

INTRODUÇÃO

DADOS, VARIÁVEIS: O QUE FAZER COM ELES?

□ Objetivo

- Apresentar introdução aos estudos empíricos relacionados com as grandes áreas do curso de RI.
- Apresentar os conceitos estatísticos necessários para uma análise descritiva dos dados
 - Particularmente, dados de economia, história e política comparada em laboratório de informática usando como ferramenta o software Excel.

Pilares do programa

3

- Iniciação à pesquisa empírica
- Definição de variáveis, levantamento de dados
- Estatística descritiva: instrumentos importantes (análise no tempo, conversão em diferentes moedas e ajustes por índices de preços)
- Conceitos de probabilidade
- Distribuições de probabilidade

Método

4

- Estudo de conceitos teóricos sempre seguido de aplicação em laboratório de informática
- Fundamental experiência na construção de base de dados
- Elaboração de um trabalho de pesquisa empírico

Avaliação

5

- Processo de avaliação constante e gradual \Rightarrow não dá para “perder o pé”
- Ajustes com o curso “à distância”
- Cálculo da média

$$MF = \left[\frac{(L_1 + L_2 + L_3)}{3} \right] \times 0,3 + [FP \times 0,1 + VP \times 0,2 + VF \times 0,8] \times 0,5 + [Exerc] \times 0,2$$

- Dica: não ir deixando o conteúdo e as atividades para “fazer depois”...

Agenda do semestre

6

Data	Compromisso
14-17/05	Lista 1 (sobre definição de variáveis e hipóteses)
Início junho	Entrega da ficha com propostas de variáveis e temas para o projeto de trabalho
07/05	Entrega da versão preliminar do trabalho
11-14/06	Lista 2 (medidas de posição e dispersão combinadas com N° índices)
09-13/07	Lista 3 (distribuição de probabilidades)
21/07	Entrega, apresentação/discussão dos trabalhos

“Contrato de trabalho”

7

- ▶ Tendo em vista que
 - ▶ Somos numerosos
 - ▶ O curso traz informações novas quanto a seu conteúdo teórico e aplicado
 - ▶ Trabalharemos à distância
 - ▶ E é extremamente interessante, se estivermos “inteiros” nas aulas

- ▶ É fundamental
 - ▶ **Estar em dia com os conceitos que vão sendo apresentados aula a aula**
 - ▶ **“Presença” às aulas**
 - ▶ Saber que dúvidas, opiniões, debates são **sempre bem vindos**
 - ▶ Fazer os exercícios à medida em vão sendo propostos

- **ÉTICA:** Como é de conhecimento geral, espera-se que todos os trabalhos produzidos pelos estudantes sejam, de fato, produzidos por eles mesmos. A apresentação de um trabalho ou de palavras, ideias, opiniões, imagens ou dados criados por outros, no todo ou em parte, como se eles fossem originalmente desenvolvidos pelo estudante para a atribuição de nota, intencionalmente ou não, constitui um **ato de plágio**.

8

Começando de verdade

Dados e variáveis

Trabalhar com dados

- Em muitas situações da atividade de pesquisa, deparamo-nos com a necessidade de analisar dados e trabalhar com eles de maneira a construir informações
- Em um primeiro momento, pode-se analisar e interpretar os dados como eles estão. Essa primeira observação nos dá idéias para avançar
- Um segundo passo se dá pela verificação de relações e elaboração de modelos nas observações empíricas disponíveis

Trabalhar com dados

10

□ Modelos

- Uma análise de dados tem por objetivo identificar um padrão ou um modelo nas observações
- Modelos são representações de fenômenos, por meio da interrelação de seus componentes (as variáveis)
- O modelo determina um comportamento previsível, regular da relação entre variáveis
- Portanto, podemos tentar explicar um fenômeno a partir do comportamento de diversas variáveis
- Os dados deste fenômeno seriam explicados pelas variáveis que se escolheu e uma parte pode também ficar não explicada (é o resíduo do modelo)

Trabalhar com dados

11

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

Exemplo: Consideremos as taxas de variação anual dos índices de preços ao consumidor em seis países latino-americanos.

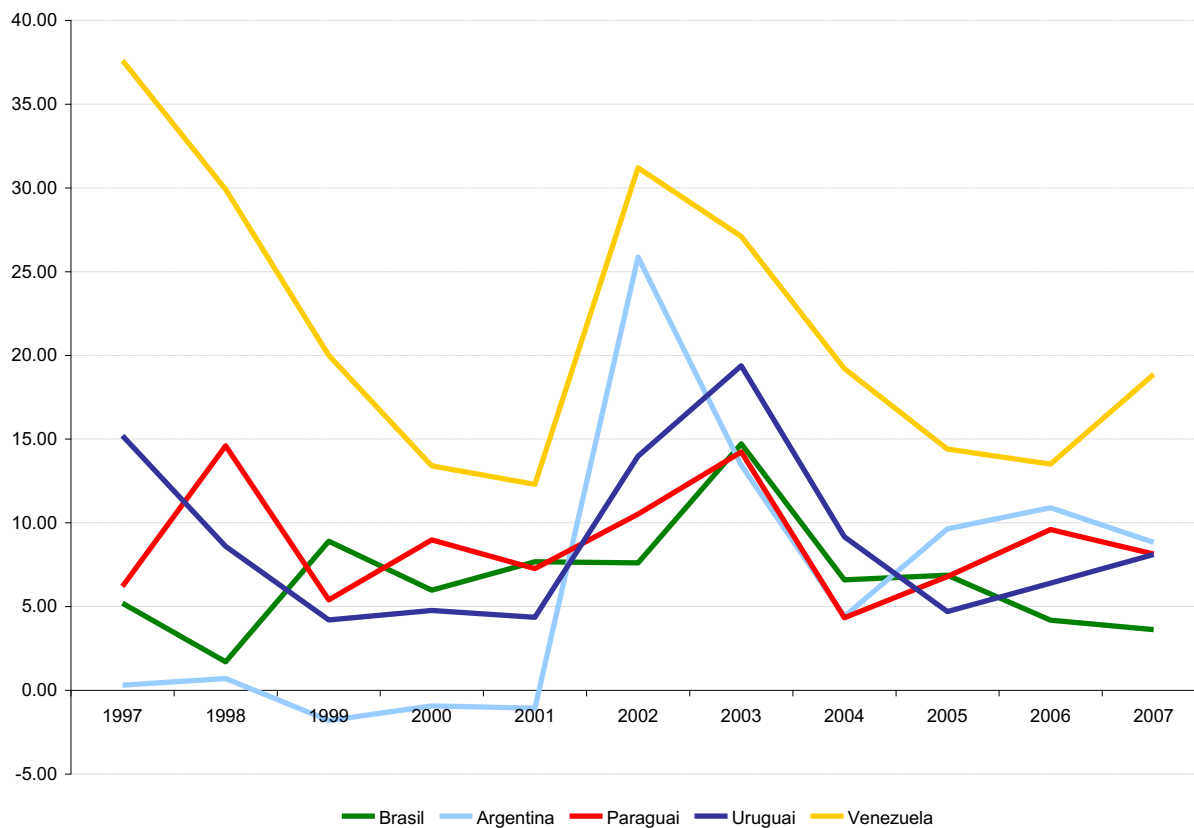
○ que podemos observar, assim, de cara?

