

# INTRODUÇÃO À ECONOMETRIA

## Análise de Regressão

### Aula 1

Escola Nacional de Administração Pública

# OBJETIVO DO CURSO

**O Aluno será capaz de fazer:**

- Regressões lineares;
- Análise de resultados de modelos simples;
- Projeção de dados;
- Usar o *software* R-studio;
- Introdução à Econometria de séries de tempo e painel.

# INTRODUÇÃO

- ▶ O que é econometria?
- ▶ Metodologia da econometria
- ▶ Regressão versus correlação
- ▶ Natureza dos dados
- ▶ Cenários econômicos
- ▶ Modelos macroeconômicos em instituições públicas
- ▶ Modelos estatísticos em instituições privadas
- ▶ Análise de Regressão

# O QUE É ECONOMETRIA?

- ▶ Econometria significa “**medida econômica**”. No entanto, o escopo da econometria é bem mais amplo:
- ▶ “Econometria pode ser definida como a ciência social em que os instrumentos da teoria econômica, matemática e inferência estatística são aplicados a **análise de fenômenos econômicos**” (Goldberger 1964).

# O QUE É ECONOMETRIA?

- ▶ “Econometria baseia-se no **desenvolvimento de métodos estatísticos** para estimar relações econômicas, testar hipóteses econômicas e avaliar programas governamentais e corporativos” (Wooldridge, 2007).

# METODOLOGIA DA ECONOMETRIA

## **1. Formulação da Teoria ou da Hipótese**

- ▶ Keynes postulou que em média, os consumidores aumentam seu consumo quando sua renda cresce, mas não tanto quanto o aumento na renda ( $PMC < 1$ ).

# METODOLOGIA DA ECONOMETRIA

## 2. Especificação do Modelo Matemático

$$Y = \beta_0 + \beta_1 X$$

$Y$  = consumo (variável dependente)

$X$  = renda (variável independente ou explicativa)

$\beta_0$  = coeficiente de intercepto

$\beta_1$  = coeficiente de declividade

- ▶ O coeficiente de declividade  $\beta_1$  mede a propensão marginal a consumir (PMC).

# METODOLOGIA DA ECONOMETRIA

## 3. Especificação do Modelo Econométrico

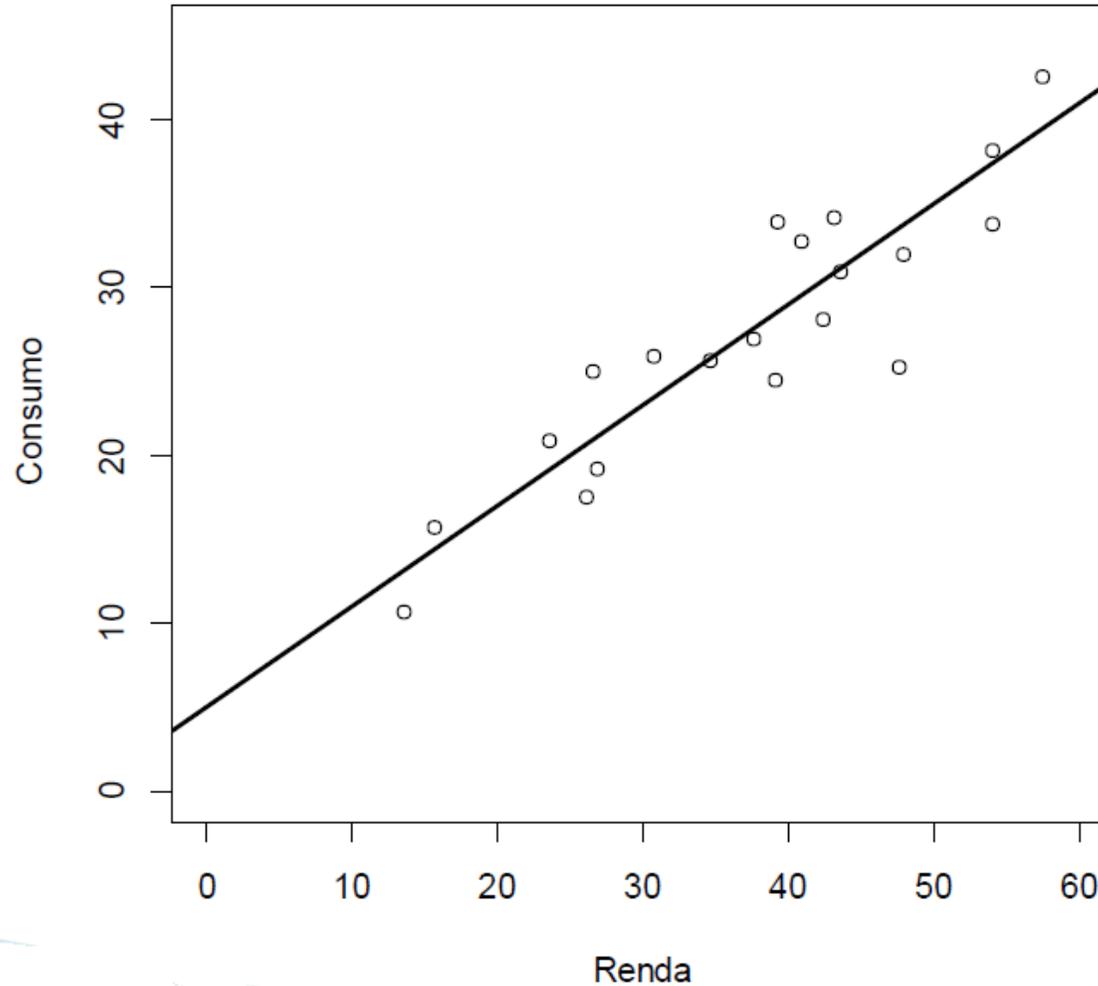
- ▶ O modelo anterior supõe uma relação exata entre consumo e renda. Além da renda, outras variáveis afetam o consumo.

$$Y = \beta_0 + \beta_1 X + u$$

- ▶ onde  $u$ , conhecido como **termo de perturbação ou erro**, é uma variável aleatória (estocástica) que representa todos os fatores que afetam o consumo, mas não são considerados explicitamente.

