

Lista de Exercícios 1

Revisão de Econometria

Parte 1 da disciplina – 50 questões

Nome: _____ Data: _____

Orientações. Apresente os cálculos, as hipóteses e os argumentos econômicos utilizados. Quando necessário, indique claramente as condições de identificação, os procedimentos de inferência e as limitações da interpretação causal. Nas questões computacionais, descreva também como implementaria a estimação em Stata ou Python.

A. Modelo linear e inferência

1. Considere o modelo salarial

$$\ln(w_i) = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i.$$

Explique a interpretação econômica de cada coeficiente. Derive o efeito marginal da experiência sobre o salário e determine como calcular o nível de experiência no qual o salário previsto é máximo.

2. Partindo do modelo matricial

$$y = X\beta + u,$$

derive o estimador de mínimos quadrados ordinários

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Indique as condições necessárias para que a solução seja única.

3. Demonstre que o estimador de mínimos quadrados ordinários é não viesado quando

$$\mathbb{E}(u \mid X) = 0.$$

Explique por que a ausência de correlação amostral entre X e os resíduos estimados não demonstra que essa condição seja satisfeita na população.

4. Em uma regressão do salário sobre escolaridade, a habilidade individual é omitida. Suponha que habilidade esteja positivamente correlacionada tanto com escolaridade quanto com salário. Utilize a fórmula do viés de variável omitida para determinar o provável sinal do viés em $\hat{\beta}_{educ}$.
5. Uma regressão estimada apresenta os seguintes resultados:

$$\widehat{\ln(w)} = 1,80 + 0,085 educ + 0,032 exper - 0,0005 exper^2.$$

Calcule: (a) o retorno aproximado de um ano adicional de escolaridade; (b) o efeito marginal de um ano adicional de experiência para alguém com 10 anos de experiência; e (c) o nível de experiência que maximiza o salário previsto.

6. Explique a diferença entre homocedasticidade e heterocedasticidade. Discuta quais propriedades do estimador de mínimos quadrados permanecem válidas sob heterocedasticidade e quais procedimentos de inferência precisam ser modificados.

7. Um pesquisador encontrou $\hat{\beta} = 0,12$, com erro-padrão convencional igual a 0,03 e erro-padrão robusto igual a 0,05. Teste $H_0 : \beta = 0$ ao nível de 5% utilizando os dois erros-padrão. Explique por que as conclusões podem diferir.

8. Em uma equação de salários, estime o modelo

$$\ln(w_i) = \beta_0 + \beta_1 \text{mulher}_i + \beta_2 \text{pos}_i + \beta_3 (\text{mulher}_i \times \text{pos}_i) + u_i.$$

Apresente a interpretação de β_1 , β_2 e β_3 . Escreva o salário previsto para cada uma das quatro combinações possíveis entre sexo e pós-graduação.

9. Discuta os problemas de utilizar R^2 como único critério de escolha entre modelos. Compare o significado do R^2 , do R^2 ajustado e de critérios de informação, como AIC e BIC.
10. Considere uma regressão com dados de trabalhadores pertencentes a diferentes empresas. Explique por que os erros dos trabalhadores de uma mesma empresa podem estar correlacionados e por que seria adequado utilizar erros-padrão agrupados no nível da empresa.

B. Endogeneidade, variáveis instrumentais e 2SLS

11. Defina endogeneidade. Apresente um exemplo relacionado ao mercado de trabalho para cada uma das seguintes fontes: (a) variável omitida; (b) simultaneidade; (c) erro de mensuração; e (d) seleção amostral.
12. Um pesquisador pretende utilizar a distância entre a residência do indivíduo e a universidade mais próxima como instrumento para escolaridade. Discuta separadamente: (a) a condição de relevância; (b) a restrição de exclusão; e (c) possíveis ameaças à validade do instrumento.
13. Considere o modelo estrutural

$$y_i = \beta_0 + \beta_1 D_i + \gamma' X_i + u_i,$$

no qual D_i é endógeno e Z_i é um instrumento. Escreva as duas etapas do estimador de mínimos quadrados em dois estágios e explique por que não se devem usar, numa implementação manual, os erros-padrão OLS comuns da segunda regressão.

14. Na primeira etapa de uma estimação por variáveis instrumentais, o coeficiente do instrumento é estatisticamente significativo, mas a estatística F dos instrumentos excluídos é baixa. Explique o problema de instrumentos fracos e suas consequências para viés, precisão e inferência.
15. Suponha que um programa de treinamento seja oferecido apenas a trabalhadores cujo índice de elegibilidade seja superior a determinado valor. Discuta como a regra de elegibilidade poderia ser utilizada como instrumento para a participação efetiva no programa.
16. Explique a interpretação do estimador de variáveis instrumentais como efeito médio local do tratamento (LATE). Defina os grupos de sempre participantes, nunca participantes, *compliers* e *defiers* e apresente a hipótese de monotonicidade.