	Brasil	Argentina	Paraguai	Uruguai	Venezuela
1997	5.20	0.30	6.20	15.20	37.60
1998	1.70	0.70	14.60	8.60	29.90
1999	8.90	-1.80	5.40	4.20	20.00
2000	5.97	-0.94	8.98	4.76	13.40
2001	7.67	-1.07	7.27	4.36	12.30
2002	7.61	25.87	10.51	13.97	31.20
2003	14.71	13.44	14.22	19.38	27.10
2004	6.60	4.42	4.33	9.16	19.20
2005	6.87	9.64	6.79	4.70	14.40
2006	4.18	10.90	9.59	6.40	13.50
2007	3.62	8.83	8.13	8.11	18.86

Fonte: CEPAL

Fica melhor para identificar padrões/tendências/ relações assim?

Taxas de variação de preços ao consumidor, % ao ano



Fonte: CEPAL

Com outros países e dados mais atuais

14

Taxas de
variação
anual de
preços ao
consumidor, %

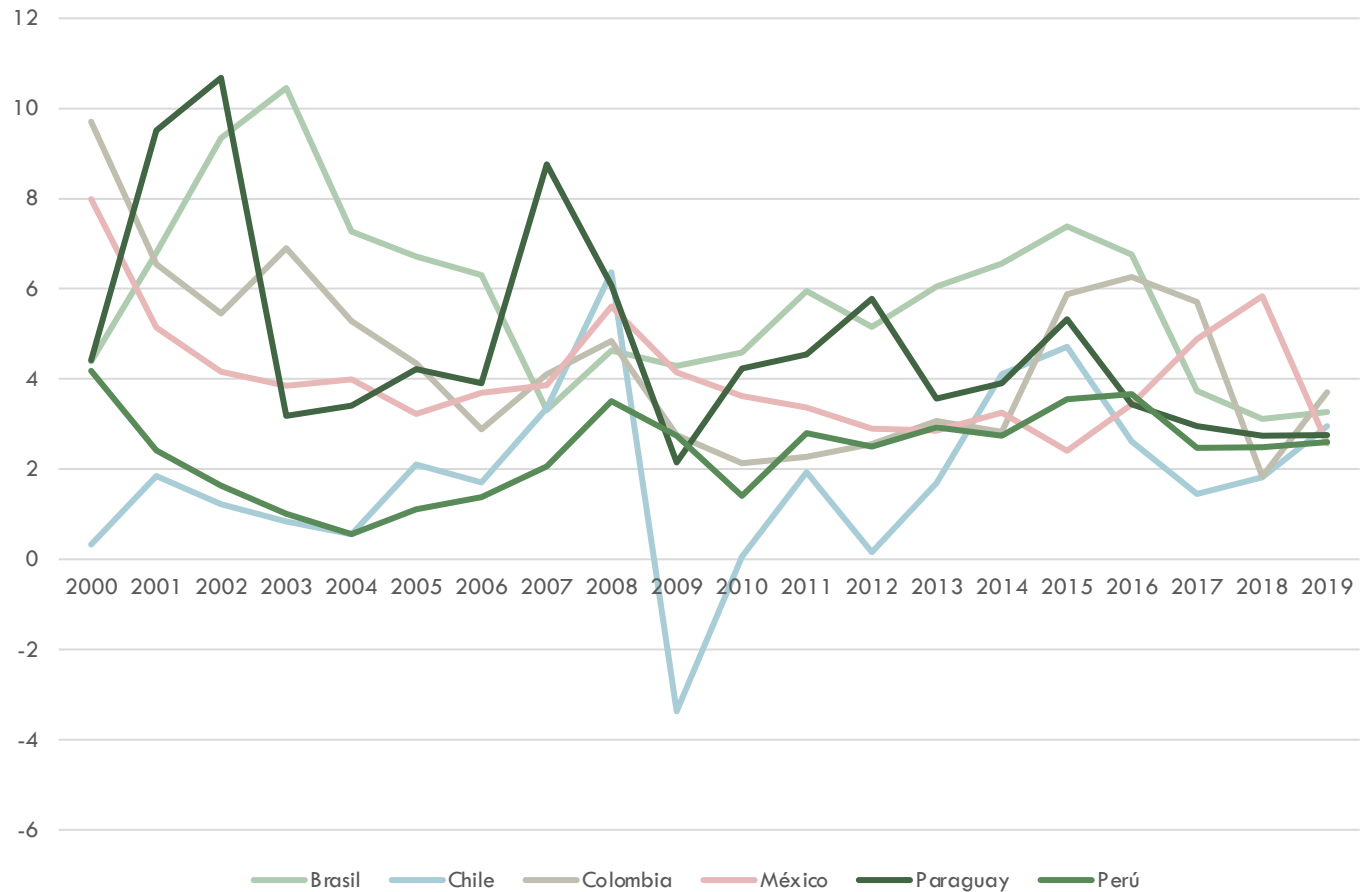
ANOS	BRASIL	CHILE	COLOMBIA	MÉXICO	PARAGUAY	PERÚ
2000	4,39	0,32	9,71	7,99	4,42	4,18
2001	6,80	1,85	6,55	5,15	9,51	2,41
2002	9,34	1,22	5,45	4,16	10,68	1,64
2003	10,46	0,84	6,90	3,85	3,18	1,00
2004	7,27	0,55	5,28	3,99	3,41	0,56
2005	6,71	2,10	4,35	3,22	4,22	1,11
2006	6,30	1,71	2,88	3,69	3,91	1,38
2007	3,30	3,35	4,10	3,87	8,75	2,07
2008	4,62	6,37	4,85	5,61	6,08	3,51
2009	4,28	-3,37	2,76	4,14	2,15	2,74
2010	4,59	0,05	2,13	3,62	4,23	1,41
2011	5,95	1,93	2,28	3,37	4,54	2,79
2012	5,16	0,15	2,55	2,90	5,78	2,49
2013	6,05	1,70	3,07	2,86	3,56	2,93
2014	6,56	4,10	2,83	3,26	3,91	2,74
2015	7,38	4,71	5,87	2,41	5,32	3,56
2016	6,76	2,62	6,26	3,44	3,44	3,67
2017	3,73	1,44	5,71	4,89	2,95	2,48
2018	3,11	1,81	1,84	5,83	2,74	2,49
2019	3,27	2,96	3,71	2,57	2,75	2,61

