

# Análise de Modelos de Multinível

Ciclo 2016-I de Seminários de Pesquisa  
Economics and Politics Research Group

Ana Collares ([anacollares@unb.br](mailto:anacollares@unb.br))

# Exemplos de populações organizadas hierarquicamente:

- Alunos dentro de salas de aula e salas de aula dentro de escolas;
- Pacientes/pessoal de saúde dentro de hospitais;
- Animais dentro de rebanhos dentro de fazendas;
- Fábricas dentro de cidades e cidades dentro de estados;
- Tratamentos “dentro” de indivíduos em diferentes ocasiões;
- Notas “dentro” de alunos.
- Dados de painel, com “ondas do survey” agrupadas por unidade de análise

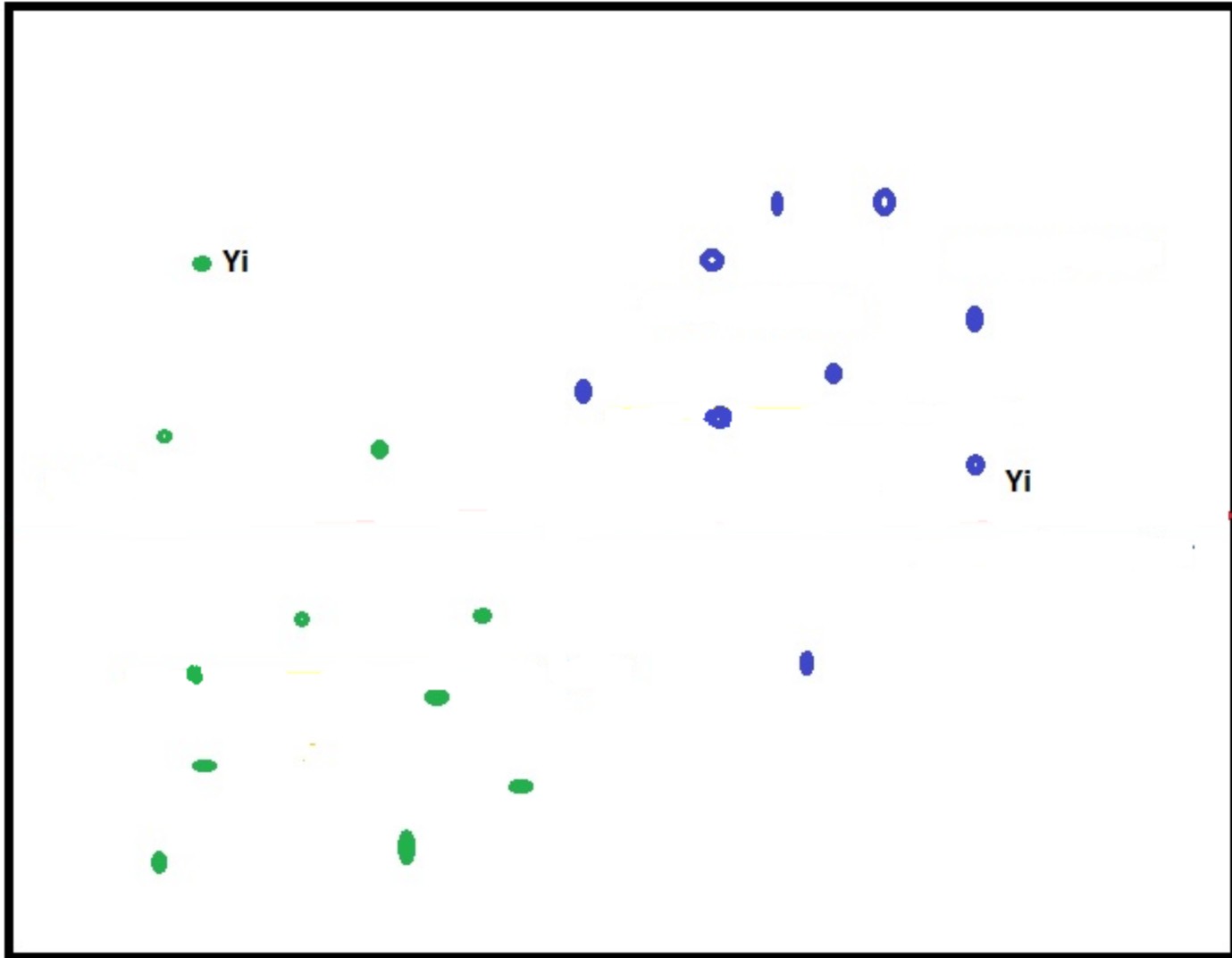
Vantagens de usar os modelos de multinível ao invés de modelos de regressão tais como o de mínimos quadrados dentro desses contextos:

- Estimar corretamente o erro padrão dos coeficientes de regressão (caso contrário estes poderiam ser subestimados por ignorar a relação de dependência dos casos pertencentes ao mesmo agrupamento);
- Analisar mais eficientemente a influência dos efeitos contextuais;
- Realizar interações cruzando níveis diferentes de análise e usando efeitos fixos ou aleatórios;
- Etc.

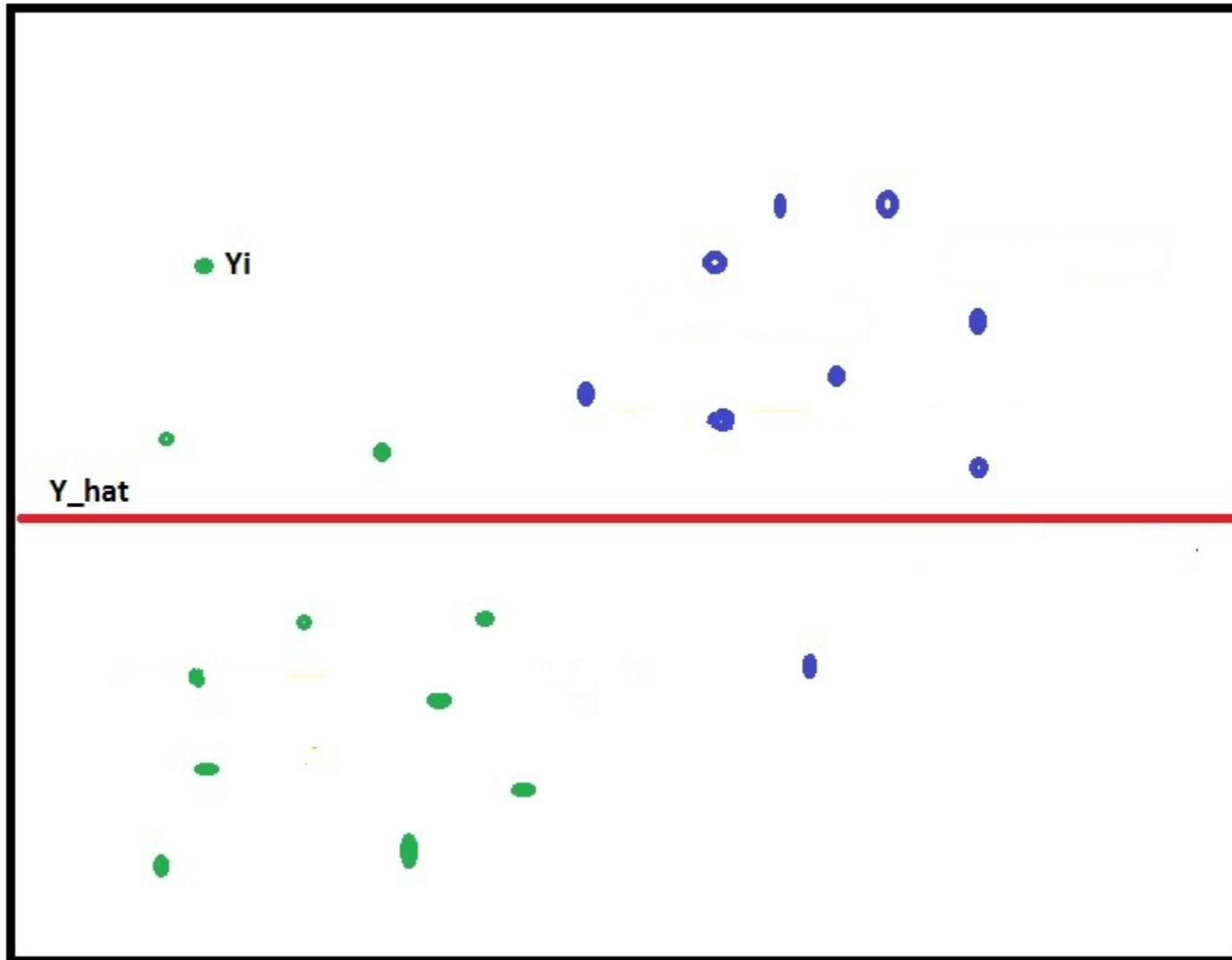
# Outras nomenclaturas na literatura

- Modelos de Multinível ou *Multilevel linear models* (pesquisa sociológica);
- Modelos de efeitos mistos, ou *Mixed-effects models* (biometria);
- Modelos de coeficiente aleatório ou *Random-coefficient regression models* (econometria);
- *Hierarchical linear models* (Raudenbush and Bryk)

