

Esta aula

▶ Plano

- ▶ Dados e variáveis
- ▶ Distribuições de Frequência
- ▶ Medidas de Posição

▶ Bibliografia

- ▶ Barrow, M. Estatística para economia, contabilidade e administração. São Paulo: Ática, 2007, Cap. I
- ▶ Morettin, P. e W. Bussab. Estatística básica. 5. ed. São Paulo: Saraiva, 2005. Caps. 1 a 3
- ▶ Wooldridge, J. M. Introductory Econometrics: A modern Approach.



Dados e Variáveis



Introdução

- ▶ Em muitas situações da atividade de pesquisa, deparamo-nos com a necessidade de aplicar os modelos teóricos de relações internacionais aos dados do mundo real
- ▶ Em um primeiro momento, pode-se analisar e interpretar os dados como eles estão. Essa primeira observação nos fornece idéias para avançar
- ▶ Um segundo passo acontece através da utilização desses dados para testar uma teoria de RI ou estimar uma relação causal entre variáveis.

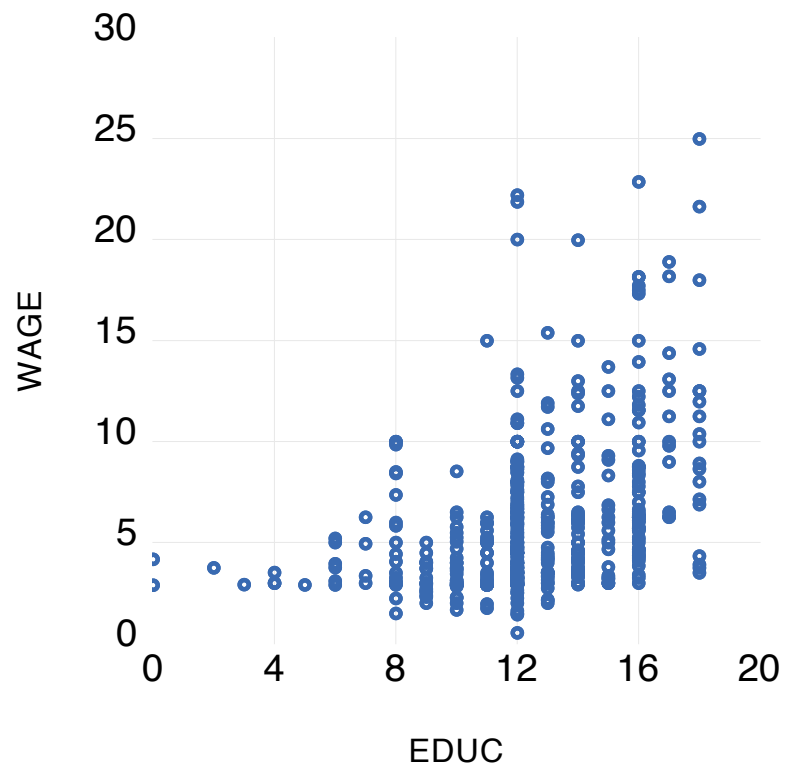
Introdução

▶ Modelos

- ▶ Modelos são explicações de fenômenos, por meio da interrelação de seus componentes (as variáveis)
- ▶ O modelo determina um comportamento previsível, regular da relação entre variáveis
- ▶ O fenômeno seria explicado pelas variáveis que se escolheu e uma parte pode também ficar não explicada (é o resíduo do modelo)

