

## 2ª Lista de Exercícios

Professor: Rafael Terra  
Inferência Causal (Econometria do Setor Público)

20 de dezembro de 2024

**Responda as questões de 1 a 6 abaixo.**

1. Use o banco da PNAD elaborado para a primeira lista de exercícios, mas inclua os outros membros do domicílio além do chefe.

- (a) Estime o seguinte modelo por OLS usando o comando *lm* do R e somente indivíduos com um número maior do que zero horas trabalhadas

$$\begin{aligned} \text{horas\_trabalho}_i = & \alpha + \beta_1 \cdot \text{PBF}_i + \beta_2 \cdot \text{idade}_i + \beta_3 \cdot \text{idade}_i^2 + \beta_4 \cdot \text{escolaridade}_i \\ & + \beta_5 \cdot \text{escolaridade}_i^2 + \beta_6 \cdot \text{sexo}_i + \beta_7 \cdot \text{negro\_pardo}_i + \beta_8 \cdot \text{casado}_i \\ & + \beta_9 \cdot \text{rural}_i + \beta_{10} \cdot \text{renda\_lchefe}_{pc_i} + \varepsilon \end{aligned} \quad (1)$$

Quais problemas você apontaria nessa estimação? Que hipóteses seriam necessárias para que o efeito estimado refletisse o verdadeiro efeito do programa?

- (b) Agora estime o mesmo modelo usando o estimador de Heckman (Heckit), em que o modelo de decisão de participação no mercado de trabalho contém variáveis *filhos\_l14*, *filhos\_m14* como determinantes exógenos da participação, além de todas as variáveis de controle mencionadas no exemplo anterior.
- (c) Interprete o coeficiente do efeito do tratamento (PBF) em específico.
- (d) Como esse modelo é melhor do que o modelo estimado por OLS no item *a*?
- (e) Interprete  $\rho$  e  $\sigma$  e o coeficiente da razão de Mills inversa. O que a sua direção e significância estatística implicam?
- (f) Em média, qual a correção do número de horas trabalhadas a mais ou a menos que devemos fazer nas horas trabalhadas da população ocupada no mercado formal

(indivíduo médio)? O que isso significa?

2. Usando o mesmo banco da PNAD, estime o efeito do tratamento sobre os tratados (ATT) usando *Propensity Score Matching*. Use o comando “machIT” com as mesmas covariadas usadas no modelo estimado por OLS, e o algoritmo de matching de 5 vizinhos mais próximos. Antes de usar o comando defina uma “semente aleatória” igual ao seu número de matrícula na UnB (Isso permite replicar exatamente os resultados cada vez que o comando for estimado novamente). Restrinja as observações ao suporte comum.
  - (a) Qual a hipótese de identificação do efeito do tratamento nesse caso? Isto é, o que precisamos supor para que o efeito estimado do tratamento sobre os tratados seja efetivamente o efeito verdadeiro?
  - (b) Estime o modelo Probit (ou Logit, ou ainda, o Modelo de Probabilidade Linear) que será usado na função MachIT. Interprete os coeficientes do efeito marginal.
  - (c) Interprete o efeito do tratamento sobre os tratados. O programa Bolsa Família aumenta ou reduz a oferta de trabalho?
  - (d) Faça um teste de diferenças de médias das covariadas entre tratados e controles antes e depois do pareamento. Você pode usar a função `summary` com os objetos criados pela função e fazer um “Love plot”, ou usar `wtd.table` (usa pesos, você pode usar uma variável com peso 1 para todos se quiser estimativas sem peso).  
**Obs. Note que eu fiz modificações nessa última função. Ela pode não funcionar em todos os contextos.**
  - (e) Mostre os histogramas dos propensity scores antes e depois do pareamento.
3. Você foi contratado por um Organismo Internacional para fazer uma Avaliação Randomizada Controlada (RCT) dos efeitos de um programa militar que estimula a disciplina e a competição entre os alunos. O programa é implementado por escola (todos os alunos se beneficiam), mas para efeito de avaliação só importam os alunos da 4ª série que participam da Prova Brasil. Use os dados pré-teste de Matemática da Prova Brasil de 2005 disponibilizados em anexo para ter uma noção das magnitudes dos desvios-padrão e da correlação intra-cluster. Responda:
  - (a) Qual deveria ser o tamanho da amostra se não há correlação intra-cluster, para uma Probabilidade de Erro tipo I igual a 5% (2, 5% em cada lado no teste bicaudal) e um Poder de teste igual a 90%, para detectar um Tamanho de Efeito Mínimo Detectável (MDES) igual a 2 pontos (escala SAEB)? Seu orçamento é ilimitado.

