

ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

(AULA 06)

Novembro e dezembro de 2018

Reinaldo Soares de Camargo

Testes de Hipóteses para Vários Parâmetros ao Mesmo Tempo

Testes de Hipóteses para Vários Parâmetros

- Considere o teste de hipótese, com hipóteses nulas e alternativas conforme abaixo:

$$H_0: \gamma_1 = a$$

$$H_A: \gamma_1 \neq a$$

- Nesse caso, estamos testando o valor para apenas um parâmetro do modelo de regressão
- Em geral, é possível testar hipóteses correspondentes a vários parâmetros simultaneamente
- Por exemplo, considere novamente o nosso modelo de regressão com variáveis *dummies* para os efeitos da macrorregiões:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \\ + \delta_1 D_{SU} + \delta_2 D_{NO} + \delta_3 D_{SE} + \delta_4 D_{NE} + \epsilon_i$$

- Gostaríamos agora de testar se todas as *dummies* para regiões são nulas. Isso equivale a dizermos que as diferenças para y_i entre as regiões são explicadas totalmente pelas demais variáveis explicativas na regressão

Testes de Hipóteses para Vários Parâmetros

- Nesse caso, as hipóteses nulas e alternativas podem ser escritas conforme abaixo:

$$H_0: \delta_1 = 0, \delta_2 = 0, \delta_3 = 0, \delta_4 = 0$$

H_A : pelos menos um dos coeficientes testados é diferente de zero

- A estatística teste é dada por:

$$F = \frac{(R_{\text{irrestrito}}^2 - R_{\text{restrito}}^2) \times (n - k - 1)}{(1 - R_{\text{irrestrito}}^2) \times m}$$

- $R_{\text{irrestrito}}^2$ é o coeficiente de determinação da regressão irrestrita (incluindo as *dummies*)
- R_{restrito}^2 é o coeficiente de determinação da regressão restrita (excluindo as *dummies*)
- n é o número de observações na amostra
- k é o número de variáveis explicativas da regressão irrestrita (incluindo as *dummies*)
- m é o número de restrições testadas; no exemplo, é o número de coeficientes das *dummies*

Testes de Hipóteses para Vários Parâmetros

- Implementação no R (programa ajuste_6.R)
- Equação do modelo “irrestrito” (com as *dummies* de regiões)

```
mod2.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita  
+ dados3$indice_gini  
+ dados3$salario_medio_mensal  
+ dados3$perc_crianças_extrem_pobres  
+ dados3$perc_crianças_pobres  
+ dados3$perc_pessoas_dom_agua_estogo_inadequados  
+ dados3$perc_pessoas_dom_paredes_inadequadas  
+ dados3$perc_pop_dom_com_coleta_lixo  
+ dados3$perc_pop_rural  
+ as.factor(dados3$Regiao))  
summary(mod2.ex)
```

Testes de Hipóteses para Vários Parâmetros

- Equação do modelo “restrito” (sem as *dummies* de regiões)

```
mod2.ex.rest <- lm(dados3$mort_infantil ~ dados3$renda_per_capita  
  + dados3$indice_gini  
  + dados3$salario_medio_mensal  
  + dados3$perc_crianças_extrem_pobres  
  + dados3$perc_crianças_pobres  
  + dados3$perc_pessoas_dom_agua_estogo_inadequados  
  + dados3$perc_pessoas_dom_paredes_inadequadas  
  + dados3$perc_pop_dom_com_coleta_lixo  
  + dados3$perc_pop_rural)  
summary(mod2.ex.rest)
```

- Testando a exclusão das variáveis *dummy*:

```
anova(mod2.ex.rest, mod2.ex, test='LRT')
```

Testes de Hipóteses para Vários Parâmetros

- Resultados:

Analysis of Variance Table

Model 1: dados3\$mort_infantil ~ dados3\$renda_per_capita + dados3\$indice_gini + dados3\$salario_medio_mensal + dados3\$perc_crianças_extrem_pobres + dados3\$perc_crianças_pobres + dados3\$perc_pessoas_dom_agua_estogo_inadequados + dados3\$perc_pessoas_dom_paredes_inadequadas + dados3\$perc_pop_dom_com_coleta_lixo + dados3\$perc_pop_rural

Model 2: dados3\$mort_infantil ~ dados3\$renda_per_capita + dados3\$indice_gini + dados3\$salario_medio_mensal + dados3\$perc_crianças_extrem_pobres + dados3\$perc_crianças_pobres + dados3\$perc_pessoas_dom_agua_estogo_inadequados + dados3\$perc_pessoas_dom_paredes_inadequadas + dados3\$perc_pop_dom_com_coleta_lixo + dados3\$perc_pop_rural + as.factor(dados3\$Regiao)

Res.Df RSS Df Sum of Sq Pr(>Chi)

1 5554 86683

2 5550 67741 4 18942 < 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Qual a conclusão a partir dos resultados do teste de hipótese? Os coeficientes das dummies de regiões são significativos conjuntamente ou não?

Testes de Hipóteses para Vários Parâmetros

- Implementação no R (alternativamente):
- Equação: (com função “linearHypothesis” do pacote “car”)

```
mod2.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita
+ dados3$indice_gini
+ dados3$salario_medio_mensal
+ dados3$perc_crianças_extrem_pobres
+ dados3$perc_crianças_pobres
+ dados3$perc_pessoas_dom_agua_estogo_inadequados
+ dados3$perc_pessoas_dom_paredes_inadequadas
+ dados3$perc_pop_dom_com_coleta_lixo
+ dados3$perc_pop_rural
+ as.factor(dados3$Regiao))
```

```
summary(mod2.ex)
```

```
linearHypothesis(mod2.ex, c("dados3$indice_gini = 0",
                             "dados3$salario_medio_mensal = 0",
                             "dados3$perc_pop_rural"))
```


Testes de Hipóteses para Vários Parâmetros

- Implementação no R (alternativamente):

Resultados:

Linear hypothesis test

Hypothesis:

$\text{dados3\$indice_gini} = 0$

$\text{dados3\$salario_medio_mensal} = 0$

$\text{dados3\$perc_pop_rural} = 0$

Model 1: restricted model

Model 2: $\text{dados3\$mort_infantil} \sim \text{dados3\$renda_per_capita} + \text{dados3\$indice_gini} +$
 $\text{dados3\$salario_medio_mensal} + \text{dados3\$perc_criancas_extrem_pobres} +$
 $\text{dados3\$perc_criancas_pobres} + \text{dados3\$perc_pessoas_dom_agua_estogo_inadequados} +$
 $\text{dados3\$perc_pessoas_dom_paredes_inadequadas} + \text{dados3\$perc_pop_dom_com_coleta_lixo} +$
 $\text{dados3\$perc_pop_rural} + \text{as.factor(dados3\$Regiao)}$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	5553	68663				
2	5550	67741	3	922.43	25.191	3.498e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Testes de Hipóteses para Vários Parâmetros

- Implementação no R:

Quando usar o “linearHypothesis” versus quando usar o “ANOVA”?

