

Controladoria e Contabilidade

Análise Multivariada Aplicada à Contabilidade

Prof. Dr. Marcelo Botelho da Costa Moraes

www.marcelobotelho.com

mbotelho@usp.br

Turma: 2º / 2016

Agenda – Aula 8/15

- Regressão com Dados em Painel
 - Tipos de Painel
 - POLS
 - Efeitos Fixos
 - Efeitos Aleatórios
 - Testes para escolha do modelo

Dados em Painel

- Dados coletados de diferentes elementos coletados ao longo do tempo
- Seção Transversal + Série Histórica
- *Panel Data* ou Dados Longitudinais

Dados em Painel – CS Empilhada

	T	Var 1	Var 2	...	Var m
Obs 1	1				
Obs 2	1				
...	...				
Obs n	1				
Obs 1	2				
Obs 2	2				
...	...				
Obs n	2				
...	...				
Obs 1	t				
Obs 2	t				
...	...				
Obs n	t				

Dados em Painel – ST Empilhada

	T	Var 1	Var 2	...	Var m
Obs 1	1				
Obs 1	2				
...	...				
Obs 1	t				
Obs 2	1				
Obs 2	2				
...	...				
Obs 2	t				
...	...				
Obs n	1				
Obs n	2				
...	...				
Obs n	t				

Painel: Dados Combinados

- **Painel Balanceado:** número de observações igual para todas as unidades de análise
- **Painel Desbalanceado:** existem dados faltantes por observação ao longo do tempo
 - Não existe diferença prática entre os dois tipos de painel
 - Cuidado com a origem dos dados faltantes, pois pode gerar viés para a análise
 - Como a quantidade de dados no painel desbalanceado é menor, a eficiência do estimador diminui

Painel: Dados Combinados

- **Painel Curto:** painel com $T < N$
 - Número de períodos no tempo é menor que número de observações na amostra
- **Painel Longo:** painel com $T > N$
 - Número de períodos no tempo é maior que número de observações na amostra
- Não existe diferença prática entre os dois tipos de painel

Dados em Painel

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

- Onde,
- y_{it} é a variável dependente
- α é o intercepto
- β é um vetor de $k \times 1$ dos parâmetros a serem estimados nas variáveis explicativas
- x_{it} é o vetor de $1 \times k$ das observações das variáveis explicativas
- $t = 1, \dots, T$
- $i = 1, \dots, N$

Vantagem dos Dados em Painel

- Primeiro, e talvez mais importante, podemos abordar uma gama mais ampla de questões e resolver problemas mais complexos, com dados em painel, do que seria possível com a série temporal pura ou dados transversais puros sozinhos

Vantagem dos Dados em Painel

- Em segundo lugar, muitas vezes é interessante examinar como as variáveis, ou as relações entre elas, muda dinamicamente (com o tempo)
- Para fazer isso usando dados de séries temporais puro, muitas vezes requerem um longo prazo de dados simplesmente para obter um número suficiente de observações para ser capaz de realizar quaisquer testes de hipóteses significativas. Mas, combinando dados de séries de secção transversal e de tempo, pode-se aumentar o número de graus de liberdade, e, portanto, a potência do teste, através do emprego de informação sobre o comportamento dinâmico de um grande número de entidades ao mesmo tempo. A variação adicional introduzida pela combinação dos dados desta forma também pode ajudar a mitigar os problemas de multicolinearidade que podem surgir se séries temporais são modelados individualmente

Vantagem dos Dados em Painel

- Em terceiro lugar, pela estruturação do modelo de forma adequada, pode-se remover o impacto de certas formas de viés nas variáveis omitidas no resultado de regressão