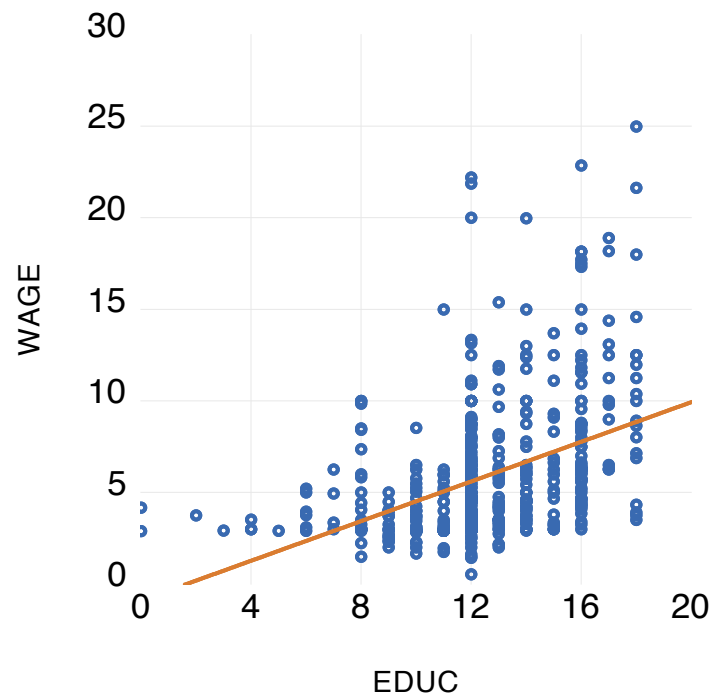
Introdução

► Modelos



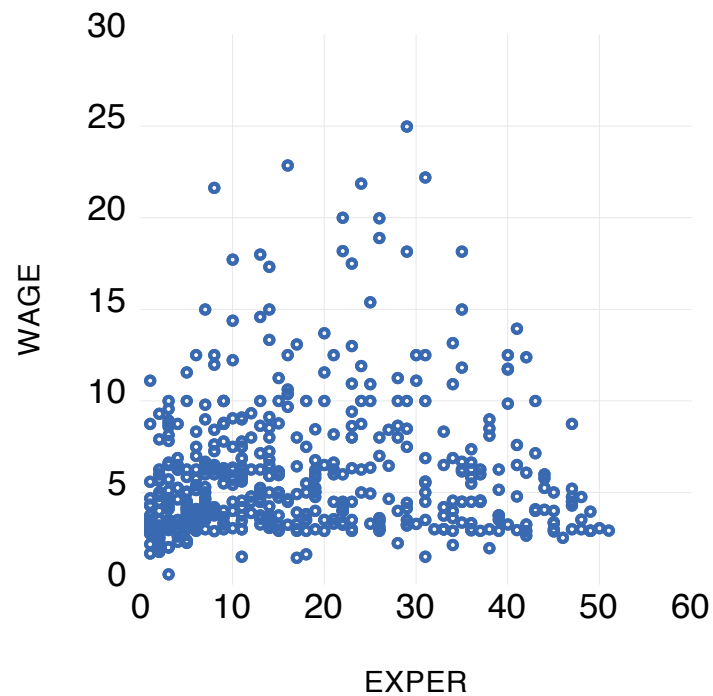
Introdução

► Modelos



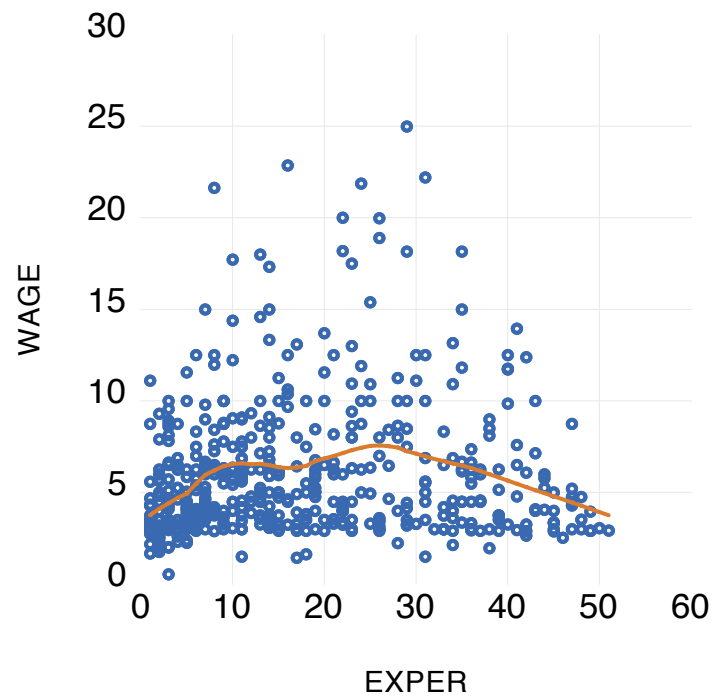
Introdução

► Modelos



Introdução

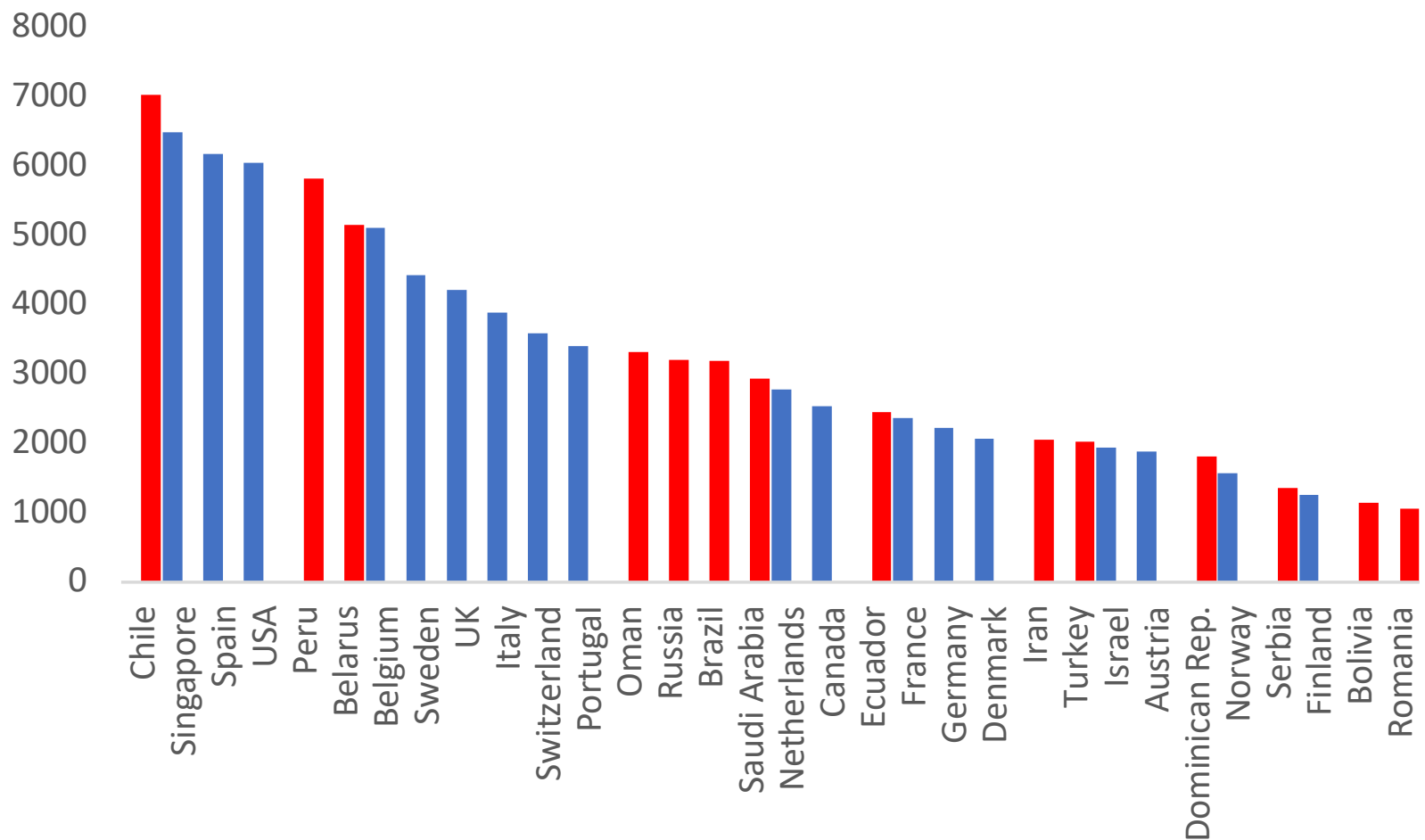
► Modelos



Introdução

- ▶ **A Estatística Descritiva busca resumir e interpretar observações empíricas disponíveis**