- Na prática, ambos os testes, quando usados para testar as mesmas hipóteses, implicam na mesma resposta. As contas são exatamente as mesmas.
- No entanto, quando a hipótese nula testada corresponder à retirada de um conjunto de termos (por exemplo, várias *dummies*) ao mesmo tempo, o comando ANOVA pode ser o mais fácil de utilizar.
 - Note que, para utilizar o comando ANOVA, precisamos rodar os dois modelos (com e sem as variáveis dos coeficientes testados), e depois aplicar o comando para comparar os dois modelos. O modelo chamado “restrito” corresponde a um caso particular do modelo chamado “irrestrito”.
- Quando a hipótese nula testada corresponder a valores específicos de alguns dos coeficientes da equação (por exemplo, $\text{coeficiente1} = 0$ e $\text{coeficiente2} = 2.2$), então o comando “linearHypothesis” pode ser o mais fácil de ser implementado.

Testes de Hipóteses para Vários Parâmetros

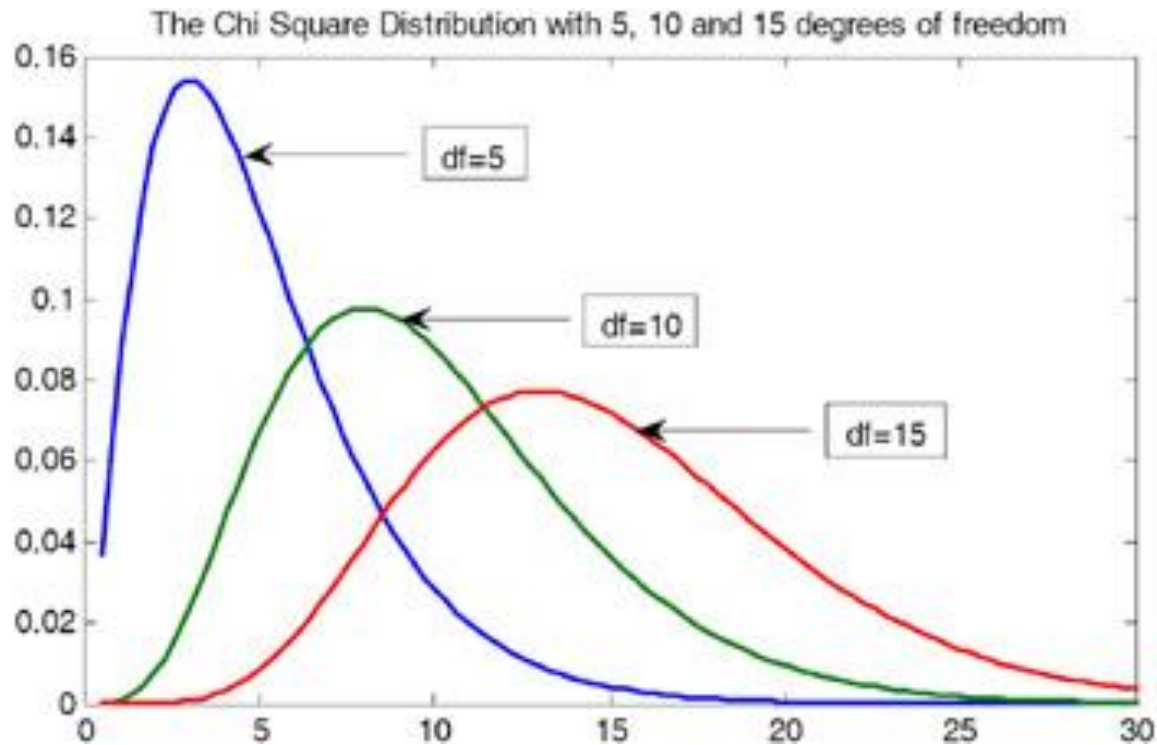
- **Exercício prático 1.** Considere o modelo de regressão abaixo. Teste as hipóteses conjuntamente:
 - Teste a hipótese conjunta: (coeficiente do índice de gini) = 0, (coeficiente do salario médio mensal) = 1, (coeficiente prec crianças pobres) = 0. Qual o p-valor do teste? Você rejeita a hipótese nula? Com que nível de significância?

```
mod2.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita  
+ dados3$indice_gini  
+ dados3$salario_medio_mensal  
+ dados3$perc_crianças_extrem_pobres  
+ dados3$perc_crianças_pobres  
+ dados3$perc_pessoas_dom_agua_estogo_inadequados  
+ dados3$perc_pessoas_dom_paredes_inadequadas  
+ dados3$perc_pop_dom_com_coleta_lixo  
+ dados3$perc_pop_rural  
+ as.factor(dados3$Regiao))
```

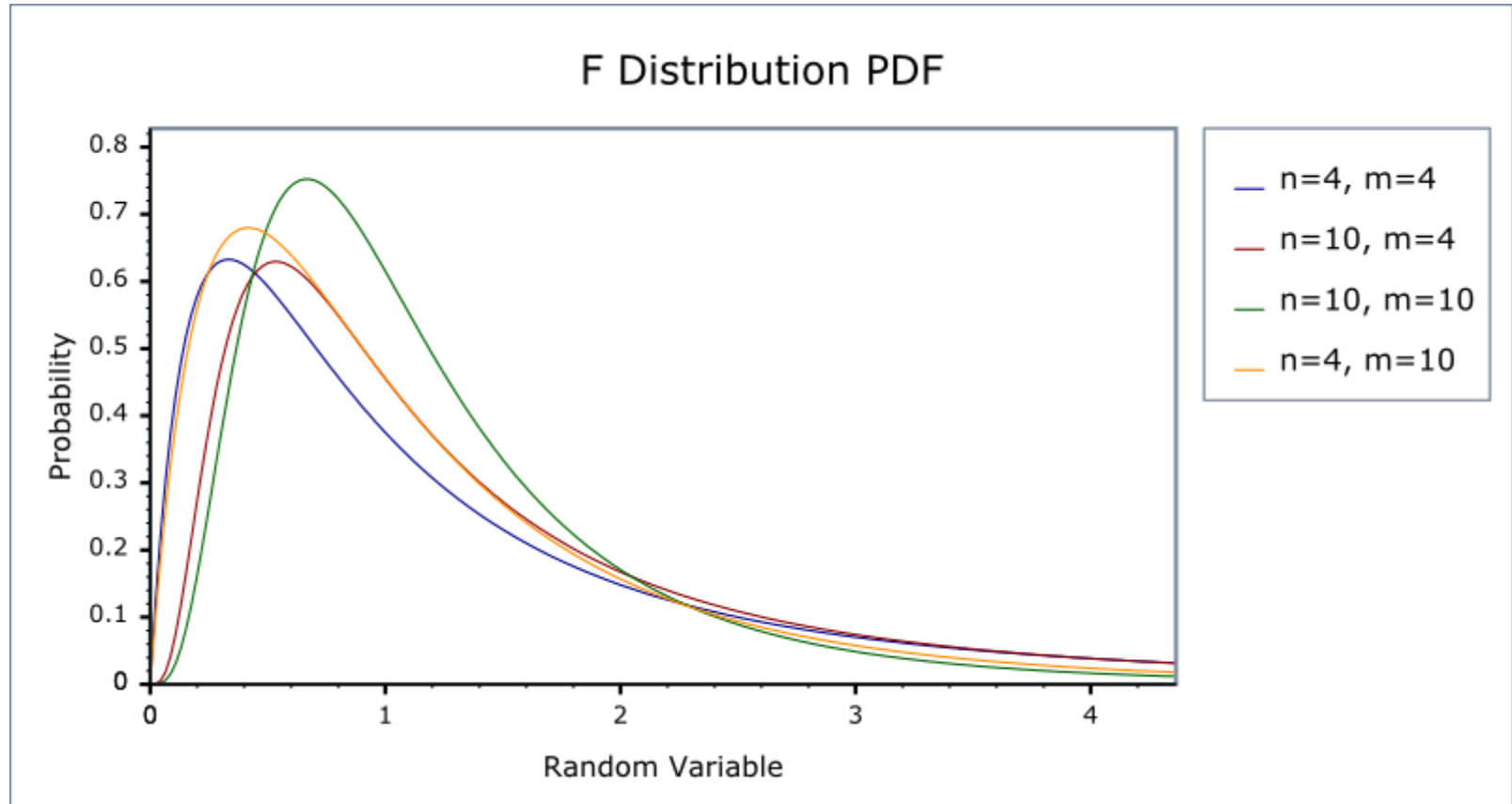
Testes de Hipóteses para Vários Parâmetros