Dados em Painel – POLS

- A maneira mais simples para lidar com tais dados seria estimar uma regressão agrupada (*pooled*), o que implicaria a estimativa de uma única equação em todos os dados em conjunto, de modo que o conjunto de dados para y é empilhado em cima de uma única coluna que contém todo o tempo em corte transversal e observações da série, e da mesma forma todas as observações sobre cada variável explicativa seriam empilhados em colunas individuais na matriz x . Então essa equação poderia ser estimado na forma usual usando MQO
- Conhecido como POLS \rightarrow *Pooled Ordinary Least Square*

Dados em Painel – POLS

- É a técnica de painel mais simples e ingênua
- Desconsidera as dimensões de tempo e espaço
- Empilha-se as observações e obtém-se a regressão
 - O principal problema é camuflar a **Heterogeneidade** que possa existir entre as variáveis. Os coeficientes estimados na equação, podem ser tendenciosos e inconsistentes (isso é tratado em Durbin-Watson)

Simplesmente empilhamos todas as observações e estimamos uma regressão “grande”, desprezando a natureza de corte transversal e de séries temporais de nossos dados (Gujarati, 2011, p. 589)

Dados em Painel – Efeitos Fixos

Combina-se todas as observações, deixando que cada unidade de corte transversal tenha sua própria variável *dummy* (intercepto)

Embora o intercepto possa diferir entre os indivíduos, o intercepto de cada indivíduo não varia com o tempo

(Gujarati, 2011, p. 590/592)

Dados em Painel – Efeitos Fixos

- Decomposição do termo de erro

$$u_{it} = \mu_i + v_{it}$$

- μ_i é um efeito específico individual e v_{it} uma perturbação residual que varia ao longo do tempo e entidades

$$y_{it} = \alpha + \beta x_{it} + \mu_i + v_{it}$$

- μ_i carrega toda a informação fixa ao longo do tempo que afeta y_{it} , por exemplo setor da empresa