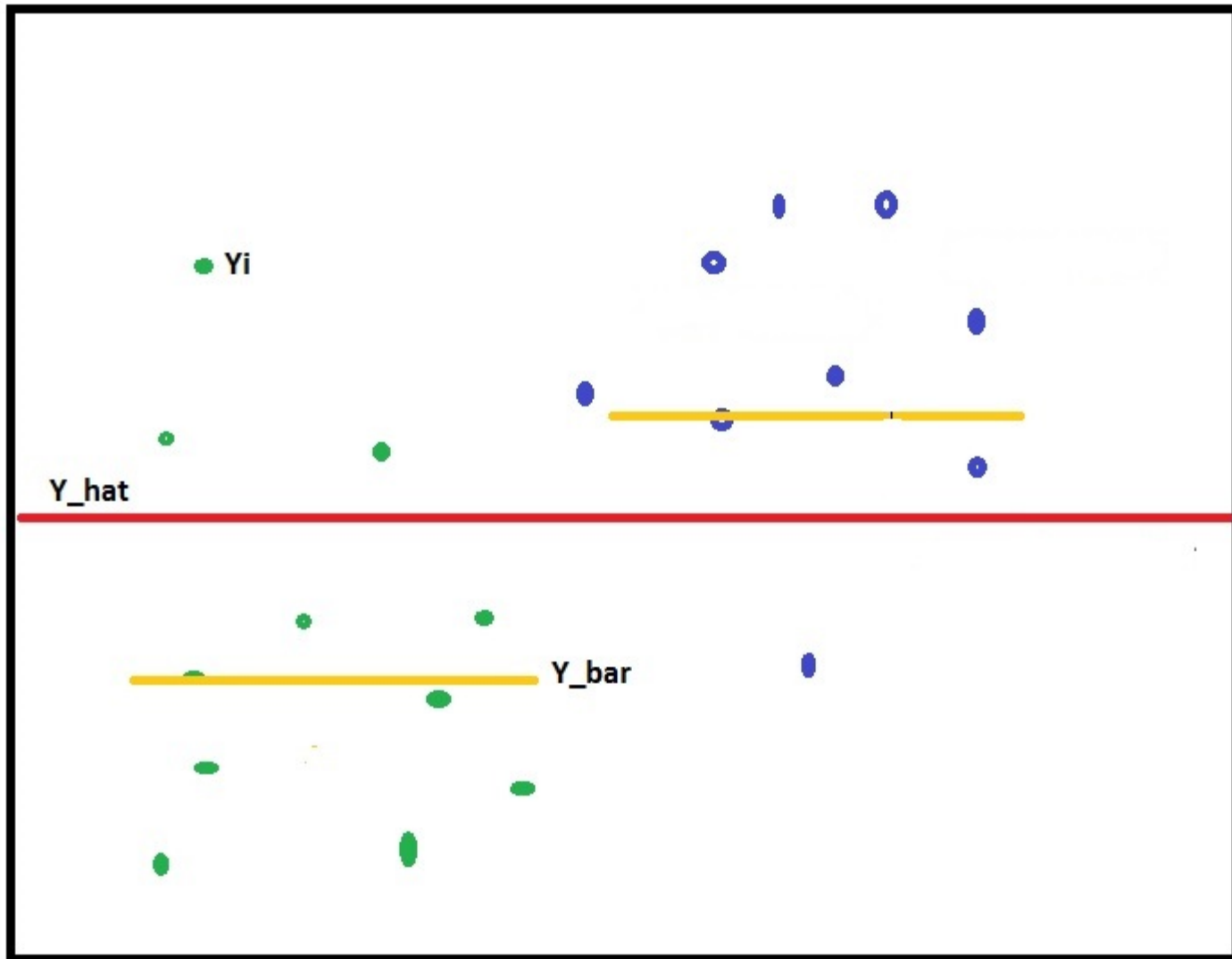
# Noções Preliminares



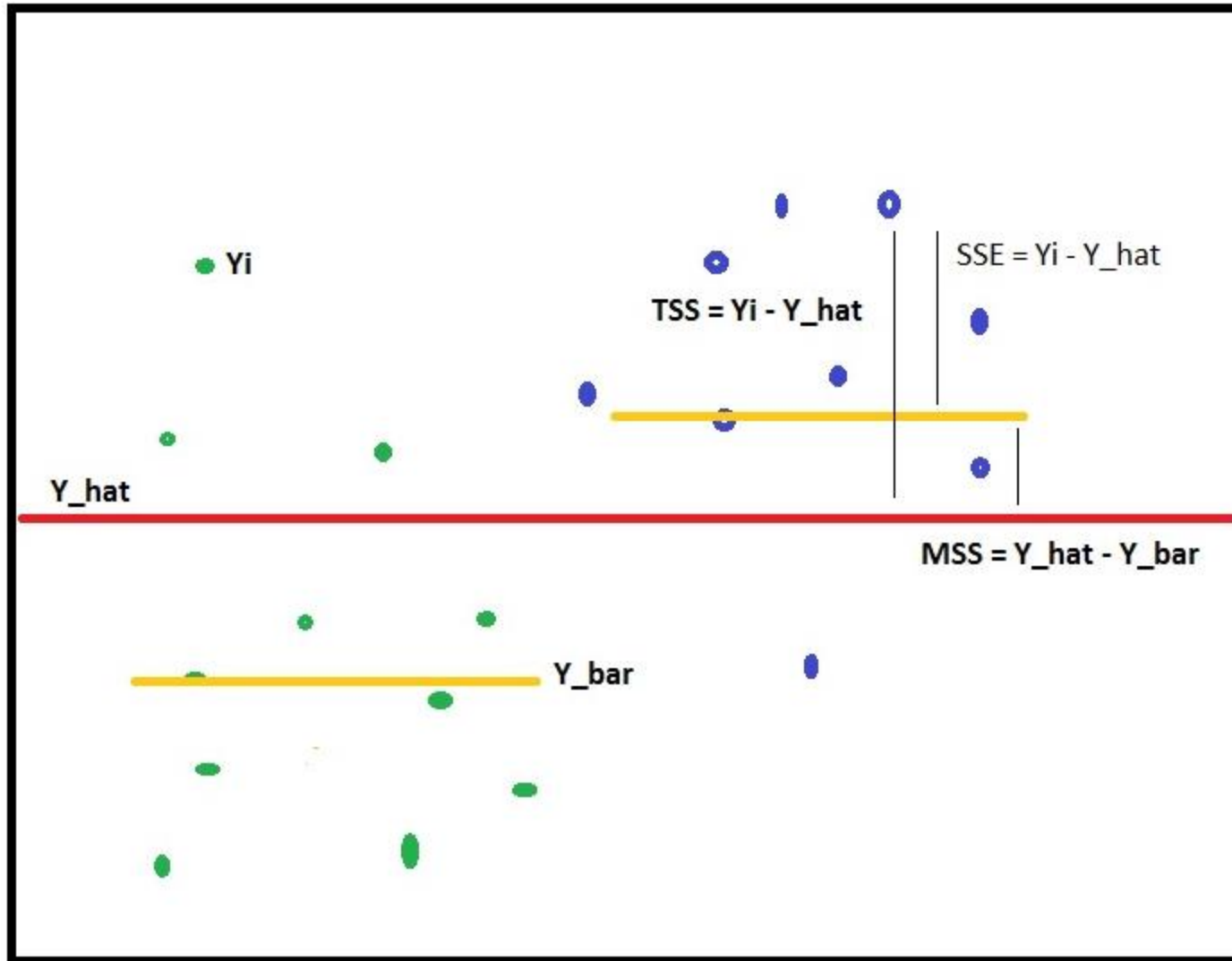
# Noções Preliminares



# Noções Preliminares

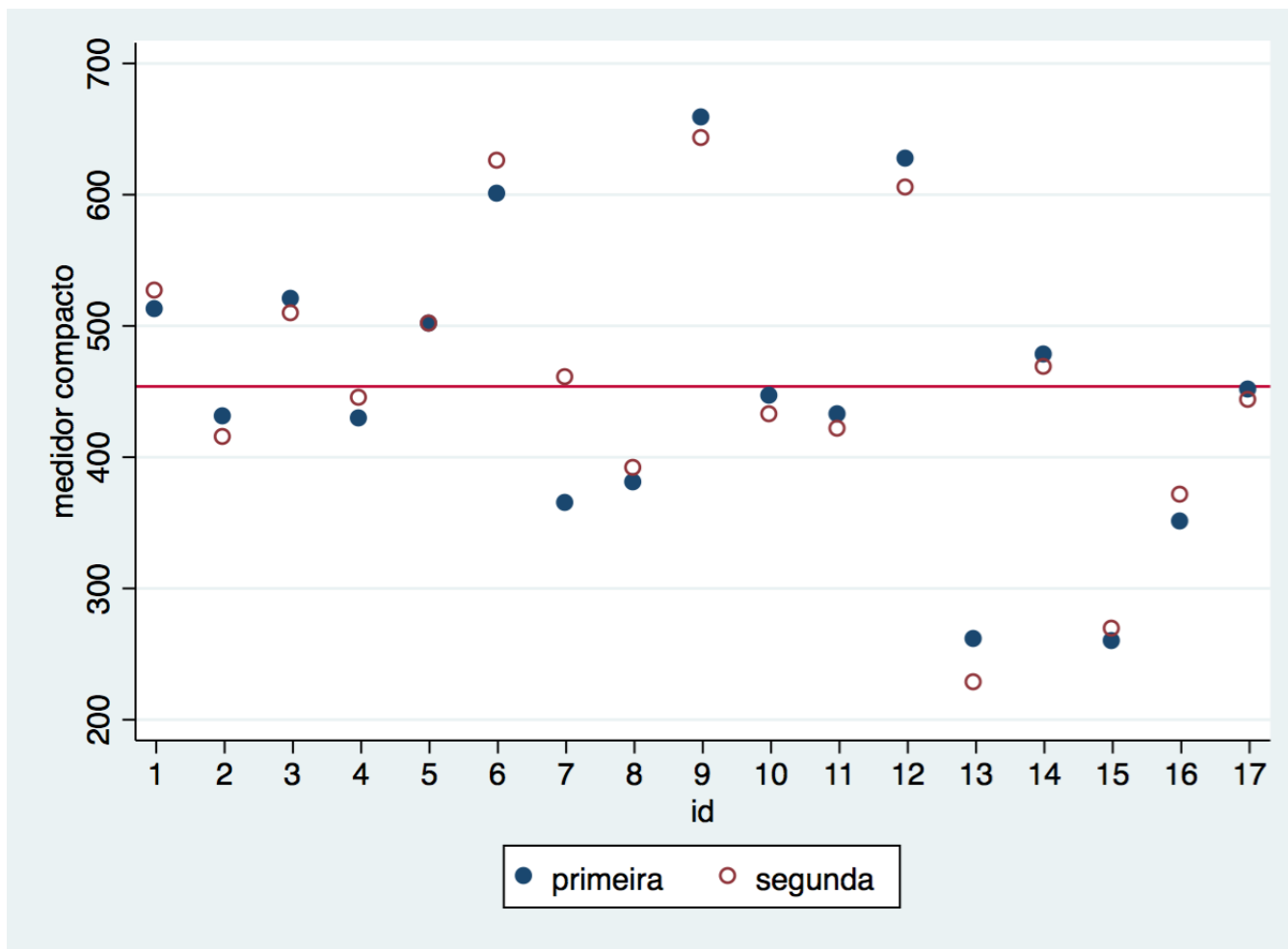


# Noções Preliminares





## Noções Preliminares\*



\* Exemplo extraído do livro "Rabe-Hesketh, S. and Skrondal, A. (2012). Multilevel and Longitudinal Modeling Using Stata (Third Edition). College Station, TX: Stata Press. Vols. I e II.