- Sob a hipótese nula, a estatística teste possui distribuição aproximadamente qui-quadrada, com número de graus de liberdade igual ao número de restrições no modelo
- Por exemplo, se estivermos testando a significância de quatro parâmetros conjuntamente, a estatística teste tem distribuição qui-quadrada com quatro graus de liberdade
- Para pequenas amostras, da mesma forma que utilizamos a distribuição t -Student, ao invés da distribuição normal padronizada, para testar múltiplos parâmetros, nós utilizamos a distribuição F ao invés da distribuição qui-quadrada
- Nesse caso, assumimos que, sob a hipótese nula, a distribuição da estatística teste é aproximadamente uma distribuição F , com número de graus de liberdade no numerador igual ao número de restrições. No denominador, o número de graus de liberdade é igual a $n-k-1$ (k é o número de parâmetros do modelo irrestrito)
- Pode-se mostrar que, quando n vai para o infinito, a distribuição F converge para uma qui-quadrada dividido pelo seu número de graus de liberdade

Testes de Hipóteses para Vários Parâmetros



Testes de Hipóteses para Vários Parâmetros



Testes de Hipóteses para Vários Parâmetros

- Considere agora o exemplo com termos quadrático e cúbico para uma das variáveis – vimos que dessa forma podemos capturar não-linearidades na relação entre a variável preditora e a variável resposta:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \beta_{(k+1)} x_{1i}^2 + \beta_{(k+2)} x_{1i}^3 + \epsilon_i$$

- Podemos testar se os termos não-lineares são necessários

$$H_0: \beta_{(k+1)} = \beta_{(k+2)} = 0$$

H_A : pelos menos um dos parâmetros é diferente de zero

- Para isso, podemos proceder da mesma forma que o exemplo anterior: (i) rodamos o modelo irrestrito; (ii) rodamos o modelo restrito; (iii) fazemos a comparação entre os dois modelos

Testes de Hipóteses para Vários Parâmetros

- Implementação no R:
- Equação do modelo “irrestrito” (com os termos quadrático e cúbico)

```
mod1b.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita
+ l(renda_per_capita^2)
+ l(renda_per_capita^3)
+ dados3$indice_gini
+ dados3$salario_medio_mensal
+ dados3$perc_crianças_extrem_pobres
+ dados3$perc_crianças_pobres
+ dados3$perc_pessoas_dom_agua_estogo_inadequados
+ dados3$perc_pessoas_dom_paredes_inadequadas
+ dados3$perc_pop_dom_com_coleta_lixo, data = dados)
summary(mod1b.ex)
```


Testes de Hipóteses para Vários Parâmetros

- Resultados do modelo irrestrito:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.519e+01	1.202e+00	29.278	< 2e-16	***
dados3\$renda_per_capita	-6.809e-02	4.012e-03	-16.975	< 2e-16	***
I(renda_per_capita^2)	6.170e-05	4.105e-06	15.030	< 2e-16	***
I(renda_per_capita^3)	-1.753e-08	1.400e-09	-12.523	< 2e-16	***
dados3\$indice_gini	1.178e+00	1.492e+00	0.789	0.429906	
dados3\$salario_medio_mensal	-9.563e-02	9.277e-02	-1.031	0.302675	
dados3\$perc_crianças_extrem_pobres	-2.928e-02	1.280e-02	-2.287	0.022230	*
dados3\$perc_crianças_pobres	6.767e-02	1.418e-02	4.772	1.87e-06	***
dados3\$perc_pessoas_dom_agua_estogo_inadequados	2.714e-02	6.010e-03	4.517	6.40e-06	***
dados3\$perc_pessoas_dom_paredes_inadequadas	2.632e-02	7.768e-03	3.389	0.000708	***
dados3\$perc_pop_dom_com_coleta lixo	2.948e-03	6.368e-03	0.463	0.643460	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.928 on 5553 degrees of freedom
Multiple R-squared: 0.6977, Adjusted R-squared: 0.6971
F-statistic: 1281 on 10 and 5553 DF, p-value: < 2.2e-16

- Os termos quadrático e cúbico são estatisticamente significantes individualmente?
- De acordo com a significância dos termos quadrático e cúbico acima, você acha que a hipótese nula $H_0: \beta_{(k+1)} = \beta_{(k+2)} = 0$ vai ser rejeitada ou aceita?

Testes de Hipóteses para Vários Parâmetros

- Implementação no R:
- Equação do modelo “restrito” (sem os termos quadrático e cúbico)

```
mod1b.ex.rest <- lm(dados3$mort_infantil ~ dados3$renda_per_capita  
+ dados3$indice_gini  
+ dados3$salario_medio_mensal  
+ dados3$perc_crianças_extrem_pobres  
+ dados3$perc_crianças_pobres  
+ dados3$perc_pessoas_dom_agua_estogo_inadequados  
+ dados3$perc_pessoas_dom_paredes_inadequadas  
+ dados3$perc_pop_dom_com_coleta_lixo, data = dados)  
summary(mod1b.ex.rest)
```

- Comparação entre os modelos:

```
anova(mod1b.ex.rest, mod1b.ex, test='LRT')
```

Testes de Hipóteses para Vários Parâmetros

- Resultados do teste de exclusão de variáveis:

Analysis of Variance Table

```
Model 1: dados3$mort_infantil ~ dados3$renda_per_capita + dados3$indice_gini +
  dados3$salario_medio_mensal + dados3$perc_crianças_extrem_pobres +
  dados3$perc_crianças_pobres + dados3$perc_pessoas_dom_agua_estogo_inadequados +
  dados3$perc_pessoas_dom_paredes_inadequadas + dados3$perc_pop_dom_com_coleta_lixo
Model 2: dados3$mort_infantil ~ dados3$renda_per_capita + I(renda_per_capita^2) +
  I(renda_per_capita^3) + dados3$indice_gini + dados3$salario_medio_mensal +
  dados3$perc_crianças_extrem_pobres + dados3$perc_crianças_pobres +
  dados3$perc_pessoas_dom_agua_estogo_inadequados +
  dados3$perc_pessoas_dom_paredes_inadequadas +
  dados3$perc_pop_dom_com_coleta_lixo
  Res.Df    RSS Df Sum of Sq  Pr(>Chi)
1     5555 90564
2     5553 85670    2    4893.6 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- De acordo com os resultados do teste acima, você rejeita a hipótese nula com nível de significância de 1%? E de 5%? E de 10%
- Qual a nossa conclusão sobre a necessidade de inclusão de termos quadrático e cúbico na equação?

Testes de Hipóteses para Vários Parâmetros

- Voltando aos resultados do modelo irrestrito:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.519e+01	1.202e+00	29.278	< 2e-16	***
dados3\$renda_per_capita	-6.809e-02	4.012e-03	-16.975	< 2e-16	***
I(renda_per_capita^2)	6.170e-05	4.105e-06	15.030	< 2e-16	***
I(renda_per_capita^3)	-1.753e-08	1.400e-09	-12.523	< 2e-16	***
dados3\$indice_gini	1.178e+00	1.492e+00	0.789	0.429906	
dados3\$salario_medio_mensal	-9.563e-02	9.277e-02	-1.031	0.302675	
dados3\$perc_crianças_extrem_pobres	-2.928e-02	1.280e-02	-2.287	0.022230	*
dados3\$perc_crianças_pobres	6.767e-02	1.418e-02	4.772	1.87e-06	***
dados3\$perc_pessoas_dom_agua_estogo_inadequados	2.714e-02	6.010e-03	4.517	6.40e-06	***
dados3\$perc_pessoas_dom_paredes_inadequadas	2.632e-02	7.768e-03	3.389	0.000708	***
dados3\$perc_pop_dom_com_coleta lixo	2.948e-03	6.368e-03	0.463	0.643460	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.928 on 5553 degrees of freedom

Multiple R-squared: 0.6977, Adjusted R-squared: 0.6971

F-statistic: 1281 on 10 and 5553 DF, p-value: < 2.2e-16

- Qual o significado do termo *F-statistic* e do respectivo *p-value* na última linha do output da regressão?
- Na verdade, essa linha corresponde um teste de hipótese conjunto de vários parâmetros

Testes de Hipóteses para Vários Parâmetros

- **Exercício prático 2.** Considere o modelo de regressão abaixo. Teste as hipóteses conjuntamente:
 - Teste a hipótese: (coeficiente do índice de gini) + 2 * (coeficiente do salário médio) = 0. Qual o p-valor do teste? Você rejeita a hipótese nula? Com que nível de significância?

(dica: use o comando `?linearHypothesis` e veja os exemplos no help dessa função)

```
mod2.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita
+ dados3$indice_gini
+ dados3$salario_medio_mensal
+ dados3$perc_crianças_extrem_pobres
+ dados3$perc_crianças_pobres
+ dados3$perc_pessoas_dom_agua_estogo_inadequados
+ dados3$perc_pessoas_dom_paredes_inadequadas
+ dados3$perc_pop_dom_com_coleta_lixo
+ dados3$perc_pop_rural
+ as.factor(dados3$Regiao))
```

Testes de Hipóteses para Vários Parâmetros

- Considere agora o modelo geral de regressão:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \epsilon_i$$

- Nesse modelo geral, as variáveis preditoras podem incluir termos quadrático, cúbico, etc., podem incluir variáveis *dummy*, e podem incluir interações entre variáveis
- O termo *F-statistic* e o respectivo *p-value* correspondem justamente à hipótese nula conjunta

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

H_A : pelo menos um dos coeficientes é diferente de zero

- Note que o intercepto β_0 não está sendo testado. Portanto, a estatística *F* nesse caso está testando um modelo com apenas o intercepto versus um modelo com o intercepto mais as variáveis preditoras

Modelos de Regressão

- **Exercício prático 3.** Considere o modelo de regressão abaixo.