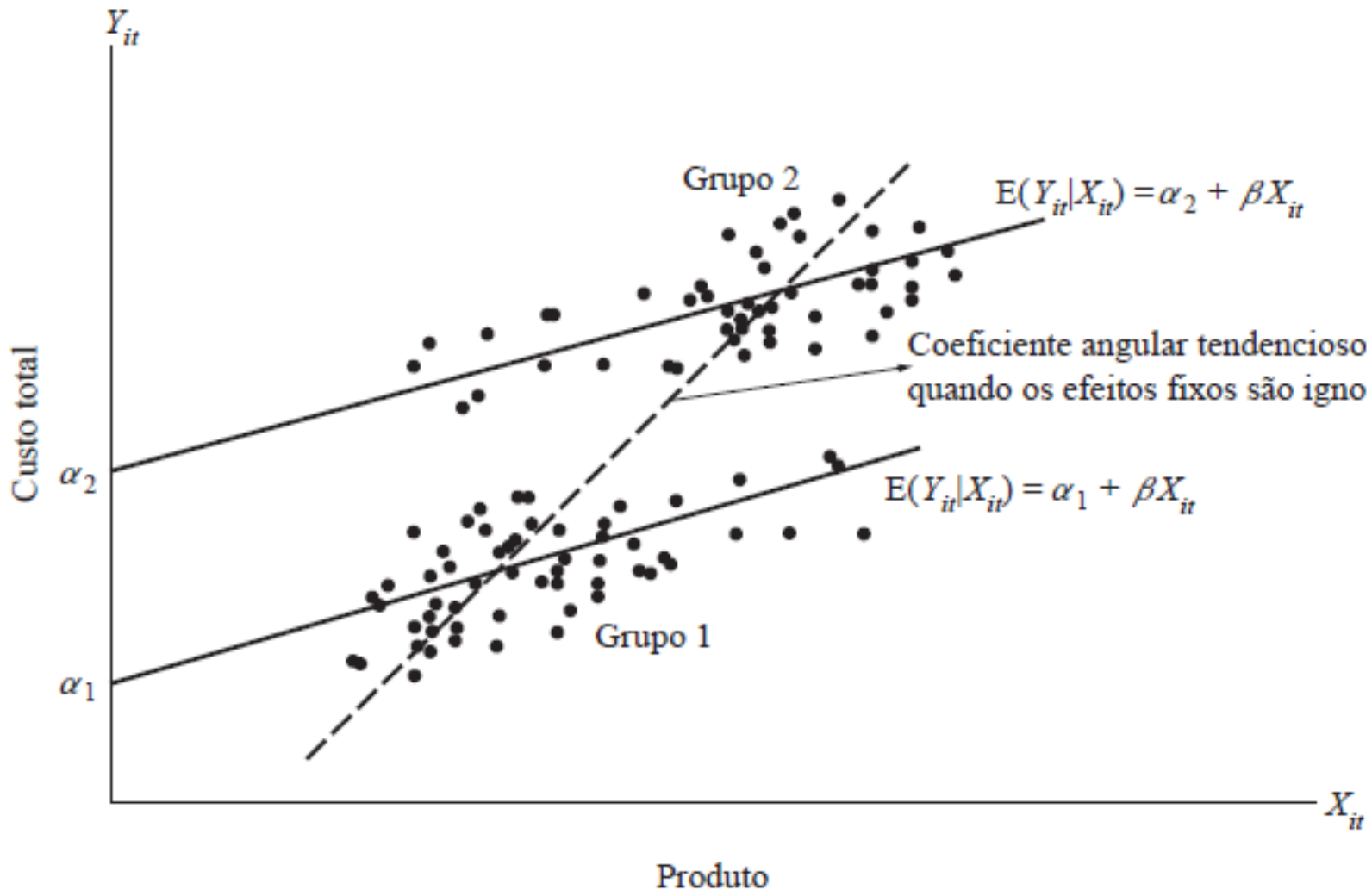
Dados em Painel – Efeitos Fixos

- Esse modelo poderia ser estimado utilizando uma abordagem com variáveis *dummy*, que seriam chamados de variável *dummy* de mínimos quadrados (LSDV)

$$y_{it} = \beta x_{it} + \mu_1 D1_i + \mu_2 D2_i + \dots + \mu_N DN_i + v_{it}$$

- Onde cada D_i (exemplo $D1$) representa o valor 1 para aquela observação e zero para todas as demais
- O intercepto (α) dada a multicolinearidade entre a variável *dummy* e o intercepto

Dados em Painel – Efeitos Fixos



Dados em Painel – Efeitos Fixos

Problemas:

- Se introduzirmos variáveis *dummies* demais, teremos um problema de falta de graus de liberdade
- Se introduzirmos variáveis *dummies* demais, existe a possibilidade de multicolinearidade
- MQVD pode não ser capaz de identificar o impacto das variáveis que não mudam ao longo do tempo
- Deve-se pensar cuidadosamente no termo de erro

Dados em Painel – Efeitos Fixos

Combinamos todas as observações, mas para cada empresa expressamos cada variável como um desvio de seu valor médio e, então, estimamos uma regressão de MQO contra esses valores ***corrigidos para a média***

(Gujarati, 2011, p. 590)

Dados em Painel – Efeitos Fixos

- Como o modelo precisaria estimar $N + k$ parâmetros, ele pode passar por uma **transformação intragrupo**