Poderíamos modelar a mensuração desse aparelho nos 17 indivíduos como dependente apenas da média mais um erro:

$$Y_i = B_1 + E_{ij}$$

Onde  $E_{ij}$  não é correlacionado nem entre sujeitos nem entre ocasiões.

Porém, cada indivíduo tem uma "média individual" que corresponde ao valor médio entre os valores das duas medições feitas para cada indivíduo.

Portanto, podemos escrever o modelo criando um componente de erro "U" que corresponde à variação das médias de cada indivíduo em relação à média global.

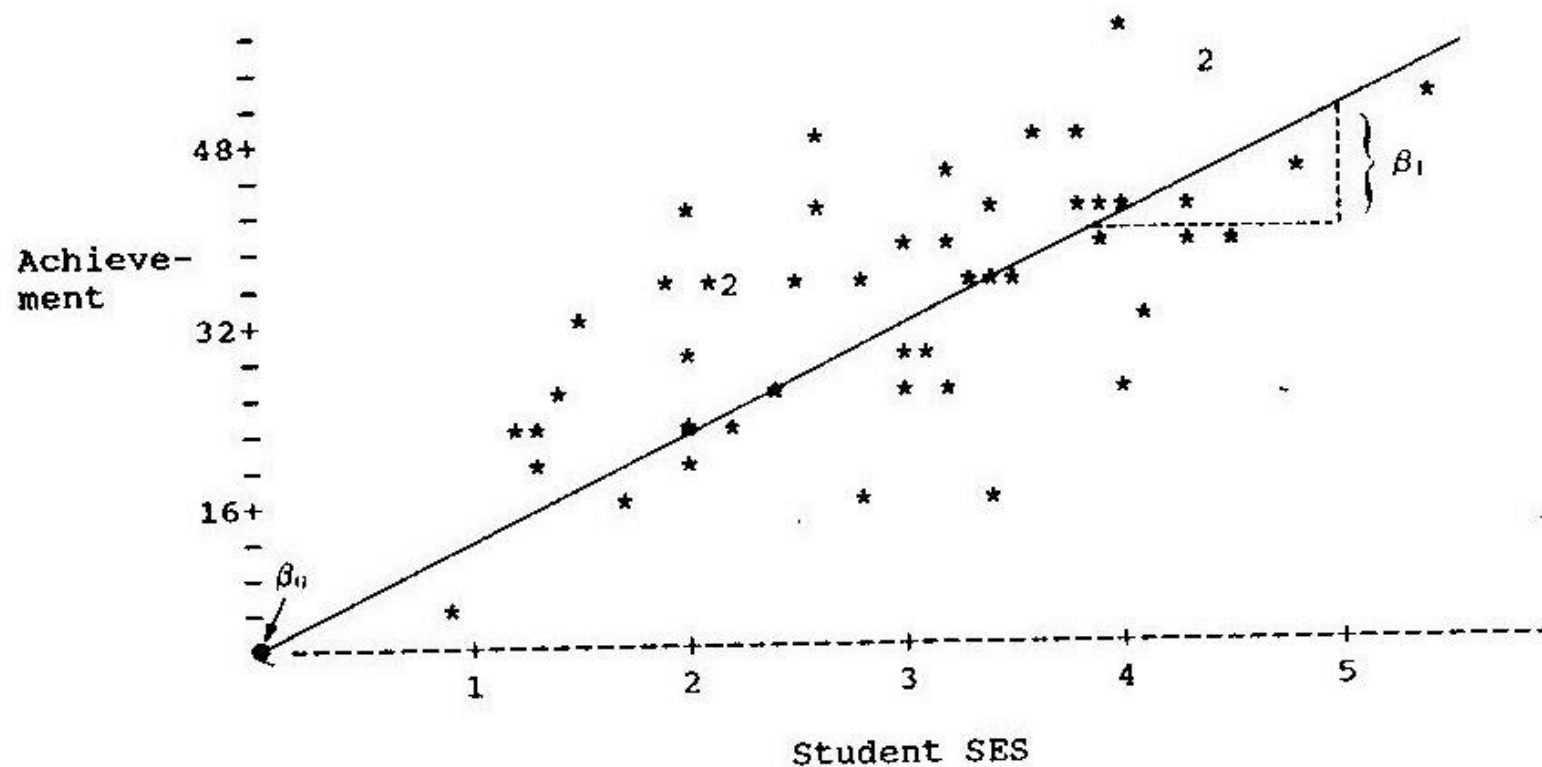
Estamos assim “partindo” o erro total entre este componente U e um componente  $e_{ij}$  correspondente à variação entre cada mensuração ou ocasião e a média do indivíduo:

$$Y_{ij} = B_1 + U_j + e_{ij}$$

1. Modelo de regressão simples com duas variáveis: desempenho em matemática (achievement) e status sócio-econômico (SES) em uma escola.\*

$$Y = B_0 + B_1X_i + r$$

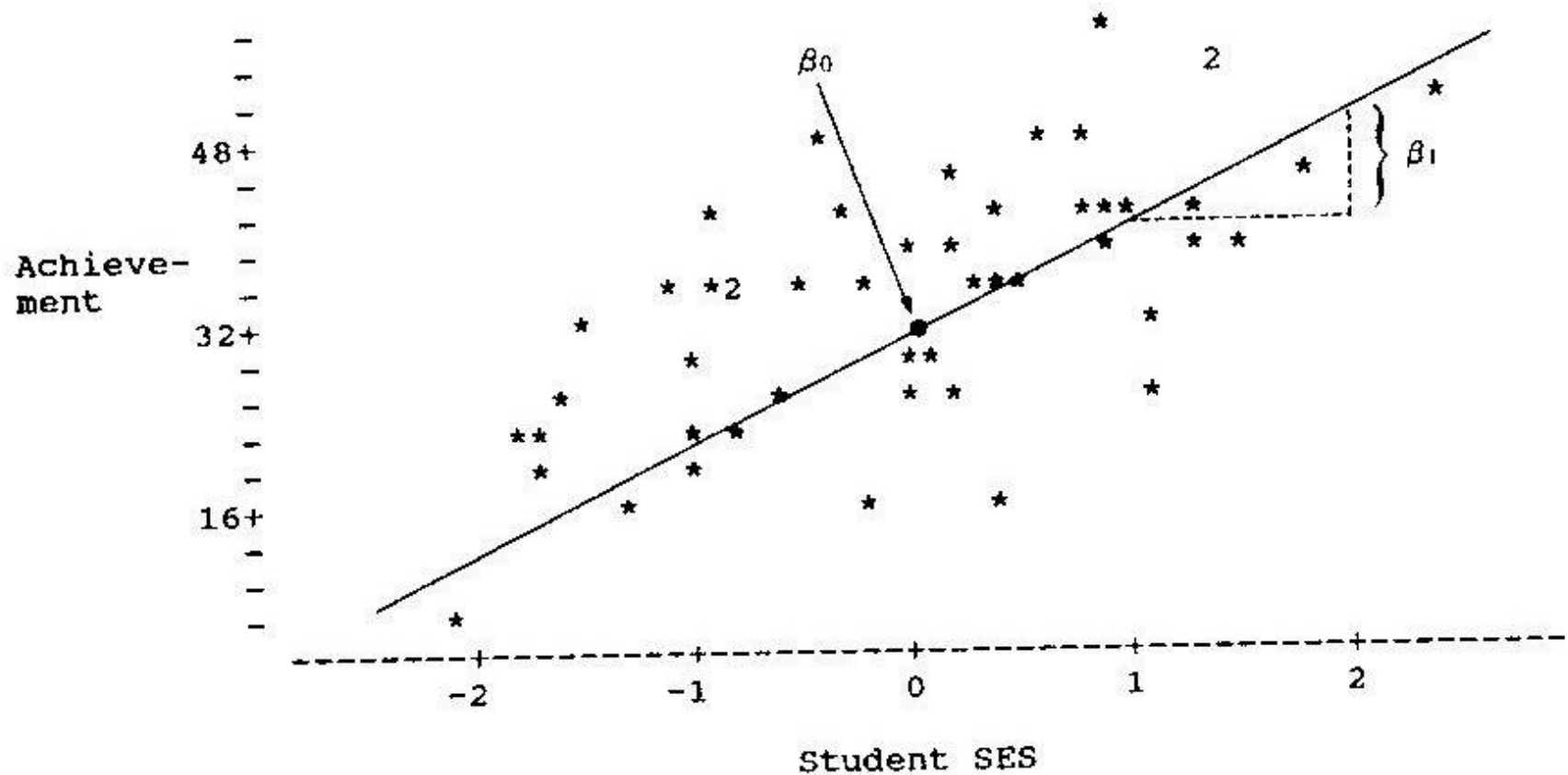
\* NOTA: Todas as lâminas seguintes desta explicação foram diretamente inspiradas em Raudenbush e Bryk, 2002.