# METODOLOGIA DA ECONOMETRIA

## Representação Gráfica



# METODOLOGIA DA ECONOMETRIA

## 4. Obtenção de Dados

- ▶ Para obter os valores numéricos de  $\beta_0$  e  $\beta_1$ , precisamos de dados.

Ano	Y	X
1982	3081.5	4620.3
1983	3240.6	4803.7
1984	3407.6	5140.1
1985	3566.5	5323.5
1986	3708.7	5487.7
1987	3822.3	5649.5
1988	3972.7	5865.2
1989	4064.6	6062
1990	4132.2	6136.3
1991	4105.8	6079.4
1992	4219.8	6244.4
1993	4343.6	6389.6
1994	4486	6610.7
1995	4595.3	6742.1
1996	4714.1	6928.4

# METODOLOGIA DA ECONOMETRIA

## 5. Estimativa do Modelo Econométrico

- ▶ **Análise de Regressão** é a principal técnica estatística utilizada para obter as estimativas dos parâmetros do modelo.

$$\hat{Y} = -184,80 + 0,7064X$$

- ▶ O coeficiente de inclinação estimado foi de aproximadamente 0.70, sugerindo que um aumento na renda real de R\$ 1 leva, em média, a um aumento na despesa de consumo de R\$ 0,70.

# METODOLOGIA DA ECONOMETRIA

## 6. Teste de Hipótese

- ▶ A propensão marginal a consumir é positiva, mas menor do que 1?
- ▶ Em outras palavras, 0.70, o valor estimado de  $\beta_1$ , é estatisticamente menor do que 1?
- ▶ Essa confirmação ou rejeição da teoria econômica com base na evidência empírica se baseia em um ramo da teoria estatística conhecido como **inferência estatística** (teste de hipótese).

# METODOLOGIA DA ECONOMETRIA

## 7. Previsão

- ▶ Para ilustrar, suponha a expectativa de um PIB real de US\$ 7269.8 bilhões em 1997. Qual a previsão de consumo em 1997?

$$\hat{Y}_{1997} = -184,0779 + 0,7064(7269.8)$$

$$\hat{Y}_{1997} = 4951,3167$$

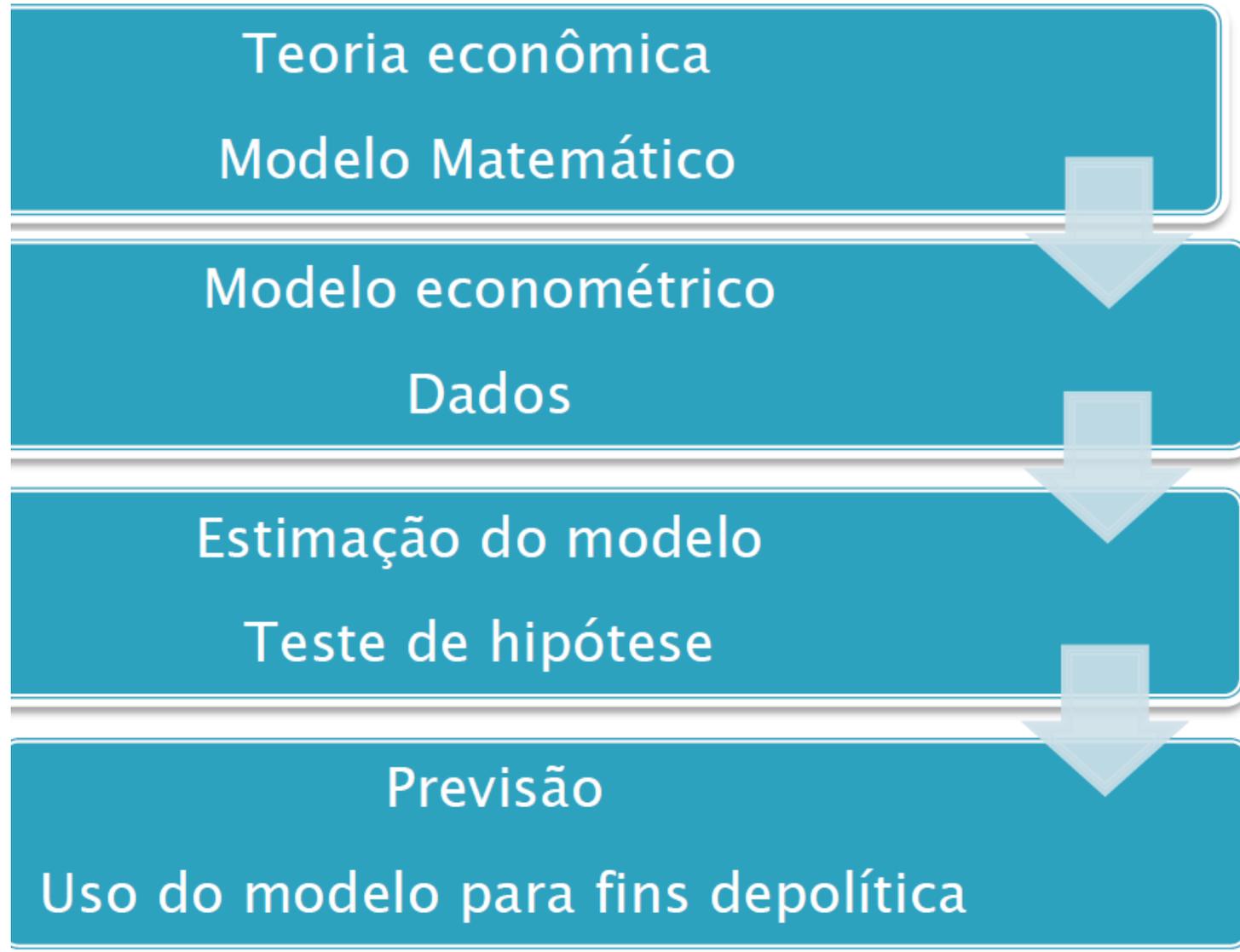
- ▶ O valor observado da despesa de consumo em 1997 foi US\$ 4913.5 bilhões. Assim, o erro de previsão foi de US\$ 37.82 bilhões, que é 0,76 por cento do PIB.