```
mod1.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita  
+ dados3$indice_gini  
+ dados3$salario_medio_mensal  
+ dados3$perc_crianças_extrem_pobres  
+ dados3$perc_crianças_pobres  
+ dados3$perc_pessoas_dom_agua_estogo_inadequados  
+ dados3$perc_pessoas_dom_paredes_inadequadas  
+ dados3$perc_pop_dom_com_coleta_lixo)
```

Questão 1: Teste a hipótese nula conjunta de que todos os coeficientes da regressão são nulos, exceto o intercepto. Qual o p-valor para esse teste? Você rejeita a hipótese nula com nível de significância de 1%? Você rejeita com nível de significância de 5%? Escreva as hipótese nula e alternativa.

Questão 2: Com base na regressão da questão anterior, aumente o modelo de regressão incluindo um termo quadrático para a renda per capita e outro para o índice Gini. Teste a significância conjunta desses dois termos quadráticos. Qual o p-valor para esse teste? Você rejeita a hipótese nula com nível de significância de 1%? Você rejeita com nível de significância de 5%?

Significância Econômica x Estatística

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).

Significância econômica x Estatística

- 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 significativa 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 significativa 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).

Exemplos de Modelos de Regressão para Avaliação de Programas

Avaliação dos Fundos Constitucionais de Desenvolvimento Regional

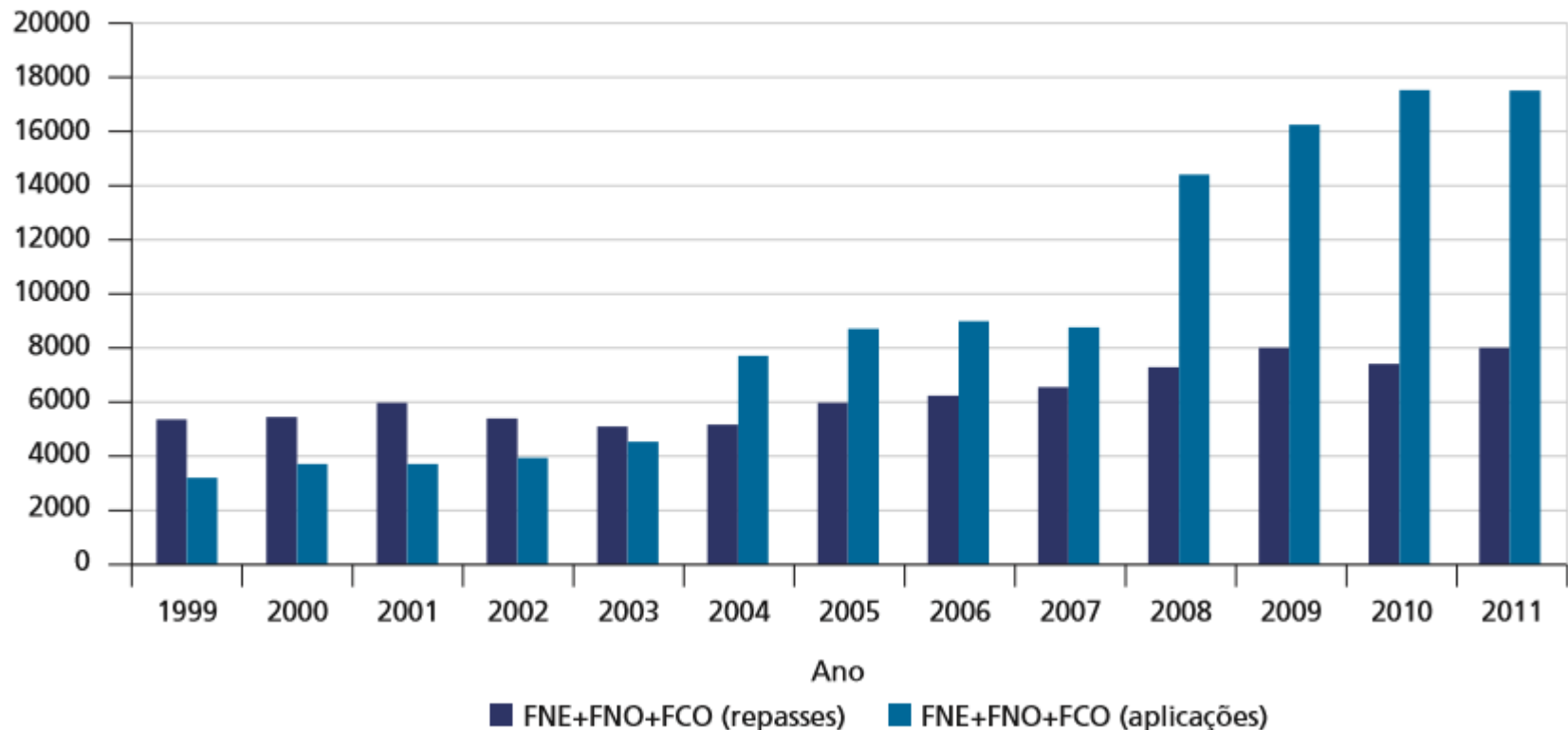
- Discussão importante dado o montante de recursos destinados via fundos constitucionais de desenvolvimento regional
- Texto: “Avaliação dos Efeitos Econômicos dos Fundos Constitucionais de Financiamento do Nordeste, do Norte e do Centro-Oeste; uma Análise por Tipologia da PNDES entre 1999 e 2011”
- Regressão com dados de painel municipal, analisando o crescimento do PIB per capita municipal, em intervalos de tempo de 3 anos
 - Inclusão de variáveis *dummies* por município (efeitos fixos) e por ano
- Variáveis indicadoras da política são a proporção dos aportes dos fundos constitucionais sobre o PIB do município no início da janela de 3 anos
- Diversas variáveis explicativas foram incluídas no modelo para controlar para o efeito de outras variáveis

Avaliação dos Fundos Constitucionais de Desenvolvimento Regional

GRÁFICO 1

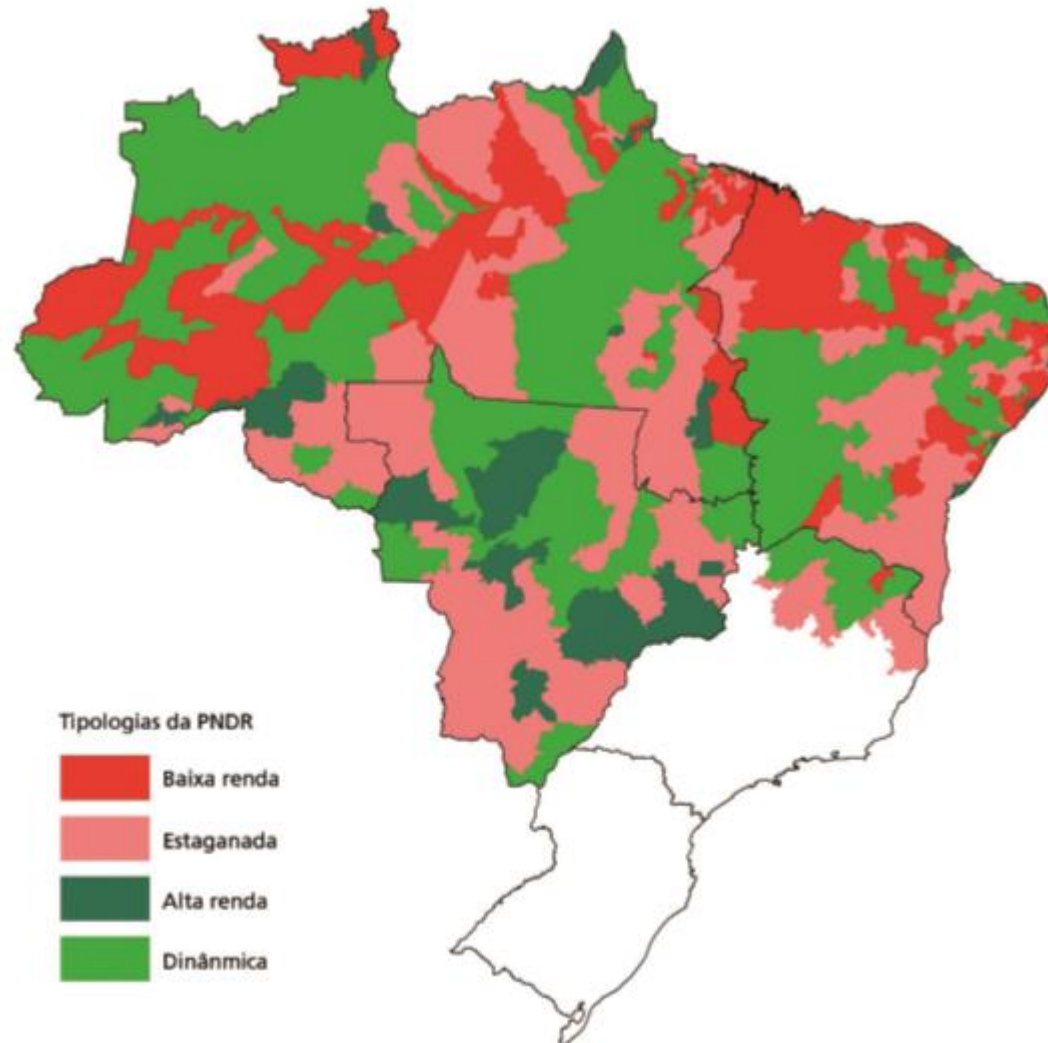
Repasses anuais do Tesouro Nacional e aplicações anuais dos recursos (1999-2011)

(R\$ em milhões, a preços constantes de 2010)



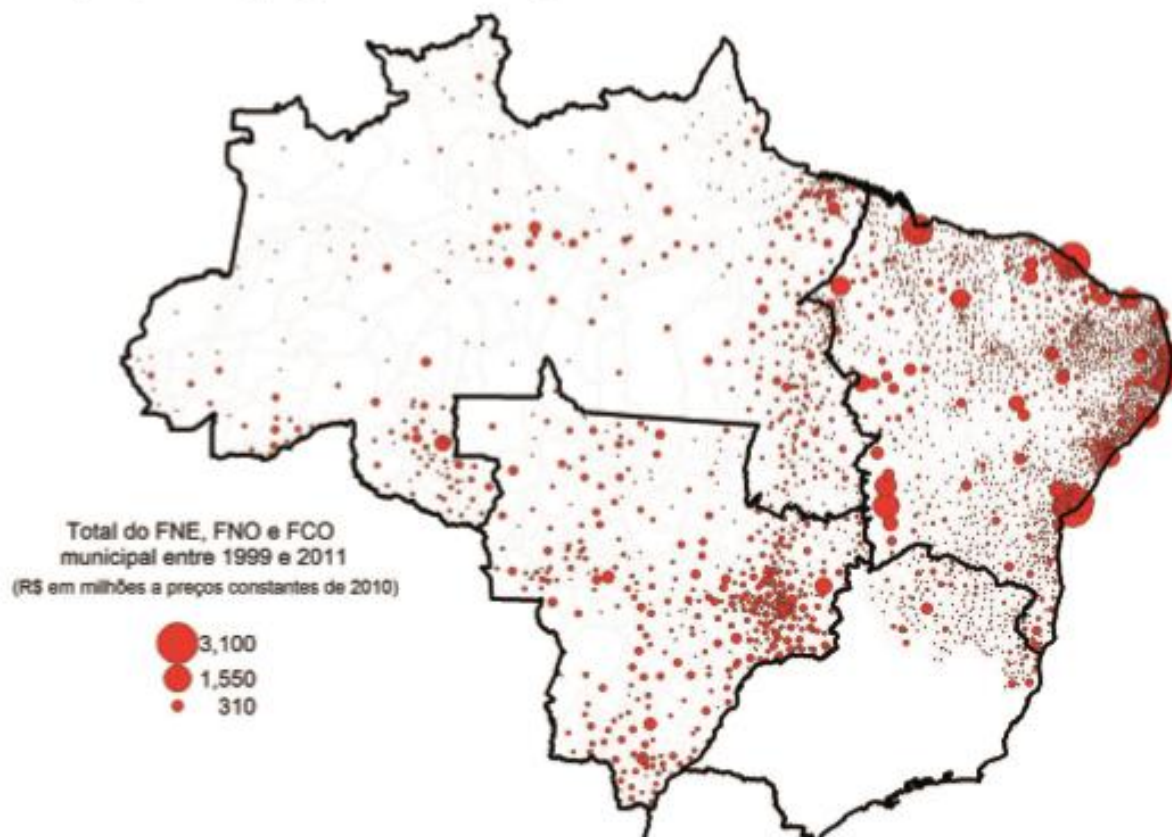
Avaliação dos Fundos Constitucionais de Desenvolvimento Regional

Tipologias da Política Nacional de Desenvolvimento Regional



Avaliação dos Fundos Constitucionais de Desenvolvimento Regional

Distribuição espacial dos recursos do FNE, do FNO e do FCO no nível municipal (1999-2011)
(Em R\$ milhões, a preços constantes de 2010)



Resultado dos efeitos do FNE sobre o crescimento médio anual do PIB *per capita* no nível municipal – método painel de efeitos fixos

Método de estimação	Variável dependente – taxa de crescimento anual média do PIB <i>per capita</i>			
	Painel efeitos fixos	Painel efeitos fixos	Painel efeitos fixos	Painel efeitos fixos
	(1)	(2)	(3)	(4)
Alta renda_Proporção do FNE início do período (1º ano) em relação ao PIB do início de cada período	0.9982** (0.0157)	0.8501** (0.0208)	Alta renda_Proporção do FNE início do período (1º + 2º ano) em relação ao PIB do início de cada período (0.5977)	-0.0122 (0.0665)
Dinâmica_Proporção do FNE início do período (1º ano) em relação ao PIB do início de cada período	0.1407*** (0.0006)	0.1225*** (0.0010)	Dinâmica_Proporção do FNE início do período (1º + 2º ano) em relação ao PIB do início de cada período (0.0000)	0.1282*** (0.0000)
Baixa renda_Proporção do FNE início do período (1º ano) em relação ao PIB do início de cada período	0.4528*** (0.0000)	0.2129*** (0.001)	Baixa renda_Proporção do FNE início do período (1º + 2º ano) em relação ao PIB do início de cada período (0.0002)	0.0934*** (0.2259)
Estagnada_Proporção do FNE início do período (1º ano) em relação ao PIB do início de cada período	0.1508** (0.0411)	-0.0191 0.7733	Estagnada_Proporção do FNE início do período (1º + 2º ano) em relação ao PIB do início de cada período (0.0000)	0.1322*** (0.0099)
