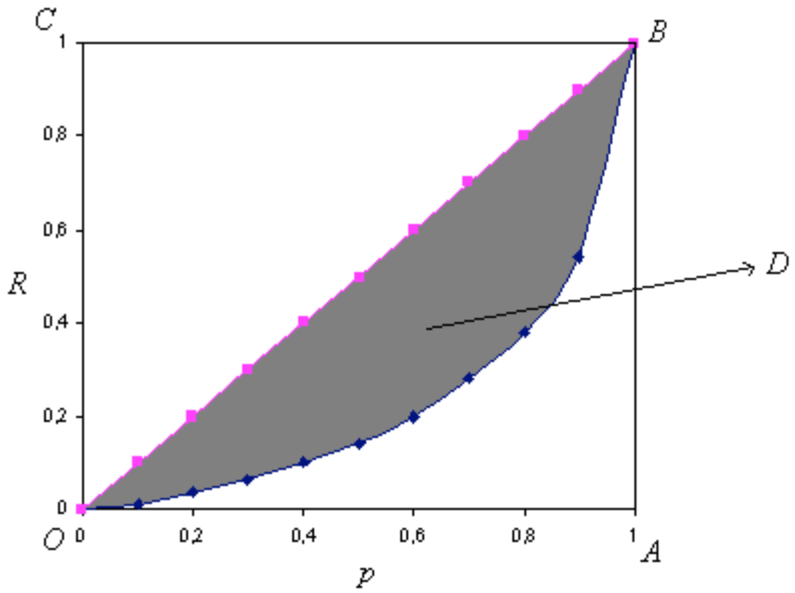
A curva de Lorenz para o caso de perfeito equilíbrio é a linha reta $R = p$, mostrada no gráfico abaixo juntamente com a curva de Lorenz do exemplo anterior.



Para o caso da distribuição de máxima desigualdade, a curva de Lorenz seria a linha reta que sai da origem no gráfico acima (ponto indicado por *O*) e vai até o ponto indicado por *A*, subindo dali até o ponto *B* (como apenas 1 indivíduo tem toda a renda da população sua posição é a última na lista ordenada de rendas, coincidindo com a linha AB).

Os dois casos extremos mencionados acima são apenas teóricos. Nenhuma população real corresponde a eles. Para qualquer população real, sua curva de Lorenz é uma curva dentro do triângulo *OAB* na figura acima. Note que quanto mais afastada estiver essa curva da linha *OB* de equilíbrio perfeito, maior será o grau de desigualdade na distribuição de renda da população.

Define-se a área entre a linha de equilíbrio perfeito e a curva de Lorenz (indicada por *D* na figura abaixo) como a área de desigualdade da população.



Observe que, como a área do triângulo *OAB* é 0,5, temos que:

$$0 \leq D \leq 0{,}5.$$

O índice de Gini (indicado por *G*) é definido como a razão entre a área de desigualdade e o máximo valor que ela pode assumir,

$$G = \frac{D}{0{,}5} = 2D.$$

Note que, pela definição,

$$0 \leq G \leq 1.$$

Quando $G = 0$ temos o caso de perfeito equilíbrio na distribuição de renda e quando $G = 1$ temos o caso de perfeito desequilíbrio na distribuição de renda.

O índice de Gini foi proposto por Corrado Gini em 1914 e é uma das principais medidas de desigualdade de renda usadas internacionalmente para avaliar populações e países.

Exercício para casa: Procure na internet pelos valores do índice de Gini para os diferentes países do mundo. Se organizarmos os valores do índice de Gini dos países em ordem crescente, em que quartil se encontra o Brasil?