# METODOLOGIA DA ECONOMETRIA

## 8. Uso do Modelo para Fins de Política

- ▶ Considere que o governo considere a equação estimada para fins de política.
- ▶ Qual o impacto de uma elevação de 1 dólar nos gastos públicos sobre o nível total de renda economia?
  - ▶  $M = 1 / 1 - PMC = 3.33$

# METODOLOGIA DA ECONOMETRIA



# METODOLOGIA DA ECONOMETRIA

- ▶ Ponto de partida
  - Formulação da questão de interesse.
  
- ▶ Os modelos econométricos podem ser usados para:
  - Analisar aspectos da teoria econômica
  - Testar os efeitos de uma política governmental
  - Auxiliar nas decisões empresariais

# METODOLOGIA DA ECONOMETRIA

- ▶ Modelos econômicos formais são, em alguns casos, o ponto inicial da análise empírica.
- ▶ É comum, contudo, o uso da teoria econômica **menos formalmente** ou mesmo o estudo do objeto de pesquisa baseado na própria **intuição** do pesquisador.

# METODOLOGIA DA ECONOMETRIA

- ▶ Em sua maioria, a análise econométrica começa pela especificação de um modelo econométrico, onde a **racionalidade econômica** e o **bom senso** serão determinantes na escolha das variáveis.

# METODOLOGIA DA ECONOMETRIA

## Exemplo

- ▶ Questão de interesse:
  - Avaliar o efeito de um treinamento sobre a produtividade do trabalhador.
  
- ▶ Modelo econômico:

$$prod = f(educ, tempo, treina)$$

- *educ* é o nível de escolaridade;
- *tempo* é a experiência do trabalhador;
- *treina* é o tempo gasto no treinamento.

# METODOLOGIA DA ECONOMETRIA

- ▶ Modelo econométrico:

$$prod = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 treina + \varepsilon,$$

- ▶ onde  $\varepsilon$ , conhecido como **termo de erro**, é uma variável aleatória (estocástica) que representa todos os fatores que afetam a produtividade, mas não são considerados explicitamente.

# METODOLOGIA DA ECONOMETRIA

- ▶ Uma vez estimados, os modelos econométricos podem ser utilizados para:
  - ▶ Testar hipóteses de interesse sobre os parâmetros
  - ▶ Realizar projeções
  - ▶ Testar teorias econômicas
  - ▶ Estudar o impacto de políticas

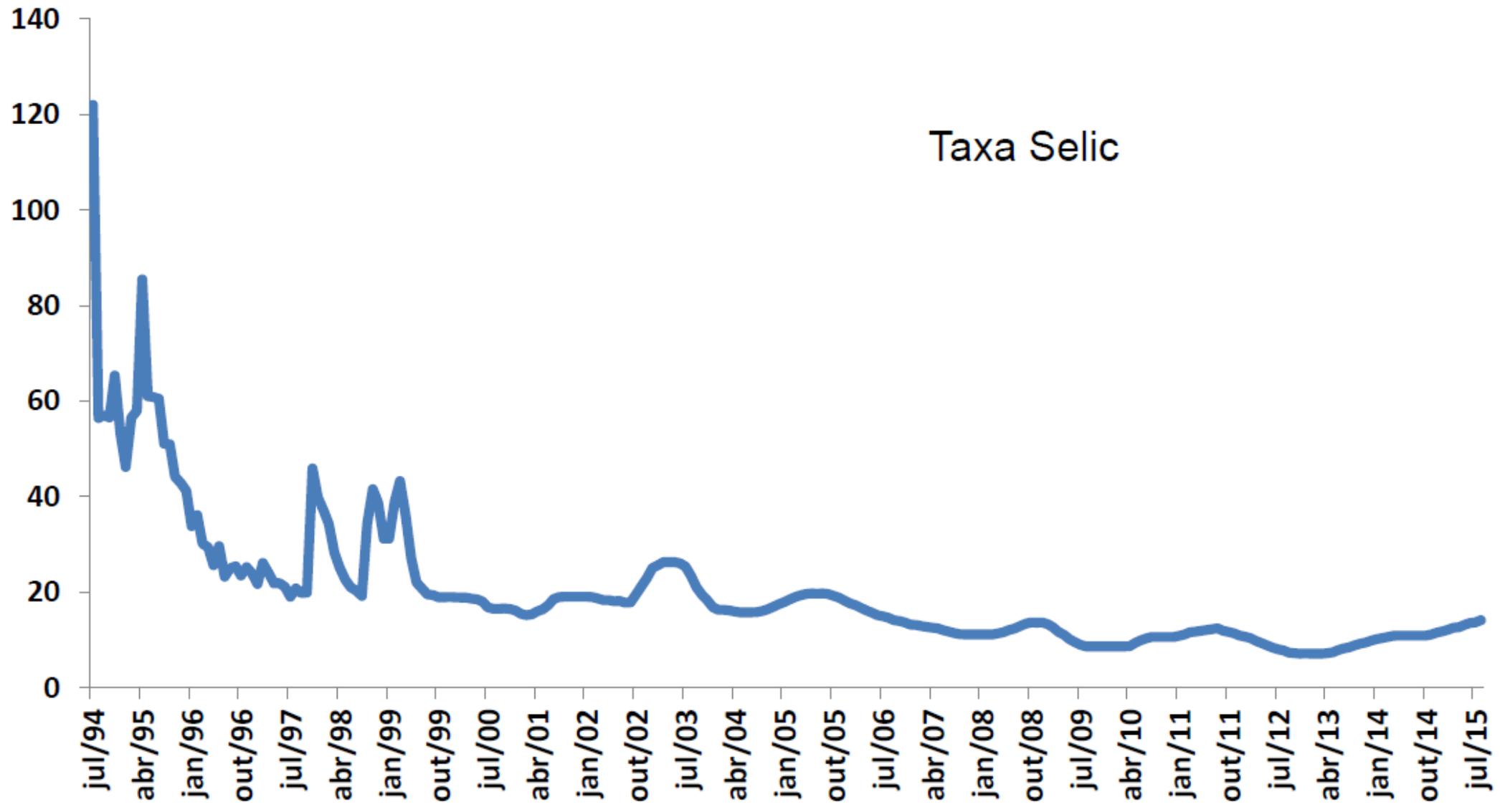
# ESTRUTURA DOS DADOS E METODOLOGIA

- ▶ **Dados em corte:** consiste em uma amostra de indivíduos, famílias, firmas, países, etc. em um determinado período de tempo.
  - Ex: dados individuais sobre salários e escolaridade dos funcionários do MPDG e do MF no mês de setembro de 2017.
- ▶ **Séries de tempo:** consiste de observações sobre uma ou várias variáveis ao longo do tempo.
  - Ex: dados da PIM, PMC e IPCA ao longo dos últimos 60 meses.