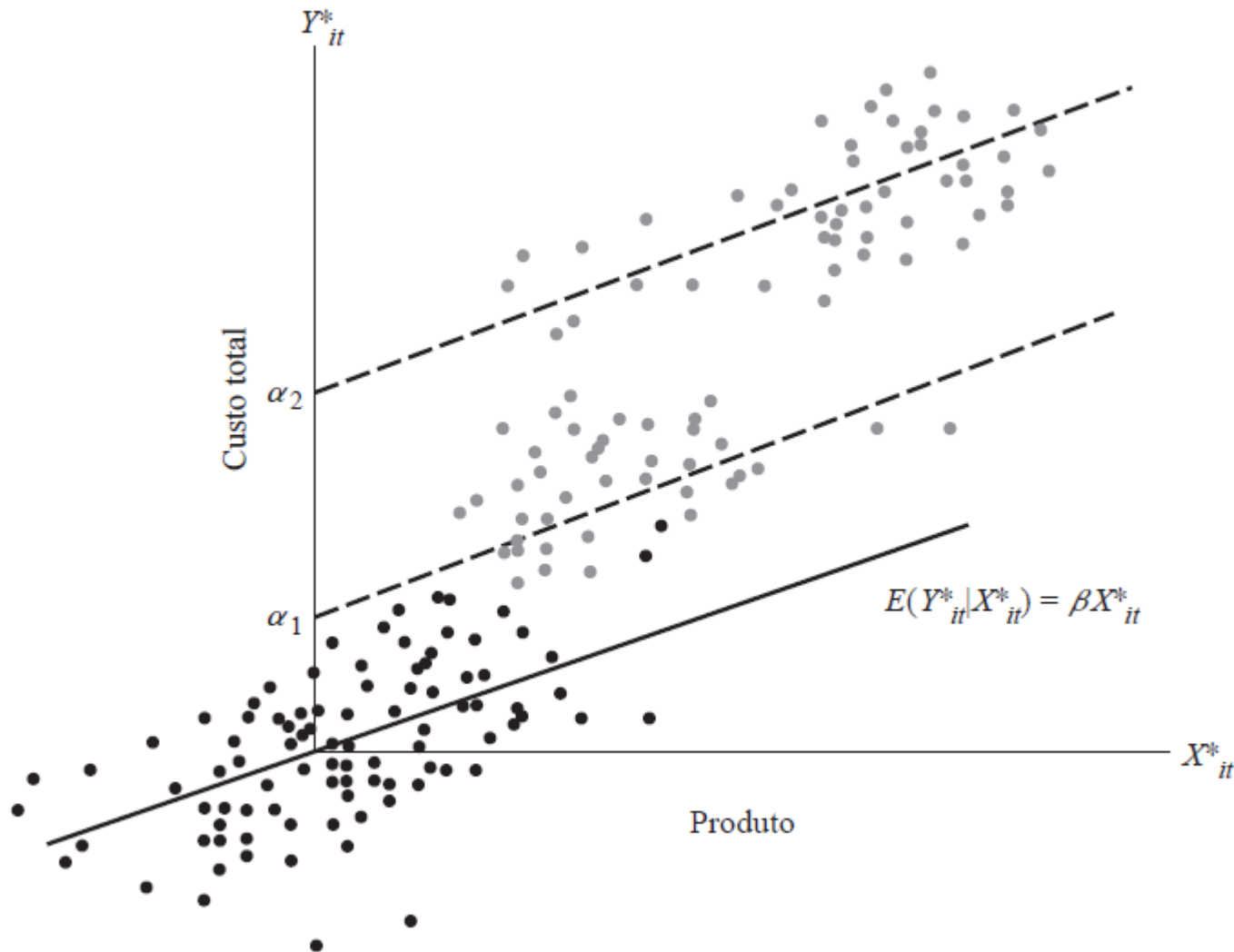
$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i$$

- Deixando os dados centrados na média, sendo reescrito como

$$\dot{y}_{it} = \beta \dot{x}_{it} + \dot{u}_{it}$$

- Com $NT - N - k$ graus de liberdade

Dados em Painel – Efeitos Fixos



Dados em Painel – Efeitos Fixos de Tempo

- É possível ter um modelo de efeitos fixos de tempo, em vez de um modelo de efeitos fixos-entidade
- Com efeitos fixos de tempo, os interceptos seriam autorizados a variar ao longo do tempo, assumindo como sendo as mesmas entidades em cada ponto no tempo

$$y_{it} = \alpha + \beta x_{it} + \lambda_t + v_{it}$$

- λ_t é um intercepto variável no tempo que captura todas as variáveis que afetam y_{it} e que variam ao longo do tempo, mas são constantes transversalmente