- ▶ **Gráficos**
 - ▶ Representar dados, variáveis graficamente é um meio muito utilizado para descrever relações entre eles
 - ▶ Facilidade de compreensão
 - ▶ Identificação de padrões

Nº Casos COVID-19 (por milhão) (JUNE-2020)



Fonte: Djankov e Panizza (2020)

Nº Casos COVID-19 (por milhão) Hoje

- ▶ Progresso da doença

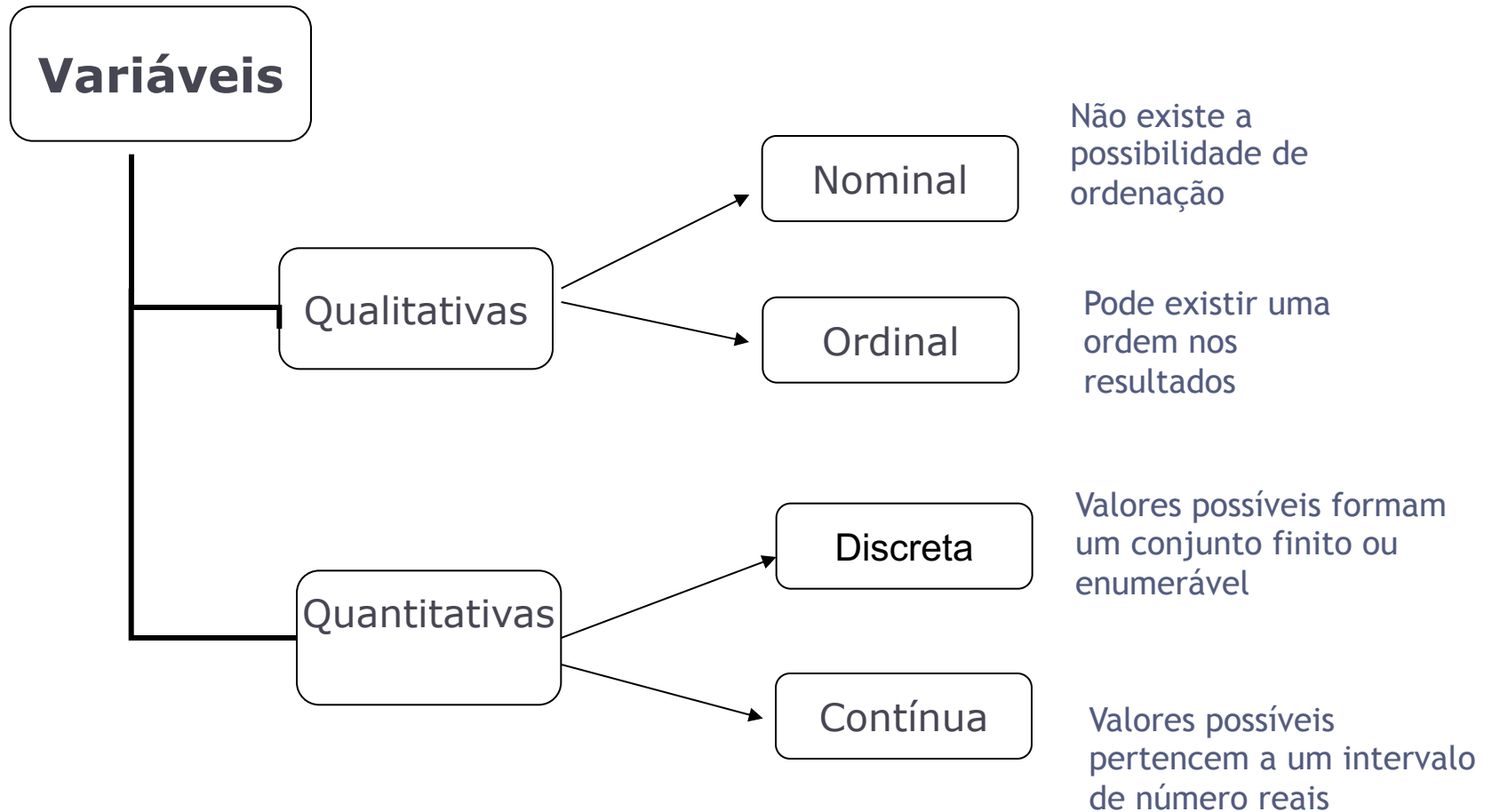
- ▶ Hoje:

<https://www.worldometers.info/coronavirus/#countries>

Dados, variáveis

- ▶ Dados são informações numéricas ou qualitativas
- ▶ As formas de coleta ou de obtenção dos dados são diversas
- ▶ Os dados podem provir de pesquisas, experimentos, etc.
- ▶ Definindo:
 - ▶ Unidade elementar: elemento de uma população (pessoa, objeto, país)
 - ▶ Variável: característica que pode assumir distintos valores por unidade elementar

Variáveis podem ser



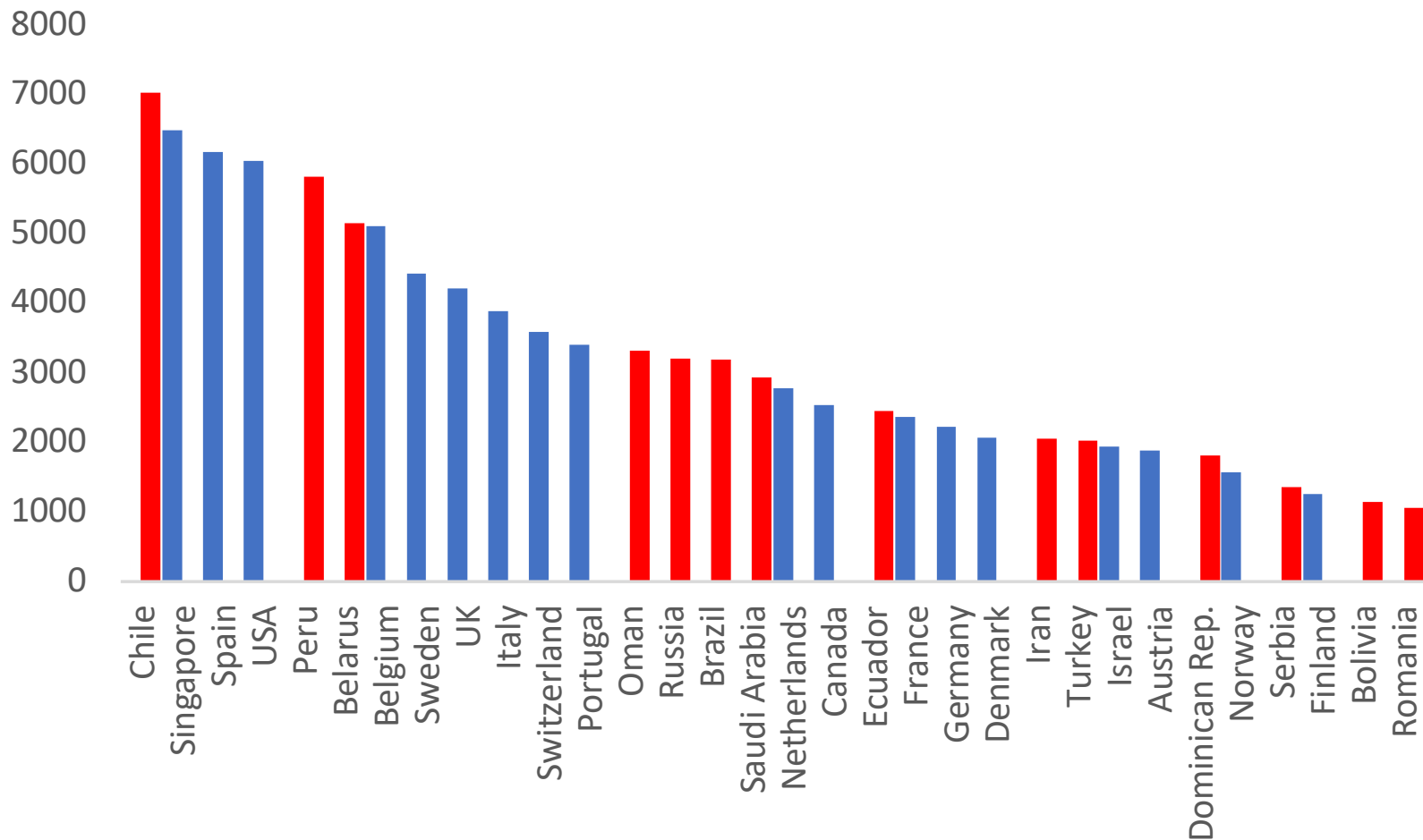
Número de variáveis

- ▶ A unidade elementar pode ter qualquer número de variáveis
 - ▶ Uma única. Exs: Exportações totais do Brasil por ano/mês; saldo de reservas internacionais de um país; número de eleitores a cada pleito
 - ▶ Duas variáveis. Exs: Exportações de petróleo e preço internacional do barril; Fluxo de divisas e taxa de câmbio
 - ▶ Três ou mais. Exs: Relação entre o saldo comercial, preços de commodities e taxa de câmbio.
- ↪ Nos casos de mais de uma variável, devem/podem ser usados métodos estatísticos para identificar relações entre elas