Fonte: CEPAL

Exemplo: Consideremos as taxas de variação anual dos índices de preços ao consumidor em seis países latino-americanos.
O que podemos observar, assim, de cara?

15

Taxas de variação anual de preços ao consumidor, %



Fonte: CEPAL

Porcentagem da renda nacional detida pelos 10% mais ricos e os 10% mais pobres da população

16

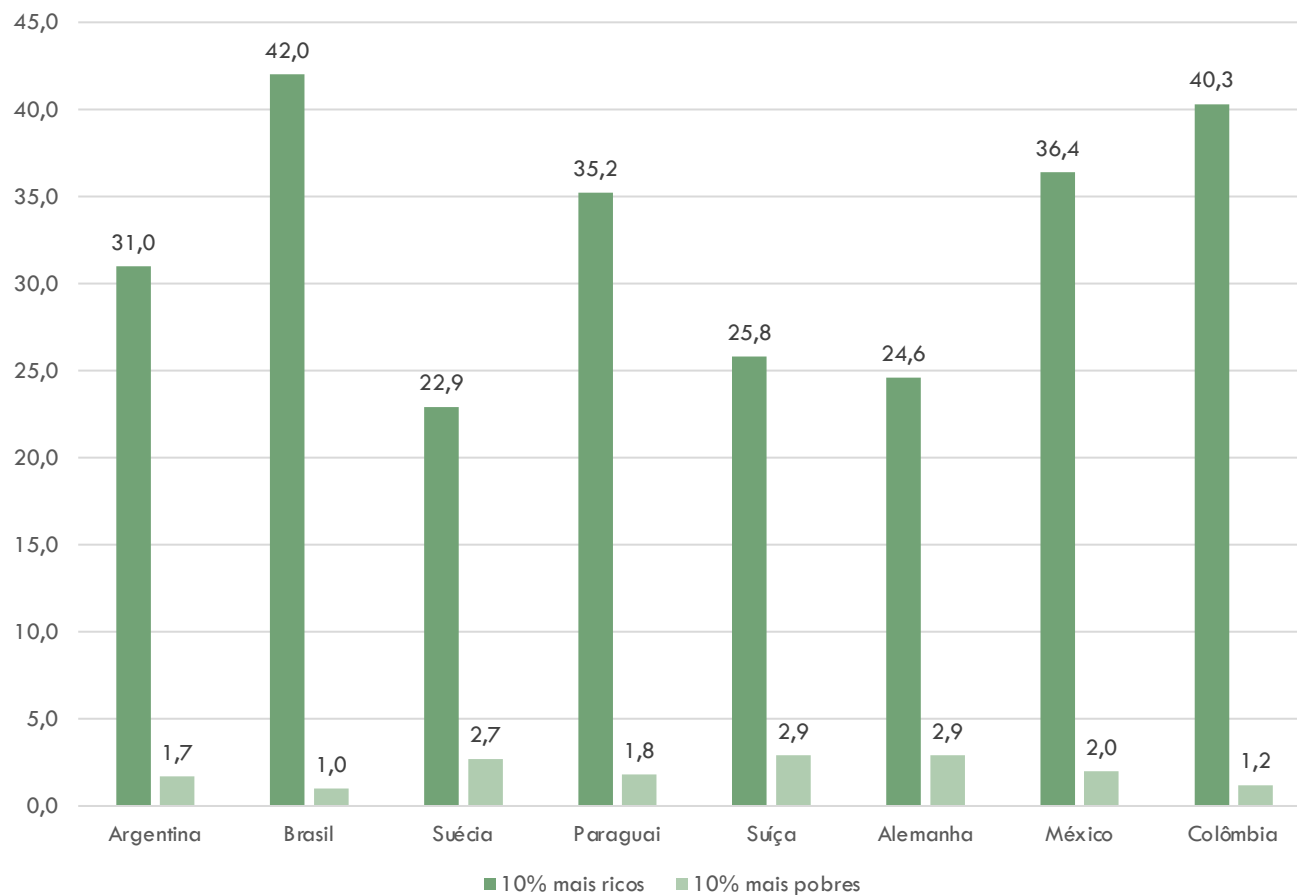
Fonte: Banco Mundial

	10% mais ricos	10% mais pobres
Argentina	31,0	1,7
Brasil	42,0	1,0
Suécia	22,9	2,7
Paraguai	35,2	1,8
Suíça	25,8	2,9
Alemanha	24,6	2,9
México	36,4	2,0
Colômbia	40,3	1,2

Interpretação? Visualização?

