AULAS 20 E 21
Análise de regressão múltipla: inferência

Ernesto F. L. Amaral

15 e 22 de abril de 2014
Avaliação de Políticas Públicas (DCP 046)

Fonte:
Wooldridge, Jeffrey M. "Introdução à econometria: uma abordagem moderna". São Paulo: Cengage Learning, 2008. pp.110-157 (capítulo 4).

TRANSFORMAÇÃO É QUESTÃO EMPÍRICA

– Os objetivos de realizar transformações de variáveis independentes e dependente são:

  – Alcançar distribuição normal da variável dependente.

  – Estabelecer correta relação entre variável dependente e independentes.

– Fazer uma transformação de salário, especialmente tomando o log, produz uma distribuição que está mais próxima da normal.

– Sempre que y assume apenas alguns valores, não podemos ter uma distribuição próxima de uma distribuição normal.

– “Essa é uma questão empírica.” (Wooldridge, 2008: 112)

EXEMPLOS DE TRANSFORMAÇÕES

– **Variável dependente em alguns modelos:**

– Lineares (MQO): deve ter nível de mensuração de razão (contínua) e distribuição normal (logaritmo reduz concentração à esquerda de variáveis com valores positivos diferentes de zero).

– Logísticos e probit: variável dicotômica.

– Multinomiais: variável nominal com mais de duas categorias.

– Poisson: variável é contagem com concentração em zero.

– **Variável independente:**

– Se for contínua, deve ter distribuição normal (logaritmo).

– Se for contínua, também podemos calcular o quadrado (x^2) e incluir junto com variável independente original (x).

– Se for categórica, buscamos distribuição uniforme entre categorias, mas nem sempre é possível (categoria de referência geralmente possui vários casos).

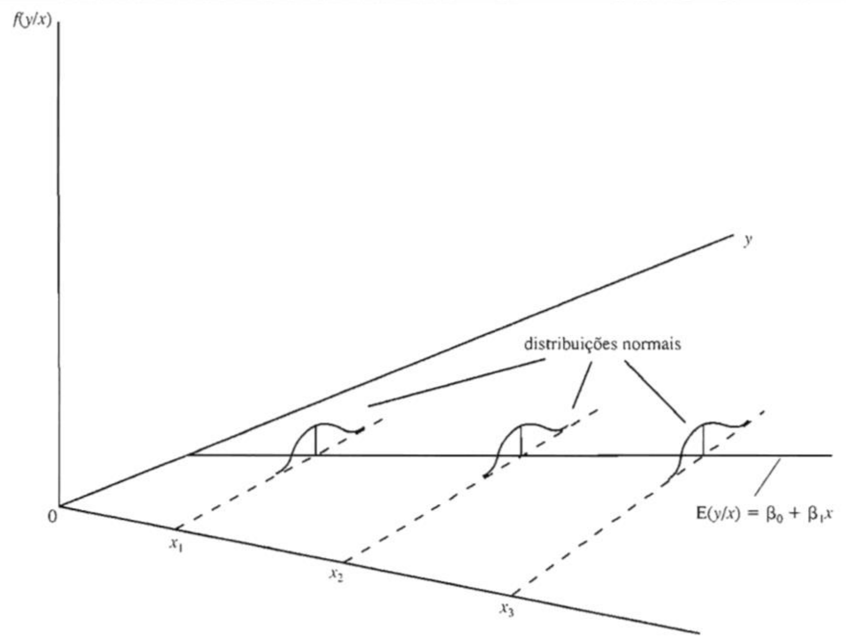
MODELO LINEAR CLÁSSICO

– As hipóteses BLUE, adicionadas à hipótese da normalidade (erro não-observado é normalmente distribuído na população), são conhecidas como hipóteses do modelo linear clássico (MLC).

– Distribuição normal homoscedástica com uma única variável explicativa:

Figura 4.1

A distribuição normal homoscedástica com uma única variável explicativa.



Fonte: Wooldridge, 2008: 111.

TESTES DE HIPÓTESE

– Podemos fazer testes de hipóteses sobre um único parâmetro da função de regressão populacional.

– Os β_j são características desconhecidas da população.

– Na maioria das aplicações, nosso principal interesse é testar a hipótese nula (H_0: $\beta_j = 0$).

– Como β_j mede o efeito parcial de x_j sobre o valor esperado de y, após controlar todas as outras variáveis independentes, a hipótese nula significa que, uma vez que x_1, x_2, ..., x_k foram considerados, x_j não tem nenhum efeito sobre o valor esperado de y.

– O teste de hipótese na regressão múltipla é semelhante ao teste de hipótese para a média de uma população normal.

– É difícil obter os coeficientes, erros-padrão e valores críticos, mas os programas econométricos (nosso amigo Stata) calculam estas estimativas automaticamente.

TESTE *t*

– A estatística *t* é a razão entre o coeficiente estimado (β_j) e seu erro padrão: ep(β_j).

– O erro padrão é sempre positivo, então a razão *t* sempre terá o mesmo sinal que o coeficiente estimado.

– Valor estimado de beta distante de zero é evidência contra a hipótese nula, mas devemos ponderar pelo erro amostral.

– Como o erro-padrão de β_j é uma estimativa do desvio-padrão de β_j, **o teste *t* mede quantos desvios-padrão estimados $\boldsymbol{\beta_j}$ está afastado de zero**.

– Isso é o mesmo que testar se a média de uma população é zero, usando a estatística *t* padrão.

– A regra de rejeição depende da hipótese alternativa e do nível de significância escolhido do teste.

– Sempre testamos hipótese sobre parâmetros populacionais, e não sobre estimativas de uma amostra particular.

p-VALORES DOS TESTES *t*

– Dado o valor observado da estatística *t*, qual é o menor nível de significância ao qual a hipótese nula seria rejeitada?

– Não há nível de significância “correto”.

– O *p*-valor é a probabilidade da hipótese nula não ser rejeitada:

  – *p*-valores pequenos são evidências contra hipótese nula.

  – *p*-valores grandes fornecem pouca evidência contra H_0.

– Se α é o nível de significância do teste, então H_0 é rejeitada se *p*-valor < α.

– H_0 não é rejeitada ao nível de 100*α%.

TESTE: HIPÓTESES ALTERNATIVAS UNILATERAIS

H_1: $\beta_j > 0$ OU H_1: $\beta_j < 0$

– Devemos decidir sobre um nível de significância (geralmente de 5%).

– Corremos o risco de rejeitar erroneamente H_0, quando ela é verdadeira, em 5% das vezes (erro tipo I igual ao α).

– Um valor suficientemente grande de *t*, com um nível de significância de 5%, é o 95º percentil de uma distribuição *t* com *n*-*k*-1 graus de liberdade (ponto c).

– **Regra de rejeição** é que H_0 é rejeitada em favor de H_1, se $t>c$ (H_1:$\beta_j>0$) ou $t<-c$ (H_1:$\beta_j<0$), em um nível específico.

– Quando os graus de liberdade da distribuição *t* ficam maiores, a distribuição *t* aproxima-se da distribuição normal padronizada.

– Para graus de liberdade maiores que 120, pode-se usar os valores críticos da distribuição normal padronizada.

REGRA DE REJEIÇÃO DE H_0 (UNILATERAL)

Figura 4.2

Regra de rejeição a 5% para a hipótese alternativa $H_1: \beta_j > 0$ com 28 *gl*.

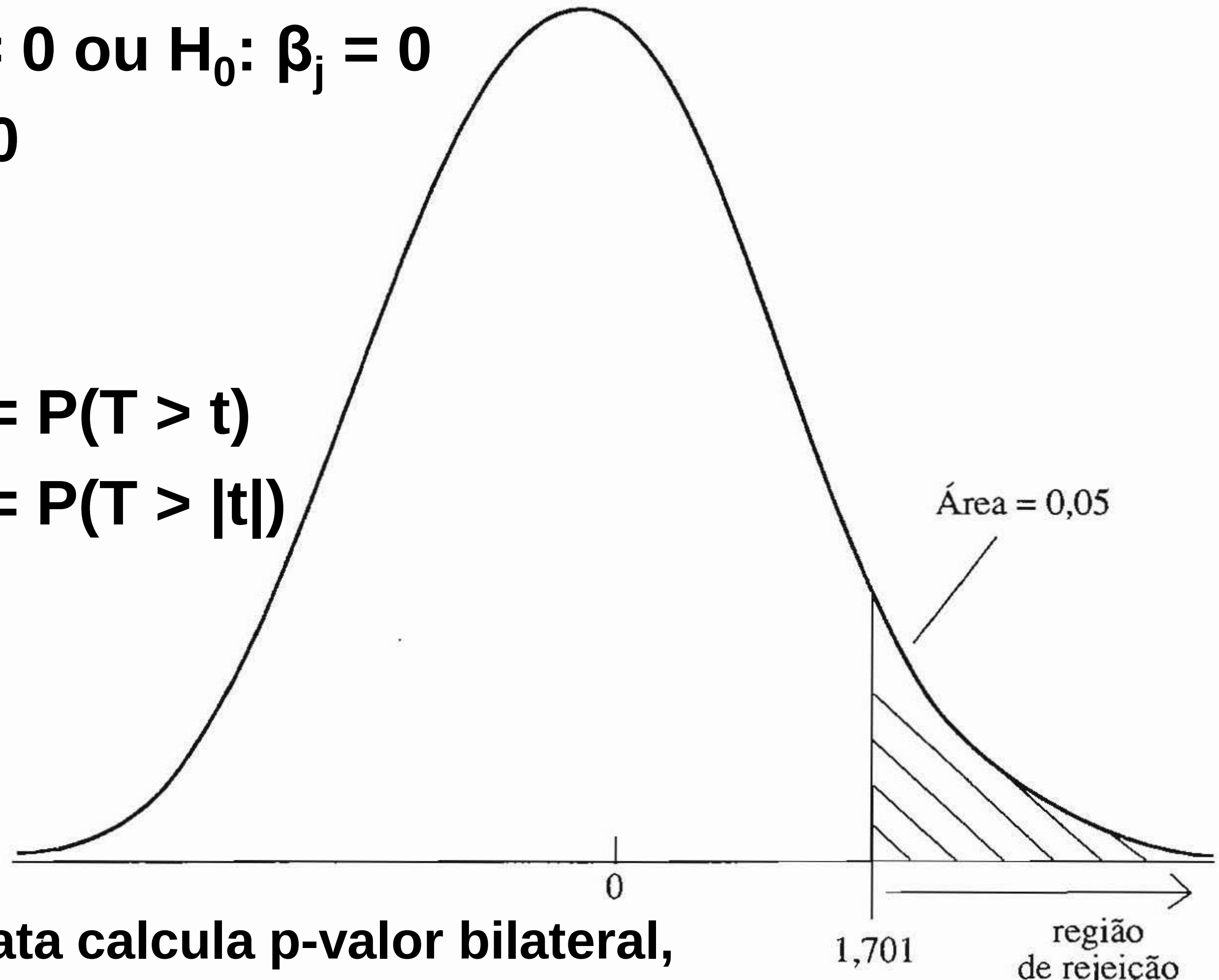
H_0: β_j <= 0 ou H_0: $\beta_j = 0$

H_1: $\beta_j > 0$

$t_{\beta j} > c$

p-valor = P(T > t)

p-valor = P(T > |t|)



Como Stata calcula p-valor bilateral, é só dividir por 2 para obter o p-valor unilateral.

Fonte: Wooldridge, 2008: 117.

REGRA DE REJEIÇÃO DE H_0 (UNILATERAL)

Figura 4.3

Regra de rejeição a 5% para a hipótese alternativa H_1: $(\beta_j) < 0$, com 18 *gl*.