C. Dados em painel

17. Considere o modelo

$$y_{it} = \beta x_{it} + c_i + u_{it}.$$

Explique a diferença entre c_i e u_{it} . Mostre por que uma regressão Pooled OLS será inconsistente quando c_i estiver correlacionado com x_{it} .

18. Derive a transformação *within* utilizada pelo estimador de efeitos fixos. Explique por que variáveis invariantes no tempo, como sexo ou local de nascimento, não podem ter seus coeficientes identificados em um modelo de efeitos fixos individuais.
19. Compare os estimadores de efeitos fixos e efeitos aleatórios. Apresente a principal hipótese adicional exigida pelo estimador de efeitos aleatórios e explique a lógica do teste de Hausman.
20. Um pesquisador possui dados anuais de salários e sindicalização de trabalhadores entre 2015 e 2025. Explique as diferenças substantivas entre: (a) efeitos fixos de indivíduo; (b) efeitos fixos de ano; (c) efeitos fixos de empresa; e (d) tendências específicas por setor.

D. Máxima verossimilhança e modelos de escolha discreta

21. Defina função de verossimilhança, log-verossimilhança, vetor score e matriz Hessiana. Explique por que a maximização da log-verossimilhança produz o mesmo estimador que a maximização da verossimilhança.
22. Para uma variável binária y_i , derive a função de verossimilhança de uma observação Bernoulli:

$$L_i = p_i^{y_i} (1 - p_i)^{1-y_i}.$$

Em seguida, escreva a log-verossimilhança para uma amostra de n observações independentes.

23. Compare os modelos de probabilidade linear, Logit e Probit. Discuta: (a) intervalo das probabilidades previstas; (b) forma funcional; (c) heterocedasticidade; (d) interpretação dos coeficientes; e (e) facilidade de estimação.
24. Em um modelo Logit, o coeficiente de escolaridade é 0,20. Calcule a razão de chances associada a um ano adicional de escolaridade. Explique por que isso não significa necessariamente que a probabilidade do evento aumente em 20 pontos percentuais.
25. Considere o modelo Probit

$$P(\text{empregado}_i = 1 \mid X_i) = \Phi(X_i' \beta).$$

Derive o efeito marginal de uma variável contínua x_k . Explique por que o efeito marginal depende dos valores das demais variáveis explicativas.

26. Diferencie: (a) efeito marginal no indivíduo médio; (b) efeito marginal médio; e (c) mudança discreta de uma variável binária. Discuta qual medida costuma ser mais informativa em uma amostra heterogênea.
27. Explique a estrutura de um modelo Logit ou Probit Ordenado para uma variável de satisfação no trabalho com categorias “baixa”, “média” e “alta”. Defina a variável latente e os pontos de corte. Qual restrição é conhecida como hipótese de linhas paralelas?
28. Explique a estrutura do Logit Multinomial para a escolha entre emprego formal, emprego informal, trabalho por conta própria e desemprego. Apresente a hipótese de independência das alternativas irrelevantes e um caso em que ela provavelmente seria inadequada.

E. Avaliação de políticas públicas

29. Utilize a notação de resultados potenciais $Y_i(1)$ e $Y_i(0)$ para definir: (a) efeito individual do tratamento; (b) efeito médio do tratamento (ATE); (c) efeito médio sobre os tratados (ATT); e (d) o problema fundamental da inferência causal.

30. Um programa de qualificação profissional foi implementado em um grupo de municípios, mas não em outro. Os salários médios foram:

	Antes	Depois
Tratados	2.000	2.500
Controles	2.100	2.300

- (a) Calcule o estimador de diferenças em diferenças. (b) Apresente a regressão correspondente. (c) Explique a hipótese de tendências paralelas. (d) Discuta como estudos de evento, testes de pré-tendência, matching e variáveis instrumentais poderiam complementar a análise.