Tipos de dados

- ▶ **Corte transversal numa data ou período (Cross-section):**
 - ▶ São observações de distintas unidades elementares (indivíduos, firmas, países) em um determinado momento. Não se considera a evolução no tempo de uma variável, mas apenas observações pontuais (Ex: cobertura de saneamento básico em vários países).
 - ▶ Pode-se agrupar distintas cross-sections. Se a mesma unidade elementar aparece nos distintos momentos, tem-se dados em painel ou longitudinais (panel data).

Cross-Section: N° Casos COVID-19 (por milhão) (JUNE-2020)



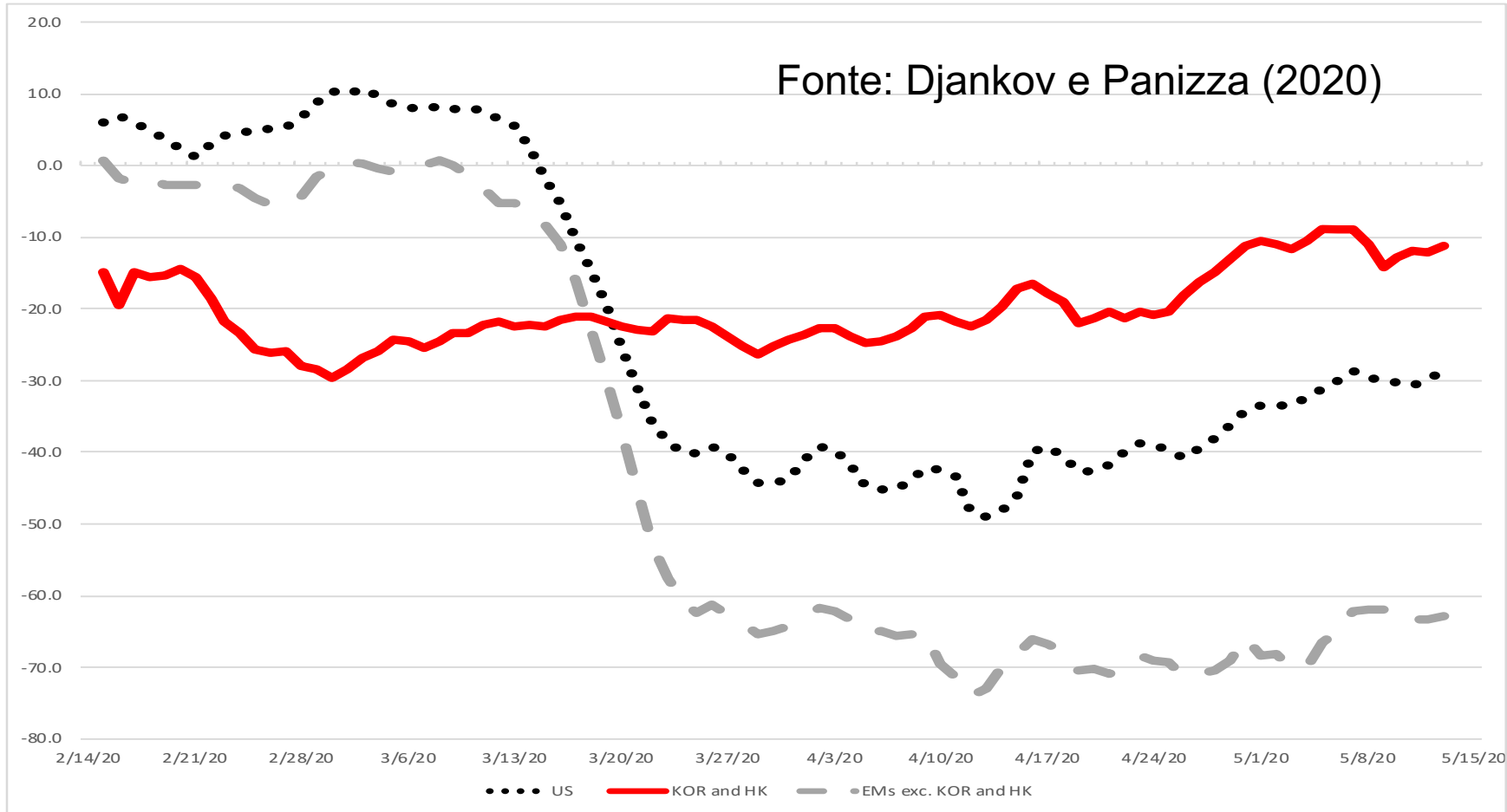
Fonte: Djankov e Panizza (2020)



Tipos de dados

- ▶ **Séries temporais:** Observações uma unidade elementar em distintos pontos do tempo. Nesse caso, deve-se considerar a presença de tendências e comportamentos sazonais