# Dados de Corte (*Cross-Section*)

Instituições	Ativo total	Depósito total	Patrimônio líquido	Lucro líquido	Nº de agências
BB	1.324.464.414	<b>469.680.089</b>	70.675.464	5.833.139	5.524
CAIXA ECONOMICA FEDERAL	1.064.674.796	<b>419.338.756</b>	26.222.766	3.702.318	3.391
ITAU	1.117.848.197	<b>310.831.489</b>	103.079.063	10.311.204	3.885
BRADESCO	883.438.773	<b>212.507.911</b>	81.588.015	7.880.304	4.665
SANTANDER	598.224.807	<b>143.887.187</b>	58.224.926	1.218.969	2.640
HSBC	167.971.558	<b>57.727.489</b>	9.732.307	-532.785	854
BANRISUL	58.933.772	<b>34.302.822</b>	5.670.031	463.545	528
BTG PACTUAL	154.593.402	<b>21.280.905</b>	14.678.867	1.561.384	9
BNDES	871.410.178	<b>16.808.363</b>	30.737.494	3.122.228	0
BANCOOB	21.889.831	<b>16.569.448</b>	786.874	30.787	3
CITIBANK	60.860.588	<b>14.278.844</b>	6.706.737	17.268	127
BCO DO NORDESTE DO BRAS	38.204.992	<b>11.925.427</b>	3.367.809	421.338	292
BCO COOPERATIVO SICREDI	28.929.333	<b>11.731.452</b>	1.062.123	52.893	5
PAN	25.722.546	<b>11.576.354</b>	3.786.214	106.107	2
SAFRA	140.675.131	<b>9.781.404</b>	8.733.555	830.183	108
BRB	12.229.693	<b>8.513.539</b>	1.165.305	45.662	115
BANESTES	14.992.118	<b>8.386.872</b>	1.061.694	65.921	134
MERCANTIL DO BRASIL	13.261.618	<b>8.355.766</b>	709.360	-75.129	189

# Dados de Séries de Tempo



# ESTRUTURA DOS DADOS E METODOLOGIA

- ▶ **Dados em painel:** corresponde a dados com séries de tempo para cada unidade da amostra.
  - Ex: receita financeira e despesa de pessoal, em frequência mensal, dos 50 maiores bancos brasileiros ao longo da última década.

# Dados em Painel

Ano	País	PIB	TC
2005	EUA	3,35	-745,78
2005	Argentina	9,2	5,28
2005	México	3,08	-5,1
2006	EUA	2,67	-800,62
2006	Argentina	8,5	7,77
2006	México	4,98	-4,5
2007	EUA	1,79	-710,3
2007	Argentina	8,7	7,35
2007	México	3,22	-8,87
2008	EUA	-0,29	-677,14
2008	Argentina	7	6,76
2008	México	1,38	-16,35
2009	EUA	-2,8	-376,55
2009	Argentina	0,7	8,41
2009	México	-4,74	-6,37
2010	EUA	2,51	-470,9
2010	Argentina	9,2	3,08
2010	México	5,2	-5,64

Fonte: FMI

PIB: Var(%) ao ano

TC: Transações correntes  
(US\$ bilhões)

# Tipos de Variáveis

- ▶ Quantitativas: podem ser medidas numericamente,
  - Exs: altura e renda.
- ▶ Qualitativas: indicam a presença de uma qualidade ou atributo. Não podem ser medidas numericamente.
  - Exs: homem ou mulher, católico ou não católico.

# Tipos de Variáveis

- Quantitativas
  - Discretas: assumem apenas valores inteiros
    - Número de filhos
  - Contínuas: podem assumir infinitos valores.
    - Renda

# Tipos de Variáveis

- Qualitativas:
  - Nominal: não existe ordenação entre as categorias.
    - Brasileiro ou não brasileiro
  - Ordinal: existe ordenação entre as categorias
    - Ensino fundamental, ensino médio ou nível superior

# ESTRUTURA DOS DADOS E METODOLOGIA

## ▶ **Microeconometria**

- A abordagem de dados em corte e em painel é utilizada frequentemente para análise de unidades econômicas individuais, tais como consumidores, firmas, famílias, etc.
- Em geral, usada para avaliação de políticas e teste de hipóteses microeconômicas nas áreas de organização industrial, economia da demografia, finanças públicas e economia do trabalho.

# ESTRUTURA DOS DADOS E METODOLOGIA

## ▶ **Macroeconometria**

- A econometria de séries de tempo geralmente é utilizada para análise, projeção e construção de cenários para variáveis macroeconômicas, tais como PIB, taxa de câmbio, inflação, taxa de juros, etc.
- A maioria dos trabalhos aplicados em series temporais supõe que essas séries são **estacionárias**.
- Ao trabalhar com dados de séries temporais, é necessário examinar sua estacionariedade.

# MODELOS MACROECONOMÉTRICOS EM INSTITUIÇÕES PÚBLICAS E INTERNACIONAIS

- ▶ Indicadores antecedentes
- ▶ Modelos Arima e modelos VAR
- ▶ Modelos semiestruturais pequenos
- ▶ Modelos semiestruturais de médio porte
- ▶ Modelos DSGE
- ▶ Modelos de longo prazo

# MODELOS MACROECONOMÉTRICOS

## CENÁRIOS ECONÔMICOS

- ▶ “As previsões são o resultado de diversos modelos, trajetórias exógenas e hipóteses, refletindo a combinação de diferentes análises econométricas, do julgamento dos membros do Copom e, até certo ponto, dos próprios membros do grupo de projeções...” (Dez Anos de Metas para Inflação no Brasil 1999-2009, Banco Central do Brasil).