**Figure 2.1.** Scatterplot Showing the Relationship Between Achievement and SES in One Hypothetical School

2. Mudando a métrica –  
centralizando o SES da escola em  
torno da média:

$$Y = B_0 + B_1(X_m - X_i) + r$$



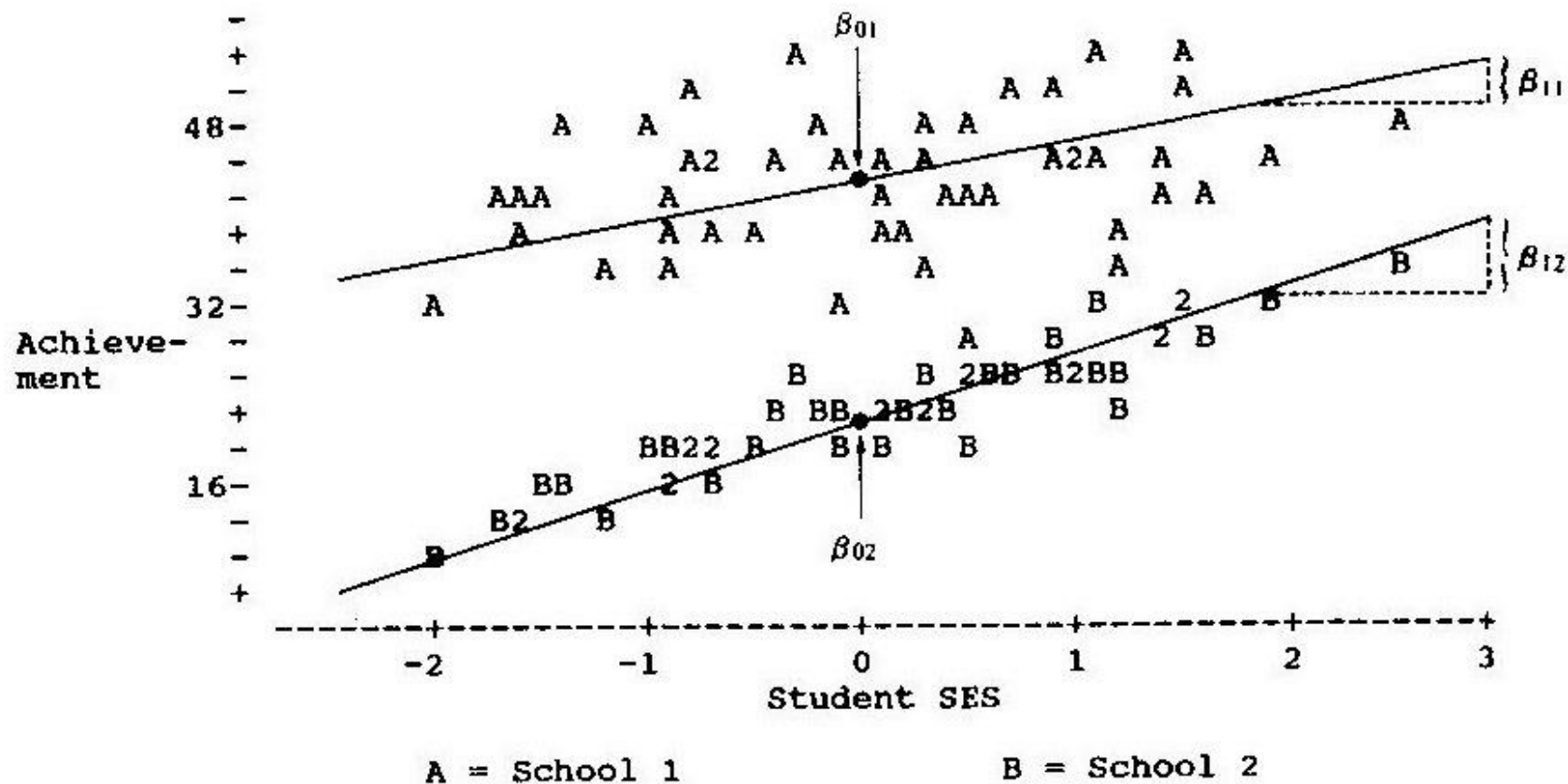
**Figure 2.2.** Scatterplot Showing the Relationship Between Achievement and SES (Centered) in One Hypothetical School

### 3. Relação entre desempenho e SES em duas escolas (centralizada):

$$Y = B0b + Bb(Xbm - Xbi) + rb$$

$$Y = B0a + Ba(Xam - Xai) + ra$$





**Figure 2.3.** Scatterplot Showing the Relationship Between Achievement and SES Within Two Hypothetical Schools

Ou podemos fazer uma regressão de mínimos quadrados com efeitos fixos de escola:

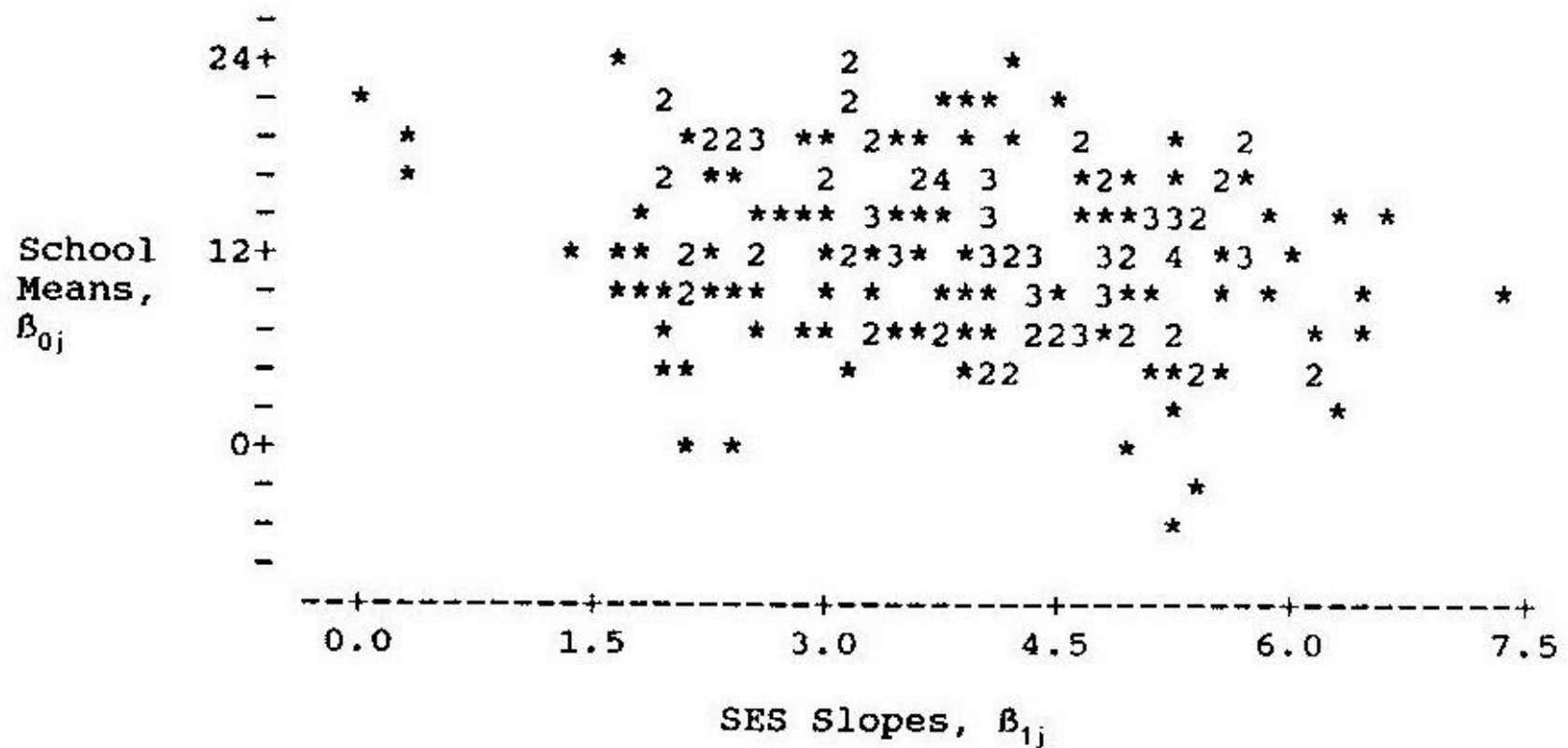
$$Y_{ab} = B_{0ab} + B_{1ab}(X_{abm} - X_{abi}) + B_2(A) + r_{ab}$$

#### 4. Relação entre desempenho e SES em “J” escolas (sendo J um número alto):

$$Y_{ij} = B_{0j} + B_{1j}(X_{ij} - X_{m \cdot j}) + r_{ij}$$

$$r_{ij} \sim N(0, S^2)$$

(Nesse caso, o sub-escrito “j” indica que cada escola tem uma curva e variância próprias. Para cada escola, sua qualidade(*effectiveness*) e equidade(*equity*) são descritas pelo par de valores ( $B_{0j}$ ,  $B_{1j}$ ).)