Série temporal: Mobilidade Urbana

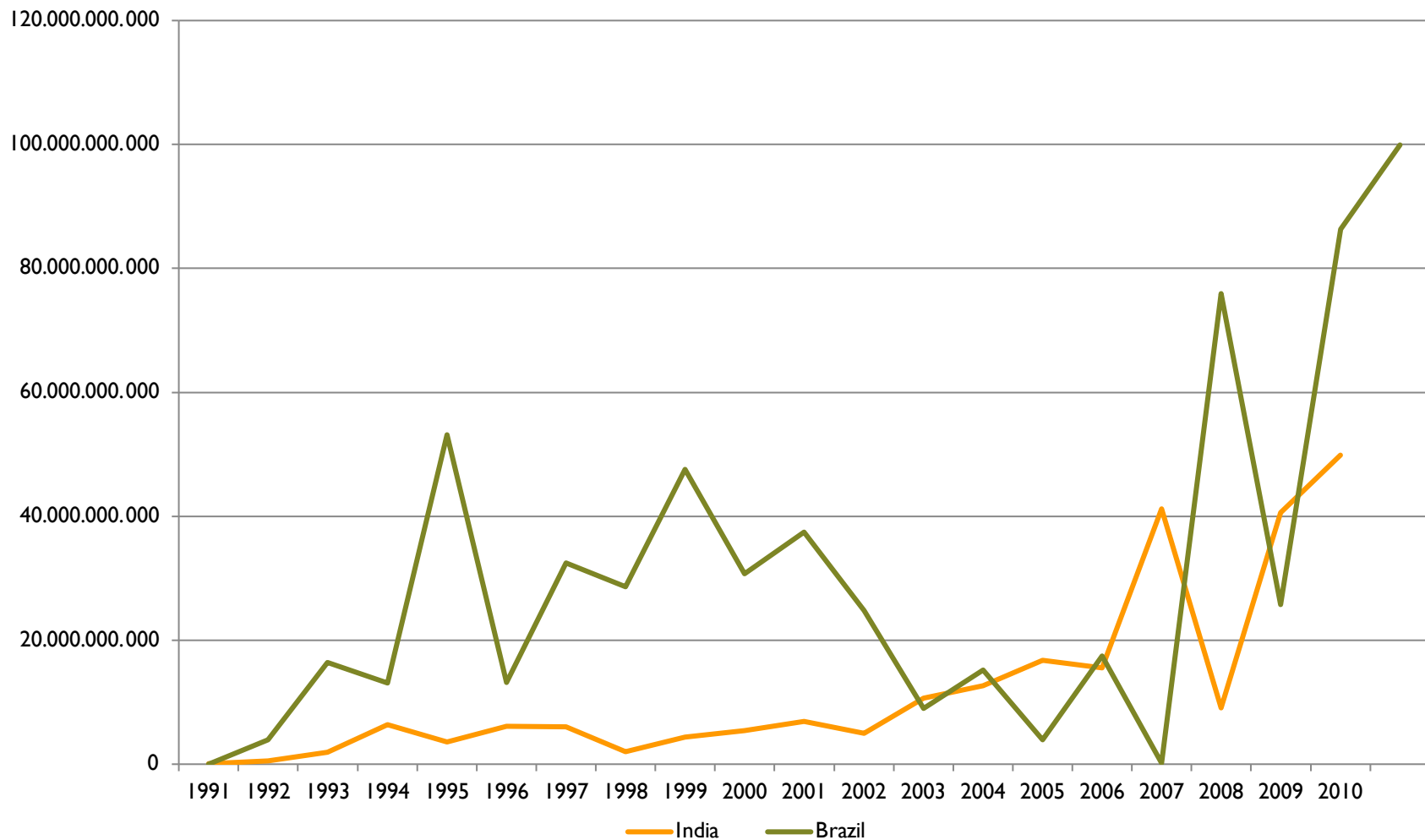


Notes: The figure plots Google Mobility Data for the retail and recreation sectors. The data measure percent change in visits to establishments in these sectors relative to the first five weeks of 2020. US is the United States; KOR and HK is the average percent change for South Korea and Hong Kong; EMs is the average percent change across the following countries: Singapore, India, Malaysia, Brazil, Chile, Colombia, Mexico, Peru, and Uruguay. Sample period 2/15/2020-5/15/2020. *Source* <https://www.google.com/covid19/mobility/>.