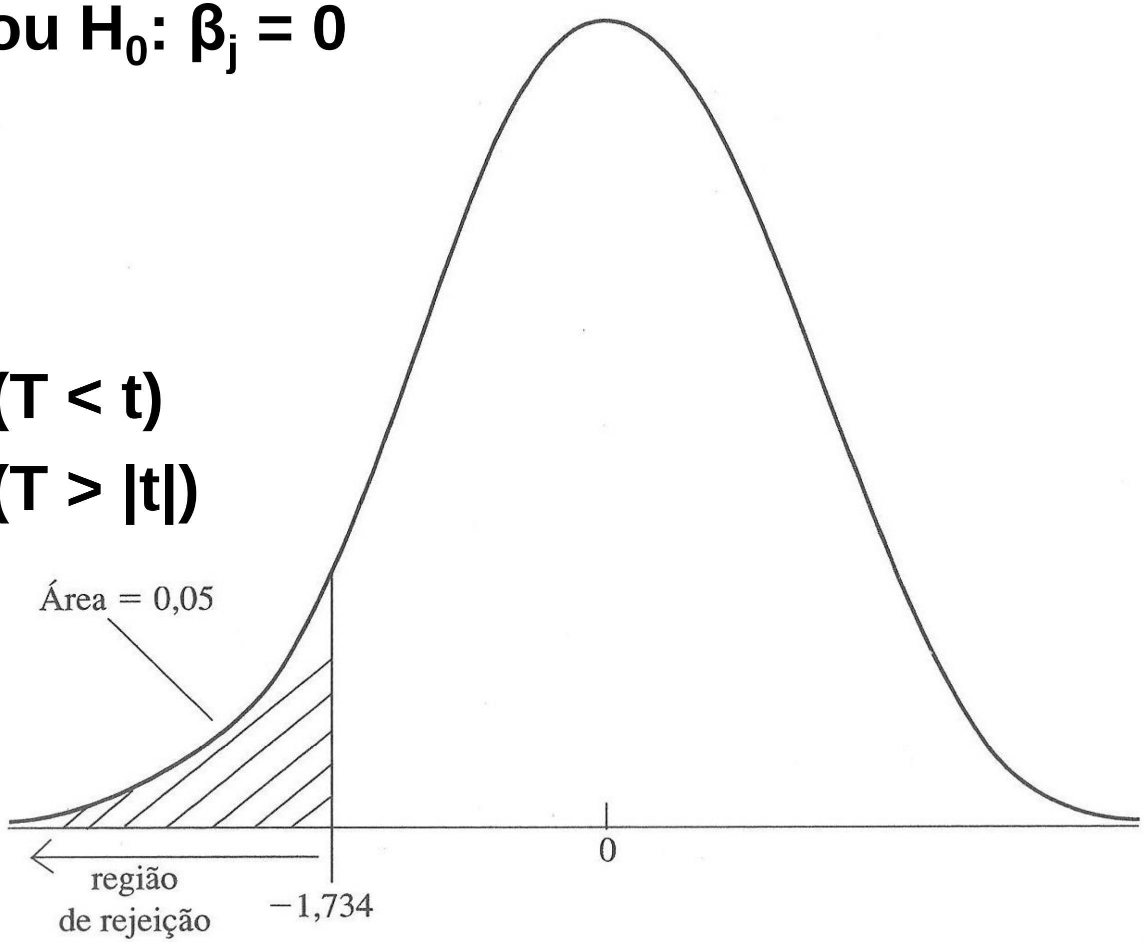
H_0: β_j >= 0 ou H_0: $\beta_j = 0$

H_1: $\beta_j < 0$

$t_{\beta j} < - c$

p-valor = P(T < t)

p-valor = P(T > |t|)



Como Stata calcula p-valor bilateral, é só dividir por 2 para obter o p-valor unilateral.

Fonte: Wooldridge, 2008: 119.

TESTE: HIPÓTESES ALTERNATIVAS BILATERAIS

$$\mathbf{H_1: \boldsymbol{\beta_j} \neq 0}$$

– Essa hipótese é relevante quando o sinal de β_j não é bem determinado pela teoria.

– Usar as estimativas da regressão para nos ajudar a formular as hipóteses nula e alternativa não é permitido, porque a inferência estatística clássica pressupõe que formulamos as hipóteses nula e alternativa sobre a população antes de olhar os dados.

– Quando a alternativa é bilateral, estamos interessados no valor absoluto da estatística t: $|t|>c$.

– Para um nível de significância de 5% e em um teste bi-caudal, c é escolhido de forma que a área em cada cauda da distribuição t seja igual a 2,5%.

– Se H_0 é rejeitada, x_j é estatisticamente significante (ou estatisticamente diferente de zero) ao nível de 5%.

REGRA DE REJEIÇÃO DE H_0 (BILATERAL)

Figura 4.4

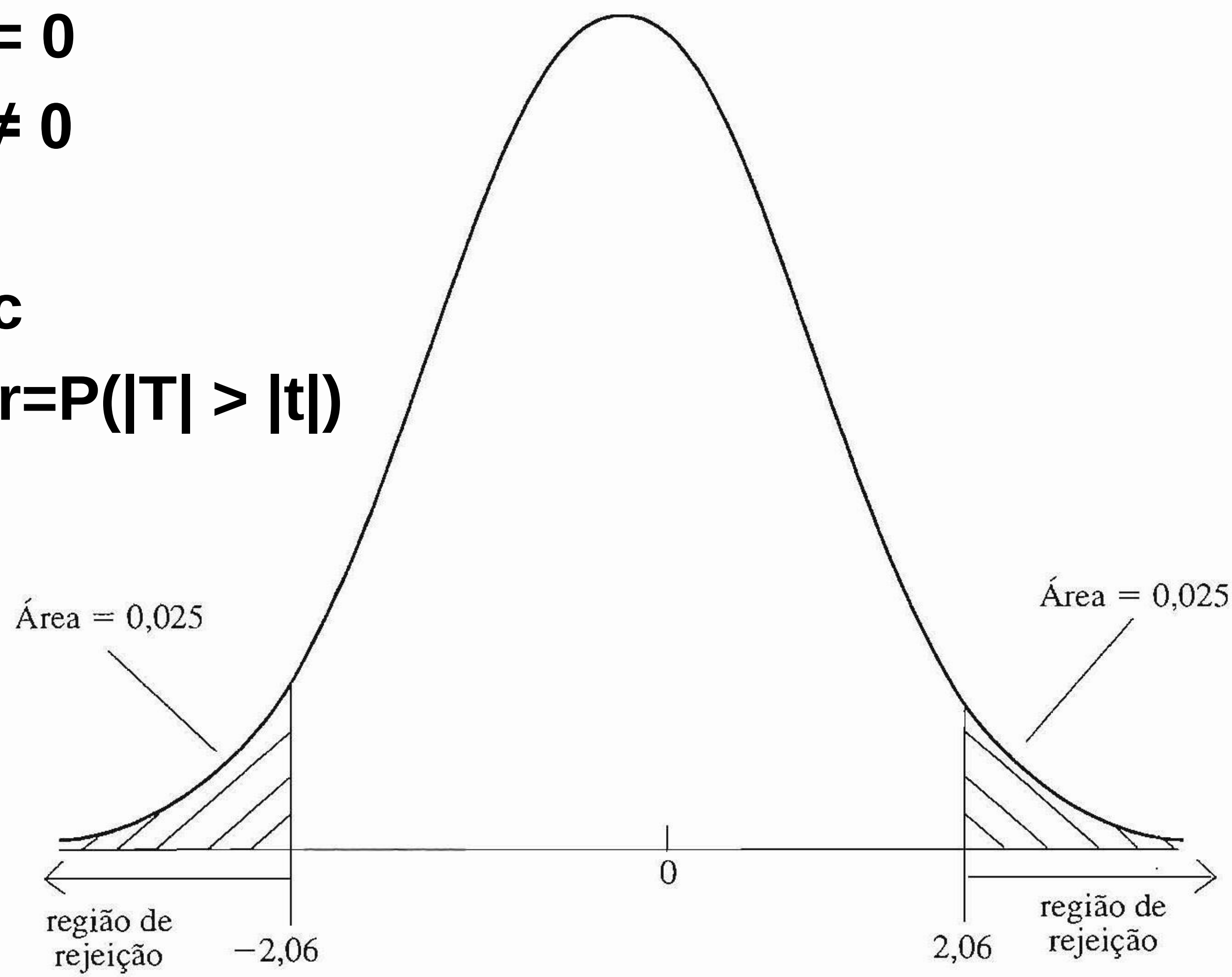
Regra de rejeição a 5% para a hipótese alternativa H_1: $\beta_j \neq 0$ com 25 *gl*.

H_0: $\beta_j = 0$

H_1: $\beta_j \neq 0$

$|t_{\beta j}| > c$

p-valor=P(|T| > |t|)



Fonte: Wooldridge, 2008: 122.

EXEMPLO DE NÃO-REJEIÇÃO DE H_0 (BILATERAL)

Figura 4.6

Obtendo o *p*-valor contra uma alternativa bilateral, quando t = 1,85 e gl = 40.

p-valor

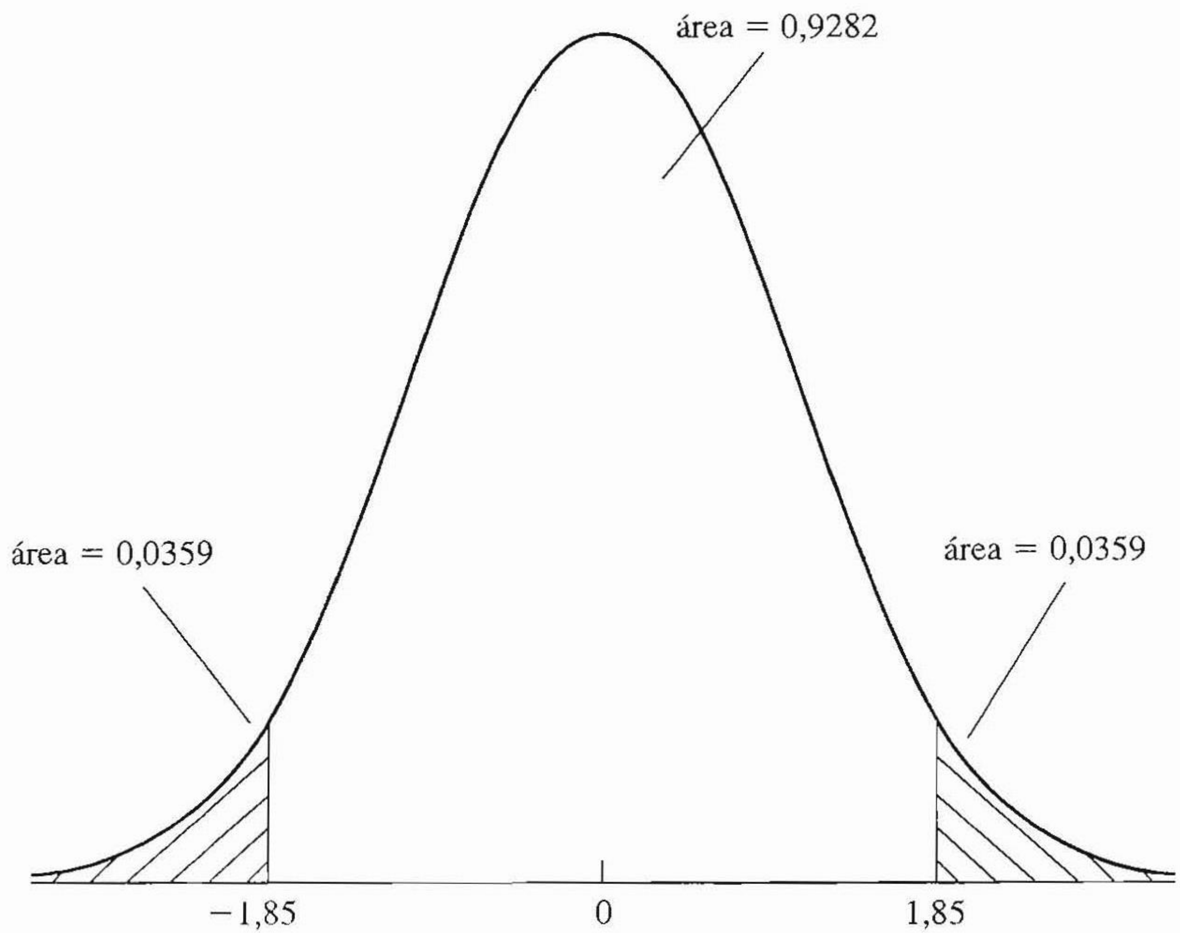
$= P(|T| > |t|)$

$= P(|T| > 1{,}85)$

$= 2P(T > 1{,}85)$

$= 2(0{,}0359)$

$= 0{,}0718$



p-valor > α

0,0718 > 0,05

$H_0 : \beta_j = 0$ não é rejeitada

Fonte: Wooldridge, 2008: 127.