**Figure 2.4.** Plot of School Means (vertical axis) and SES Slopes (horizontal axis) for 200 Hypothetical Schools

Modelo hipotético para predizer estes resultados:

$$B_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$B_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

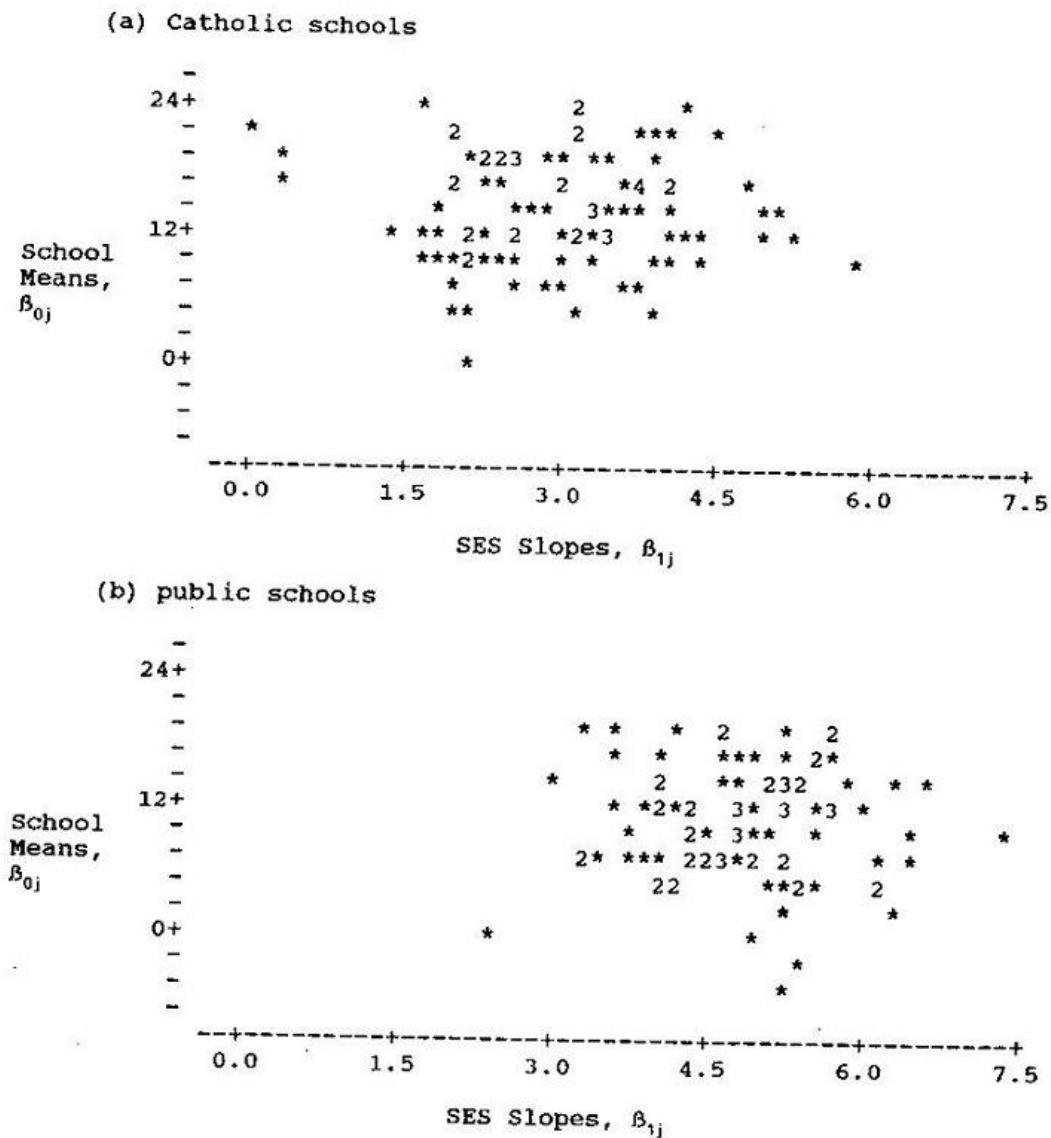
$W$  = tipo de escola (1 = particular; 0 = pública)

Note que o modelo inicial era:

$$Y_{ij} = B_{0j} + B_{1j}(X_{ij} - X_{m.j}) + r_{ij}$$

Substituindo equações (1) e (2) na equação (3) temos:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}(X_{ij} - X_{m.j}) + \gamma_{11}W_j(X_{ij} - X_{m.j}) + u_{0j} + u_{ij}(X_{ij} - X_{m.j}) + r_{ij}$$



**Figure 2.5.** Plot of School Means (vertical axis) and SES Slopes (horizontal axis) for 100 Hypothetical Catholic Schools and 100 Hypothetical Public Schools

$\gamma_{00}$  – desempenho médio para escolas públicas

$\gamma_{01}$  – diferença média entre escolas públicas e católicas

$\gamma_{10}$  – coeficiente de regressão médio em escolas públicas

$\gamma_{11}$  – diferença média entre os coeficientes de regressão de escolas públicas e católicas

$u_{0j}$  – efeito único da escola  $j$  no desempenho médio quando tipo de escola é constante

$u_{1j}$  – efeito único da escola  $j$  no coeficiente de regressão quando  $W$  é constante



## Exemplo 1

**Tabela 1**  
**Caracterização das Variáveis Usadas no Estudo\***

Nível da Característica	Descrição da Variável	Código da Variável
Estudante	Sexo: 0 - Homem 1 - Mulher	SEXO
	Raça: 0 - Branco ou Amarelo 1 - Outras	RAÇA
	Atitude em relação à escola: Escala TRI	ATITUDE
	Atraso: defasagem escolar em anos	ATRASSO
	Proficiência: desempenho: Escala TRI	PROFICT
Família	Recursos culturais: Escala TRI	CULTURAL
	Recursos econômicos: Escala TRI	ECONO
	Participação dos pais: Escala TRI	PAIS
	Estrutura familiar: 0 - Ausência de um pai 1 - Presença de ambos	FAMÍLIA
Escola	Homogeneidade socioeconômica: análise fatorial	ESCOLA
	Localização: 0 – Interior 1 – Capital	LOCAL

Fonte: Síntese dos dados do Saeb, elaborada pelos autores.

\* As variáveis construídas pela TRI ou análise fatorial assumem valores entre -3 e 3.

O modelo final utilizado é:

Modelo de nível 1: aluno

$$\text{ZPROFICT} = B_0 + B_1 * (\text{ZSEXO}) + B_2 * (\text{ZRAÇA}) + B_3 * (\text{ZCULTO}) + B_4 * (\text{ZATRASSO}) + B_5 * (\text{ZATITUDE}) + R$$

Modelo de nível 2: escola

$$B_0 = G_{00} + G_{01} * (\text{ZESCOLA}) + U_0$$

**Tabela 3**  
**Estimativas dos Parâmetros do Modelo e Respectivos Desvios-Padrão**

Termos do Modelo	Notação	Coeficiente	Desvio Padrão	Razão-T	Graus de Liberdade	Valor-P
INTERCEPTO	G00	-0,075	0,010	-7,44	1.690	0,000
ZESCOLA	G01	0,302	0,011	26,56	1.690	0,000
ZSEXO	B1	-0,161	0,015	-10,82	30.346	0,000
ZRAÇA	B2	0,052	0,010	5,20	30.346	0,000
ZCULTO	B3	0,076	0,014	5,31	30.346	0,000
ZATRASSO	B4	-0,184	0,012	-14,57	30.346	0,000
ZATITUDE	B5	0,147	0,013	11,33	30.346	0,000
ZFAMÍLIA	B6	0,075	0,010	7,01	30.346	0,000

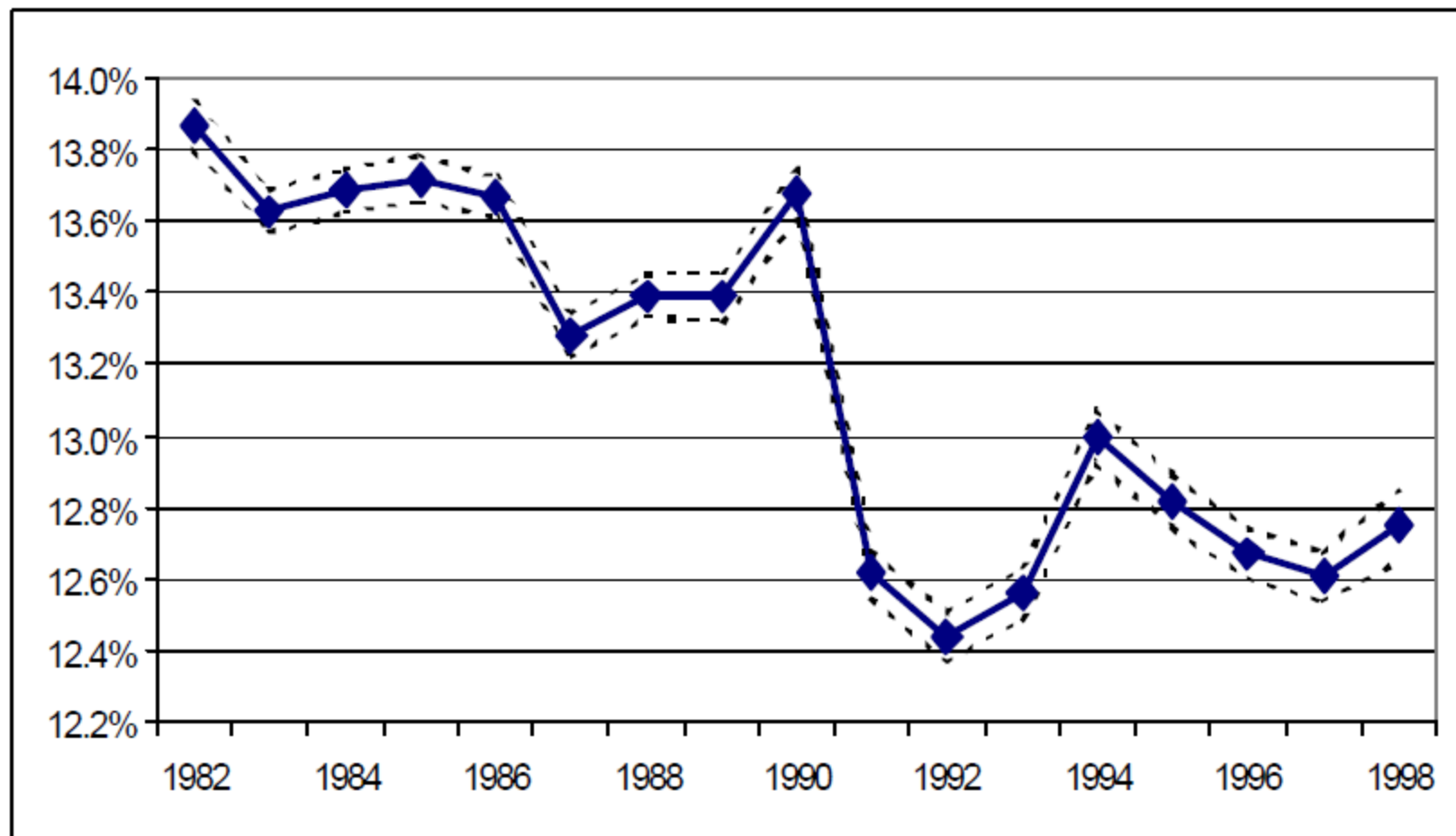
Fonte: Resultado de modelos estatísticos ajustados pelos autores.