# MODELOS MACROECONOMÉTRICOS

## CENÁRIOS ECONÔMICOS

- ▶ As previsões do trimestre corrente (*nowcasting*) são fundamentadas, basicamente, na avaliação de indicadores econômicos atuais, em expectativas do mercado e em resultados de modelos autorregressivos” (Dez Anos de Metas para Inflação no Brasil 1999-2009, Banco Central do Brasil).

# MODELOS MICROECONOMÉTRICOS

## APLICAÇÕES

- ▶ Estimação de funções de produção e de custo
- ▶ Funções de demanda
- ▶ Retorno da renda à educação
- ▶ Avaliação de eficiência

# Análise de Regressão

## Introdução

# Análise de Regressão

## Origem Histórica

- ▶ O termo *regressão* foi introduzido por Francis Galton.
- ▶ Galton verificou que, embora houvesse uma tendência de pais altos terem filhos altos e de pais baixos terem filhos baixos, a altura média dos filhos de pais de uma dada altura tende a se descolar ou “*regredir*” para a altura média da população.

# Análise de Regressão

## Interpretação Atual

- ▶ *A Análise de regressão* ocupa-se do estudo da dependência de uma variável, a *variável dependente*, em relação a uma ou mais variáveis, as *variáveis explicativas ou independentes*.

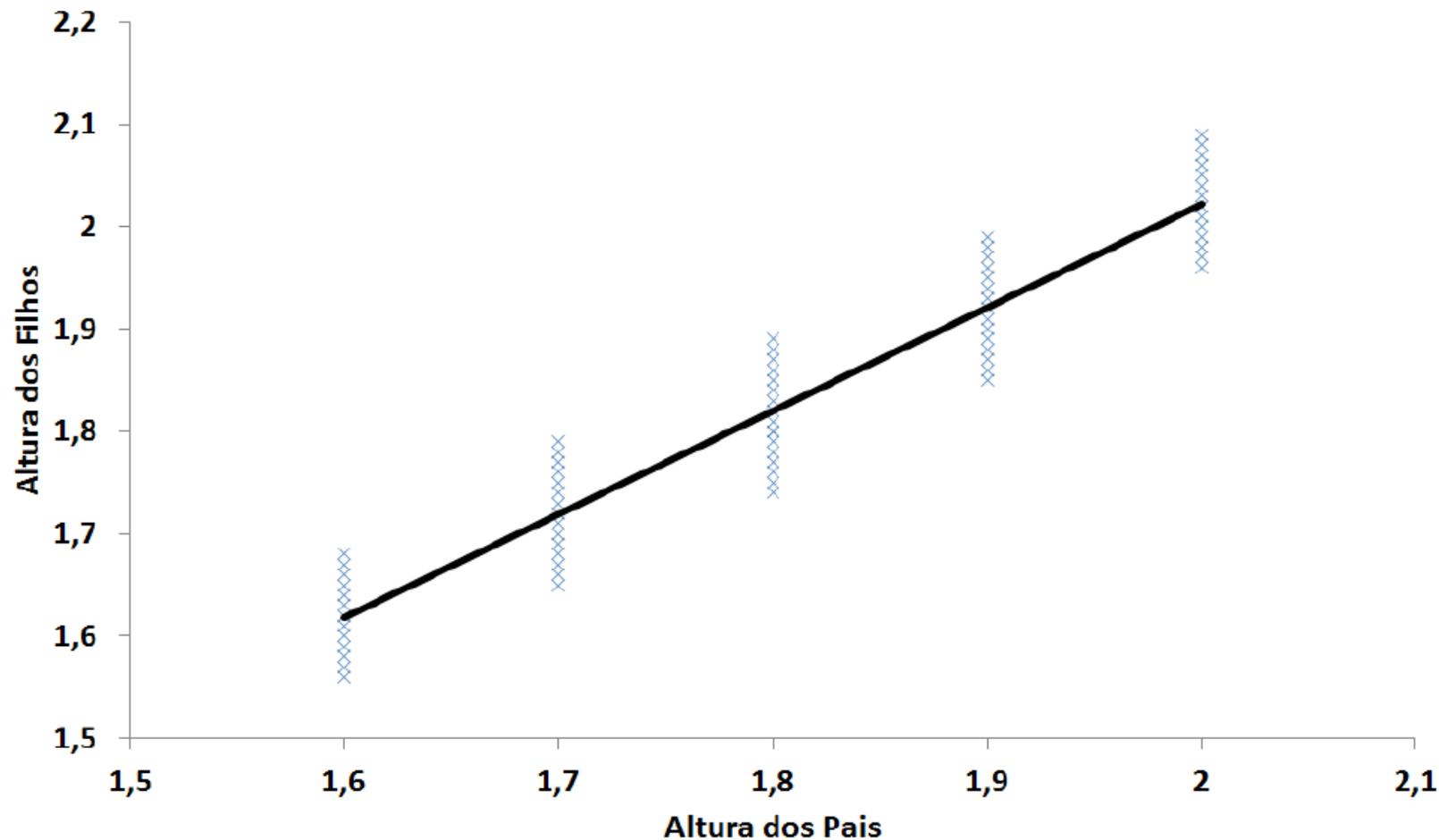
# Análise de Regressão – Exemplo

## Altura dos filhos vs. Altura dos pais

- ▶ Considerando o estudo de Galton, nosso maior interesse seria descobrir como varia a altura média dos filhos, dada a altura dos pais.

# Análise de Regressão – Exemplo

## Altura dos filhos vs. Altura dos pais



# Análise de Regressão

## Regressão versus Correlação

- ▶ Na análise de *correlação*, o principal objetivo é medir a intensidade ou o *grau de associação linear* entre duas variáveis.
- ▶ Na análise de *regressão*, nós estamos interessados em estimar ou *prever o valor médio* de uma variável com base nos valores das variáveis explicativas.

# Analise de Regressão

## Regressão Simples vs. Regressão Múltipla

- ▶ *Regressão simples*: quando o problema tem por objetivo prever ou estudar o comportamento da variável dependente a partir de uma única variável independente:
- ▶ *Regressão múltipla*: quando o objetivo é estudar o comportamento da variável dependente a partir da dinâmica de mais de uma variável dependente.

