



Instruções

- Essa lista de exercícios é parte da avaliação do curso Avaliação de Políticas Públicas e corresponde a 30% da nota.
- A lista é individual.
- Cada aluno deve entregar um arquivo pdf com suas respostas e um arquivo de script do R com os comandos que usou para resolver as questões.
- No script, cada questão deve ser indicada por #QNÚMERO (ex.: # Q1, # Q2) e sub-itens devem ser indicados por #LETRA (ex.: #a, #b, #c) ou #LETRA.NÚMERO (ex. #a.1, #a.2).
- A lista deve ser entregue por [e-mail](#) com o assunto “Lista APP 2024 SEU NOME” até às 23:59 do dia 31/12/2024.

Questões

Q1. O banco de dados **PIAA_2017-2018.xlsx** contém informações desidentificadas sobre notas dos alunos e frequência na monitoria da disciplina Matemática I no departamento de economia nos anos 2017 e 2018. A tabela a seguir lista as variáveis contidas no banco de dados.

Variável	Definição
id	Identificador do aluno
Semestre	Semestre no qual a disciplina foi cursada
Nota	Nota do aluno
Faltas	Número de faltas na disciplina
Situação	Situação final do aluno na disciplina
Presença	Número de seções de monitoria que o aluno foi.
Período	Número de períodos matriculado
Sexo	Masculino ou feminino

- Usando o pacote “TableOne” do R, faça um teste de balanceamento da amostra de acordo com uma variável que indica se o aluno foi a mais do que 4 seções de monitoria (tratamento). Faça esse teste sobre as seguintes variáveis: Faltas, Sexo Masculino, variáveis dummies para os semestres e variável dummy para alunos do primeiro período. O que você conclui sobre o balanceamento dos alunos entre tratados e não tratados?
- Faça uma regressão da nota dos alunos contra uma variável indicativa de que o aluno foi a mais do que 4 seções de monitoria (tratamento). Interprete o resultado. O que seria necessário para que esse resultado possa ser interpretado como um efeito causal? Essas condições são válidas para esses dados? Por quê? Em seguida, repita essa regressão adicionando ao modelo as variáveis



Faltas, variável dummy para Sexo Masculino, variáveis dummies para o semestre no qual a disciplina foi cursada e variável dummy para alunos do primeiro período. O que acontece com o valor do coeficiente da presença na monitoria em relação ao modelo do item “a”? Essa estimativa seria mais próxima de um efeito causal? Por quê? Para as regressões, use os desvios padrões robustos a heterocedasticidade.

- c. Usando o pacote “MatchIt”, faça o pareamento dos tratados e controles usando os seguintes critérios: pareamento exato, vizinho mais próximo por distância de Mahalanobis e vizinho mais próximo por escore de propensão usando o modelo logit. Para cada um deles, verifique o balanceamento entre tratados e controles e verifique se o pareamento foi bem-sucedido.
- d. Para cada amostra pareada, faça uma regressão da nota dos alunos contra uma variável indicativa de que o aluno foi a mais do que 4 seções de monitoria (tratamento). Usando o pacote “Sargazer”, organize os resultados das cinco regressões em uma tabela arrumada, interprete-os e compare os resultados da amostra pareada com o que você encontrou nas regressões sem pareamento. O que você conclui sobre o efeito da monitoria sobre as notas dos alunos?

Q2. Para essa questão, use o banco de dados **minwage.xlsx**. Ele contém informações coletadas por [Card & Krueger \(1994\)](#) para restaurantes de fast food nos estados de New Jersey (NJ) e Pennsylvania (PA) em duas rodadas de entrevistas: em Março e Novembro/Dezembro de 1992. Em Abril daquele ano, NJ aumentou seu salário mínimo de \$4,25 para \$5,05 por hora. Em um artigo bastante discutido, Card e Krueger usaram esse experimento natural para avaliar o efeito de um aumento do salário-mínimo sobre o emprego (um dos exemplos clássicos de controle de preços usados em livros texto de economia). Você vai usar esse banco de dados para replicar parte do estudo de Card e Krueger.