TESTES DE OUTRAS HIPÓTESES SOBRE β_j

– Poderíamos supor que uma variável dependente (log do número de crimes) necessariamente será relacionada positivamente com uma variável independente (log do número de estudantes matriculados na universidade).

– A hipótese alternativa testará se o aumento de 1% nas matrículas aumentará o crime em mais de 1%:

$$H_0: \beta_j = 1$$

$$H_1: \beta_j > 1$$

– t = (estimativa - valor hipotético) / (erro-padrão)

– Neste exemplo, $t = (\beta_j - 1) / ep(\beta_j)$

– Observe que adicionar 1 na hipótese nula, significa subtrair 1 no teste *t*.

– Rejeitamos H_0 se $t > c$, em que c é o valor crítico unilateral.

DECISÃO SOBRE HIPÓTESES

Hipóteses	***p*** **<** ***α***	***p*** **>** ***α***
Hipótese nula (H_0)	Rejeita	Não rejeita
Hipótese alternativa (H_1)	Aceita	Não aceita

- ***p-valor***: é a probabilidade de não rejeitar a hipótese nula.
- Como Stata calcula *p-valor* bilateral, divida por 2 para obter *p-valor* unilateral.

Nível de significância (*α*)	**Nível de confiança (NC)**
0,10 (10%)	90%
0,05 (5%)	95%
0,01 (1%)	99%
0,001 (0,1%)	99,9%

ERROS TIPO I E TIPO II

– Ao testar H_0, chegamos a uma conclusão de rejeitá-la ou de deixar de rejeitá-la.

– Tais conclusões podem estar corretas ou erradas.

		Estado verdadeiro da natureza	
		A hipótese nula é verdadeira	A hipótese nula é falsa
Decisão	Decidimos rejeitar a hipótese nula.	**Erro tipo I** (rejeitar uma hipótese nula verdadeira) α	Decisão Correta
	Deixamos de rejeitar a hipótese nula	Decisão Correta	**Erro tipo II** (deixar de rejeitar uma hipótese nula falsa) β

– α: probabilidade de erro tipo I (probabilidade de rejeitar hipótese nula quando ela é verdadeira).

– β: probabilidade de erro tipo II (probabilidade de deixar de rejeitar hipótese nula quando ela é falsa).

INTERVALOS DE CONFIANÇA

– Os intervalos de confiança (IC), ou estimativas de intervalo, permitem avaliar uma extensão dos valores prováveis do parâmetro populacional, e não somente estimativa pontual:

  – Valor inferior: β_j - c*ep(β_j)

  – Valor superior: β_j + c*ep(β_j)

– A constante *c* é o 97,5º percentil de uma distribuição t_{n-k-1}.

– Quando n-k-1>120, podemos usar a distribuição normal para construir um IC de 95% (c=1,96).

– Se amostras aleatórias fossem repetidas, então valor populacional estaria dentro do IC em 95% das amostras.

– Esperamos ter uma amostra que seja uma das 95% de todas amostras em que estimativa de intervalo contém beta.

– Se a hipótese nula for $H_0:\beta_j=a_j$, H_0 é rejeitada contra $H_1:\beta_j\neq a_j$, ao nível de significância de 5%, se a_j não está no IC.

SIGNIFICÂNCIA ECONÔMICA X ESTATÍSTICA

– É importante levar em consideração a magnitude das estimativas dos coeficientes, além do tamanho das estatísticas *t*.

– A **significância estatística** de uma variável x_j é determinada completamente pelo tamanho do teste *t*.

– A **significância econômica** (ou significância prática) da variável está relacionada ao tamanho e sinal do coeficiente beta estimado.

– Colocar muita ênfase sobre a significância estatística pode levar à conclusão falsa de que uma variável é importante para explicar *y* embora seu efeito estimado seja moderado.

– Com amostras grandes, os erros-padrão são pequenos, o que resulta em significância estatística.

– Erros-padrão grandes podem ocorrer por alta correlação entre variáveis independentes (multicolinearidade).

DISCUTINDO AS SIGNIFICÂNCIAS

– Verifique a **significância econômica**, lembrando que as unidades das variáveis independentes e dependente mudam a interpretação dos coeficientes beta.

– Verifique a **significância estatística**, a partir do teste *t* de cada variável.

– Se: (1) sinal esperado e (2) teste *t* grande, a variável é significante economicamente e estatisticamente.

– Se: (1) sinal esperado e (2) teste *t* pequeno, podemos aceitar *p*-valor maior, quando amostra é pequena (mas é arriscado, pois pode ser problema no desenho amostral).

– Se: (1) sinal não esperado e (2) teste *t* pequeno, variável não significante economicamente e estatisticamente.

– Se: (1) sinal não esperado e (2) teste *t* grande, é problema sério em variáveis importantes (falta incluir variáveis ou há problema nos dados).

DESCRIÇÃO DOS RESULTADOS DA REGRESSÃO

– Informar os **coeficientes** estimados de MQO (betas).

– Interpretar **significância econômica** (prática) dos coeficientes da variáveis fundamentais, levando em consideração as unidades de medida.

– Interpretar **significância estatística**, ao incluir erros-padrão entre parênteses abaixo dos coeficientes (ou estatísticas *t*, ou *p*-valores, ou asteriscos).

  – Erro padrão é preferível, pois podemos: (1) testar hipótese nula quando parâmetro populacional não é zero; (2) calcular intervalos de confiança.

– Informar o **R-quadrado**: (1) grau de ajuste; (2) cálculo de F.

– **Número de observações** usado na estimação (n).

– Apresentar resultados em **equações** ou **tabelas** (indicar variável dependente, além de independentes na 1ª coluna).

– Mostrar **SQR** e **erro-padrão** (Root MRE), mas não é crucial.

EXEMPLOS COM PNAD DE MINAS GERAIS DE 2007

– O banco de dados de pessoas possui informações de:

  – Anos de escolaridade (anest).

  – Idade (idpia).

  – Idade ao quadrado (idquad).

  – Raça preta/parda (negra), em comparação com branca.

  – Sexo (mulher).

  – Rendimento no trabalho principal (renpri).

  – Logaritmo do rendimento no trabalho principal (lnrenpri).

EXEMPLO 1: PNAD DE MINAS GERAIS DE 2007

– Variável dependente: rendimento em reais

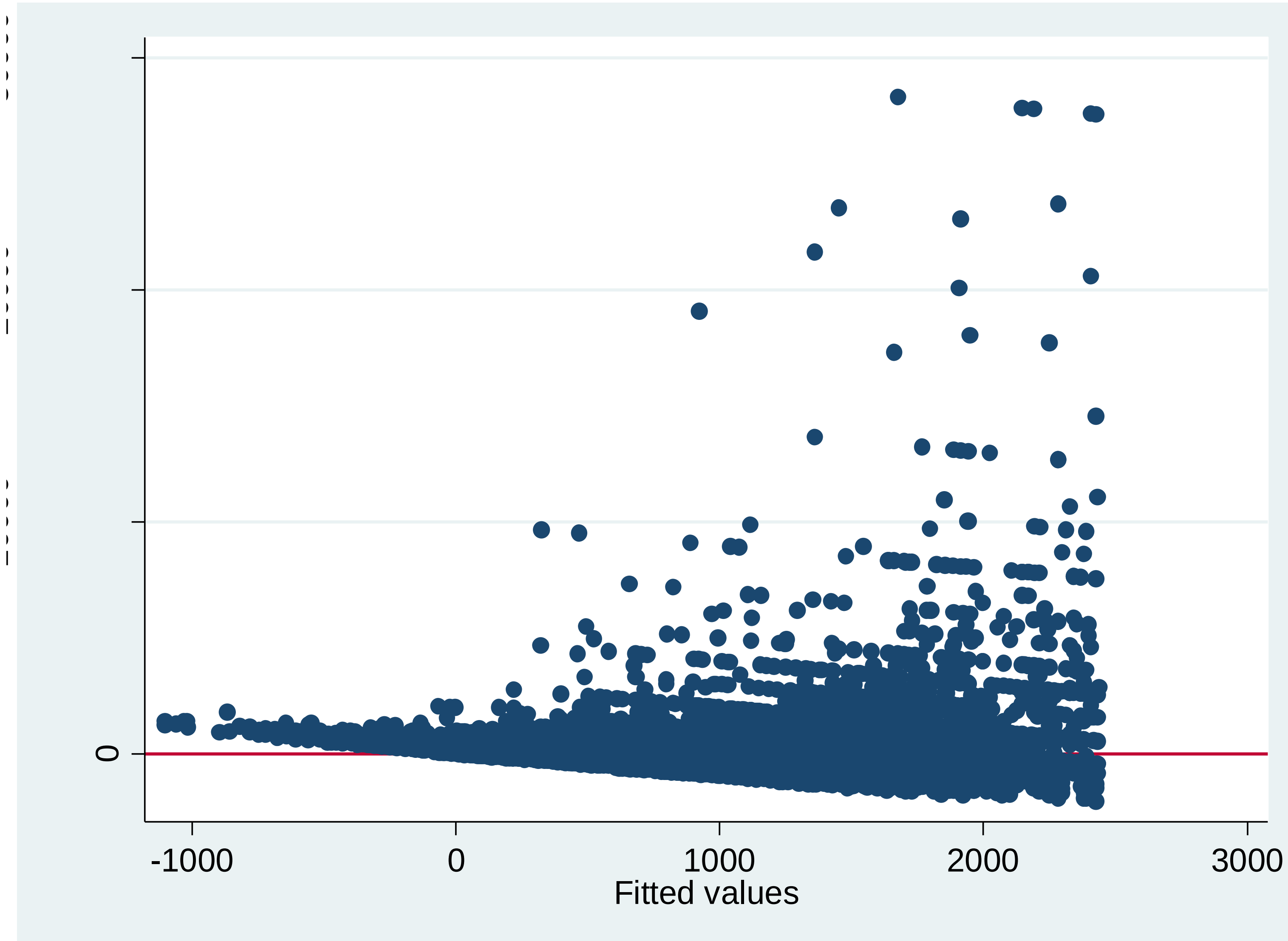
```
. reg renpri idpia idquad anest negra mulher [aweight=v4729]
(sum of wgt is   8,4198e+06)

      Source |       SS       df       MS              Number of obs =   15620
-------------+------------------------------           F(  5, 15614) =  747,58
       Model |  4,8305e+09     5   966097989           Prob > F      =  0,0000
    Residual |  2,0178e+10 15614  1292306,36           R-squared     =  0,1932
-------------+------------------------------           Adj R-squared =  0,1929
       Total |  2,5009e+10 15619  1601162,78           Root MSE      =  1136,8

------------------------------------------------------------------------------
      renpri |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       idpia |   56,91032   4,607963    12,35   0,000     47,87817    65,94246
      idquad |   -,402666   ,0601803    -6,69   0,000    -,5206263   -,2847056
       anest |   117,3971   2,375815    49,41   0,000     112,7402    122,0539
       negra |  -176,1501   18,78247    -9,38   0,000     -212,966   -139,3343
      mulher |  -461,3267   18,80184   -24,54   0,000    -498,1805    -424,473
       _cons |  -1315,827   86,21179   -15,26   0,000    -1484,812   -1146,841
------------------------------------------------------------------------------
```

EXEMPLO 1: PNAD DE MINAS GERAIS DE 2007

– Resíduos por rendimento predito em reais:



EXEMPLO 2: PNAD DE MINAS GERAIS DE 2007

– Variável dependente: logaritmo do rendimento

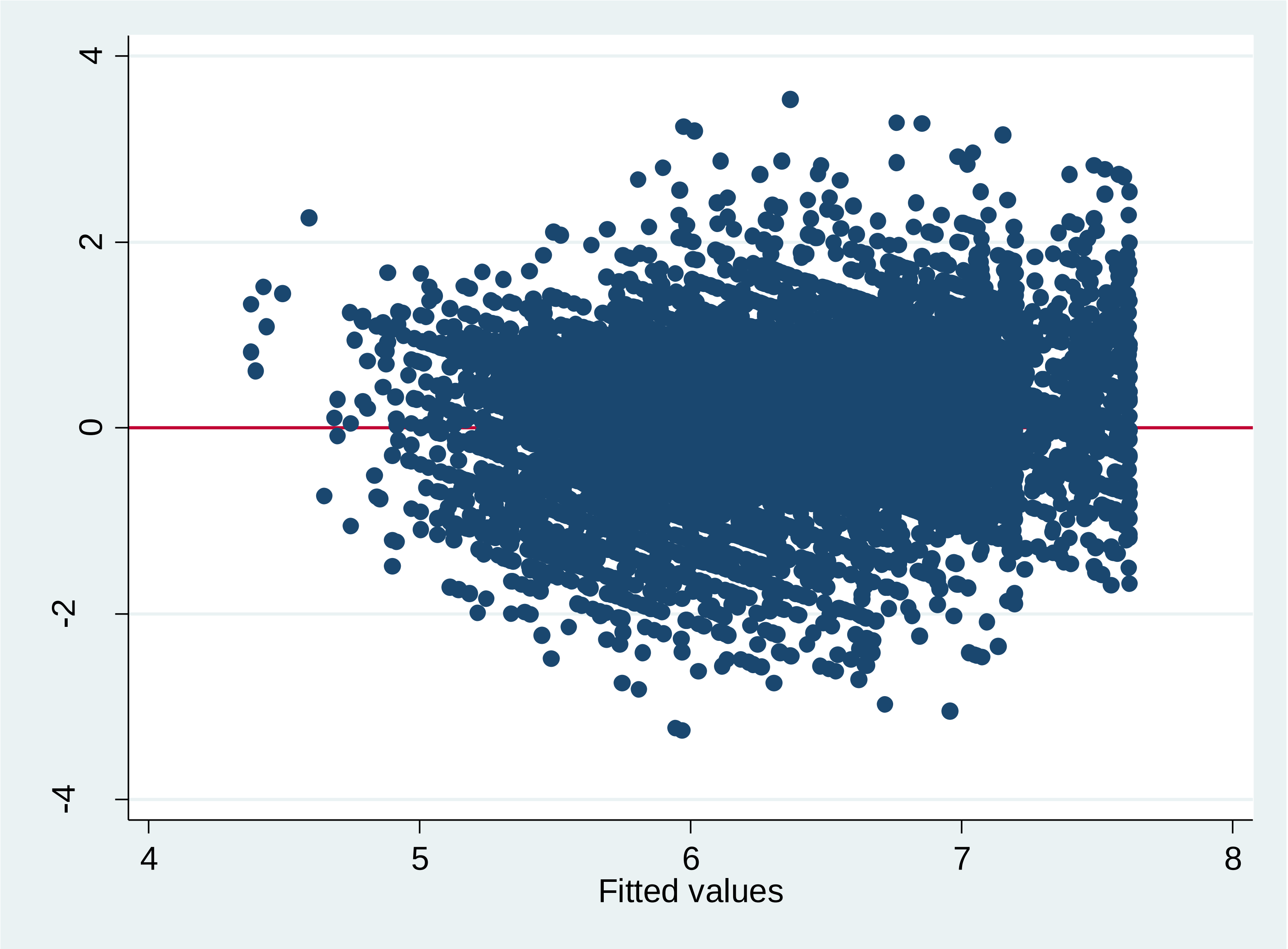
```
. reg lnrenpri idpia idquad anest negra mulher [aweight=v4729]
(sum of wgt is   8,4198e+06)

      Source |       SS       df       MS              Number of obs =   15620
-------------+------------------------------           F(  5, 15614) = 1987,41
       Model |  4332,2922     5  866,458439            Prob > F      =  0,0000
    Residual | 6807,29008 15614   ,43597349            R-squared     =  0,3889
-------------+------------------------------           Adj R-squared =  0,3887
       Total | 11139,5823 15619  ,713207137            Root MSE      =  ,66028

------------------------------------------------------------------------------
    lnrenpri |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       idpia |    ,088968   ,0026764    33,24   0,000     ,0837219    ,0942141
      idquad |  -,0008933    ,000035   -25,56   0,000    -,0009618   -,0008248
       anest |   ,1067622   ,0013799    77,37   0,000     ,1040573     ,109467
       negra |  -,1368042   ,0109094   -12,54   0,000    -,1581878   -,1154205
      mulher |  -,5440937   ,0109206   -49,82   0,000    -,5654994    -,522688
       _cons |   3,805854   ,0500742    76,00   0,000     3,707703    3,904005
------------------------------------------------------------------------------
```

EXEMPLO 2: PNAD DE MINAS GERAIS DE 2007

– Resíduos por logaritmo do rendimento predito:



Coeficientes estimados por modelos de mínimos quadrados ordinários para explicação do logaritmo do rendimento no trabalho principal (variável dependente), Minas Gerais, 2007.

Variáveis independentes	**Modelo 1**	**Modelo 2**	**Modelo 3**	**Modelo 4**
Constante	4,5830***	3,6660***	3,7810***	3,8060***
	(0,0590)	(0,0532)	(0,0539)	(0,0501)
Idade	0,0858***	0,0831***	0,0832***	0,0890***
	(0,0033)	(0,0029)	(0,0029)	(0,0027)
Idade ao quadrado	-0,0010***	-0,0008***	-0,0008***	-0,0009***
	(4,31e-05)	(3,78e-05)	(3,76e-05)	(3,50e-05)
Anos de escolaridade		0,0996***	0,0956***	0,1070***
		(0,0014)	(0,0015)	(0,0014)
Cor/raça				
Branca			ref.	ref.
Negra (preta e parda)			-0,1360***	-0,1370***
			(0,0117)	(0,0109)
Sexo				
Homem				ref.
Mulher				-0,5440***
				(0,0109)
R^2	0,0643	0,2860	0,2920	0,3890
R^2 ajustado	0,0640	0,2850	0,2920	0,3890
Observações	15.620	15.620	15.620	15.620

Obs.: Erros padrão em parênteses.
* Significativo ao nível de confiança de 90%; ** Significativo ao nível de confiança de 95%; *** Significativo ao nível de confiança de 99%.
Fonte: Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2007.

TESTE *F*: TESTE DE RESTRIÇÕES DE EXCLUSÃO

– Testar se um grupo de variáveis não tem efeito sobre a variável dependente.

– A hipótese nula é que um conjunto de variáveis não tem efeito sobre y (β_3, β_4 e β_5, por exemplo), já que outro conjunto de variáveis foi controlado (β_1 e β_2, por exemplo).

– Esse é um exemplo de restrições múltiplas.

– H_0: $\beta_3=0$, $\beta_4=0$, $\beta_5=0$.

– H_1: H_0 não é verdadeira.

– Quando pelo menos um dos betas for diferente de zero, rejeitamos a hipótese nula.

ESTATÍSTICA *F* (OU RAZÃO *F*)

– Precisamos saber o quanto SQR aumenta, quando retiramos as variáveis que estamos testando.

– Modelo restrito terá β_0, β_1 e β_2.

– Modelo irrestrito terá β_0, β_1, β_2, β_3, β_4 e β_5.

– A estatística F é definida como:

$$F \equiv \frac{(SQR_r - SQR_{ir})/q}{SQR_{ir}/(n-k-1)}$$

– SQRr é a soma dos resíduos quadrados do modelo restrito.

– SQRir é a soma dos resíduos quadrados do modelo irrestrito.

– *q* é o número de variáveis independentes retiradas (neste caso temos três: β_3, β_4 e β_5), ou seja, $q=gl_r-gl_{ir}$.

REGRAS DE REJEIÇÃO DE *F*

– O valor crítico (*c*) depende de:

   – Nível de significância (10%, 5% ou 1%, por exemplo).

   – Graus de liberdade do numerador ($q=gl_r-gl_{ir}$).

   – Graus de liberdade do denominador (n-k-1).

   – Quando os gl do denominador chegam a 120, a distribuição *F* não é mais sensível a eles (usar gl=∞).

– Uma vez obtido *c*, rejeitamos H_0, em favor de H_1, ao nível de significância escolhido se: $F > c$.

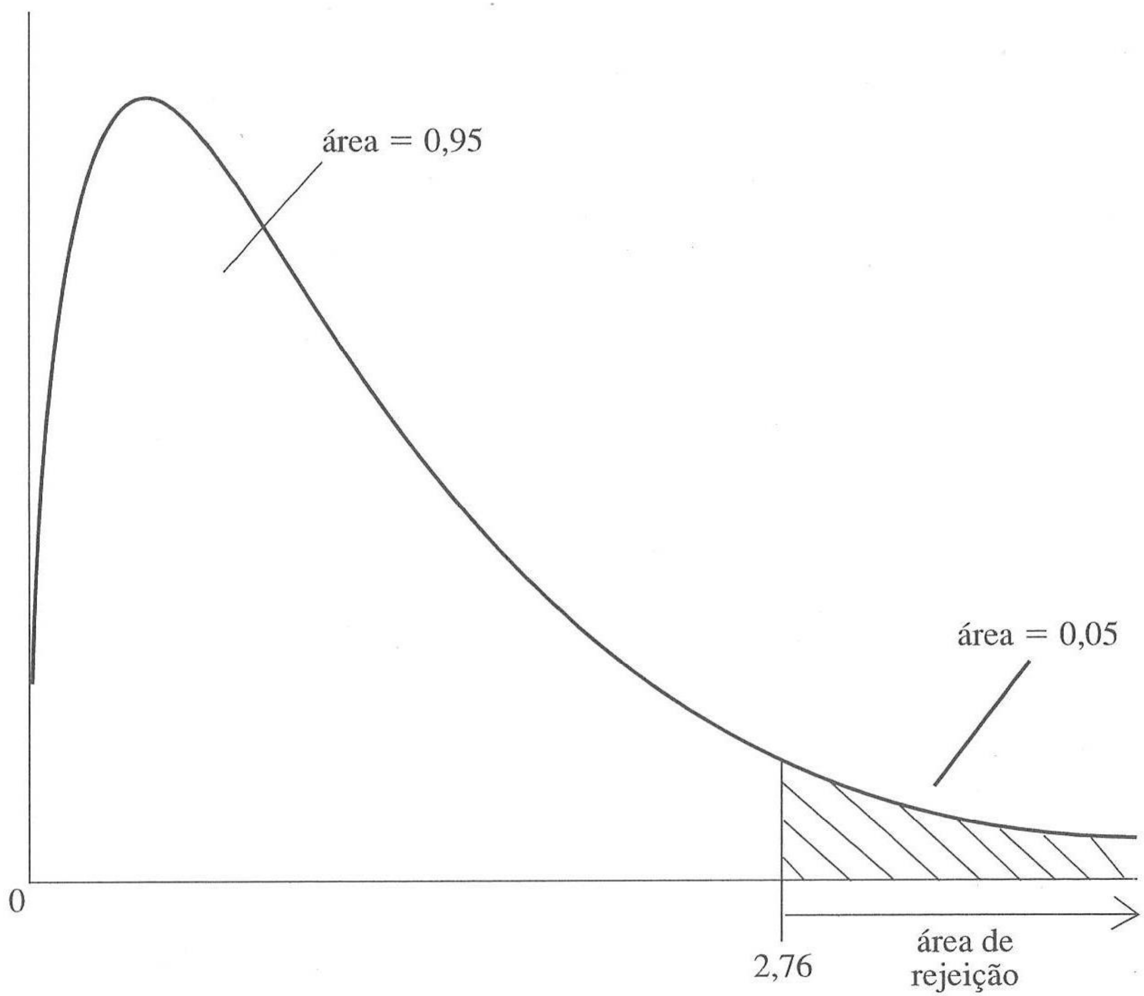
– Se H_0 ($\beta_3=0, \beta_4=0, \beta_5=0$) é rejeitada, β_3, β_4 e β_5 são **estatisticamente significantes conjuntamente**.