# Análise de Regressão

## Regressão Simples

- ▶ Muitas aplicações econométricas começam com a seguinte premissa:
- ▶ *y e x são duas variáveis*, que representam alguma população, e estamos interessados em “*explicar y em termos de x,*” ou em “*estudar como y varia com mudanças em x.*”
- ▶ O modelo de regressão linear simples pode ser usado para estudar a relação entre duas variáveis.

# Regressão Simples

## Exemplo Hipotético

- ▶ Imagine um país com uma população total de 60 famílias. Estamos interessados em estudar a relação entre a despesa de consumo familiar semanal ( $Y$ ) e a renda familiar semanal disponível ( $X$ ).
- ▶ Mais especificamente, estamos interessados em prever o nível médio de consumo ( $Y$ ), dada a renda da família ( $X$ ).

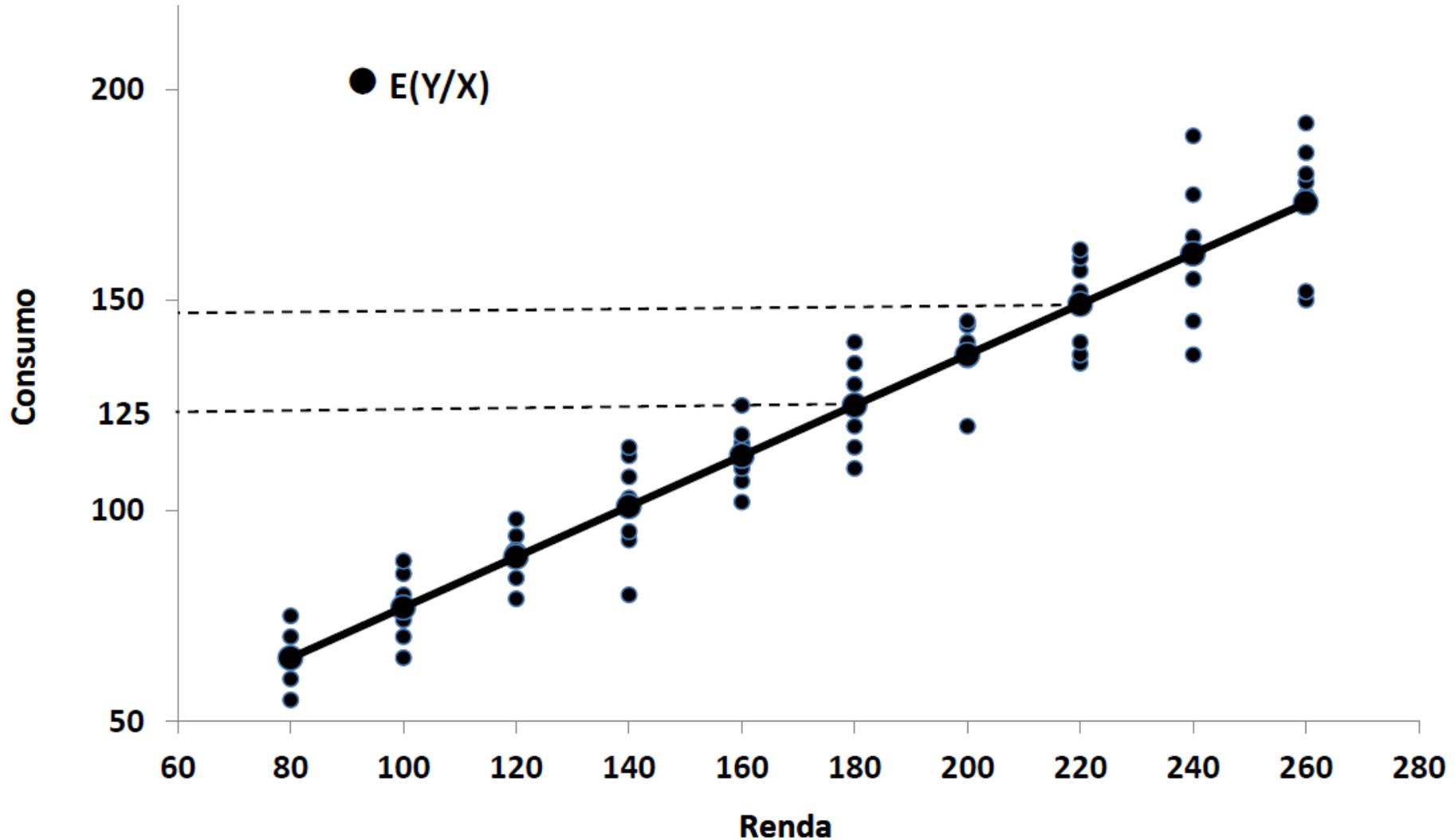
# Regressão Simples

## Exemplo Hipotético – Dados

Renda (X)	→	80	100	120	140	160	180	200	220	240	260
Despesa de consumo (Y)		55	65	79	80	102	110	120	135	137	150
↓		60	70	84	93	107	115	136	137	145	152
		65	74	90	95	110	120	140	140	155	175
		70	80	94	103	116	130	144	152	165	178
		75	85	98	108	118	135	145	157	175	180
			88		113	125	140		160	189	185
					115				162		192
Média - $E(Y/X)$		65	77	89	101	113	125	137	149	161	173

# Regressão Simples

## Exemplo Hipotético: Consumo vs. Renda



# Regressão Simples

## Função de Regressão da População

- ▶ Os pontos ampliados na figura anterior mostram a média condicional de  $Y$  (*despesa de consumo*) correspondentes aos vários valores de  $X$  (renda).
- ▶ Se nós ligarmos os valores médios condicionais, nós obtemos a *curva ou função de regressão da população (FRP)*. De forma mais simples, a regressão de  $Y$  sobre  $X$ .

# Regressão Simples

## Função de Regressão da População

- ▶ A **função de regressão da população**, então, é o lugar geométrico das médias ou expectativas condicionais das variáveis dependentes, para os valores fixados da variável ou variáveis explicativas.

# Regressão Simples

## Função de Regressão da População

- ▶ Pelo que examinamos anteriormente, fica claro a média condicional  $E(Y/X_i)$  é uma função de  $X_i$ .