## Exemplo 2

# Retornos à educação no Brasil

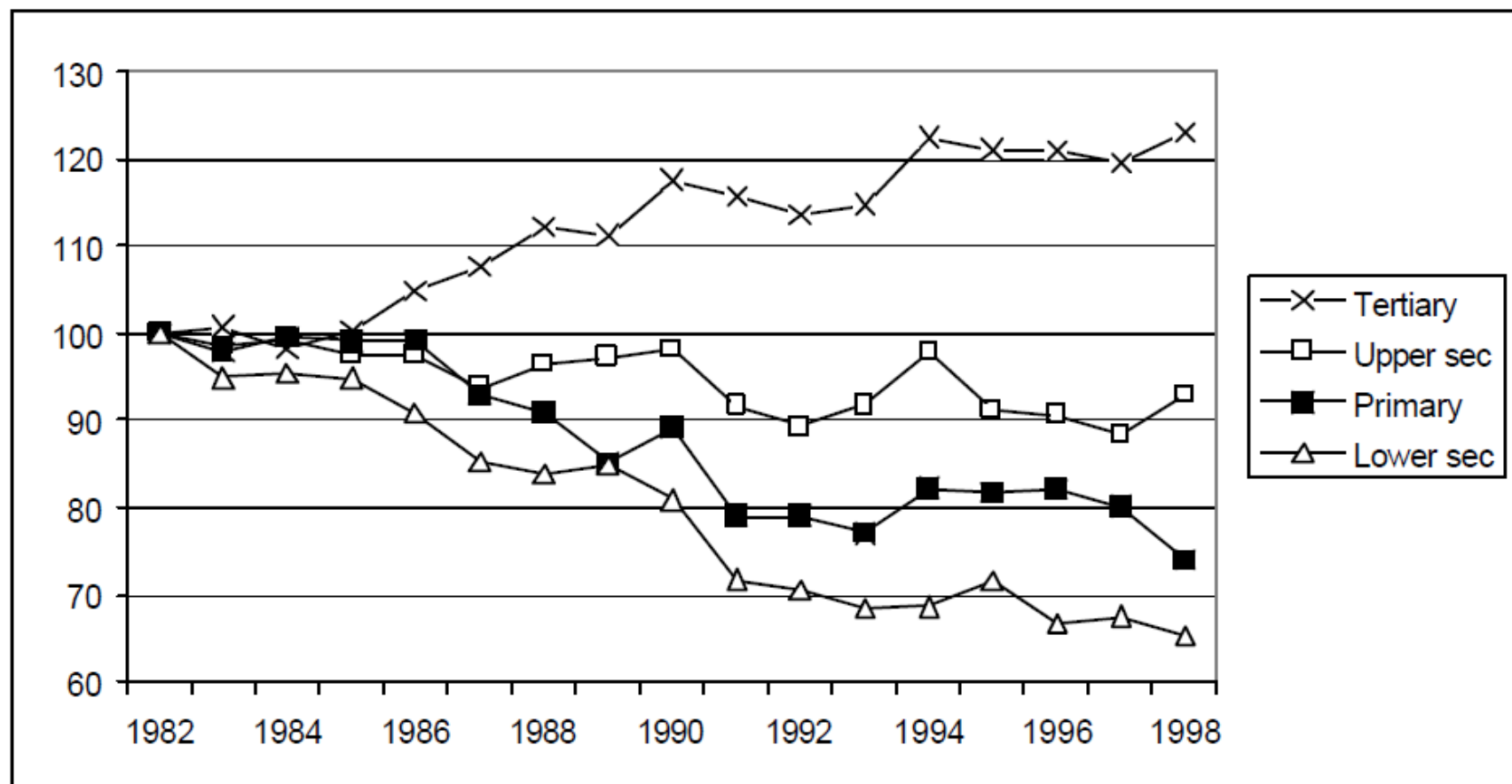
- Bloom, Nielsen e Werner (2001) – renda por nível educacional, Brasil 1982 a 1998.
- Declínio da renda da educação primária e secundária, e aumento da renda da educação superior no período.
- Conclusão: há uma demanda no mercado brasileiro por trabalhadores com nível educacional mais alto.

**Figure 5 The returns to one additional year of schooling**



*Source: authors' calculation based on PME-data*

**Figure 7 Indexed returns by level of education (1982=100)**



*Source: authors' calculation based on PME-data.*

Note: The coefficients are indexed returns to schooling estimated from earnings-regression with control variables.



# Retornos à educação no Brasil

- Schwartzman (2004) – Distribuição dos trabalhadores com ensino superior por setor econômico. Dados PNAD 1992 e 2001.
- Aumento se concentra nos setores onde a renda é mais baixa.
- Conclusão do autor: a expansão do ensino superior no Brasil se dá não pela demanda do mercado mas pelo aspecto credencialista desse mercado.

# Distribuição dos trabalhadores com ensino superior em cada setor da economia

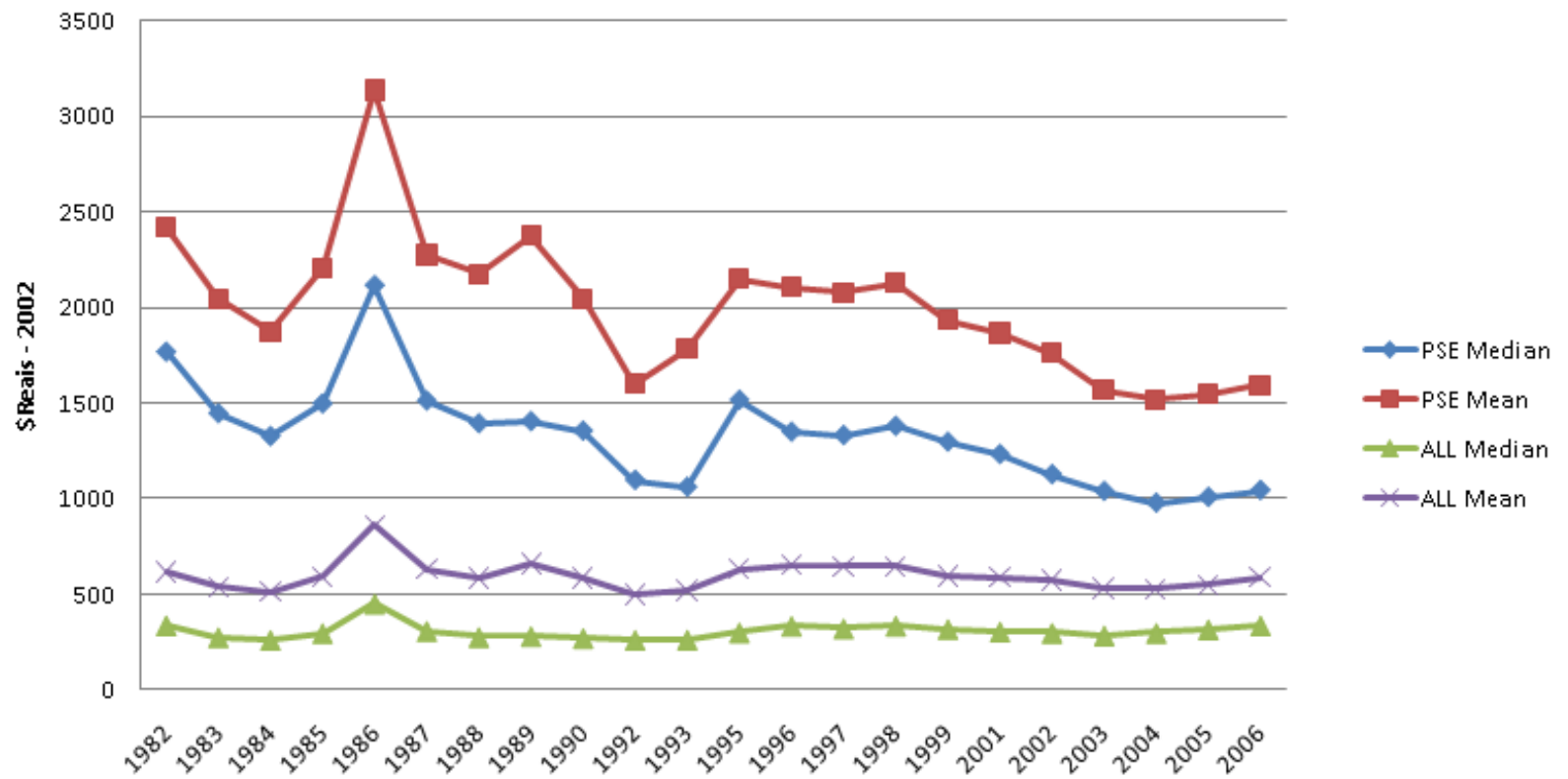
Occupations	1992	2001	change 1992-2001	mean monthly income, 2001(*)
Agriculture	2.0%	1.4%	-0.6%	2,034.45
Industry	12.1%	9.2%	-2.9%	2,525.28
Construction	2.2%	1.6%	-0.6%	2,586.38
Other industrial activities	2.0%	1.3%	-0.8%	2,326.33
Commerce	9.5%	11.2%	1.6%	1,812.59
Services to business	4.0%	6.1%	2.1%	1,380.19
Public administration	10.4%	14.2%	3.8%	2,369.89
Transportation and communications	2.6%	2.7%	0.2%	2,178.56
Social activities, education	33.8%	34.3%	0.5%	1,637.14
Public administration	13.1%	12.2%	-0.9%	2,324.58
others	8.2%	5.8%	-2.4%	2,266.40
Total econ. active	5,142,229	6,396,257	24%	1,989.96

Fonte: Schwartzman, 2004.

O impacto do incremento de  
trabalhadores com diploma de ensino  
superior no mercado de trabalho  
brasileiro.