17

Porcentagem da renda nacional detida pelos 10% mais ricos e 10% mais pobres da população



Fonte: Banco Mundial

Dados, variáveis

18

- Dados são informações numéricas ou qualitativas
- As formas de coleta ou de obtenção dos dados são diversas e estes podem ser ajustados, se necessário (em relação a seu formato ou outros vários aspectos)
- Os dados podem provir de pesquisas, experimentos, etc.
- Definindo:
 - Unidade elementar: elemento de uma população (pessoa, objeto, coisa)
 - Dado: resultado de investigação, pesquisa ou cálculos
 - Variável: característica que pode assumir distintos valores por unidade elementar

Variáveis podem ser

Qualitativas

```
graph LR; Qualitativas --> Nominal; Qualitativas --> Ordinal;
```

Nominal

Não existe a possibilidade de ordenação

Ordinal

Pode existir uma ordem nos resultados

Quantitativas

```
graph LR; Quantitativas --> Discreta; Quantitativas --> Contínua;
```

Discreta

Valores possíveis formam um conjunto finito ou enumerável ou números

Contínua

Valores possíveis pertencem a um intervalo de número reais e resultam de mensuração

Número de variáveis

20

- 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. Ex: Exportações de petróleo e preço internacional do barril; Fluxo de divisas e taxa de câmbio
 - Três ou mais. Ex: Relação entre o saldo comercial e preços de commodities e taxa de câmbio; Estrutura da pauta de importações, origem, produtos, preços, etc.
- ↪ Nos casos de mais de uma variável, devem/podem ser usados métodos estatísticos além dos descritivos para identificar relações entre elas

Escala de medição dos dados

21

- **Escala nominal.** Usam-se valores numéricos apenas para dar nome ou classificar uma categoria. Essa escala serve para categorizar indivíduos. Os números, nesse caso, servem apenas para identificar características e perdem suas propriedades normais. Ex: atribuir números da determinados países (1=Alemanha; 2=Bélgica; 3=França; 4=Reino Unido, etc)
- **Escala ordinal.** Os valores dão nome e ordem a um objeto. O ordenamento não garante a observância de todas as propriedades dos números. Ex: Intensidade de exercício físico (1= 60% freq. cardíaca; 2= 80% freq. cardíaca; 3= 100% freq. cardíaca, etc. $\Rightarrow 3 > 1$ mas $1 + 2 \neq 3$)

Escala de medição dos dados

22

- **Escala de intervalos.** Valores em uma escala de intervalos iguais permitem ordenar as medições e explicar as diferenças entre as observações a partir de um zero arbitrário. Ex: entre o dia 25 e o dia 20 passaram 5 dias.
- **Escala proporcional (ou escala razão).** Valores nesta escala eliminam a limitação do intervalo, estabelecendo um zero da própria categoria, o zero absoluto. Os valores são comparados em relação uns aos outros. Ex: embalagem de arroz de 1kg e embalagem de 5 kg. Uma embalagem é cinco vezes maior (=mais pesada) que a outra

Tipos de variáveis

23

- Séries temporais. Observações são dados de uma variável em distintos pontos do tempo.
- Corte transversal numa data ou período. Não se considera a evolução no tempo de uma variável, mas observações pontuais que podem ou não ser comparadas com outras observações. Aqui podem ser consideradas amostras colhidas em determinado momento ou uma variável referente a distintas unidades elementares (Ex: cobertura de saneamento básico em vários países)

Gráficos

24

- Vimos que a representação gráfica de variáveis pode ser uma forma rápida e concisa de observar seu comportamento
- Existem inúmeras possibilidades para essa representação, de acordo com o tipo de variável
- O objetivo central é obter clareza e conseguir fazer os dados “falarem e serem bem entendidos” por meio de gráficos

Exemplo de representação gráfica

25

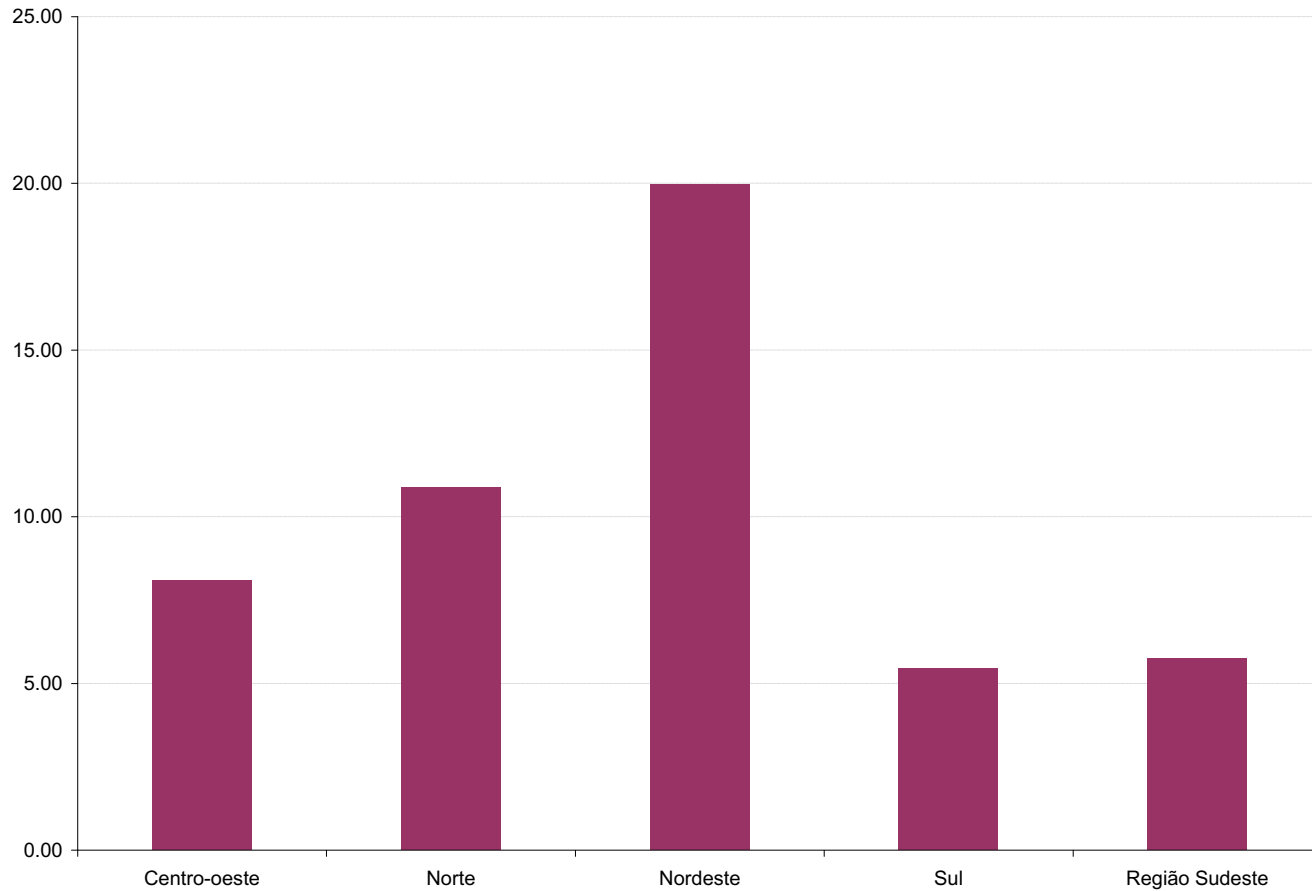
Distribuição regional do percentual de pessoas de 15 ou mais anos de idade que não sabem ler nem escrever um bilhete simples, Brasil, %