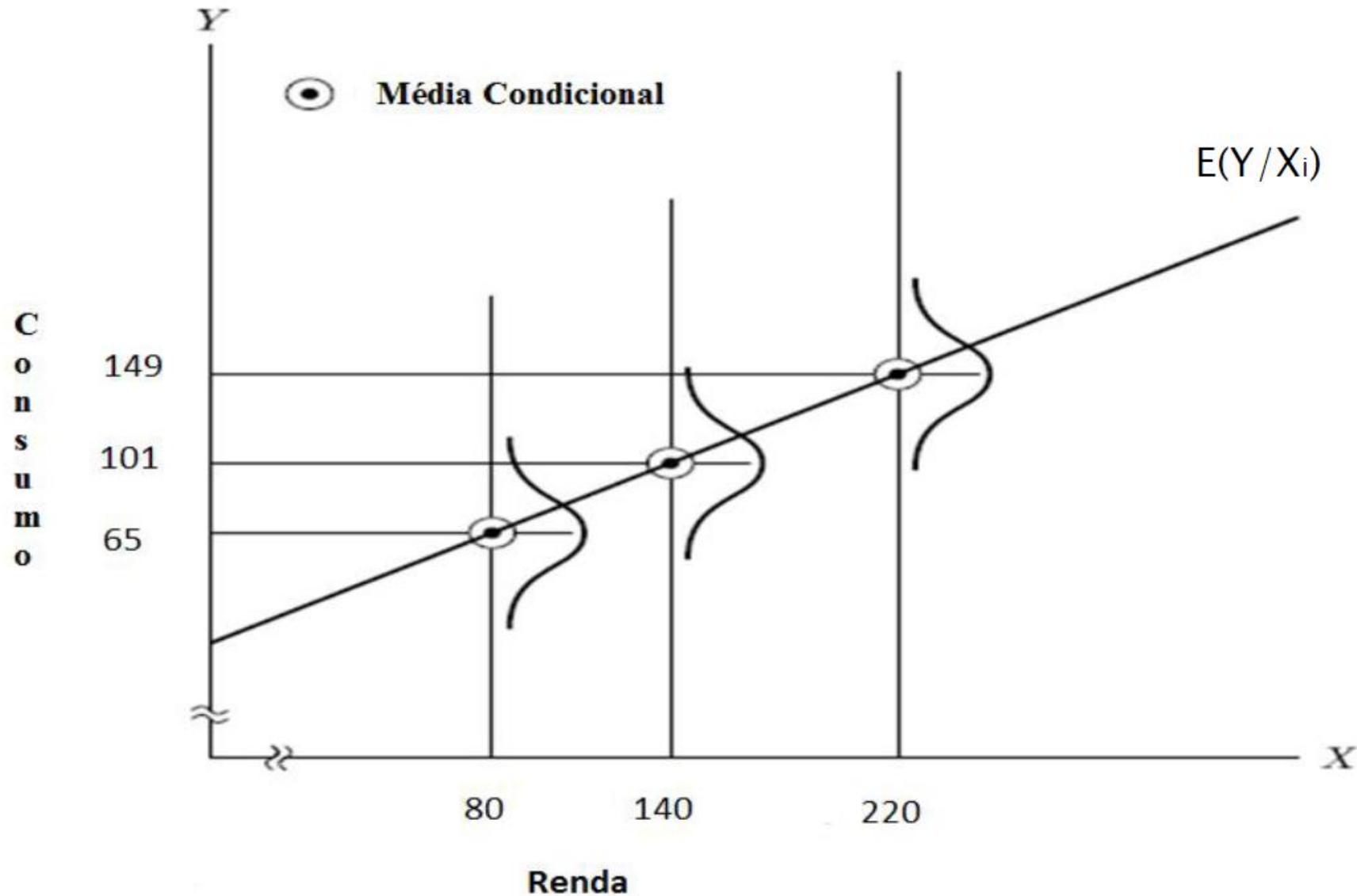
$$E(Y / X_i) = f(X_i)$$

- ▶ Considerando que  $f(X)$  é linear, obtemos a **função de regressão linear da população**:

$$E(Y / X_i) = \beta_0 + \beta_1 X_i$$

# Regressão Simples

## Função de Regressão da População



# Análise de Regressão

## Modelo de Regressão Linear Simples

- ▶ Qual a relação entre o valor do consumo **individual** ( $Y_i$ ) e um dado nível de renda ( $X_i$ )?
- ▶ A figura anterior indica que o consumo individual se situa em torno do consumo médio:

$$Y_i = E(Y / X_i) + u_i$$

- ▶ onde  $u_i$  é um variável aleatória não observável que pode assumir valores positivos e negativos.

# Análise de Regressão

## Modelo de Regressão Linear Simples

- ▶ Supondo que  $E(Y/X)$  é linear, temos:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- ▶ Tecnicamente,  $u_i$  é conhecido como **termo de erro estocástico** ou **perturbação estocástica**.
- ▶ A equação acima define o que é conhecido como **modelo de regressão linear simples**.

# Análise de Regressão

## Modelo de Regressão Linear Simples

$y$	$x$
Variável dependente	Variável independente
Variável explicada	Variável explicativa
Variável resposta	Variável controle
Variável prevista	Variável preditora
Regressando	Regressor

# Modelo de Regressão Linear Simples

## Exemplo

- ▶ Estudo do efeito de um fertilizante sobre a produção de soja:

$$produção = \beta_0 + \beta_1 fertilizante + u,$$

- ▶ Onde o termo  $u$  capta os efeitos de outros fatores como a qualidade da terra e as chuvas.
- ▶ O coeficiente  $\beta_1$  mensura o efeito do fertilizante, mantendo os outros fatores fixos:  $\Delta produção = \beta_1 \Delta fertilizante$ .

# Modelo de Regressão Linear Simples

## Exemplo

- ▶ Um modelo relacionando salário à escolaridade:

$$\text{Salário} = \beta_0 + \beta_1 \text{educ} + u$$

- ▶ O coeficiente  $\beta_1$  mensura o retorno à educação: a variação no salário, dado outro ano de escolaridade.
- ▶ O termo  $u$  é capta os efeitos de outros fatores como experiência, habilidade, etc.