– Se H_0 ($\beta_3=0, \beta_4=0, \beta_5=0$) não é rejeitada, β_3, β_4 e β_5 são **conjuntamente não significantes**.

CURVA DA DISTRIBUIÇÃO *F*

Figura 4.7

O valor crítico de 5% e a região de rejeição em uma distribuição $F_{3,60}$.



Fonte: Wooldridge, 2008: 142.

RELAÇÃO ENTRE ESTATÍSTICAS *F* E *t*

– A estatística *F* para testar a exclusão de uma única variável é igual ao quadrado da estatística *t* correspondente.

– As duas abordagens levam ao mesmo resultado, desde que a hipótese alternativa seja bilateral.

– A estatística *t* é mais flexível para testar uma única hipótese, porque pode ser usada para testar alternativas unilaterais.

– As estatísticas *t* são mais fáceis de serem obtidas do que o teste *F*.

FORMA R-QUADRADO DA ESTATÍSTICA *F*

– O teste *F* pode ser calculado usando os R-quadrados dos modelos resitrito e irrestrito.

– É mais fácil utilizar números entre zero e um (R^2) do que números que podem ser muito grandes (SQR).

– Como $SQR_r = SQT(1 - R_r^2)$, $SQR_{ir} = SQT(1 - R_{ir}^2)$ e:

$$F \equiv \frac{(SQR_r - SQR_{ir})/q}{SQR_{ir}/(n-k-1)}$$

– ... os termos SQT são cancelados:

$$F \equiv \frac{\left(R_{ir}^2 - R_r^2\right)/q}{\left(1 - R_{ir}^2\right)/(n-k-1)}$$

CÁLCULO DOS *p*-VALORES PARA TESTES *F*

$$p\text{-valor} = P(\mathcal{F} > F)$$

– O *p*-valor é a probabilidade de observarmos um valor de F pelo menos tão grande ($\mathcal{F}$) quanto aquele valor real que encontramos (F), dado que a hipótese nula é verdadeira.

– **Um *p*-valor pequeno é evidência para rejeitar H_0**, porque a probabilidade de observarmos um valor de F tão grande quanto aquele para o qual a hipótese nula é verdadeira é muito baixa.

– **Um *p*-valor alto é evidência para NÃO rejeitar H_0**, porque a probabilidade de observarmos um valor de F tão grande quanto aquele para o qual a hipótese nula é verdadeira é muito alta.

TESTE *F* PARA SIGNIFICÂNCIA GERAL DA REGRESSÃO

– No modelo com k variáveis independentes, podemos escrever a hipótese nula como:

– H_0: $x_1, x_2, ..., x_k$ não ajudam a explicar y.

– H_0: $\beta_1 = \beta_2 = ... = \beta_k = 0$.

– Modelo restrito: $y = \beta_0 + u$.

– Modelo irrestrito: $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$.

– Número de variáveis independentes retiradas (q = graus de liberdade do numerador) é igual ao próprio número de variáveis independentes (k):

$$F \equiv \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

– Mesmo com R^2 pequeno, podemos ter teste F significante para o conjunto, por isso não podemos olhar somente o R^2.

EXEMPLOS DE VALORES PREDITOS EM GRÁFICOS

IMPACTO ECONÔMICO DA RELIGIÃO

– Unidade de análise: quatro grupos de idade (15-24, 25-34, 35-49, 50-64) e três grupos de escolaridade (0-4, 5-8, 9+) geram doze grupos de idade-escolaridade.

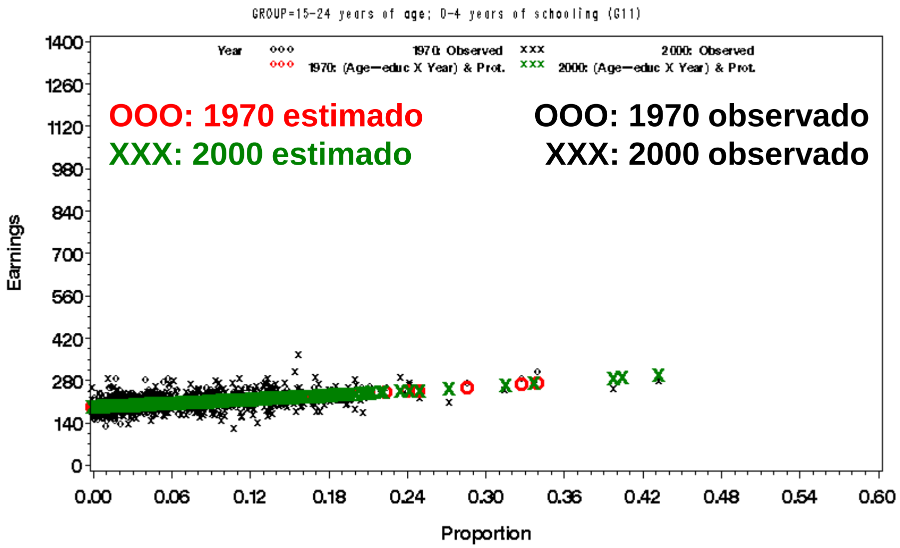
– Dados: informações para 502 microrregiões e quatro anos censitários (1970, 1980, 1991, 2000).

– Variável dependente: logaritmo da renda média do grupo de idade-escolaridade em cada microrregião e ano.

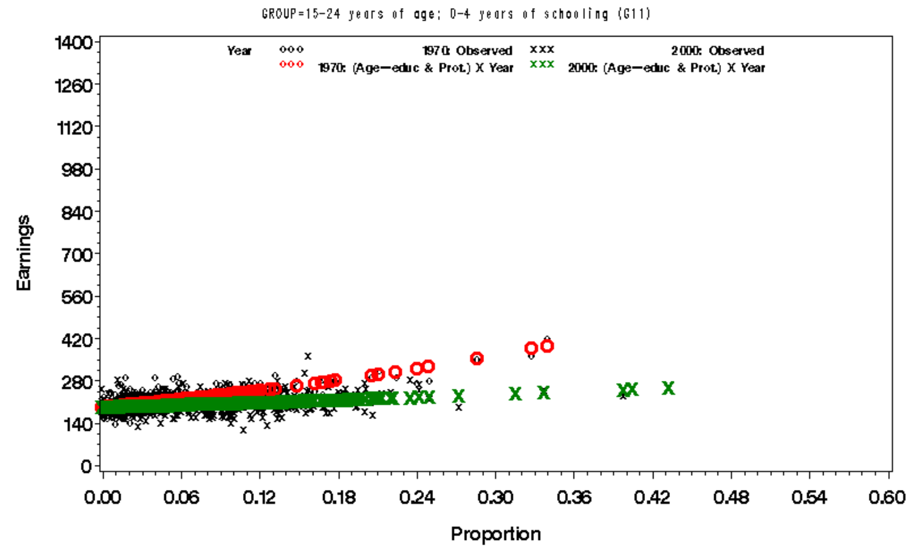
– Variáveis independentes: variáveis dicotômicas dos grupos de idade-escolaridade, proporção de protestantes em cada grupo de idade-escolaridade, efeitos fixos de microrregião e ano censitário.