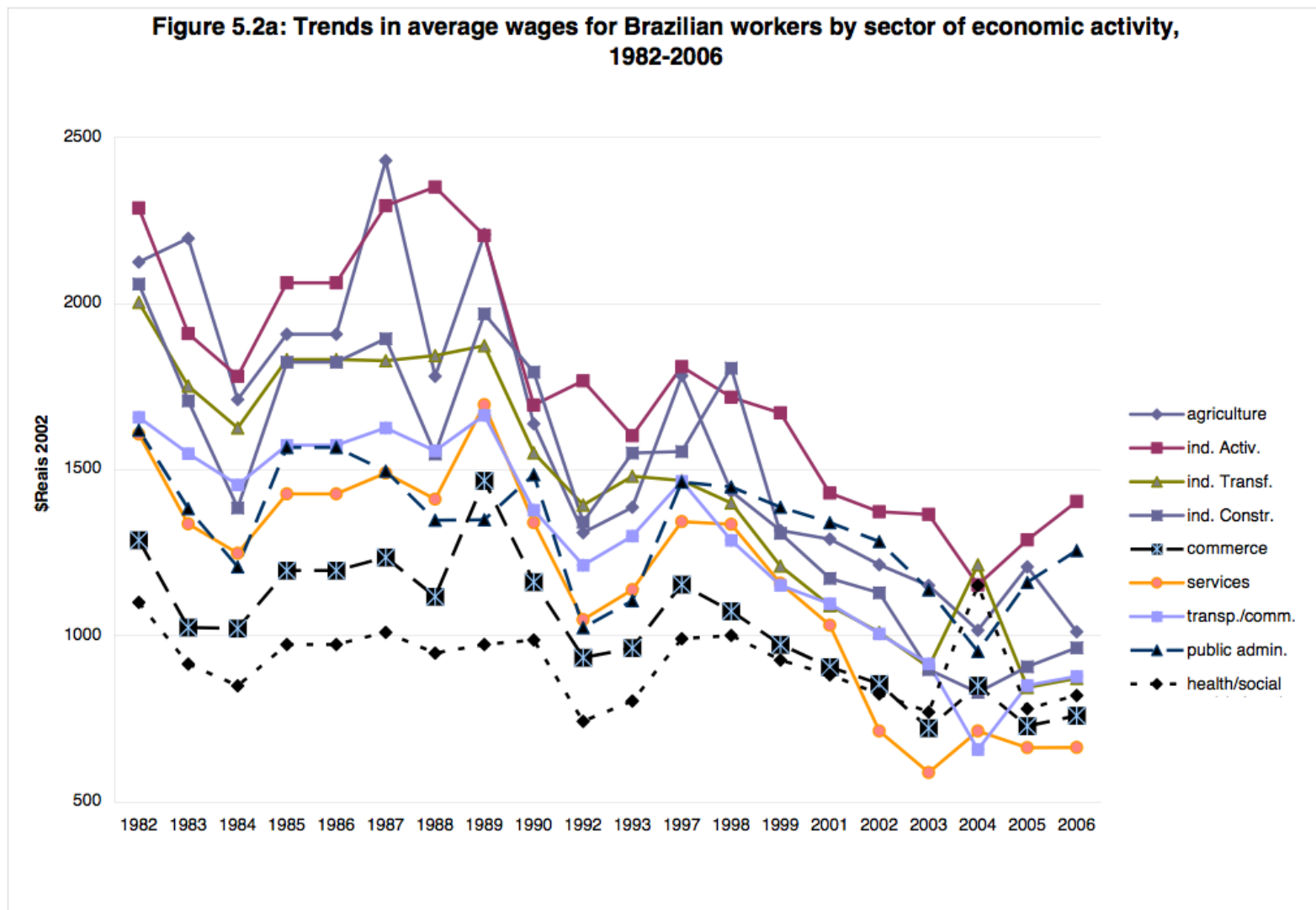
## Evolução da renda do trabalho no Brasil, 1982-2006.

**Figure 5.3: Average and median wages in Brazil, for all educational levels and for postsecondary graduates only - 1982-2006**



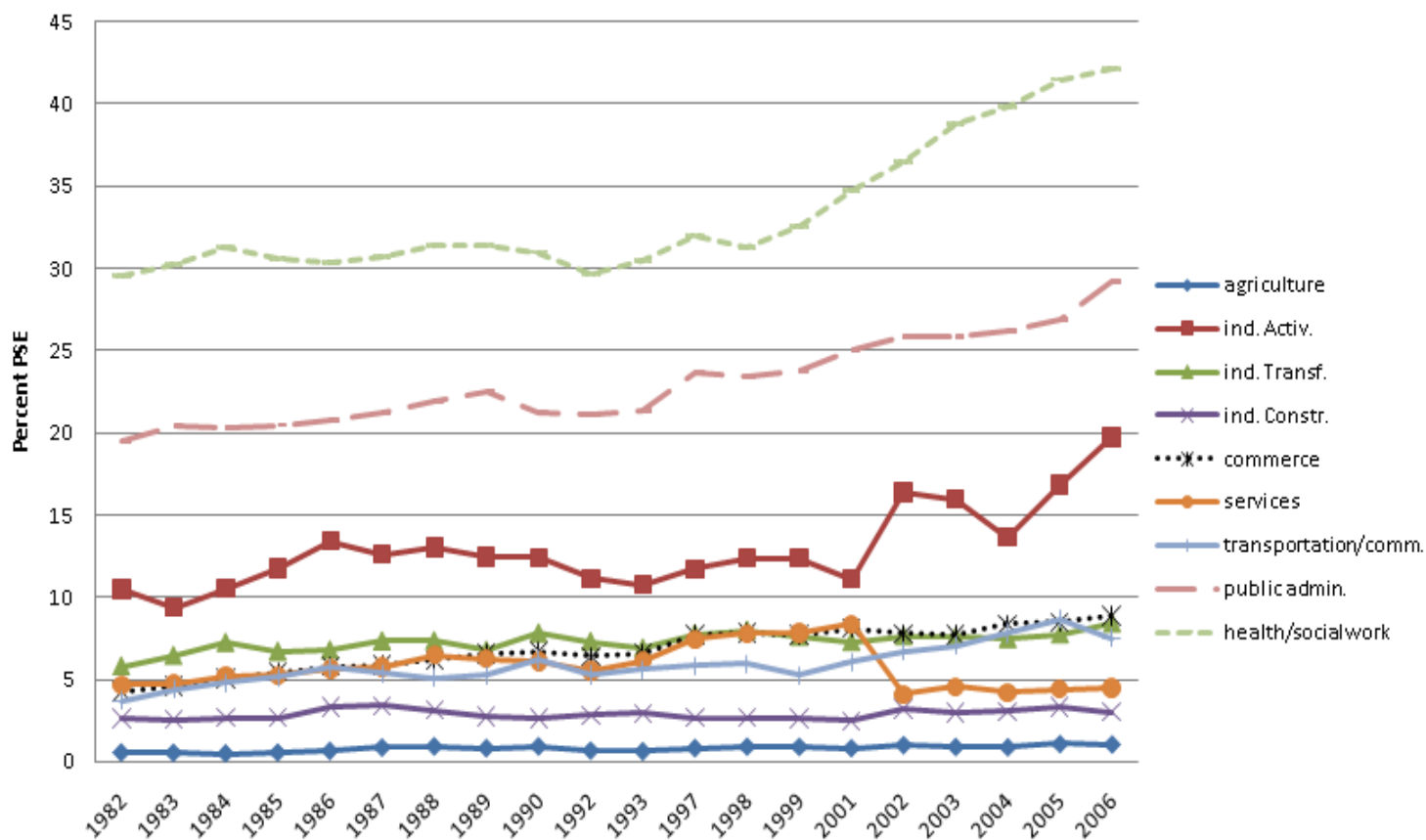
Source: PNAD/IBGE

# Variação da renda média do trabalhador por setor econômico



# Variação do percentual de graduados no ensino superior por setor de atividade econômica.

Figure 5.5a: Changes in the percent of postsecondary graduates by sector of economic activity in Brazil, 1982 to 2006.



Source: PNAD/IBGE

# Amostra

- PNAD – 1982 a 2006 (todos os anos disponíveis exceto 1995 e 1996), agregados em 3 décadas: 1980, 1990 e 2000.
- Modelos rodados separadamente para cada década.
- Homens de 25 a 40 anos, com renda positiva proveniente de salário, que trabalham 20 hrs por semana ou mais e possuem no mínimo educação secundária.

# Variáveis

- Variável resposta: renda proveniente do emprego principal (log).

- Variáveis de interesse:

A = Trabalhador possui educação superior (*versus* ensino médio);

B = Porcentagem de trabalhadores com ensino superior por setor da economia.

C = Interação  $A * B$ .



# Variáveis

- Variáveis de controle:

Sexo, raça, idade, idade ao quadrado, região do país, ano (linear) índice de GINI anual.

- Peso populacional da PNAD corrigido (nível 1).

# Modelo

- Regressão hierárquica com dois níveis:

Nível 1: Trabalhadores

Nível 2: Setor da economia \* ano

N nível 1: (1980 = 128.356; 1990 = 97.753; 2000 = 142.635)

N nível 2 = (1980 = 72; 1990 = 46; 2000 = 53)

# Modelo 1 (M1)

$$\text{Log}(Wages) = B_0 + B_{1j}PSE + B_{ij}X_i + e_{ij}$$

$$B_{0j} = \gamma_{00} + \gamma_{01}\%PSE + u_{0j}$$

## Modelo 2 (M2)

$$\text{Log}(Wages) = B_0 + B_{1j}PSE + B_{ij}X_i + e_{ij}$$

$$B_{0j} = \gamma_{00} + \gamma_{01}\%PSE + u_{0j}$$

$$B_{1j} = \gamma_{01}\%PSE + u_{1j}$$

Table 5.4: The Influence of the Supply of Postsecondary Graduates on Income Returns to Postsecondary Degrees.