	2007
Centro-oeste	8.08
Norte	10.89
Nordeste	19.98
Sul	5.45
Região Sudeste	5.75

Fonte: IPEA

Uma forma de representar

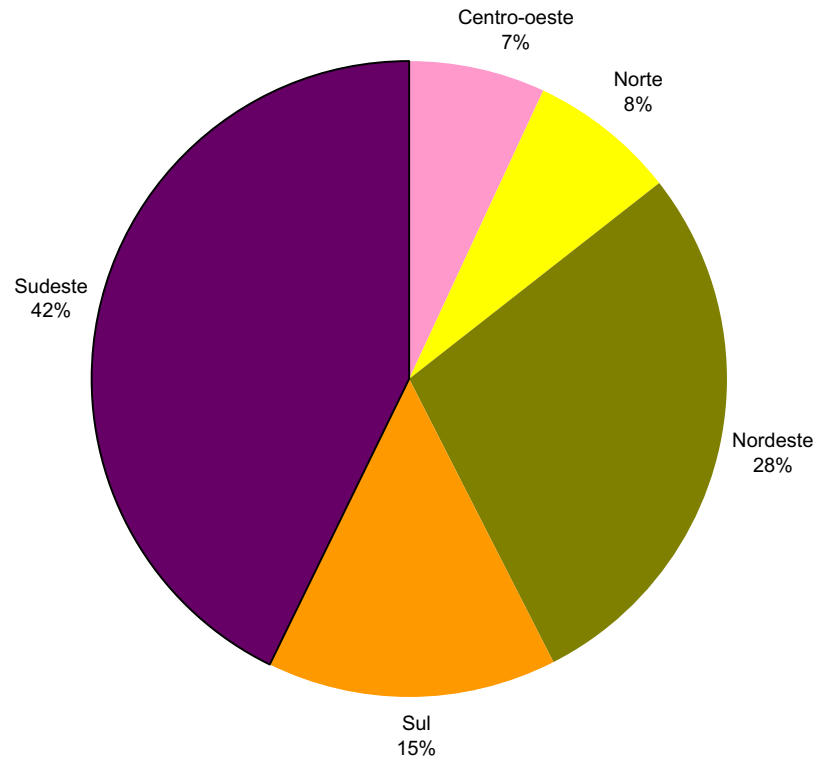
26



Fonte: IPEA

Distribuição regional da população brasileira

27



Fonte: IBGE, Censo 2000

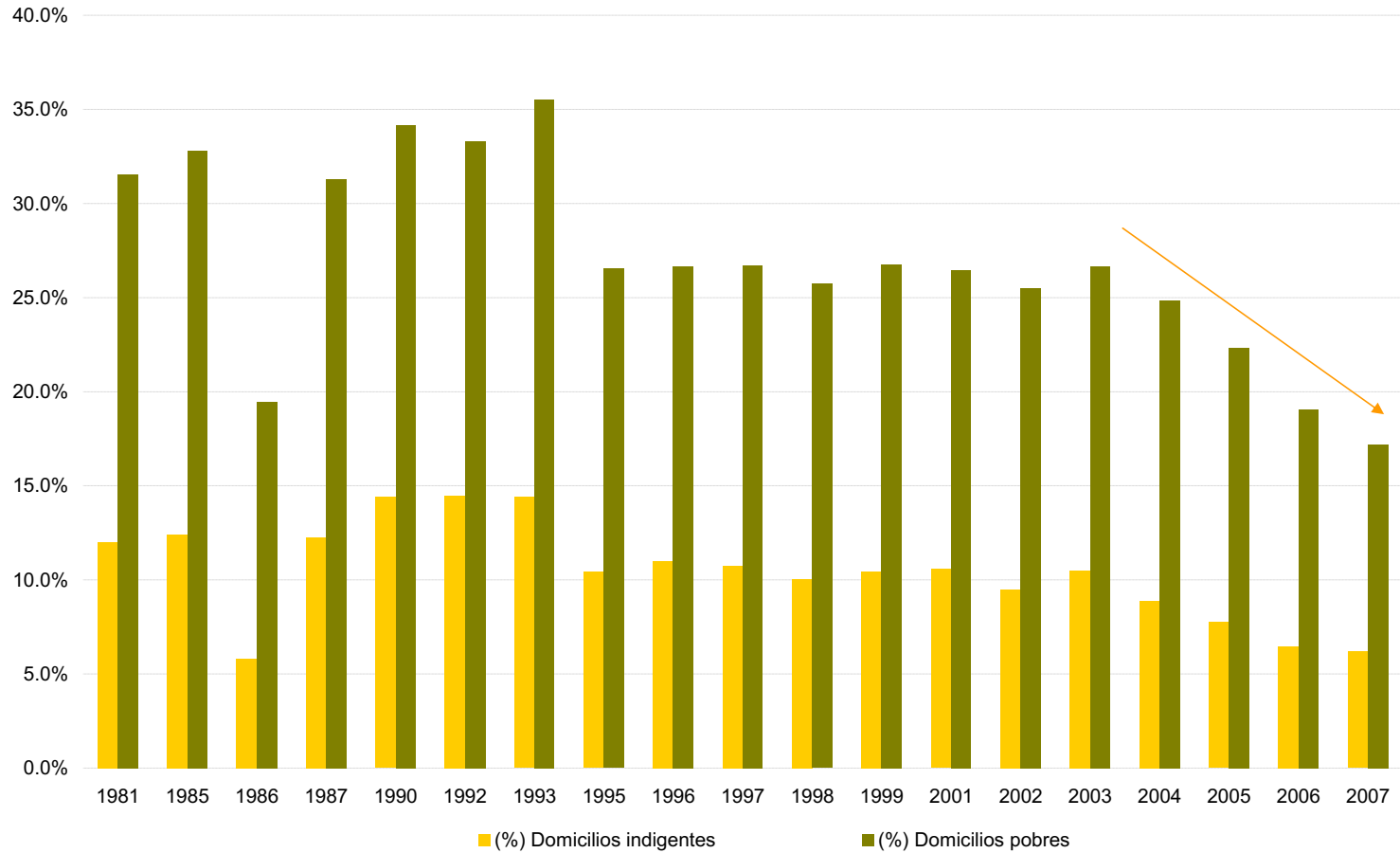
Outro exemplo...