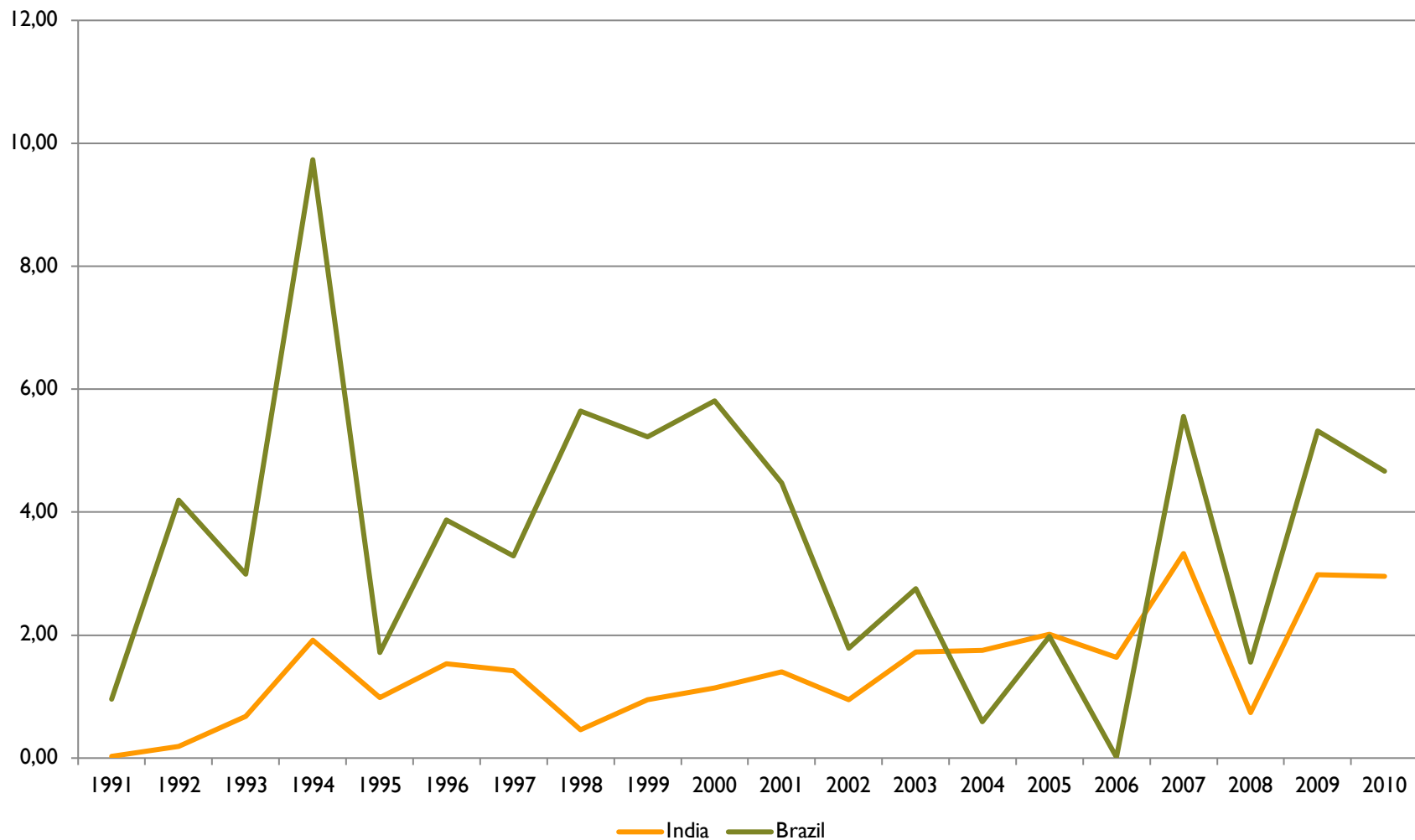
Outro aspecto se refere à unidade que escolhemos para apresentar um dado

- ▶ Quando queremos comparar mais de um país, ou estado, ou região, precisamos fazer com que sejam, de fato, comparáveis
 - ▶ Se um país tem uma economia maior, não podemos comparar com outros menores?
 - ▶ O valor menor ‘desaparece’ no gráfico?
 - ▶ O melhor a fazer, é criar um novo indicador, que descreva uma relação e possa ser comparável
 - ▶ Vejamos o exemplo de duas economias emergentes grandes, mas difíceis de comparar em alguns aspectos: Brasil e Índia

Fluxos de capitais privados, em US\$ correntes



Fluxos de capitais privados, em % do PIB





Distribuições de frequência



Distribuições de frequências

- ▶ O pesquisador está sempre interessado em estudar o comportamento de uma variável, verificando a ocorrência de suas possíveis realizações
- ▶ Por isso, ele pode começar a organizar os dados de forma a que eles lhe digam alguma coisa.
- ▶ A primeira delas, é ver com que frequência acontecem (aparecem), ou as variáveis assumem determinados valores

Tabelas de frequência

- ▶ Quando queremos estudar a distribuição de valores que assume uma variável, podemos agrupar estes valores em intervalos
- ▶ A distribuição de frequência é um agrupamento de dados em classes, ou intervalos, para os quais se observa o número de observações em cada classe

Tabelas de frequência e dados quantitativos discretos

- ▶ Frequência (absoluta) do valor de uma variável é o número de repetições desse valor
- ▶ Relacionando os valores que assume uma variável e suas frequências respectivas, temos a **distribuição de frequências absolutas**
- ▶ **Frequência relativa** do valor de uma variável é obtida dividindo sua frequência absoluta pelo tamanho da amostra
⇒ **distribuição de frequências relativas**
- ▶ **Frequência acumulada** de uma variável é a soma das freq. absolutas e relativas desde a frequência relativa ao valor inicial da variável

Elementos de uma distribuição de frequência

- ▶ **Classes:** caso as colunas da tabela de distribuição de frequência contenham muitos valores elencados, podemos reduzir a quantidade desses valores elencados agrupando-os em intervalos.
- ▶ Esses agrupamentos de valores num intervalo de abrangência são chamados de **classes**