IDADE 15-24 / ESCOLARIDADE 0-4

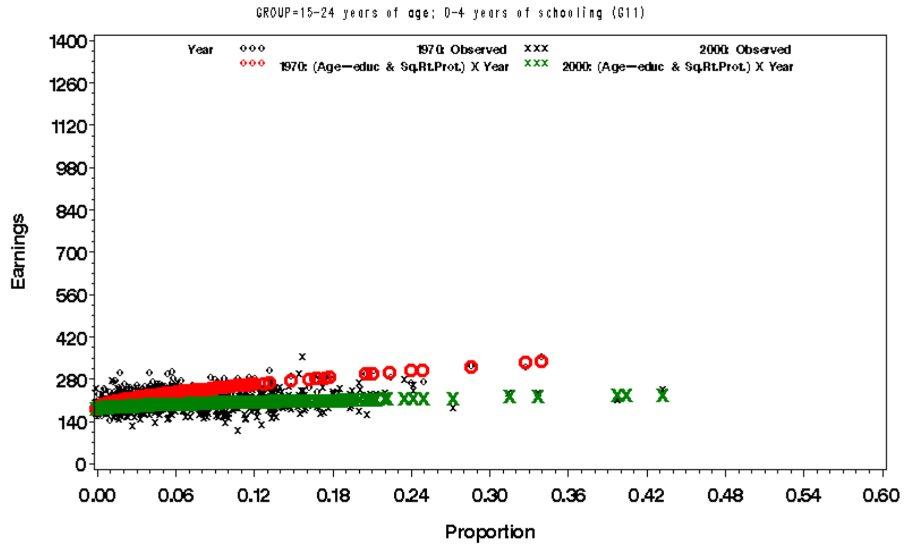
Prop. protestantes



Prop. protestantes * Ano

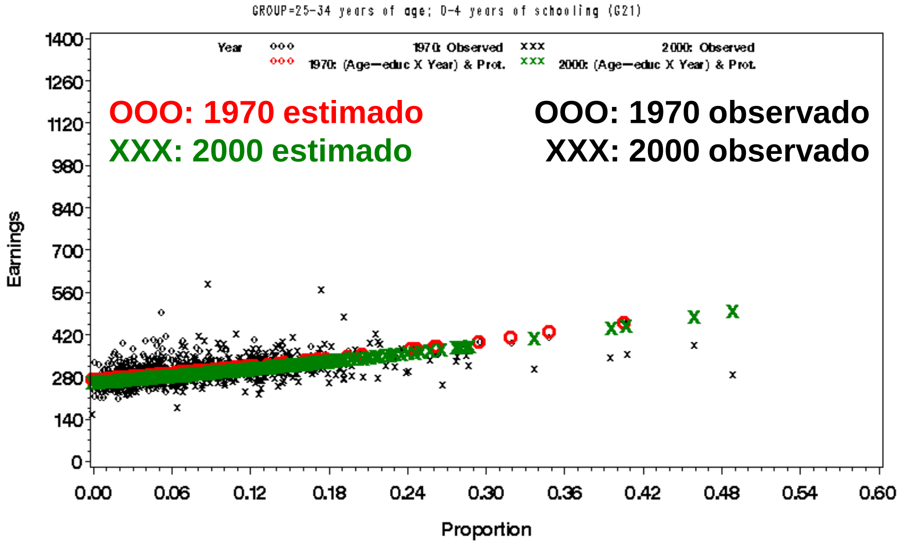


Raiz quadrada(Prop. protestantes) * Ano

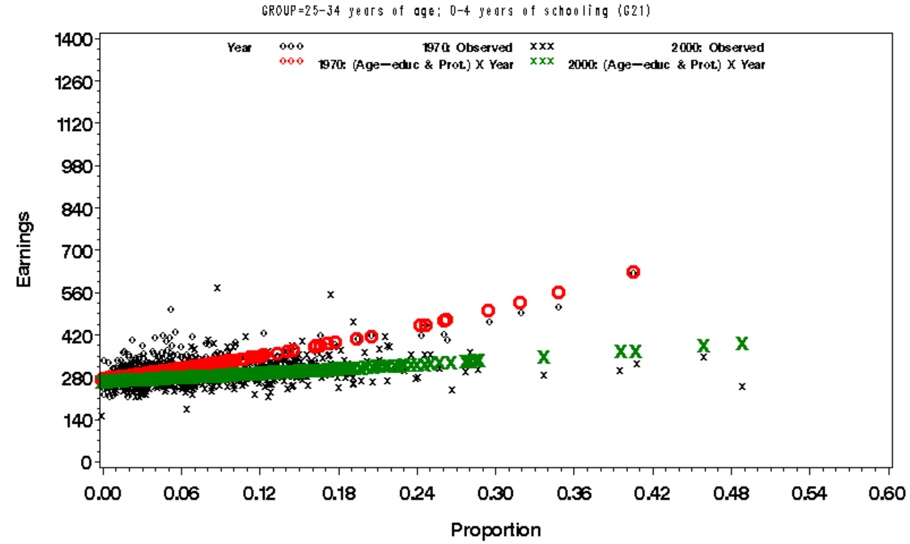


IDADE 25-34 / ESCOLARIDADE 0-4

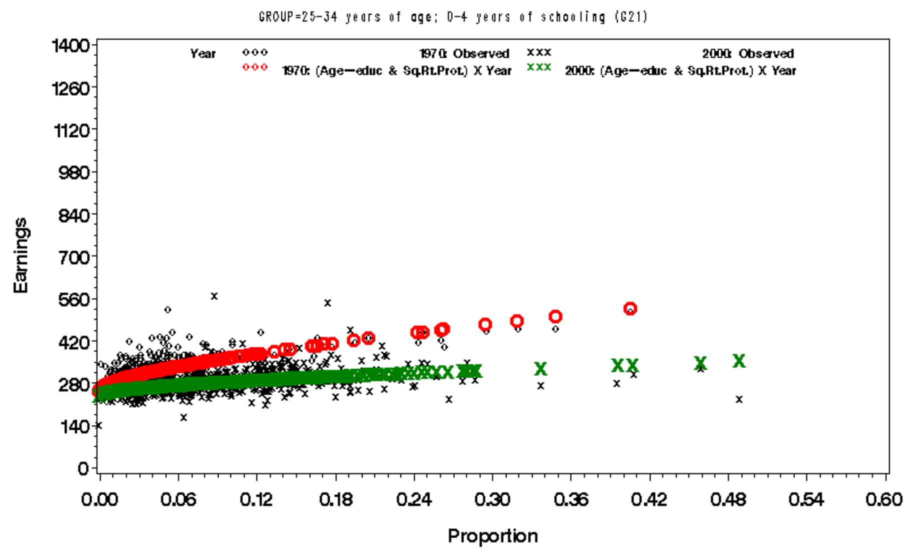
Prop. protestantes



Prop. protestantes * Ano

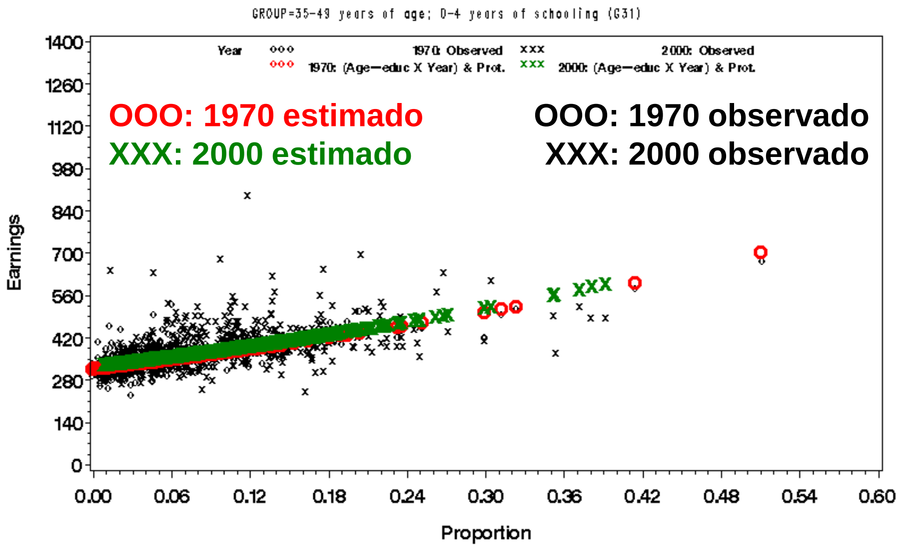


Raiz quadrada(Prop. protestantes) * Ano

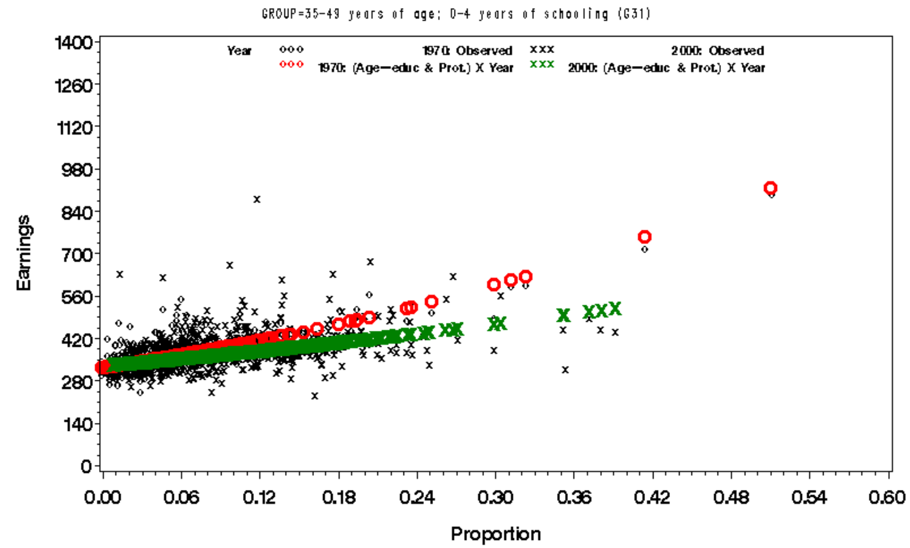


IDADE 35-49 / ESCOLARIDADE 0-4

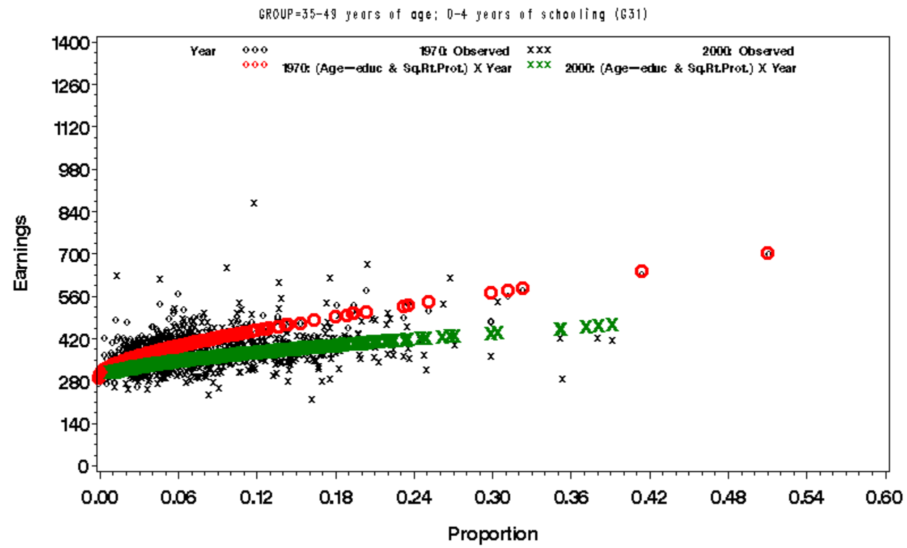
Prop. protestantes



Prop. protestantes * Ano

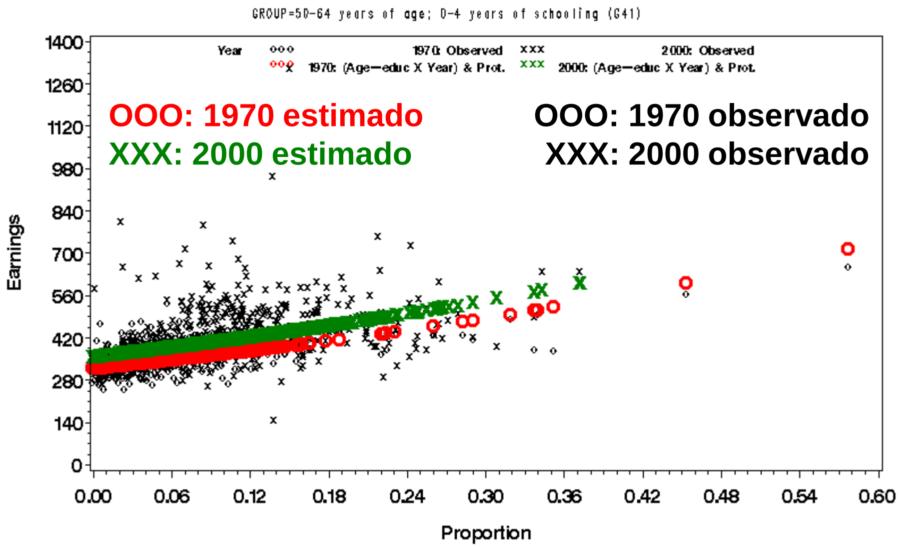


Raiz quadrada(Prop. protestantes) * Ano

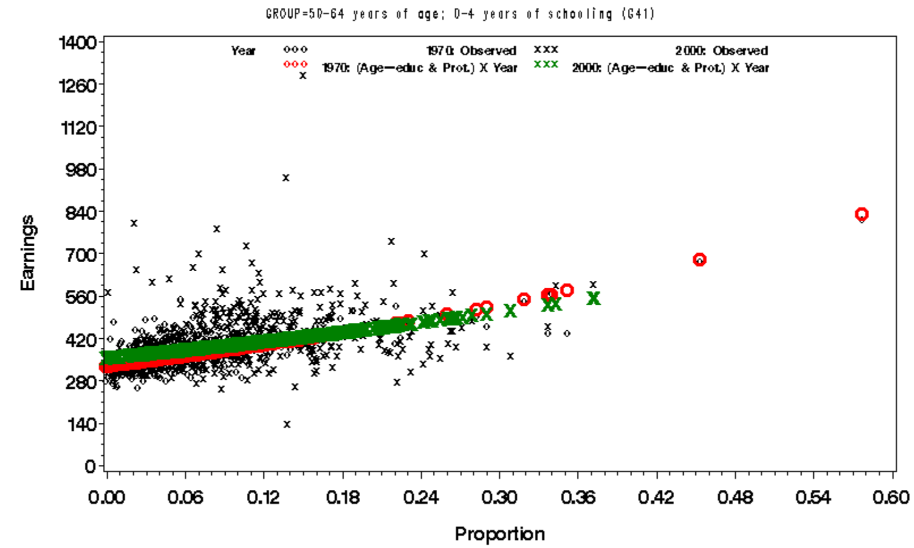


IDADE 50-64 / ESCOLARIDADE 0-4

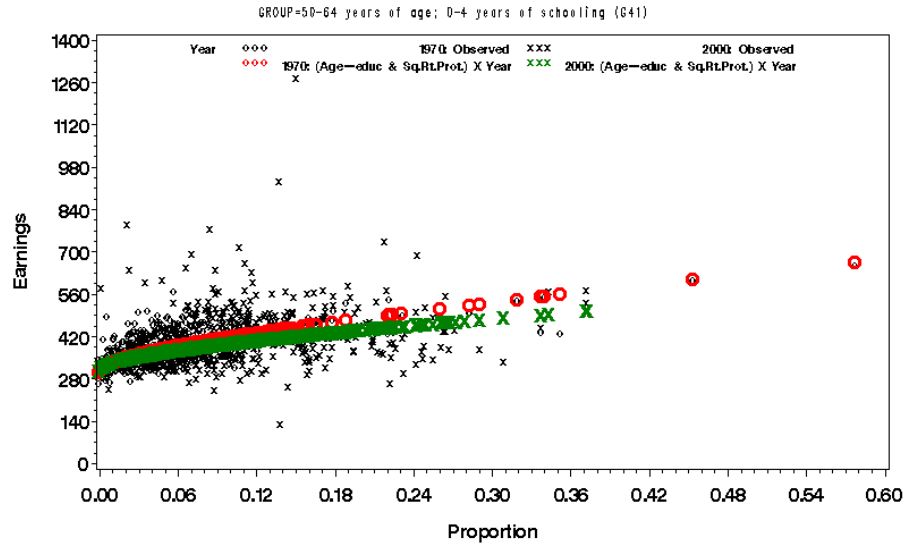
Prop. protestantes



Prop. protestantes * Ano



Raiz quadrada(Prop. protestantes) * Ano



DIFERENCIAIS DE FECUNDIDADE POR ESCOLARIDADE

– Unidade de análise: mulheres de 10 a 45 anos em quatro Estados brasileiros (Piauí, Pernambuco, Espírito Santo, Rio Grande do Sul) e quatro Estados mexicanos (Guerrero, Veracruz, Nuevo León, Tamaulipas).