F. Questões avançadas

31. Considere o modelo $y = X_1\beta_1 + X_2\beta_2 + u$. Enuncie e demonstre o teorema de Frisch–Waugh–Lovell. Mostre como ele permite interpretar $\hat{\beta}_2$ como o coeficiente da regressão dos resíduos de y em X_1 sobre os resíduos de X_2 em X_1 . Relacione o resultado à ideia de “manter constantes” as demais covariadas.
32. Um modelo contém quatro coeficientes de interesse, $\beta_2, \dots, \beta_5$. Formule um teste conjunto de $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ usando uma matriz de restrições R . Derive a estatística de Wald robusta e explique por que quatro testes t separados não são equivalentes ao teste conjunto.
33. Suponha que a variância condicional seja $\text{Var}(u_i | x_i) = \sigma^2 h_i$, com $h_i > 0$ conhecido. Derive o estimador de mínimos quadrados ponderados. Em seguida, explique como construir um FGLS quando h_i é desconhecido e discuta por que erros-padrão robustos podem ser preferíveis se o modelo da variância estiver mal especificado.
34. Uma política foi implementada em apenas 12 estados, com tratamento definido no nível estadual. Discuta por que a inferência assintótica usual com erros agrupados pode ser pouco confiável. Compare correções de graus de liberdade, *wild cluster bootstrap* e testes de randomização como alternativas.
35. Há uma variável endógena e três instrumentos excluídos. Explique as condições de ordem e de posto para identificação. Formule o teste de sobreidentificação de Hansen–Sargan e mostre por que sua rejeição não identifica automaticamente qual instrumento é inválido.
36. Em uma aplicação, o pesquisador dispõe de 80 instrumentos construídos a partir de interações, mas apenas 500 observações. Explique o problema de muitos instrumentos, seus efeitos sobre 2SLS e testes de sobreidentificação, e compare as estratégias de redução do conjunto instrumental, LIML e *jackknife IV*.
37. Compare 2SLS e função de controle em um modelo no qual o desfecho é binário e o tratamento é endógeno. Escreva as etapas de uma abordagem de resíduos incluídos e discuta as hipóteses adicionais necessárias para interpretar causalmente os parâmetros de um Probit com tratamento endógeno.
38. Considere um painel dinâmico

$$y_{it} = \rho y_{i,t-1} + \beta x_{it} + c_i + u_{it}.$$

Explique o viés de Nickell do estimador de efeitos fixos quando T é pequeno. Apresente a lógica dos estimadores Arellano–Bond e sistema GMM, incluindo instrumentos internos, testes de autocorrelação e risco de proliferação de instrumentos.

39. Estados adotam uma política em anos distintos. Explique por que o estimador de efeitos fixos bidirecionais pode produzir médias ponderadas problemáticas quando os efeitos são heterogêneos ao longo do tempo ou entre coortes. Descreva uma estratégia moderna de diferenças em diferenças por grupos e períodos.
40. Especifique um estudo de evento com períodos relativos à adoção da política. Explique a normalização necessária, a interpretação dos coeficientes anteriores e posteriores ao tratamento e as limitações de usar a não rejeição das pré-tendências como prova de identificação.
41. Distinga desenhos de regressão descontínua *sharp* e *fuzzy*. Para ambos, apresente o parâmetro identificado, as condições de continuidade, os testes de manipulação e balanceamento e os principais trade-offs na escolha da largura de banda.
42. Um programa é atribuído segundo um escore de propensão estimado por Logit. Explique as hipóteses de independência condicional e suporte comum. Proponha procedimentos para avaliar balanceamento e sobreposição e discuta por que um bom ajuste preditivo do Logit não garante um bom desenho de matching.
43. Derive o estimador por ponderação pelo inverso da probabilidade de tratamento para o ATE. Em seguida, explique a propriedade de dupla robustez de um estimador que combina ponderação e regressão do resultado, indicando em que sentido ele continua consistente quando apenas um dos dois modelos está corretamente especificado.
44. Compare regressão quantílica e OLS na estimação de equações salariais. Formule a função objetivo da regressão quantílica, explique a interpretação de $\beta(\tau)$ e discuta por que diferenças entre quantis não devem ser automaticamente interpretadas como efeitos sobre indivíduos situados permanentemente nos mesmos pontos da distribuição.
45. Considere que salários são observados apenas para indivíduos que trabalham. Apresente o modelo de seleção de Heckman, derive o papel da razão inversa de Mills e discuta a importância de uma restrição de exclusão para distinguir identificação funcional de identificação econômica.
46. Em um painel com variável dependente binária, compare: modelo de probabilidade linear com efeitos fixos, Logit de efeitos fixos e Probit de efeitos aleatórios. Discuta parâmetros incidentais, perda de unidades sem variação no desfecho e hipóteses sobre heterogeneidade não observada.
47. Para escolhas entre emprego formal, informalidade e desemprego, proponha uma estrutura de Logit Aninhado que relaxe parcialmente a hipótese IIA. Defina os ninhos, interprete o parâmetro de dissimilaridade e explique como avaliaria se a estrutura proposta é coerente com o comportamento observado.
48. Mostre a relação entre matriz de informação, Hessiana esperada e variância assintótica do estimador de máxima verossimilhança sob especificação correta. Depois, derive a forma “sanduíche” da variância quase máxima verossímil sob possível má especificação.
49. Uma reforma altera simultaneamente elegibilidade e valor de um benefício. Proponha uma estratégia empírica capaz de separar, ao menos conceitualmente, o efeito da margem extensiva do efeito da margem intensiva. Discuta ameaças de antecipação, transbordamento, composição e equilíbrio geral.
50. Elabore o protocolo de replicação de um artigo empírico sobre mercado de trabalho usando PNAD Contínua ou RAIS. O protocolo deve incluir: definição da amostra, dicionário de variáveis, tratamento de valores ausentes, pesos, especificação principal, erros-padrão, testes de robustez,

versões de software, organização de códigos e critérios para distinguir replicação exata de extensão substantiva.