Tabelas de frequência com dados contínuos

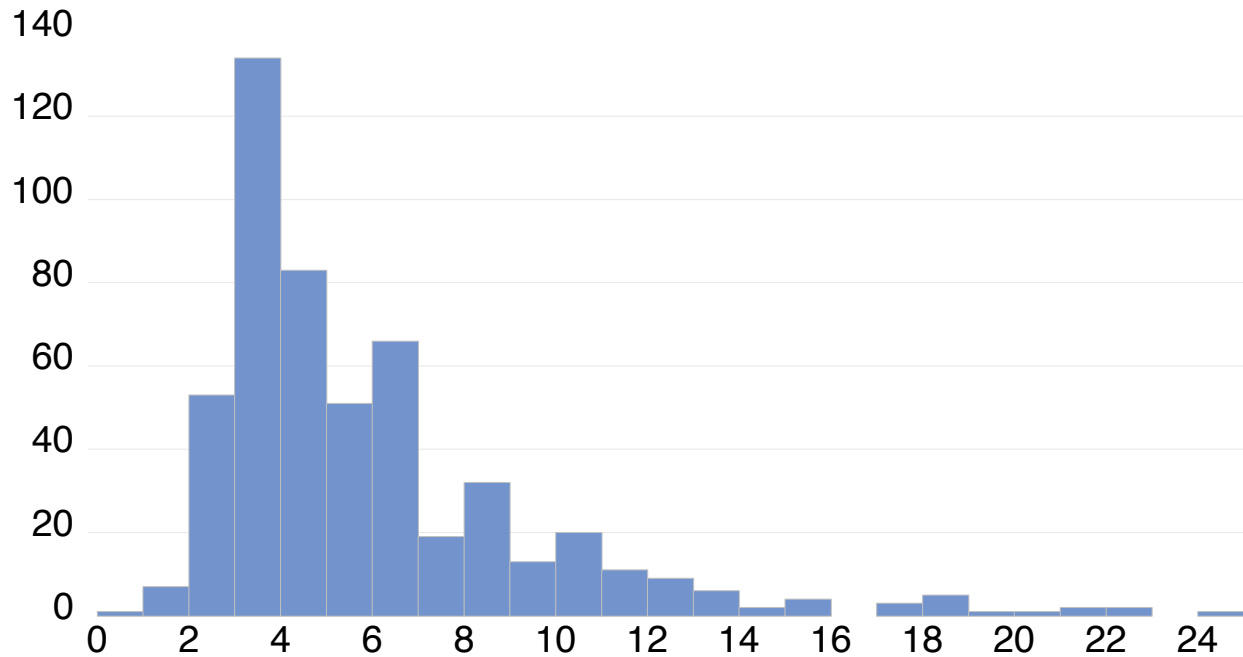
- ▶ Quando não se trabalha com valores inteiros (variáveis discretas), fica inviável determinar o número de vezes que um valor ocorre
- ▶ Por isso, o interessante é trabalhar com classes de valores
 - ▶ Definir a quantidade, limites e amplitude das classes
 - ▶ Muitas vezes, a escolha de intervalos e número de classes pode ser arbitrária e depender da sensibilidade do pesquisador
 - ▶ Com um pequeno número de classes, pode-se perder informação e um número grande dificulta o resumo dos dados

Tabelas de frequência com dados contínuos (Exemplo: Salários)

Tabulation of WAGE
Date: 04/01/21 Time: 11:37
Sample: 1 526
Included observations: 526
Number of categories: 5

Value	Count	Percent	Cumulative Count	Cumulative Percent
[0, 5)	278	52.85	278	52.85
[5, 10)	181	34.41	459	87.26
[10, 15)	48	9.13	507	96.39
[15, 20)	13	2.47	520	98.86
[20, 25)	6	1.14	526	100.00
Total	526	100.00	526	100.00

Gráfico de frequência com dados contínuos



Medidas de posição

O que são essas medidas?

- ▶ Tabelas de frequência, gráficos e um ordenamento dos dados são instrumentos poderosos para resumir essas informações sobre o comportamento de uma variável
- ▶ Mas, muitas vezes, precisamos resumir de forma ainda mais concisa e encontrar um ou poucos valores que digam muito sobre uma série de dados, que sejam representativos dela
- ▶ Usamos medidas de ordenamento ou de posição quando queremos resumir e analisar uma amostra ou a população toda

Medidas de posição central

- ▶ Em geral, utilizam-se três medidas principais
 - ▶ **Moda:** é a realização mais frequente do conjunto de valores observados.
 - ▶ **Mediana:** é a realização que ocupa a posição central na série, quando os dados estão organizados em ordem crescente.
 - ▶ **Média aritmética:** como bem sabemos, é a soma dos valores observados dividida pelo número de observações.

Formalizando os conceitos

- ▶ Se x_1, x_2, \dots, x_n são os valores da variável X , a média aritmética pode ser escrita
- ▶ Se tivermos n observações de X , das quais n_1 são iguais a x_1 , n_2 são iguais a x_2 , a média pode ser escrita

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

- ▶ Se $f_i = n_i/n$ for a frequência relativa da observação x_i , então

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n} = \frac{1}{n} \sum_{i=1}^n n_i x_i$$

$$\bar{x} = \sum_{i=1}^k f_i x_i$$



Formalizando os conceitos

- ▶ Considerando as observações ordenadas em ordem crescente e sendo a menor observação $x_{(1)}$, etc., até $x_{(n)}$
- ▶ Assim: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n-1)} \leq x_{(n)}$
- ▶ A mediana pode ser definida por:

$$\text{md}(X) = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{se } n \text{ ímpar} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{se } n \text{ par} \end{cases}$$

Gráfico das Frequências Absolutas e Medidas de Posição