- (b) Ache também o tamanho da amostra se somente 80% dos designados ao tratamento são compliers. O ajuste no MDES nesse caso é feito multiplicando o MDES por  $\frac{1}{\%compliers}$ .
- (c) E se observamos 10% de atrito aleatório entre os compliers? Que ajuste (Inflação) precisa ser feito na amostra inicial para identificar o MDES de 2 pontos com o poder de 90% e  $\Pr(\text{erro tipo I})=5\%$ ?
- (d) Suponha que o seu orçamento agora é de R\$200.000. O custo médio para acompanhar um estudante do grupo de controle é R\$100,00 e o custo para acompanhar um estudante do grupo de tratamento é R\$50,00. A restrição é *binding* para o MDES e poder especificados no item “a”? Caso seja, qual MDES poderíamos identificar dados o Poder e a  $\Pr(\text{Erro tipo I})$  definidos no item “a”. Se quisermos um MDES de 2 pontos, qual seria o poder do teste tudo mais constante?
- (e) Use os dados fornecidos para verificar a correlação intra-cluster por escola. Em uma randomização por cluster (sem limitações orçamentárias), quais são os números de clusters/escolas de tratamento e de controle suficientes para detectar um MDES igual a 2 pontos com uma Probabilidade de Erro tipo I igual a 5% e um Poder de teste igual a 90%?
- (f) Para o exemplo anterior, calcule o Efeito Desenho, que dá a dimensão do tamanho da amostra adicional da randomização por grupos frente a uma randomização feita no nível individual para detectar um MDES de 2 pontos com poder igual a 90% e  $\Pr(\text{erro tipo I})=5\%$ .
- (g) Considerando a randomização por escola, e o orçamento limitado de R\$200.000, com custo de R\$100,00 para acompanhar um estudante do grupo de controle, e igual a R\$50,00 para acompanhar um estudante do grupo de tratados, que ajustes teremos que fazer no MDES tudo mais constante? E se quisermos manter o MDES, que ajuste deve ser feito no Poder do teste?
4. Considere os tamanhos de amostra obtidos quando as escolas constituem os clusters, como no exercício anterior. Considere que não haja limitação orçamentária, mas você também não quer desperdiçar recursos, por isso irá trabalhar com uma amostra suficiente para lhe garantir um poder de 90%, uma  $\Pr(\text{Erro tipo I})=5\%$  e um MDES igual a 2. Defina a semente aleatória igual ao seu número de matrícula da Unb. Faça o seguinte:
- (a) Randomize as escolas tratadas e de controle usando um banco de escolas e crie uma

variável “treated”. Junte essa variável com os microdados individuais. Faça um teste de balanceamento do desempenho dos alunos em matemática (pré-tratamento) entre tratados e controles.

- (b) Faça um teste de balanceamento (entre tratados e controles) da média de alunos em matemática por UF.
  - (c) Volte ao banco por escolas e randomize entre UFs de forma estratificada e após definidas as escolas tratadas junte esse banco de status de tratamento com os microdados individuais. Faça novamente um teste de balanceamento do desempenho dos alunos em matemática por UF. Melhorou o balanceamento?
5. Explique as motivações de cada tipo de desenho de experimentos. Quando cada desenho é mais adequado? Quando é menos adequado?
- (a) Aleatorização Simples,
  - (b) Escalonamento do Tratamento (Phase-In),
  - (c) Randomização Intra-Grupo (Within-Group),
  - (d) Excesso de Inscrições (Oversubscription) e
  - (e) Encouragement Design (Desenho de Incentivos).
6. O estado do Texas expandiu a capacidade operacional de suas prisões em 1993. Por três anos - 1993, 1994 e 1995 - a capacidade operacional expandiu em 35% por ano, fazendo dobrar a capacidade ao final do período. Essa expansão é um experimento natural para verificarmos os efeitos sobre o encarceramento de negros (Variável dependente). Use dados dos estados americanos no arquivo “texas13.dta”.
- a Proponha um modelo com uso de controle sintético (suas potenciais covariadas) e justifique.
  - b Obtenha o gráfico “path.plot” de tendência do Texas e do Texas sintético em termos de taxa de encarceramento de negros. Interprete.
  - c Obtenha o gráfico “gaps.plot” de diferença entre Texas e Texas sintético em termos de taxa de encarceramento de negros. Interprete.
  - d Obtenha o gráfico de gaps de placebos usando as funções “generate.placebos” e “plot\_placebos”. Interprete.
  - e Calcule a razão entre o Erro de Previsão Quadrático Médio (MSPE) pós e pré tra-

- tamento para o Texas e Placebos usando “mspe.test” – inclusive com o p-valor–e “mspe.plot”. Interprete.
- f Calcule o efeito médio do tratamento sobre os tratados usando a função “multiple.synth” e “plac.dist” para gerar o gráfico.
- g Calcule o intervalo de confiança do “gap” plot usando a função “SCM.CS” de Firpo & Possebom.

Tabela 1: Variáveis do Banco

statefip	id do estado
year	ano
bmprison	presos negros
wmprison	presos brancos
bmpop	população de homens negros
wmpop	população de homens brancos
alcohol	consumo de alcool
income	renda
ur	unemployment rate
poverty	percentual de pobres
black	percentual de negros
perc1519	percentual de adolescentes
aidscapita	taxa de contaminação por HIV
state	estado
bmprate	taxa de encarceramento negros
wmprate	taxa de encarceramento brancos