- Evolução do percentual de domicílios pobres e indigentes no total dos domicílios brasileiros
- Aqui não estamos querendo representar um ponto fixo no tempo, mas uma evolução histórica
- Existem, também, inúmeras formas de fazê-lo

	(%) Domicílios indigentes	(%) Domicílios pobres
1981	12.0%	31.5%
1985	12.4%	32.8%
1986	5.8%	19.5%
1987	12.3%	31.3%
1990	14.4%	34.1%
1992	14.5%	33.3%
1993	14.4%	35.5%
1995	10.4%	26.5%
1996	11.0%	26.7%
1997	10.8%	26.7%
1998	10.0%	25.7%
1999	10.4%	26.8%
2001	10.6%	26.4%
2002	9.5%	25.5%
2003	10.5%	26.7%
2004	8.9%	24.8%
2005	7.8%	22.3%
2006	6.5%	19.1%
2007	6.2%	17.2%

Uma possibilidade seria:

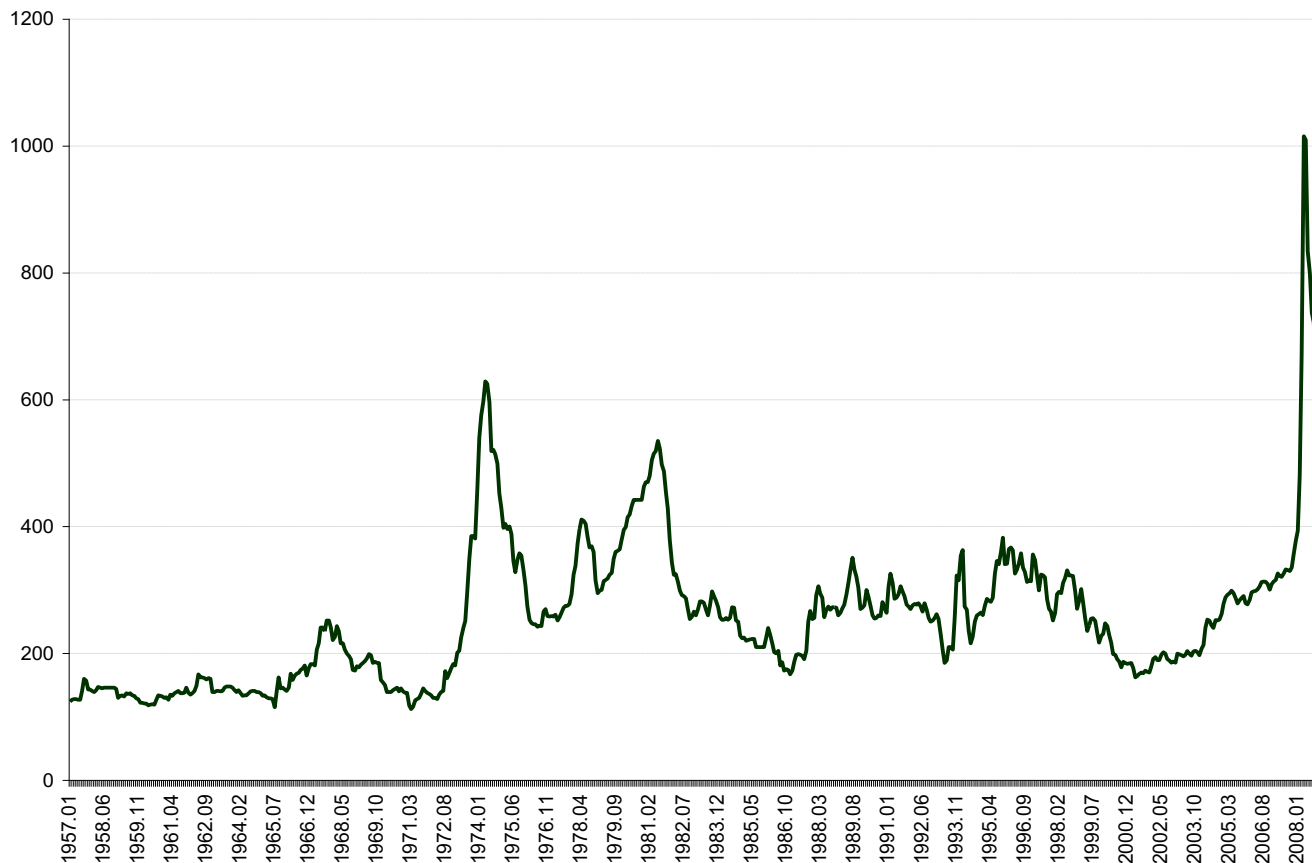
29



Fonte: PNAD-IBGE

Se quisermos estudar/mostrar uma série temporal longa

□ Cotação internacional do arroz (Bangkok), em US\$, 1957-2008



Fonte: Fundo Monetário Internacional, International Financial Statistics (FMI/IFS) , coletado em IPEADATA