No que segue, variáveis cujo nome que termina em “2” se referem a segunda rodada da pesquisa. As variáveis **fte** e **fte2** se referem a emprego equivalente em horário integral, ou seja, a soma do número de empregados em horário integral com metade do número de empregados que trabalham em meio expediente, excluindo gerentes, **dfte** se refere a mudança em fte entre a primeira e a segunda entrevista ($dfte = fte2 - fte$); **dw** se refere a mudança no salário inicial dos funcionários entre a primeira e a segunda entrevistas; **state** é uma variável dummy para lojas localizadas em NJ e **sample** é uma variável dummy que assume o valor 1 se dados de salário e emprego estavam disponíveis na primeira e segunda entrevista. Na análise a seguir, você deve usar apenas as observações para as quais **sample** é igual a 1.

- a. Calcule o salário inicial médio (**wage_st**) separadamente para restaurantes em NJ e PA, em cada uma das rodadas de entrevistas.
 1. Calcule a diferença nos salários médios para cada estado entre a primeira e a segunda entrevista;
 2. Calcule a diferença entre as diferenças para NJ e PA que você calculou acima.
 3. Qual a interpretação dessa estimativa de diferença em diferenças para o efeito sobre os salários? Sob que condições essa conta fornece uma estimativa válida do aumento do salário-mínimo sobre os salários nos restaurantes de fast food?
 4. Interprete seu resultado.



- b. Repita o mesmo exercício de (a) para a variável **fte**. Qual o impacto do aumento do salário-mínimo sobre o emprego nos restaurantes de NJ?
- c. A metodologia de diferenças em diferenças (DD) também pode ser implementada por meio da seguinte regressão:

$$Y_{ist} = \alpha + \beta TREAT_{is} + \gamma POST_t + \delta_{DD}(TREAT_{is} \times POST_t) + \epsilon_{ist},$$

onde Y_{ist} representa emprego no restaurante i , no estado s e período t , $TREAT_{is}$ é um indicador para a área de tratamento (NJ ou restaurantes de baixo salário em NJ), $POST_t$ é um indicador do período de tratamento (Novembro/Dezembro) e $TREAT_{is} \times POST_t$ é a interação entre essas duas dummies. Note que a regressão usa dados para *restaurantes individuais*, ao invés de dados para o estado, como em (a) e (b).

1. Estime a regressão acima para salários e emprego. Como as estimativas diferem dos resultados que você encontrou em (a) e (b)?
 2. A implementação por regressão permite que você inclua controles adicionais. Estime as regressões para salários e emprego incluindo uma variável se indicativa de que a loja é própria ou franquia (**owned**) e dummies para as cadeias (**chain**) de restaurantes.
 3. Coloque os resultados que você obteve em 1 e 2 lado a lado na mesma tabela. Faça uma tabela para salário e outra para emprego.
 4. Os seus resultados mudaram quando incluiu as dummies para restaurantes? Essa mudança era esperada? Explique por quê.
- d. Uma alternativa a comparação entre restaurantes em NJ e PA seria comparar restaurantes em NJ que pagam salários altos vs. restaurantes que pagam salários mais baixos antes do aumento do salário-mínimo. Restrinja sua amostra para os restaurantes de NJ apenas.
1. Você esperaria que as suposições para a metodologia DD sejam mais fáceis de serem defendidas na comparação de restaurantes em NJ do que na comparação de restaurantes em NJ vs. restaurantes em PA?
 2. Construa uma variável que indique os restaurantes que pagam menos que \$5 antes do aumento do salário-mínimo. Use uma regressão para calcular a estimativa DD do efeito do aumento do salário-mínimo sobre emprego e salários. Qual impacto você encontra para cada uma dessas variáveis usando os restaurantes de NJ apenas?
 3. Compare as estimativas obtidas com o que você obteve anteriormente na parte (c). Os resultados são muito diferentes?
- e. Repita a regressão em (d) usando agora os restaurantes em PA.
1. Compare os resultados que você encontrou para PA com os resultados que você encontrou para NJ.
 2. Faça um teste estatístico para a hipótese que o coeficiente para a variável de baixo salário tenha o mesmo valor em NJ e PA.
 3. Por que verificar se o aumento do salário-mínimo em NJ teve impacto em PA pode ser uma maneira de confirmar que a metodologia produz resultados sensatos? O que você pode concluir com essa comparação?