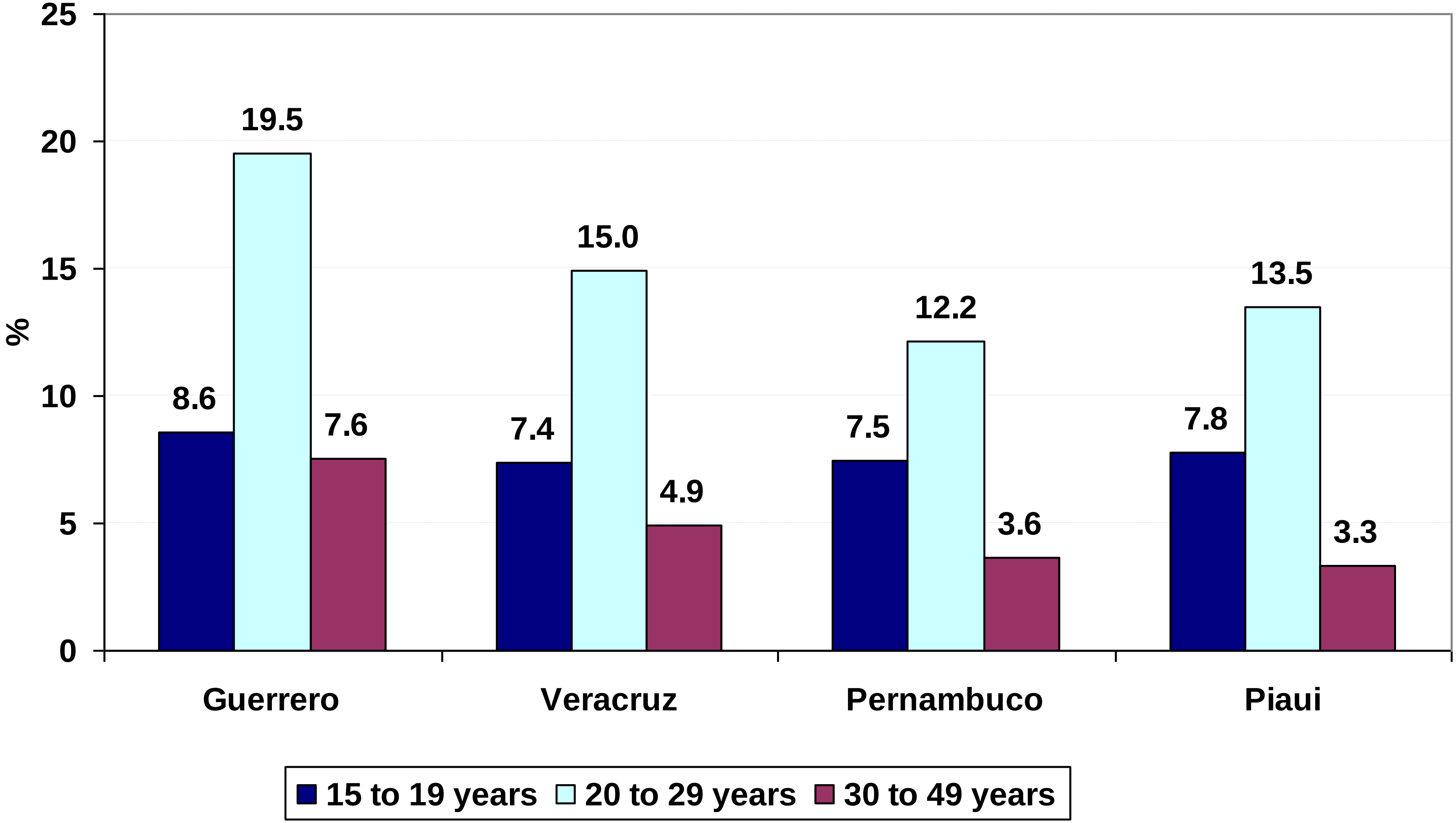
– Dados: censos demográficos de 2000 dos dois países.

– Variável dependente: informação se teve filho nascido vivo no último ano (variável binária).

– Variáveis independentes: idade, idade ao quadrado, grupos de escolaridade (0-2, 3-6, 7-9, 10+), origem indígena e características do domicílio.

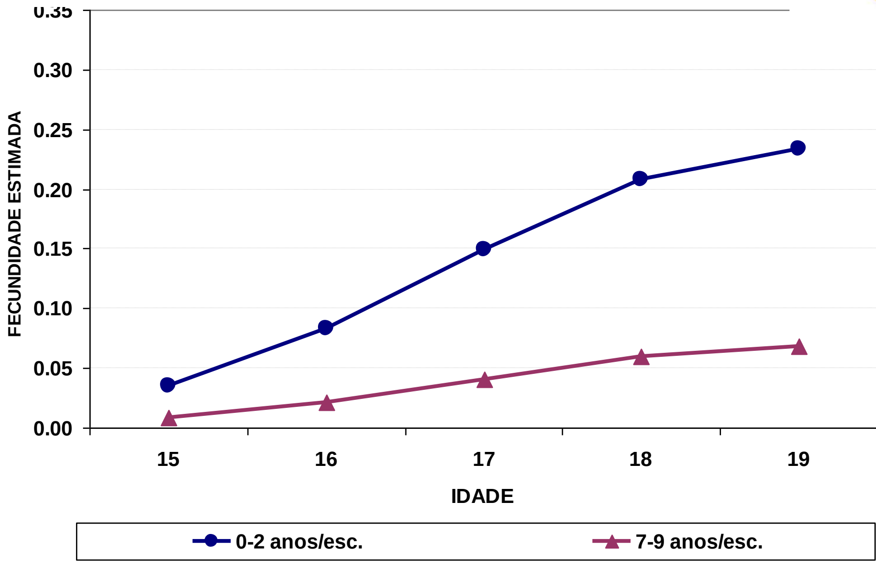
– Modelo logístico para três grupos de idade (10-19, 20-29, 30-49) e para cada Estado de residência.