Dados em Painel – Efeitos Fixos de Tempo

- Da mesma forma pode ser descrito como

$$y_{it} = \beta x_{it} + \lambda_1 D1_t + \lambda_2 D2_t + \dots + \lambda_T DT_t + v_{it}$$

- E passar por uma **transformação intragrupo**

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i$$

- Deixando os dados centrados na média, sendo reescrito como

$$\dot{y}_{it} = \beta \dot{x}_{it} + \dot{u}_{it}$$

Dados em Painel – Efeitos Fixos de Tempo

- Finalmente podemos juntar os dois modelos

$$y_{it} = \beta x_{it} + \mu_1 D1_i + \mu_2 D2_i + \dots + \mu_N DN_i + \lambda_1 D1_t + \lambda_2 D2_t + \dots + \lambda_T DT_t + v_{it}$$

- Entretanto número de parâmetros a serem estimados passa a ser $k + N + T$ e a transformação intragrupo neste modelo de duas vias seria mais complexa

Dados em Painel – Efeitos Aleatórios

- Uma alternativa ao modelo de efeitos fixos descrito é o modelo de efeitos aleatórios, que também é conhecido como o modelo de erro de componentes. Tal como acontece com efeitos fixos, a abordagem de efeitos aleatórios propõe diferentes termos de intercepto para cada entidade e, novamente, esses interceptos são constantes ao longo do tempo, com as relações entre as variáveis explicativas e explicada assumindo ser a mesmo tanto de forma transversal como temporalmente

Dados em Painel – Efeitos Aleatórios

- No entanto, a diferença é que, sob o modelo de efeitos aleatórios, os interceptos de cada unidade de corte transversal são assumidos como surgindo a partir de uma intersecção comum α (que é o mesmo para todas as unidades de corte transversal e ao longo do tempo), além de uma variável aleatória ϵ_i que varia na seção transversal, mas é constante ao longo do tempo
- ϵ_i mede o desvio aleatório de termo intercepto de cada entidade do intercepto “global” α

Dados em Painel – Efeitos Aleatórios

- Podemos escrever o modelo de efeitos aleatórios como

$$y_{it} = \alpha + \beta x_{it} + \omega_{it}$$
$$\omega_{it} = \epsilon_i + v_{it}$$

- Onde,
- x_{it} é um vetor $1 \times k$ das variáveis explicativas, mas ao contrário do modelo de efeitos fixos, não há variáveis do tipo *dummy* para capturar a heterogeneidade (variação) na dimensão transversal, isso ocorre via o termo ϵ_i

Dados em Painel – Efeitos Aleatórios

- Note-se que esta estrutura requer os pressupostos de que o novo termo de erro transversal ϵ_i , tem média zero, seja independente do termo de erro de observação individual (v_{it}), tenha constante variância σ_ϵ^2 e seja independente das variáveis explicativas (x_{it})
- O método MQO não é consistente para estimativa dos parâmetros, então utiliza-se o GLS

Dados em Painel – Efeitos Aleatórios

- A transformação envolvida neste procedimento é para subtrair uma média ponderada de y_{it} ao longo do tempo (isto é, apenas parte da média, em vez de toda a média como foi o caso para a estimativa de efeitos fixos)

- Uma “quase subtração”

$$y_{it}^* = y_{it} - \theta \bar{y}_i \quad \text{e} \quad x_{it}^* = x_{it} - \theta \bar{x}_i$$

- Onde \bar{y}_i e \bar{x}_i são as médias ao longo do tempo de y_{it} e x_{it} respectivamente

Dados em Painel – Efeitos Aleatórios

- θ é uma função da variância da observação do termo de erro, σ_v^2 , e da variância do termo de erro específico da entidade, σ_ϵ^2
- Onde \bar{y}_i e \bar{x}_i são as médias ao longo do tempo de y_{it} e x_{it} respectivamente

$$\theta = 1 - \frac{\sigma_v}{\sqrt{T\sigma_\epsilon^2 + \sigma_v^2}}$$

- Essa transformação será precisa o quanto for necessário para garantir que não existem correlações cruzadas nos termos de erro, mas, felizmente, ele pode ser automaticamente implementada por pacotes de software