Ln (PIB <i>per capita</i> no início de cada período)	-0.1693*** (0.0000)	-0.2944*** (0.0000)	Ln (PIB <i>per capita</i> no início de cada período) (0.0000)	-0.1681*** (0.0000)
Ln (anos médios de escolaridade no início de cada período, Rais)	0.0670*** (0.0000)	-0.0103** (0.0138)	Ln (anos médios de escolaridade no início de cada período, Rais) (0.0000)	0.0653*** (0.0091)
Ln (densidade populacional no início de cada período)	0.0926*** (0.0000)	-0.1280*** (0.0000)	Ln (densidade populacional no início de cada período) (0.0000)	0.0886*** (0.0000)
Efeitos fixos	Sim	Sim	Efeitos fixos	Sim
<i>Dummy</i> de tempo	Não	Sim	<i>Dummy</i> de tempo	Sim
Número de observações (municípios)	5.946	5.946		5.946
R2 ajustado	0.1739	0.3368		0.1779
				0.3403

Resultado dos efeitos do FNO sobre o crescimento médio anual do PIB *per capita* no nível municipal – método painel de efeitos fixos

Método de estimação	Variável dependente = taxa de crescimento anual média do PIB <i>per capita</i>				
	Painel efeitos fixos	Painel efeitos fixos	Painel efeitos fixos	Painel efeitos fixos	
	(1)	(2)	(3)	(4)	
Alta renda_Proporção do FNO início do período (1º ano) em relação ao PIB do início de cada período	0.0866	-0.0115	Alta renda_Proporção do FNO início do período (1º + 2º ano) em relação ao PIB do início de cada período	0.2020*	0.1898*
	(0.5902)	(0.9406)		(0.0520)	(0.0566)
Dinâmica_Proporção do FNO início do período (1º ano) em relação ao PIB do início de cada período	0.1115**	0.0891**	Dinâmica_Proporção do FNO início do período (1º + 2º ano) em relação ao PIB do início de cada período	0.1167***	0.0981***
	(0.0147)	(0.0419)		(0.0017)	(0.0058)
Baixa renda_Proporção do FNO início do período (1º ano) em relação ao PIB do início de cada período	0.0721	0.0509	Baixa renda_Proporção do FNO início do período (1º + 2º ano) em relação ao PIB do início de cada período	0.0267	0.0194
	(0.1582)	(0.2987)		(0.1539)	(0.2795)
Estagnada_Proporção do FNO início do período (1º ano) em relação ao PIB do início de cada período	-0.0077	-0.0064	Estagnada_Proporção do FNO início do período (1º + 2º ano) em relação ao PIB do início de cada período	-0.0041	-0.0064
	(0.5416)	(0.5952)		(0.7329)	(0.5835)
Ln (PIB per capita no início de cada período)	-0.2100***	-0.2711***	Ln (PIB <i>per capita</i> no início de cada período)	-0.2087***	-0.2696***
	(0.0000)	(0.0000)		(0.0000)	(0.0000)
Ln (anos médios de escolaridade no início de cada período, Rais)	0.0418***	-0.0171	Ln (anos médios de escolaridade no início de cada período, Rais)	0.0439***	-0.0150
	(0.0015)	(0.2117)		(0.0008)	(0.2722)
Ln (densidade populacional no início de cada período)	0.0459*	-0.1205***	Ln (densidade populacional no início de cada período)	0.0412	-0.1234***
	(0.0690)	(0.0000)		(0.1019)	(0.0000)
Efeitos fixos	Sim	Sim	Efeitos fixos	Sim	Sim
<i>Dummy</i> de tempo	Não	Sim	<i>Dummy</i> de tempo	Não	Sim
Número de observações (municípios)	1.347	1.347		1.347	1.347
R2 ajustado	0.253	0.3037		0.257	0.3074

Resultado dos efeitos do FCO sobre o crescimento médio anual do PIB *per capita* no nível municipal – método painel de efeitos fixos

Método de estimação	Variável dependente = taxa de crescimento anual média do PIB <i>per capita</i>				
	Painel efeitos fixos	Painel efeitos fixos		Painel efeitos fixos	Painel efeitos fixos
	(1)	(2)		(3)	(4)
Alta renda_Proporção do FCO início do período (1º ano) em relação ao PIB do início de cada período	0.1068**	0.0457	Alta renda_Proporção do FCO início do período (1º + 2º ano) em relação ao PIB do início de cada período	0.1321***	0.0517*
	(0.0150)	(0.2716)		(0.0000)	(0.0980)
Dinâmica_Proporção do FCO início do período (1º ano) em relação ao PIB do início de cada período	0.2910***	0.0890	Dinâmica_Proporção do FCO início do período (1º + 2º ano) em relação ao PIB do início de cada período	0.0544	-0.0190
	(0.0011)	(0.2983)		(0.1316)	(0.5788)
Baixa renda_Proporção do FCO início do período (1º ano) em relação ao PIB do início de cada período	-	-	Baixa renda_Proporção do FCO início do período (1º + 2º ano) em relação ao PIB do início de cada período	-	-
	-	-		-	-
Estagnada_Proporção do FCO início do período (1º ano) em relação ao PIB do início de cada período	0.0579	-0.0326	Estagnada_Proporção do FCO início do período (1º + 2º ano) em relação ao PIB do início de cada período	0.0187	-0.0451**