# Análise de Regressão

## O Termo de Erro Aleatório

- ▶ A **perturbação estocástica** é um substituto para todas as variáveis omitidas do modelo.
- ▶ Porque não incluir o maior número de variáveis possíveis?
  - Imprecisão da teoria;
  - Indisponibilidade dos dados;
  - Variáveis periféricas;
  - Erros de medidas;
  - Princípio da parcimônia;

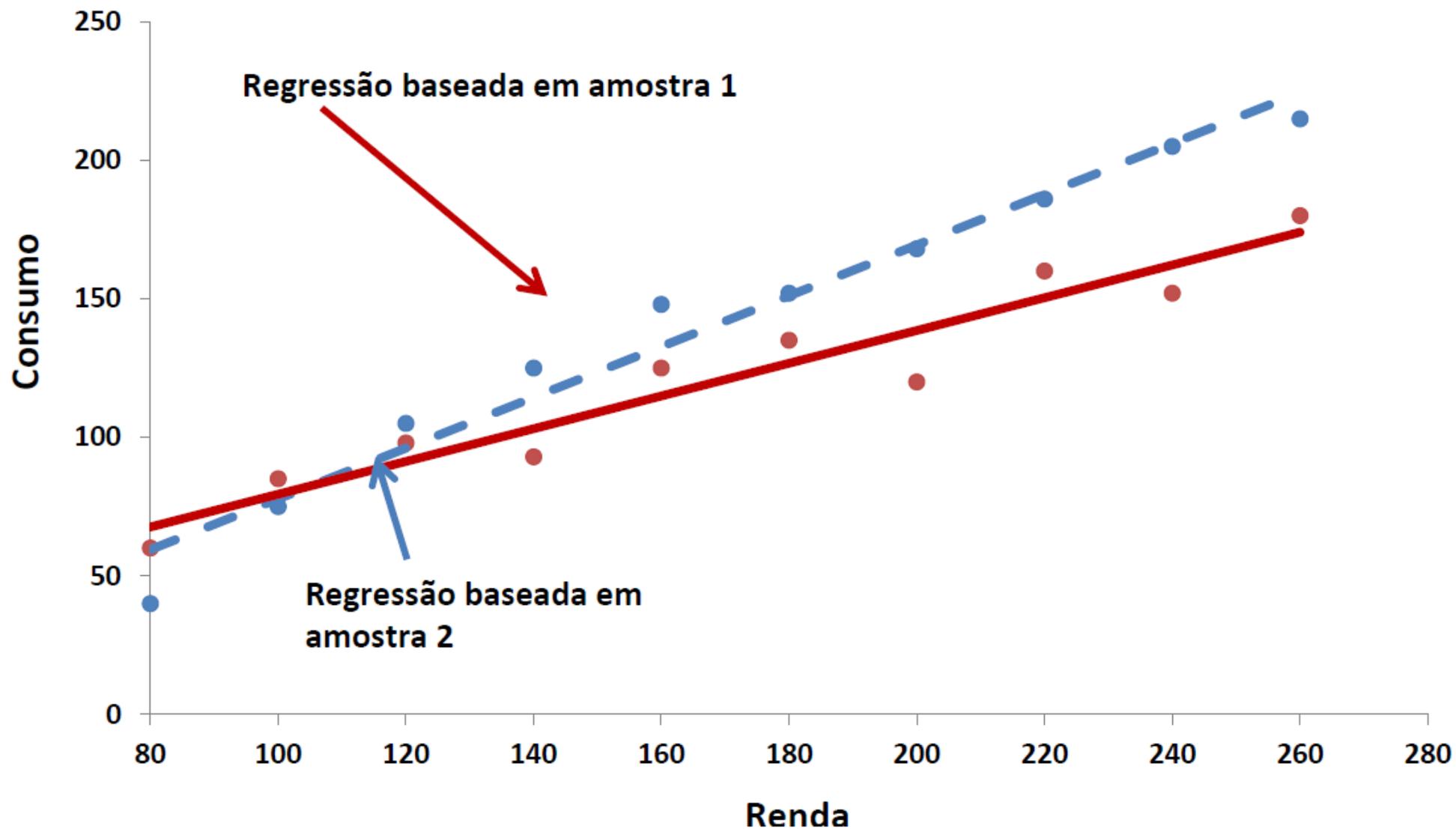
# Regressão Linear Simples

## Função de Regressão Amostral

- ▶ Na prática, a informação que temos é uma amostra de valores  $Y$  correspondentes a alguns  $X$ s fixos.
- ▶ Assim, analogamente à FRP, podemos definir o conceito de **função de regressão amostral** (FRA) para representar a reta de regressão da amostra.

# Regressão Linear Simples

## Função de Regressão Amostral



# Regressão Linear Simples

## Função de Regressão Amostral

- ▶ A contrapartida amostral da FRP é:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,$$

- ▶ onde
- ▶  $\hat{Y}_i$  é o estimador de  $E(Y/X)$
- ▶  $\hat{\beta}_0$  é o estimador de  $\beta_0$
- ▶  $\hat{\beta}_1$  é o estimador de  $\beta_1$

# Regressão Linear Simples

## Função de Regressão Amostral

- ▶ A função de regressão amostral também pode ser expressa em termos estocásticos:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i$$

- ▶ onde  $\hat{u}_i$  representa o termo de resíduo. Conceitualmente  $\hat{u}_i$  pode ser considerado uma estimativa de  $u_i$ , sendo introduzido na FRA pela mesma razão que foi adicionado na FRP.

# Regressão Linear Simples

## Função de Regressão Amostral

- ▶ A questão central é, então, como poderemos criar uma regra ou um método que fará com que a FRA seja tão próxima quanto possível da FRP?
- ▶ Em outras palavras, como estimar  $\hat{\beta}_0$  e  $\hat{\beta}_1$  de modo que eles sejam tão próximos quanto possível dos verdadeiros  $\beta_0$  e  $\beta_1$ , mesmo que nunca venhamos a conhecer seus verdadeiros valores?

# Regressão Linear Simples

## Função de Regressão Amostral

- ▶ A questão central é, então, como poderemos criar uma regra ou um método que fará com que a FRA seja tão próxima quanto possível da FRP?
- ▶ Em outras palavras, como estimar  $\hat{\beta}_0$  e  $\hat{\beta}_1$  de modo que eles sejam tão próximos quanto possível dos verdadeiros  $\beta_0$  e  $\beta_1$ , mesmo que nunca venhamos a conhecer seus verdadeiros valores?