Dados em Painel – Efeitos Aleatórios

- Assim como para o modelo de efeitos fixos, com os efeitos aleatórios também não é conceitualmente mais difícil permitir a variação de tempo do que é para permitir a variação transversal. No caso de variação no tempo, um erro específico do termo de período de tempo é incluído

$$y_{it} = \alpha + \beta x_{it} + \omega_{it}$$
$$\omega_{it} = \epsilon_t + v_{it}$$

- Semelhante ao caso anterior, variando tanto na transversal quanto no tempo

Efeito Fixo ou Aleatório?

- Costuma-se dizer que o modelo de efeitos aleatórios é mais apropriado quando as entidades da amostra pode ser considerada como tendo sido selecionada aleatoriamente da população, mas um modelo de efeito fixo é mais plausível quando as entidades incluídas na amostra constituem efetivamente toda a população

Efeito Fixo ou Aleatório?

- Tecnicamente, a transformação envolvida no procedimento GLS sob a abordagem de efeitos aleatórios não irá remover as variáveis explicativas que não variam ao longo do tempo, e, portanto, o seu impacto em y_{it} podem ser enumerados
- Além disso, uma vez que há menos parâmetros a serem estimados com o modelo de efeitos aleatórios (sem *dummies* ou transformação intragrupo para executar) e, portanto, graus de liberdade são guardados, o modelo de efeitos aleatórios deve produzir estimativa mais eficiente do que a abordagem de efeitos fixos

Efeito Fixo ou Aleatório?

- No entanto, a abordagem de efeitos aleatórios tem uma grande desvantagem que decorre do fato de que é válida somente quando o componente do termo de erro ω_{it} é não correlacionado com todas as variáveis explicativas
- Esta suposição é mais rigorosa do que o correspondente no caso de efeitos fixos, porque com efeitos aleatórios que, portanto, exigem tanto ϵ_t como ν_{it} para ser independente de todos os x_{it}
- Isso também pode ser visto como uma questão de saber se todas as variáveis não observadas omitidas (que foram autorizados pela tendo diferentes intercepto para cada entidade) não estão correlacionados com as variáveis explicativas incluídas. Se eles não são correlacionadas, uma abordagem de efeitos aleatórios pode ser utilizado, caso contrário, o modelo de efeitos fixos é preferível

Escolha entre os Modelos

- **Adaptação do Teste de Chow**

- Teste para escolher entre efeito fixo e *pols*
- H_0 : Utilizar *pols*

- **Teste Breusch-Pagan**

- Teste para escolher entre efeito aleatório e *pols*
- H_0 : Utilizar *pols*

- **Teste de Hausman**

- Para escolher entre efeito fixo e efeito aleatório
- H_0 : Utilizar efeito aleatório

Sequência de Procedimentos – Dados em Painel

- Obter o modelo POLS
- Obter o modelo EF
 - Utilizar adaptação do teste de Chow para escolher entre POLS e EF
- Obter o modelo EA
 - Utilizar teste de Breusch-Pagan para escolher entre POLS e EA
 - Utilizar teste de Hausman para escolher entre EF e EA

Sequência de Procedimentos – Dados em Painel

- Em caso de empate?
 - Os resultados sugerem que não existem diferenças significativas entre os modelos, portanto, qualquer um deles pode explicar adequadamente os resultados
- No final das contas, como escolher:
 - Análise entre amostra e população
 - O que se deseja analisar, efeitos dentro das empresas (efeitos fixos) ou indiscriminados? **A argumentação teórica pode ser mais convincente que os resultados práticos dos testes**

Obrigado pela Atenção!!!

Até a próxima aula

mbotelho@usp.br

www.marcelobotelho.com