Q3. Para essa questão, use o banco de dados **Guns.xlsx**. Uma descrição detalhada dos dados está contida no arquivo **Guns_Description.pdf**. Alguns estados dos EUA promulgaram leis que permitem que os cidadãos carreguem armas escondidas. Essas leis, conhecidas como “shall-issue laws”, instruem as autoridades locais a emitirem uma permissão de armas ocultas a todos os requerentes que sejam cidadãos, sejam mentalmente competentes e não tenham sido condenados por crime doloso (alguns estados têm algumas restrições adicionais). Os proponentes argumentam que, se mais pessoas portarem armas ocultas, o crime diminuirá, porque os criminosos são dissuadidos de atacar outras pessoas. Oponentes argumentam que o crime aumentará por causa do uso acidental ou espontâneo da arma. Nessa questão, você usará os dados de [Ayres & Donohue \(2003\)](#) para estimar o efeito das leis de armas ocultas em crimes violentos.

- a. Estime uma regressão de $\ln(\text{vio})^1$ contra shall e uma regressão de $\ln(\text{vio})$ contra shall, `incarc_rate`, `density`, `avginc`, `pop`, `pb1064`, `pw1064` e `pm1029`.
 - i. Interprete o coeficiente de shall na segunda regressão. Essa estimativa pode ser considerada um efeito causal? Por que?
 - ii. As variáveis de controle adicionadas na segunda regressão mudam muito a magnitude do efeito das leis de armas ocultas na primeira regressão? Mudam a significância estatística do coeficiente estimado? Você espera que essa estimativa se aproxima mais de um efeito causal do que a anterior?
- b. Os resultados se alteram se você adicionar efeitos fixos para estados e períodos de tempo (TWFE)? Qual dos resultados é mais crível para estimar um efeito causal e por quê?
- c. Usando o modelo de estudos de evento, tente fornecer alguma evidência da validade da hipótese de tendências paralelas entre tratados e controles.
- d. Usando a nomenclatura de Callaway e Sant’Anna (2021), quantos grupos de tratados existem no banco de dados?
- e. Usando o método de Callaway e Sant’Anna (2021) estime os diferentes Efeitos Médios do Programa para Grupo-Período. Existe evidência de efeitos heterogêneos por grupo-período?
- f. A partir da resposta anterior, obtenha o efeito médio agregado total e compare esse resultado como o que você obteve na letra “a”.
- g. Repita a análise usando $\ln(\text{rob})$ e $\ln(\text{mur})$ no lugar de $\ln(\text{vio})$. Coloque seus resultados em tabelas arrumadas.
- h. Baseado na sua análise, que conclusões você tiraria a respeito dos efeitos das leis de armas ocultas sobre as taxas de criminalidade? Use uma linguagem clara e acessível, de maneira que até o deputado [Marcel Van Hatten](#) (Novo/RS) entenda.

Q4. Para essa questão, use o banco de dados **CARD.xlsx**. Uma descrição detalhada dos dados está contida no arquivo **CARD_Description.pdf**. [Card \(1995\)](#) estimou o retorno da educação para

¹ \ln é o logaritmo natural da variável.



homens jovens no ano de 1976 usando como instrumento para educação uma variável *dummy* indicando se a pessoa havia crescido próximo a uma faculdade com cursos de graduação de 4 anos. Nessa questão você vai repetir alguns passos da análise e realizar algumas extensões.