Obs: Preço do arroz em dólar americano (US\$) por tonelada métrica

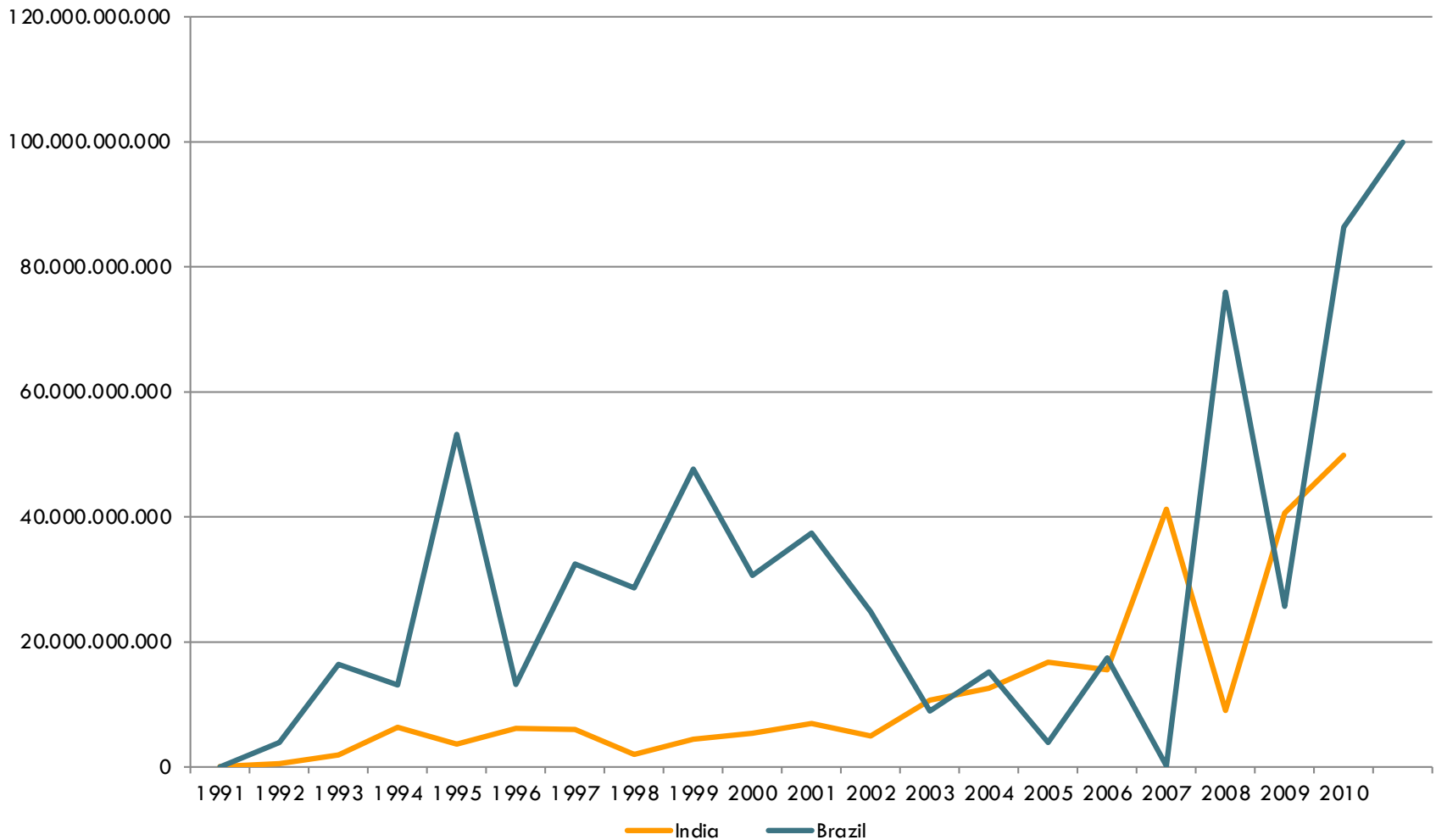
Outro aspecto se refere à unidade que escolhemos para apresentar um dado