Outcome: Log of Income from main job

Decade	1980s	1980s	1990s	1990s	2000s	2000s
Model	<u>M1</u>	<u>M2</u>	<u>M1</u>	<u>M2</u>	<u>M1</u>	<u>M2</u>
Intercept	<b>6.674</b> ** (0.026)	<b>6.674</b> ** (0.025)	<b>6.436</b> ** (0.027)	<b>6.436</b> ** (0.027)	6.159 ** (0.026)	6.159 ** (0.026)
<u>Slopes</u>						
RSE	0.714 ** (0.012)	0.714 ** (0.011)	<b>0.719</b> ** (0.023)	<b>0.719</b> ** (0.023)	0.785 ** (0.028)	0.787 ** (0.028)
<u>Level 2 Variables</u>						
%PSE	-0.451 (0.264)	-0.456 (0.262)	<b>-0.163</b> (0.246)	<b>-0.223</b> (0.207)	0.468 * (0.206)	0.615 ** (0.188)
PSE*%PSE		0.378 ** (0.121)		0.181 (0.189)		-0.384 ** (0.142)
<u>Level 1 Variables</u>						
Male	<b>0.541</b> ** (0.011)	<b>0.541</b> ** (0.011)	<b>0.472</b> ** (0.015)	<b>0.472</b> ** (0.015)	0.422 ** (0.015)	0.422 ** (0.015)
Age	<b>0.123</b> ** (0.007)	<b>0.124</b> ** (0.007)	<b>0.062</b> ** (0.007)	<b>0.062</b> ** (0.007)	0.080 ** (0.005)	0.080 ** (0.005)
Age Squared	<b>-0.001</b> ** (0.000)	<b>-0.001</b> ** (0.000)	<b>-0.001</b> ** (0.000)	<b>-0.001</b> ** (0.000)	-0.001 ** (0.000)	-0.001 ** (0.000)
South	<b>-0.001</b> (0.012)	<b>-0.001</b> (0.012)	<b>-0.067</b> ** (0.010)	<b>-0.067</b> ** (0.010)	-0.060 ** (0.010)	-0.060 ** (0.010)
North	<b>0.005</b> (0.012)	<b>0.005</b> (0.012)	<b>-0.098</b> ** (0.017)	<b>-0.098</b> ** (0.017)	-0.136 ** (0.012)	-0.136 ** (0.012)
Northeast	<b>-0.300</b> ** (0.016)	<b>-0.300</b> ** (0.016)	<b>-0.388</b> ** (0.012)	<b>-0.388</b> ** (0.012)	-0.356 ** (0.012)	-0.356 ** (0.012)
Central-West	<b>0.082</b> ** (0.014)	<b>0.082</b> ** (0.014)	-0.011 (0.022)	-0.011 (0.022)	0.032 (0.025)	0.032 (0.025)
Black	<b>-0.342</b> ** (0.031)	<b>-0.342</b> ** (0.031)	<b>-0.245</b> ** (0.024)	<b>-0.245</b> ** (0.024)	-0.181 ** (0.015)	-0.181 ** (0.015)
Brown	-0.220 ** (0.012)	-0.220 ** (0.012)	<b>-0.220</b> ** (0.012)	<b>-0.220</b> ** (0.012)	-0.185 ** (0.007)	-0.185 ** (0.007)
Year (linear)	<b>-0.016</b> (0.014)	<b>-0.016</b> (0.014)	0.028 (0.017)	0.028 (0.017)	0.053 (0.067)	0.053 (0.067)
GINI index	<b>-3.696</b> (2.204)	<b>-3.758</b> (2.192)	15.943 * (7.665)	15.944 * (7.670)	10.147 (8.329)	10.144 (8.296)
** p < .01						
* p < .05						
N (level 1)	128356		97753		142635	
N (level 2)	72		46		53	
Sigma^2	0.48864		0.49917		0.40459	
Tao (level 2)	0.05		0.03		0.04	
Tao (PSE slope)	0.01		0.02		0.04	

Bold Characters: Coefficient is significantly different from the decade immediately to the right (2000 has no bold values)

Sample (PNAD, 1982-2006): Employees 25-40 years old, with at least secondary ed. and a positive income.

Table 3-4: The Influence of the Supply of Postsecondary Graduates on Income Returns to Postsecondary Degrees.

Outcome: Log of Income from main job

Decade	1980s	1980s	1990s	1990s	2000s	2000s
Model	M1	M2	M1	M2	M1	M2
Intercept	<b>6.674</b> ** (0.026)	<b>6.674</b> ** (0.025)	<b>6.436</b> ** (0.027)	<b>6.436</b> ** (0.027)	6.159 ** (0.026)	6.159 ** (0.026)
<u>Slopes</u>						
PSE	0.714 ** (0.012)	0.714 ** (0.011)	<b>0.719</b> ** (0.023)	<b>0.719</b> ** (0.023)	0.785 ** (0.028)	0.787 ** (0.028)
<u>Level 2 Variables</u>						
%PSE	-0.451 (0.264)	-0.456 (0.262)	<b>-0.163</b> (0.246)	<b>-0.223</b> (0.207)	0.468 * (0.206)	0.615 ** (0.188)
PSE*%PSE		0.378 ** (0.121)		0.181 (0.189)		-0.384 ** (0.142)
<u>Level 1 Variables</u>						
Male	<b>0.541</b> ** (0.011)	<b>0.541</b> ** (0.011)	<b>0.472</b> ** (0.015)	<b>0.472</b> ** (0.015)	0.422 ** (0.015)	0.422 ** (0.015)
Age	<b>0.123</b> ** (0.007)	<b>0.124</b> ** (0.007)	<b>0.062</b> ** (0.007)	<b>0.062</b> ** (0.007)	0.080 ** (0.005)	0.080 ** (0.005)
Age Squared	<b>-0.001</b> ** (0.000)	<b>-0.001</b> ** (0.000)	<b>-0.001</b> ** (0.000)	<b>-0.001</b> ** (0.000)	-0.001 ** (0.000)	-0.001 ** (0.000)
South	<b>-0.001</b> (0.012)	<b>-0.001</b> (0.012)	<b>-0.067</b> ** (0.010)	<b>-0.067</b> ** (0.010)	-0.060 ** (0.010)	-0.060 ** (0.010)
North	<b>0.005</b> (0.012)	<b>0.005</b> (0.012)	<b>-0.098</b> ** (0.017)	<b>-0.098</b> ** (0.017)	-0.136 ** (0.012)	-0.136 ** (0.012)
Northeast	<b>-0.300</b> ** (0.016)	<b>-0.300</b> ** (0.016)	<b>-0.388</b> ** (0.012)	<b>-0.388</b> ** (0.012)	-0.356 ** (0.012)	-0.356 ** (0.012)
Central-West	<b>0.082</b> ** (0.014)	<b>0.082</b> ** (0.014)	-0.011 (0.022)	-0.011 (0.022)	0.032 (0.025)	0.032 (0.025)
Black	<b>-0.342</b> ** (0.031)	<b>-0.342</b> ** (0.031)	<b>-0.245</b> ** (0.024)	<b>-0.245</b> ** (0.024)	-0.181 ** (0.015)	-0.181 ** (0.015)
Brown	-0.220 ** (0.012)	-0.220 ** (0.012)	<b>-0.220</b> ** (0.012)	<b>-0.220</b> ** (0.012)	-0.185 ** (0.007)	-0.185 ** (0.007)
Year (linear)	<b>-0.016</b> (0.014)	<b>-0.016</b> (0.014)	0.028 (0.017)	0.028 (0.017)	0.053 (0.067)	0.053 (0.067)
GINI index	<b>-3.696</b> (2.204)	<b>-3.758</b> (2.192)	15.943 * (7.665)	15.944 * (7.670)	10.147 (8.329)	10.144 (8.296)
** p < .01						
* p < .05						
N (level 1)	128356		97753		142635	
N (level 2)	72		46		53	
Sigma^2	0.48864		0.49917		0.40459	
Tao (level 2)	0.05		0.03		0.04	
Tao (PSE slope)	0.01		0.02		0.04	

Bold Characters: Coefficient is significantly different from the decade immediately to the right (2000 has no bold values)

Sample (PNAD, 1982-2006): Employees 25-40 years old, with at least secondary ed. and a positive income.

# Resultados

Decade	1980s		1980s		1990s		1990s		2000s		2000s	
	<u>M1</u>		<u>M2</u>		<u>M1</u>		<u>M2</u>		<u>M1</u>		<u>M2</u>	
<b>Intercept</b>	6.674**		6.674**		6.436**		6.436**		6.159**		6.159**	
	(0.026)		(0.025)		(0.027)		(0.027)		(0.026)		(0.026)	
<b><u>Slopes</u></b>												
<b>PSE</b>	0.714**		0.714**		0.719**		0.719**		0.785**		0.787**	
	(0.012)		(0.011)		(0.023)		(0.023)		(0.028)		(0.028)	
<b><u>Level 2 Variables</u></b>												
<b>% PSE</b>	-0.451		-0.456		-0.163		-0.223		0.468*		0.615**	
	(0.264)		(0.262)		(0.246)		(0.207)		(0.206)		(0.188)	
<b>PSE* % PSE</b>			0.378**				0.181				-0.384**	
			(0.121)				(0.189)				(0.142)	

# Resultados

- A vantagem comparativa (relativa ao ensino médio) do diploma de curso superior permanece.
- Nos setores onde a renda das ocupações de ensino superior é mais alta, essa renda caiu mas a diferença de renda  $\text{EnsSup} - \text{EnsMed}$  permaneceu alta (efeito preço).
- Nos setores onde a renda das ocupações de ensino superior é mais baixa, esta permaneceu estável, mas a diferença entre a renda  $\text{EnsSup} - \text{EnsMed}$  caiu.



Esse paradoxo já foi observado nos EUA pós segunda guerra, quando a economia desacelerou, o número de college graduates continuou crescendo mas seus salários não caíram.

De acordo com Smith (1986), isso aconteceu porque educação é apenas um parâmetro comparativo. A vantagem observada no modelo apresentado se refere à renda comparada com a de trabalhadores com segundo grau. Ou seja, no geral, a renda do trabalhador com ensino superior pode ter caído, mas se comparada ao trabalhador com apenas ensino médio ela permaneceu estável.