- a. Quais as condições que a proximidade de uma faculdade com cursos de graduação de 4 anos deve satisfazer para ser um instrumento válido para educação? Qual dessas condições pode ser testada empiricamente e qual o procedimento para isso?
- b. Estime equações para o efeito da educação sobre o logaritmo dos rendimentos usando MQO e incluindo as variáveis: **educ**, **exper**, **expersq**, **black**, **south** e **smsa**. A seguir, adicione ao primeiro modelo as variáveis indicadoras das regiões (**reg661** – **reg668**) e **smsa66**. Em terceiro lugar, adicione ao segundo modelo as variáveis **fatheduc** e **motheduc**. Por fim, adicione ao terceiro modelo as variáveis **momdad14** e **sinmon14**. Organize os quatro modelos em uma tabela exibindo apenas os coeficientes das variáveis do primeiro modelo e indicando as variáveis incluídas nos demais modelos. Interprete os resultados obtidos para a variável **educ**, mostrando o que significa essa magnitude e o quão diferente são as estimativas nas diferentes especificações.
- c. Qual crítica pode ser feita em relação às estimativas para o efeito da educação estimado na equação b? Essa estimativa será viesada? Se for, em que direção deve ocorrer esse viés?
- d. Para cada modelo estimado na letra b, estime o modelo correspondente à forma reduzida. Organize os quatro modelos em uma tabela exibindo apenas os coeficientes das variáveis do primeiro modelo e indicando as variáveis incluídas nos demais modelos. Qual variável dos modelos da letra b foi substituída? Interprete os resultados obtidos para essa variável incluída, mostrando o que significa essa magnitude e o quão diferente são as estimativas nas diferentes especificações.
- e. Para cada modelo estimado na letra b, estime o modelo correspondente ao primeiro estágio. Organize os quatro modelos em uma tabela exibindo apenas os coeficientes das variáveis do primeiro modelo e indicando as variáveis incluídas nos demais modelos. Quais variáveis dos modelos da letra b foram substituídas? Interprete os resultados obtidos para essa variável incluída, mostrando o que significa essa magnitude e o quão diferente são as estimativas nas diferentes especificações.
- f. Para cada modelo estimado na letra b, estime agora usando a variável **nearc4** como instrumento para educação. Organize os quatro modelos em uma tabela exibindo apenas os coeficientes das variáveis do primeiro modelo e indicando as variáveis incluídas nos demais modelos. Interprete os resultados obtidos para a variável **educ**, mostrando o que significa essa magnitude e o quão diferente são as estimativas nas diferentes especificações. Compare essas estimativas às obtidas na letra b, comentando se a diferença obtida é consistente com a sua resposta na letra c.
- g. Aponte e discuta brevemente ao menos duas razões pelas quais a hipótese de exclusão da proximidade de uma faculdade de 4 anos da equação de salários pode ser violada.



- h. Faça uma regressão da variável **IQ** contra a variável **nearc4** para verificar se existe correlação entre o coeficiente de IQ da pessoa com a proximidade de uma faculdade. O que você verifica? O que isso significa para a hipótese de exclusão do instrumento **nearc4**?
- i. Inclua no modelo da letra h as variáveis regionais (**smas66** e **reg661 – reg668**). **QI** e **nearc4** apresentam correlação nessa especificação? Tendo em vista esse achado, o que você conclui sobre a importância de incluir as variáveis regionais de 1966 na equação para o logaritmo do salário?

ESCOLHA UMA DAS DUAS QUESTÕES A SEGUIR.

Q5. Esta questão é motivada pelo artigo "[The Persistent Effects of Peru's Mining Mita](#)" de Melissa Dell (2010). A autora busca compreender os efeitos persistentes da Mita, um sistema extrativista de trabalho forçado implementado pelo império espanhol no Peru e na Bolívia. Dell avalia se essa instituição, que foi encerrada há duzentos anos, ainda tem efeitos sobre o consumo e a renda das famílias nos dias de hoje. O artigo está disponível no link; familiarize-se com ele antes de começar. Você precisará usar o arquivo **mitaData.csv** para realizar o exercício a seguir.

- a. Dell utiliza um design de regressão descontínua para comparar os resultados entre pessoas que vivem em distritos que participaram da antiga Mita e aquelas que vivem em distritos que não participaram. Discuta brevemente a estratégia de identificação do efeito causal ressaltando:
 - i. Por que a comparação dos distritos que tiveram o sistema de Mita com os que não tiveram é inadequada?
 - ii. Qual a hipótese de identificação usada pela autora?
 - iii. Qual a equação estimada e qual variável corresponde a estratégia usada pela autora?
- b. Agora vamos reproduzir o resultado principal de Dell. Dadas as variáveis de longitude e latitude x e y , construa x^2 , y^2 , xy , x^3 , y^3 , x^2y e xy^2 . Faça uma regressão do logaritmo do consumo equivalente dos domicílios (2001) (**lhhequiv**) em relação à variável **Mita** (**pothuan_mita**), todos os termos polinomiais, elevação (**elv_sh**), inclinação média (**mean_slope**), número de bebês (**infants**), crianças (**children**), adultos (**adults**) e efeitos fixos dos segmentos de fronteira (**fe4_1**, **bfe4_2**, **bfe4_3**). [Agrupe os erros padrão por distrito](#). Execute a regressão de 3 maneiras: primeiro, para observações em que a distância até a fronteira da Mita (**d_bnd**) seja menor que 100 km; depois, quando for menor que 75 km; e, por último, quando for menor que 50 km. Reporte os resultados em uma única tabela e comente sobre os efeitos encontrados, considerando um nível de significância de 5%.
- c. Execute as mesmas regressões de antes, mas, em vez de termos polinomiais em longitude e latitude, utilize um polinômio cúbico na distância até Potosí (**dspot**). Ou seja, inclua a primeira, a segunda e a terceira potência dessa variável nas regressões. Novamente, agrupe os erros padrão por distrito e execute a regressão de 3 maneiras: primeiro, para observações em que a distância até a fronteira da Mita (**d_bnd**) seja menor que 100 km;