	(0.1257)	(0.3709)		(0.3783)	(0.0309)
Ln (PIB <i>per capita</i> no início de cada período)	-0.2090***	-0.2188***	Ln (PIB <i>per capita</i> no início de cada período)	-0.2054***	-0.2204***
	(0.0000)	(0.0000)		(0.0000)	(0.0000)
Ln (anos médios de escolaridade no início de cada período, Rais)	0.1318***	0.0103	Ln (anos médios de escolaridade no início de cada período, Rais)	0.1247***	0.0076
	(0.0015)	(0.6049)		(0.0000)	(0.7008)
Ln (densidade populacional no início de cada período)	0.0430	-0.0632**	Ln (densidade populacional no início de cada período)	0.0413	-0.0717**
	(0.1175)	(0.0231)		(0.1328)	(0.0105)
Efeitos fixos	Sim	Sim	Efeitos fixos	Sim	Sim
Dummy de tempo	Não	Sim	Dummy de tempo	Não	Sim
Número de observações (municípios)	1.338	1.338		1.338	1.338
R2 ajustado	0.2574	0.2257		0.2578	0.2287

Efeito do SIMPLES Federal sobre a Geração de Empregos

- Texto: “O SIMPLES Federal e a Geração de Empregos na Indústria”, Carlos Henrique Courseuil e Rodrigo de Moura
- Lei número 9.317, de dezembro de 1996 – simplificação tributária, aplicação de alíquotas reduzidas, visando potencializar o desempenho dos estabelecimentos alvos
- Empregam regressão de descontinuidade para identificar o efeito do SIMPLES sobre o número de empregos das firmas no Brasil
 - Empresas com faturamento anual abaixo de um valor de corte são elegíveis ao imposto
- Dados da Pesquisa Industrial Anual (PIA) do IBGE
- O estudo olha os impactos do SIMPLES em dois instantes do tempo:
 - 1997, quando o SIMPLES foi implementado (valor de corte R\$ 720.000 anual)
 - 1999, quando houve um aumento da receita máxima que torna a firma elegível (valor de corte R\$ 1.200.000 anual)

Tabela 3. Estimativas do modelo FD – PIA 1996/97

Variável Dependente						
Variação do Número de Empregados entre 1996 e 1997						
	Janela (em milhares)					
Regressores	300	250	200	150	100	50
Simples (T_{i97})	1,69 <i>0,80</i>	1,96 <i>0,85</i>	1,64 <i>0,79</i>	1,29 <i>0,76</i>	1,86 <i>0,87</i>	1,80 <i>1,20</i>
ΔF_{i96}	1E-08 <i>5E-09</i>	2E-08 <i>5E-09</i>	9E-09 <i>5E-09</i>	8E-09 <i>4E-09</i>	7E-10 <i>3E-09</i>	-7E-07 <i>2E-06</i>
$D_{i97} (F_{i96} - c)$	-1E-06 <i>4E-06</i>	-1E-06 <i>7E-06</i>	9E-06 <i>7E-06</i>	-1E-06 <i>7E-06</i>	-1E-05 <i>2E-05</i>	-2E-06 <i>5E-05</i>
Constante	-0,88 <i>1,01</i>	-1,18 <i>1,12</i>	-1,80 <i>1,14</i>	-0,88 <i>1,29</i>	-2,17 <i>1,77</i>	-2,59 <i>2,22</i>
Nº de obs.	3499	2880	2312	1728	1165	603

Nota: Foram incluídos também como regressores dummies para setores de atividade e para as Unidades Federativas.

Simples (T_{i97}) = 1 se a firma optou pelo SIMPLES em 1997 e 0 se não optou;

ΔF_{i96} é a variação do faturamento bruto de 1995 para 1996

$(F_{i96} - c) = (\text{Receita Bruta em 1996}) - 720000$;

Erro Padrão em *itálico*.

Tabela 4. Estimativas do modelo FD – PIA 1998/99

Variável Dependente

Varição do Número de
Empregados entre 1998 e 1999

Janela (em milhares)

Regressores	300	250	200	150	100	50
Simple (T ₁₉₉)	2,77 <i>1,34</i>	2,24 <i>1,14</i>	2,82 <i>1,15</i>	0,98 <i>1,07</i>	0,69 <i>1,25</i>	2,52 <i>2,11</i>
ΔF_{198}	3E-09 <i>1E-06</i>	6E-08 <i>1E-06</i>	1E-06 <i>8E-07</i>	6E-08 <i>8E-07</i>	5E-07 <i>1E-06</i>	2E-07 <i>1E-06</i>
D ₁₉₉ (F ₁₉₈ – c)	-6E-06 <i>3E-06</i>	-6E-06 <i>5E-06</i>	-2E-05 <i>7E-06</i>	4E-06 <i>1E-05</i>	7E-06 <i>2E-05</i>	3E-05 <i>5E-05</i>
Constante	0,64 <i>2,72</i>	0,66 <i>2,65</i>	-0,83 <i>2,07</i>	0,00 <i>4,03</i>	-5,54 <i>2,56</i>	-6,31 <i>3,30</i>
Nº de obs.	2111	1734	1394	1054	740	393

Nota: Foram incluídos também como regressores dummies para setores de atividade e para as Unidades Federativas.

Simple (T₁₉₉) = 1 se a firma optou pelo SIMPLES em 1999 e 0 se não optou;

ΔF_{198} é a variação do faturamento bruto de 1997 para 1998

(F₁₉₈ – c) = (Receita Bruta em 1998) - 1200000;

Erro Padrão em *itálico*.

Impacto do PBF sobre Indicadores Educacionais

- Texto: “Avaliação do Impacto do Programa Bolsa Família sobre Indicadores Educacionais”, Julio Alfredo Romero e Ana Maria Hermeto
- Estuda o efeito do PBF sobre indicadores educacionais das crianças de 7 a 14 anos, utilizando regressão de descontinuidade
 - O benefício básico (de R\$ 50,00) era pago a famílias consideradas extremamente pobres (aquelas com renda mensal de até R\$ 50,00 por pessoa);