MULHERES COM FILHO NASCIDO VIDO NO ÚLTIMO ANO, MÉXICO E BRASIL - 2000

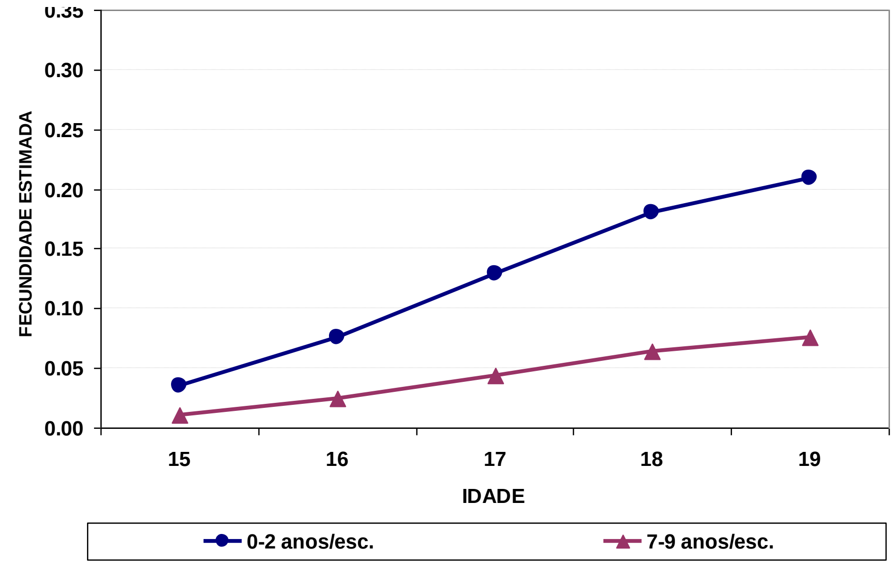


MULHERES DE 15-19 ANOS

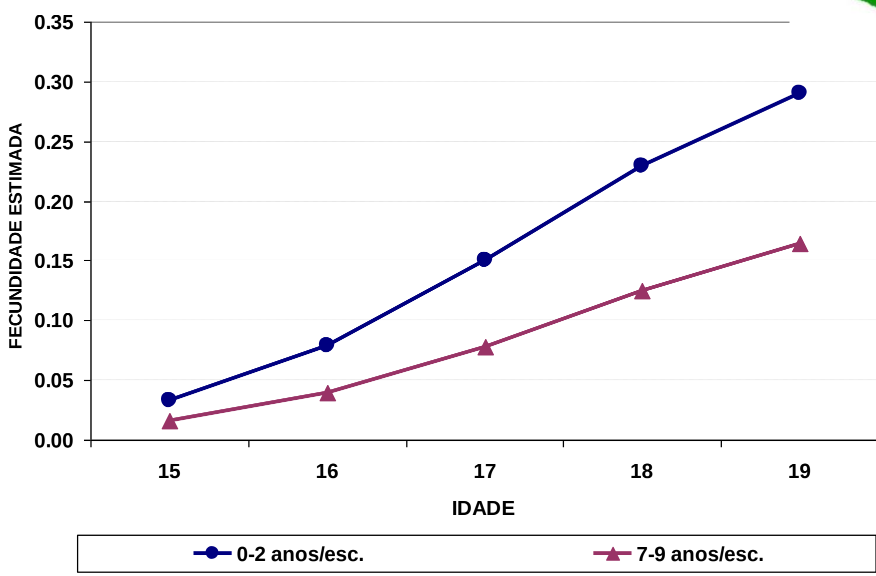
PIAUÍ - BRASIL



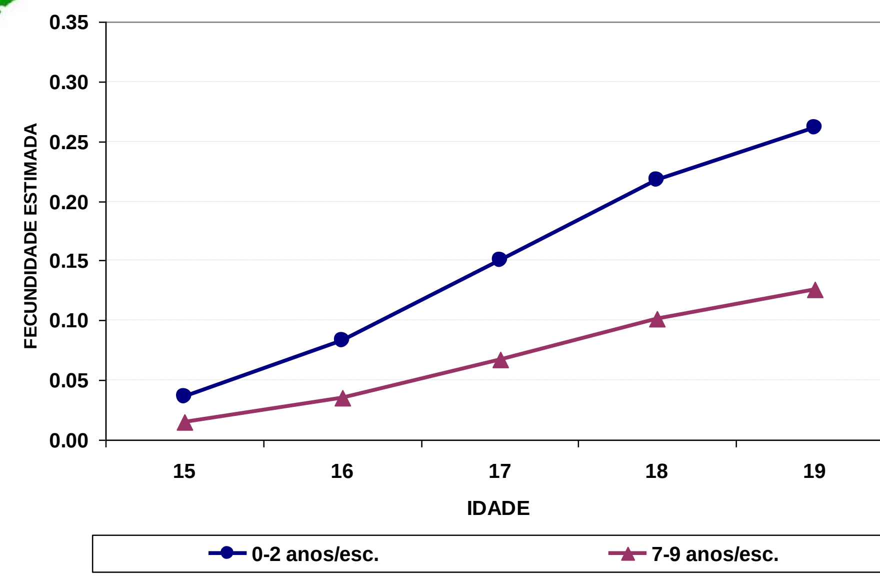
PERNAMBUCO - BRASIL



GUERRERO - MÉXICO



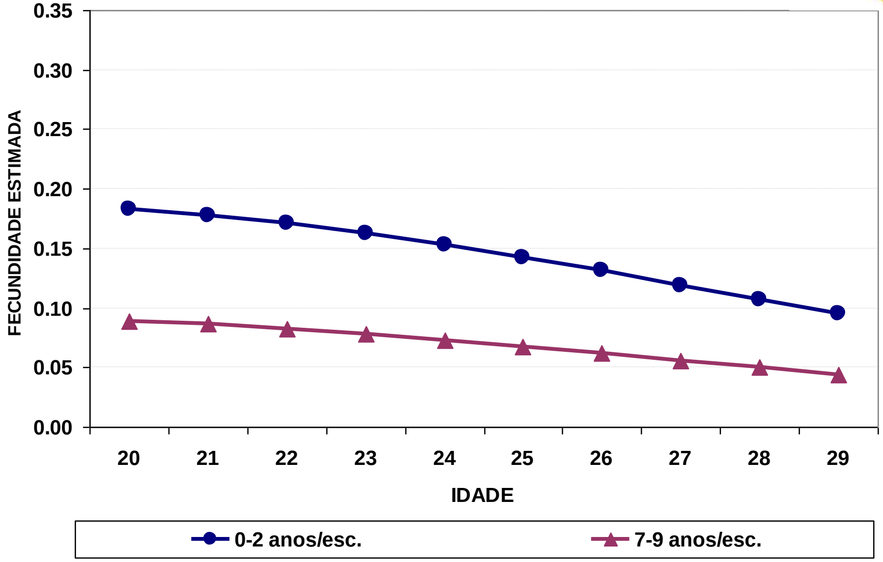
VERACRUZ - MÉXICO



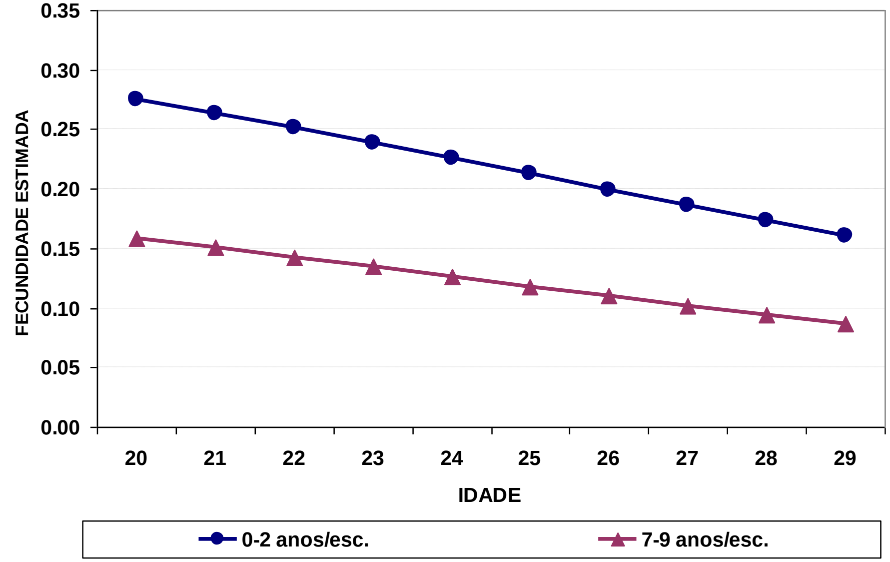
MULHERES DE 20-29 ANOS

MULHERES COM 3 FILHOS OU MAIS

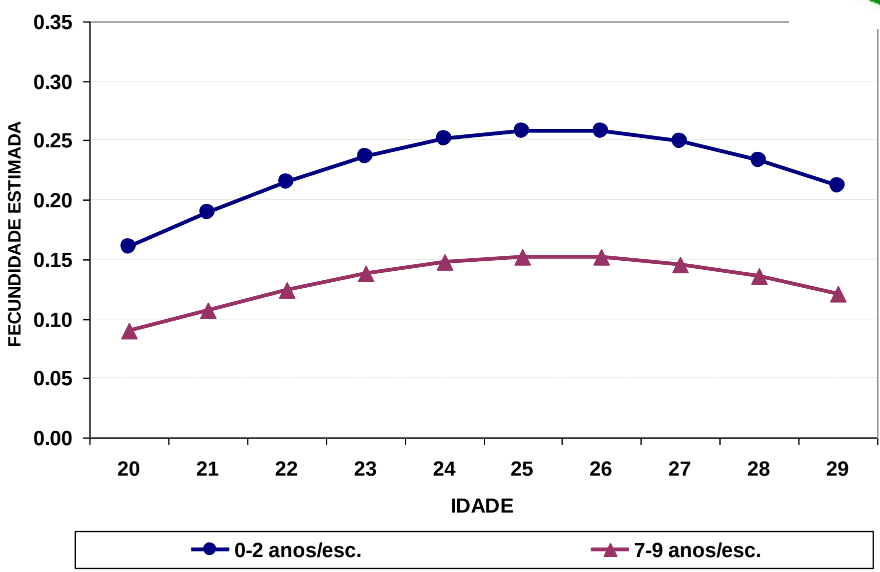
PIAUÍ - BRASIL



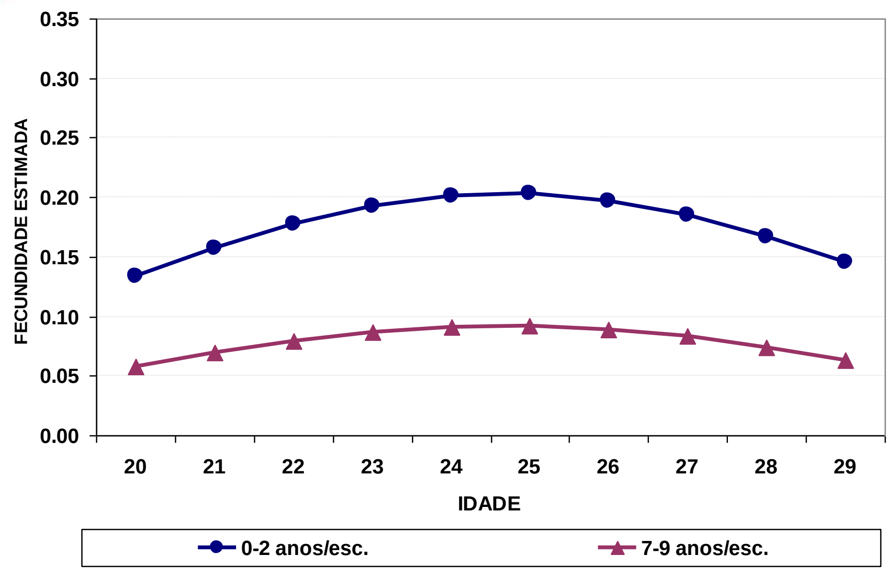
PERNAMBUCO - BRASIL



GUERRERO - MÉXICO

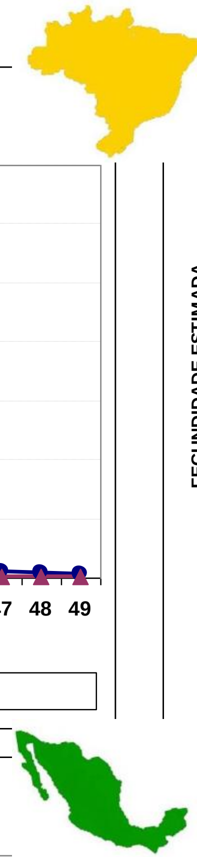


VERACRUZ - MÉXICO

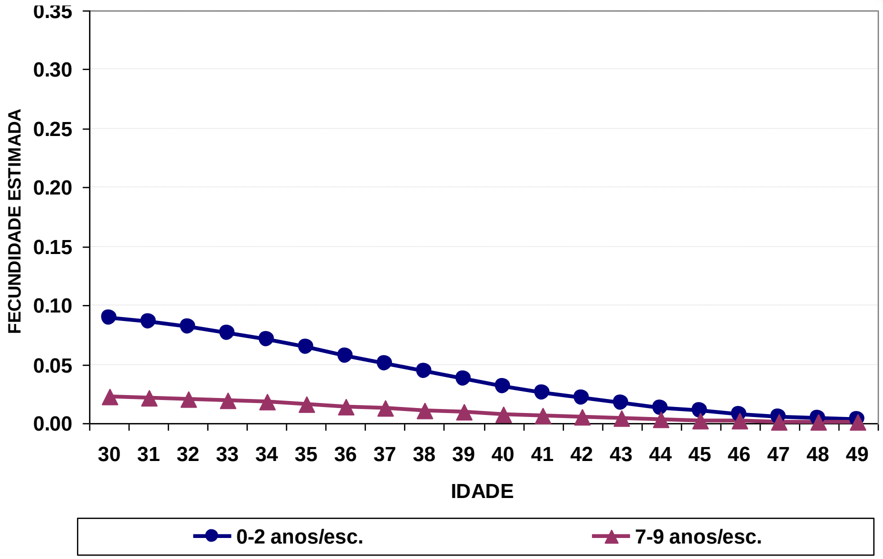


MULHERES DE 30-49 ANOS

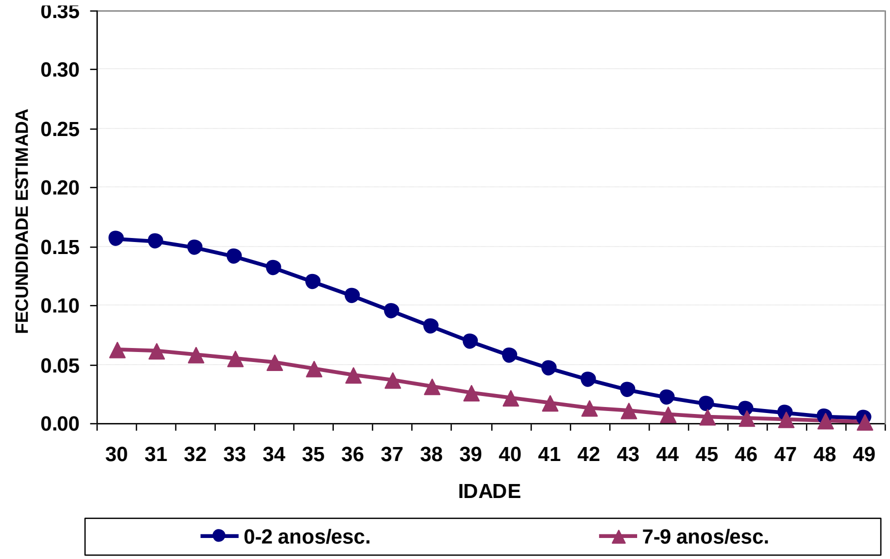
MULHERES COM 3 FILHOS OU MAIS



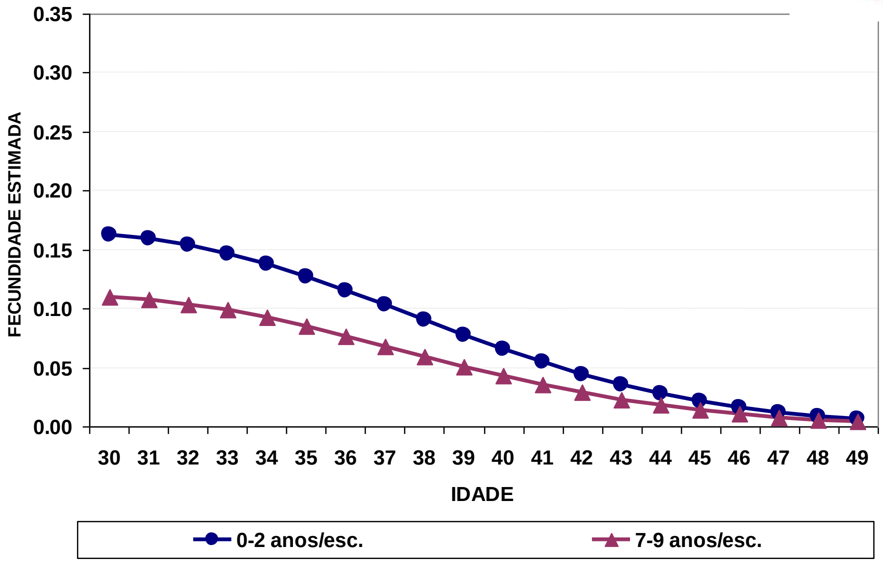
PIAUÍ - BRASIL



PERNAMBUCO - BRASIL



GUERRERO - MÉXICO



VERACRUZ - MÉXICO