31

- 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

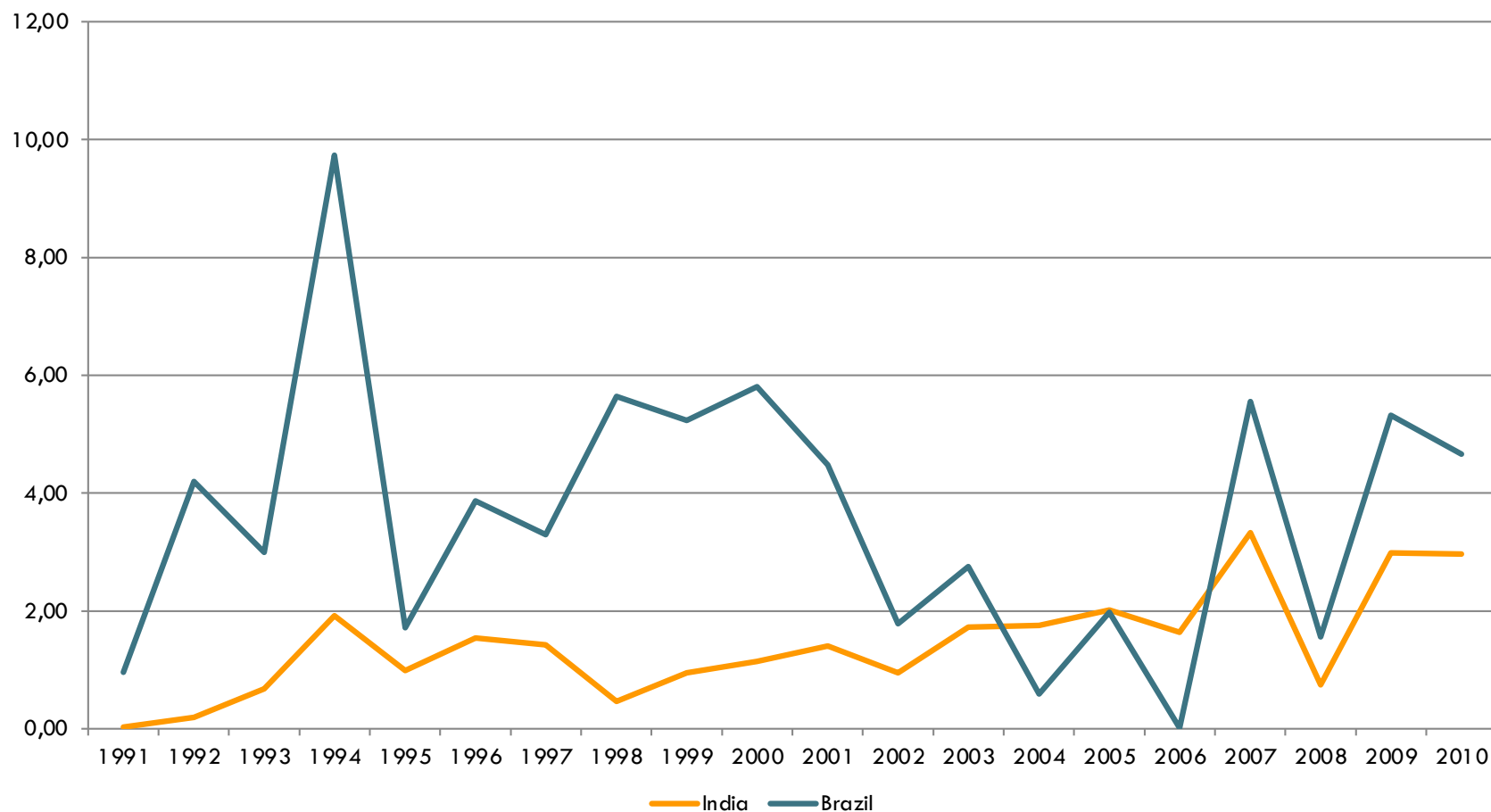
32



Fonte: Banco Mundial

Fluxos de capitais privados, em % do PIB

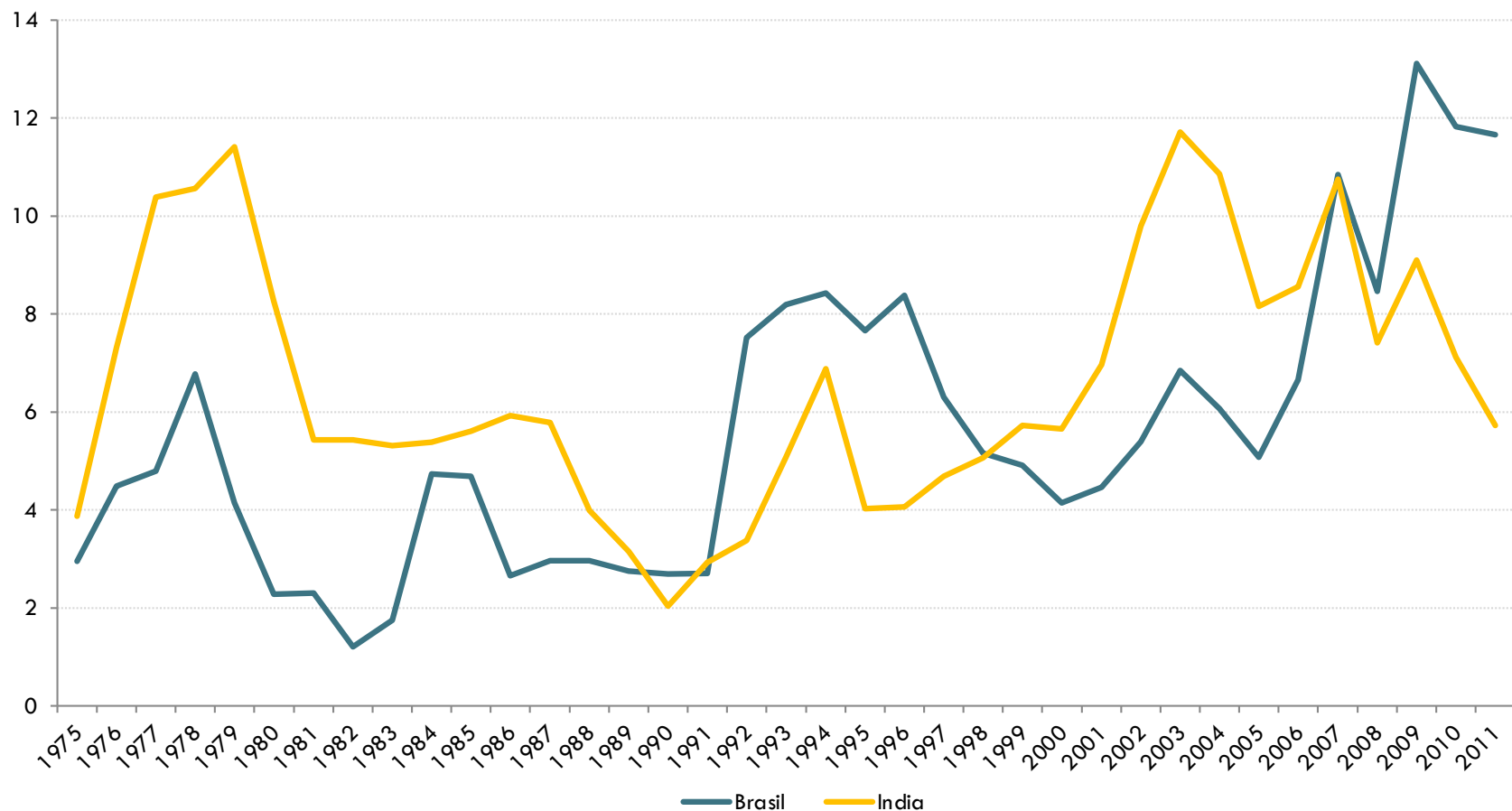
33



Fonte: Banco Mundial

Reservas internacionais em meses de importações

34



Fonte: Banco Mundial

Distribuições de frequências

35

- 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

África, países selecionados

Participação no PIB (%) de ingressos líquidos de IDE

36

	1999
Angola	28.92
Benin	1.31
Burkina Faso	0.39
Cameroon	0.44
Central African Republic	1.23
Chad	0.98
Comoros	0.52
Congo, Rep.	0.23
Cote d'Ivoire	3.12
Ethiopia	1.40
Gambia, The	3.56
Ghana	0.22
Guinea	1.82
Guinea-Bissau	1.37
Kenya	0.13
Lesotho	18.69
Madagascar	1.56
Malawi	3.31
Mali	0.74
Mauritania	0.21
Mozambique	9.65
Niger	0.74
Nigeria	2.87
Rwanda	0.09
Senegal	1.26
Sierra Leone	0.15
Sudan	3.82
Tanzania	2.09
Togo	2.14
Uganda	3.46
Zambia	5.17
Zimbabwe	1.05

IDE: Investimento direto estrangeiro

Fonte: FMI

Distribuições de frequências

- A tabela anterior apresenta uma amostra de 32 países
- Estabelecendo intervalos com o para % dos fluxos de ingresso líquido de IDE, teríamos a seguinte distribuição

	Frequência n_i	Proporção f_i	Porcentagem $100f_i$
Entre 0 e 1%	12	0.3750	37.50
Entre 1,01 e 2%	8	0.2500	25.00
Entre 2,01 e 5%	8	0.2500	25.00
Entre 5,01 e 15%	2	0.0625	6.25
Acima de 15,01%	2	0.0625	6.25
Total	32	1.0000	100.00

- Construimos aqui uma tabela de frequência para uma variável contínua, que é a participação do IDE no PIB.
- Para isso, estabelecemos intervalos aleatórios e observamos quantos países se situavam em cada intervalo