 - O benefício variável era pago às famílias pobres, com renda mensal de até R\$ 100,00 por pessoa
 - As regressões de descontinuidade consideraram dois cortes: R\$ 50,00 e R\$ 100,00
- Dados da pesquisa de Avaliação de Impacto do Bolsa Família (AIBF), de 2005, junto com o Cadastro Único
- Impactos satisfatórios sobre indicadores de curto prazo após a implantação do PBF: redução de evasão escolar de mulheres e aumento da aprovação para homens no Nordeste

Impacto do PBF sobre Indicadores Educacionais

TABELA 1 – Variáveis dependentes: Indicadores para avaliar os diferenciais do PBF na educação para crianças entre 7 e 14 anos de idade

Variáveis	Descrição
Não deixaram de ir à escola no último mês (ou o complemento deste)	Proporção de meninas e meninos no domicílio que não deixaram de ir à escola no último mês.
Evasão ou abandono	Proporção de meninas e meninos no domicílio que evadiram do sistema de ensino entre 2004 e 2005.
Progressão	Proporção de meninas e meninos no domicílio que foram aprovados entre 2004 e 2005.
Alocação entre trabalho e estudo	Proporção de meninas e meninos no domicílio que declararam só estudar atualmente, vis-à-vis aqueles que declararam só trabalhar, trabalhar e estudar e não trabalhar nem estudar.
Retenção	Proporção de meninas e meninos que foram reprovados entre 2004 e 2005.

Impacto do PBF sobre Indicadores Educacionais

TABELA 2 – Variáveis independentes: variáveis utilizadas na especificação dos modelos de regressão descontínua para avaliar os diferenciais do PBF na educação

Atributos do chefe de família:

Raça do chefe de família	Branca Não Branca
Sexo do chefe de família	Masculino Feminino
Escolaridade do chefe de família	Até 3 anos de estudos* Até 4 anos de estudos* Até 7 anos de estudos*
Idade do chefe de família	Menor e igual há 50 anos Mais que 50 anos
Altura em metros do chefe de família	Medida em metros (mts)
Escolaridade da mãe do chefe de família	Mãe alfabetizada Mãe não alfabetizada
Tempo de permanência do chefe de família no município	Menos de 10 anos* Menos de 5 anos*
Tempo de permanência do chefe de família na área rural.	Viveu até os 14 anos Não viveu até os 14 anos

Impacto do PBF sobre Indicadores Educacionais

Características da família:

Número de membros da família

Crianças entre 0 a 3 anos de idade

Crianças entre 0 a 6 anos de idade

Crianças mulheres 7 a 14/ criança 0 a 14 anos

Casal com filhos até 14 anos

Presença de pessoas de 60 anos ou mais

Número de membros no domicílio

Proporção de crianças de 0 a 3 anos

Proporção de crianças de 0 a 6 anos

Proporção crianças mulheres 7 a 14/
crianças 0 a 14

O Casal tem filhos até 14 anos

O Casal não tem filhos até 14 anos

Há pessoa de 60 anos e mais no domicílio

Há pessoa menor de 60 anos no domicílio.

Características do domicílio:

Qualidade de domicílio¹

Área de residência do domicílio

Região de residência do domicílio

Qualidade inferior*

Qualidade média*

Urbana

Rural

Nordeste*

Norte – Centro Oeste*

*Para cada uma destas categorias foi construída uma variável *dummy*.

¹ Esta variável foi gerada através do método Grade of Membership (GOM), com três categorias para a qualidade das condições dos domicílios, classificadas em: muito boa, regular e ruim.

Impacto do PBF sobre Indicadores Educacionais

TABELA 5: Estimação da regressão descontínua dos indicadores para avaliar os diferenciais do PBF na educação de crianças de 7 a 14 anos. Brasil e Regiões, 2005

Variáveis/Regiões	Ponto de corte					
	Até R\$100.00			Até R\$50.00		
	Total	Homem	Mulheres	Total	Homem	Mulheres
a) Crianças que evadiram a escola em 2004 (evasão)						
Brasil			-0,015**			-0,017*
Nordeste	-0,026*					
Norte/centro-oeste				-0,023*		
b) Crianças que foram aprovados a escola entre 2004 e 2005						
Brasil						
Nordeste					0,283*	
Norte/centro-oeste						
c) Crianças que repetiram a escola entre 2004 e 2005 (repetência)						
Brasil						
Nordeste	-0,097*				-0,290*	
Norte/centro-oeste						
d) Crianças que só estudavam em 2005						
Brasil				-0,218***		
Nordeste			-0,134***			
Norte/centro-oeste						

Fonte: AIBF e CadÚnico, 2005.

* valor significativo a 10%; ** valor significativo a 5%; *** valor significativo a 1%.

Obrigado.