depois, quando for menor que 75 km; e, por último, quando for menor que 50 km. Reporte os resultados em uma única tabela e comente sobre os efeitos encontrados, considerando um nível de significância de 5%.

Q6. Essa questão replica a análise contida no artigo [“Evaluating the Effect of Homicide Prevention Strategies in São Paulo, Brazil: A Synthetic Control Approach”](#) de Danilo Freire (2010). O autor busca mensurar o efeito causal agregado de políticas públicas sobre as taxas de homicídios em São Paulo. O artigo está disponível no link; familiarize-se com ele antes de começar. Você precisará usar o arquivo **df.Rdata** para realizar o exercício a seguir.

Descrição da base: O painel de dados contém informações de 1990 a 2009, para os 26 estados e o Distrito Federal. Os seguintes estados constam na base, mas não são utilizados na análise: Bahia e Sergipe.

- Formato: dataframe - painel.
- 1 unidade tratada: nome do estado: São Paulo.
- 24 regiões de controle.
- 20 anos (1990:2009).

Identificadores:

- abbreviation: sigla do estado
- code: código do estado
- state: nome do estado
- year: ano

Variável dependente: homicide.rates (número de homicídios por 100.000 habitantes),

Preditores: 1990:1998:

- homicide.rates: taxa de homicídios (número de homicídios por 100.000 habitantes);
- state.gdp.capita: PIB per capita estadual (em reais de 2010);
- state.gdp.growth.percent: crescimento do PIB estadual (percentual, em reais de 2010);
- population.projection.ln: logaritmo natural da população (projeções anuais da população do estado);
- proportion.extreme.poverty: população vivendo em extrema pobreza (percentual da população do estado que não atinge a ingestão mínima de 2.000 calorias por dia);



- years.schooling.imp: anos de escolaridade (número médio de anos de instrução formal em estabelecimentos de ensino - homens e mulheres com 25 anos ou mais). As observações ausentes foram imputadas com interpolação linear;
- gini.imp: índice de Gini (medida de desigualdade que varia de 0 a 1, onde 0 é o mais igual e 1 o mais desigual). As observações ausentes foram imputadas com interpolação linear.

Para organizar os dados, use os seguintes parâmetros na função `dataprep()`:

- `predictors.op = "mean";`
- `time.predictors.prior = c(1990:1998);`
- `time.optimize.ssr = c(1990:1998);`
- `time.plot = c(1990:2009);`
 - a. Discuta brevemente a estratégia de identificação empregada pelo autor ressaltando suas vantagens em relação a estudos de caso comparativos qualitativos. Indique também uma crítica a abordagem seguida pelo autor.
 - b. Obtenha os pesos para as covariadas (preditores) e para os estados doadores para construção do São Paulo sintético. Organize os resultados como na tabela 1 do artigo.
 - c. Elabore uma tabela que compara os valores dos preditores pré-tratamento para a unidade tratada, a unidade sintética de controle e todas as unidades na amostra (Tabela 2, no artigo). Comente sobre a semelhança entre SP real e seu controle em cada uma das dimensões.
 - d. Elabore dois gráficos ilustrativos dos resultados: um mostrando a evolução da trajetória da taxa de homicídios em SP e em seu controle (figura 2 do artigo) e outro mostrando o gap entre a taxa de homicídios em SP e seu controle (figura 3 do artigo) e comente brevemente os resultados.
 - e. Escolha o estado não tratado com maior peso no controle sintético e faça um teste de placebo tentando encontrar algum efeito da política sobre esse estado. Comente seu resultado ressaltando sua implicação para a análise.
 - f. Aumente essa análise para todos os demais estados e coloque os resultados em um gráfico ilustrativo (figura 6 do artigo) fazendo o teste de permutação sugerido por [Abadie Diamond e Hainmueller \(2010\)](#). Comente brevemente os resultados.