Lista de Exercícios 3

Testes de hipótese, mínimos quadrados generalizados e clusterização

Econometria I · PIMES/UFPE

Justifique todas as respostas.

---

1. **Metodologia de teste, estatística t e intervalos de confiança.**

   (a) Defina os **erros do tipo I e II**, o **tamanho** e o **poder** de um teste. Explique o compromisso entre tamanho e poder e o que significa, na prática, fixar o nível em 5%.

   (b) Para $H_0 : \beta_k = \beta_k^0$, construa a estatística $t = (\hat{\beta}_k - \beta_k^0)/\mathrm{se}(\hat{\beta}_k)$ e descreva a sua distribuição sob H_0: exata sob normalidade dos erros (A6) e assintótica em geral. O que, em cada caso, garante que a distribuição não dependa dos parâmetros desconhecidos?

   (c) Mostre a **dualidade teste–intervalo**: o intervalo de confiança de 95% para β_k é exatamente o conjunto de valores β_k^0 **não rejeitados** por um teste t bicaudal de 5%. Explique por que a leitura correta de "IC de 95%" é uma afirmação sobre o **procedimento** (cobertura em amostras repetidas), e não sobre a probabilidade de β_k estar no intervalo observado.

2. **Estimação restrita e o teste F: duas vistas.**

   Considere q restrições lineares $\mathbf{R}\beta = \mathbf{r}$, com $\mathbf{R}$ de dimensão $q \times k$ e posto q.

   (a) *(Vista da distância — Wald.)* Construa a estatística de Wald

   $$W = (\mathbf{Rb} - \mathbf{r})' \left[\mathbf{R}\, \widehat{\mathrm{Var}}(\mathbf{b})\, \mathbf{R}'\right]^{-1} (\mathbf{Rb} - \mathbf{r}),$$

   e explique por que ela mede o **afastamento** entre a estimativa irrestrita e a restrição, padronizado pela incerteza. Qual é a sua distribuição assintótica sob H_0?

   (b) *(Vista do ajuste — SSR.)* Defina o estimador de MQO **restrito** (minimiza a soma de quadrados dos resíduos sujeito a $\mathbf{Rb} = \mathbf{r}$) e escreva o teste F em função das somas de quadrados restrita e irrestrita,

   $$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_u)/q}{\mathrm{SSR}_u/(n-k)}.$$

   Prove que impor a restrição **só pode aumentar** a SSR, de modo que $F \geq 0$.

   (c) Enuncie a **identidade** que liga as duas vistas: sob homocedasticidade (e normalidade, para a versão exata), as duas construções dão a **mesma** conclusão, com $W/q = F$. O que cada vista torna transparente — uma sobre a estimativa, a outra sobre o ajuste?

3. **A trindade Wald–LR–LM e restrições não-lineares.**

   (a) Descreva os três testes: **Wald** (usa apenas o modelo **irrestrito**), **razão de verossimilhança (LR)** (compara os **dois** ajustes) e **multiplicador de Lagrange / escore (LM)** (usa apenas o modelo **restrito**). Explique por que são **assintoticamente equivalentes** sob H_0 (os três $\xrightarrow{d} \chi^2_q$) e qual a conveniência prática de cada um — qual modelo cada um exige estimar.

   (b) Para uma restrição **não-linear** $h(\beta) = 0$, use o **método Delta** para obter a variância assintótica de $h(\hat{\beta})$ e construa o teste de Wald correspondente, $W = h(\hat{\beta})^2 / \widehat{\mathrm{Var}}[h(\hat{\beta})]$. Explique por que o Wald não-linear **não é invariante** à forma algébrica como a restrição é escrita.

   (c) Explique por que o teste de **quebra estrutural** (teste de Chow, com dois subgrupos ou subperíodos) é um caso particular do teste F de restrições lineares: quais restrições, e sobre quais parâmetros, estão sendo testadas?

4. **Mínimos quadrados generalizados.**

   Considere $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ com $\mathbb{E}[\varepsilon \mid \mathbf{X}] = \mathbf{0}$ e $\mathbb{E}[\varepsilon\varepsilon' \mid \mathbf{X}] = \sigma^2\Omega$, com Ω **conhecida** e positiva definida.

   (a) Derive o estimador GLS $\mathbf{b}_{\mathrm{GLS}} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}$ e mostre que ele é o estimador de MQO aplicado aos dados transformados $\Omega^{-1/2}\mathbf{Y}$ e $\Omega^{-1/2}\mathbf{X}$. Por que a transformação "restaura" as hipóteses clássicas?

   (b) Mostre que $\mathbf{b}_{\mathrm{GLS}}$ é não-viesado condicionalmente em $\mathbf{X}$ e que $\mathrm{Var}[\mathbf{b}_{\mathrm{GLS}} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$. Enuncie o resultado de eficiência correspondente — o análogo do teorema de Gauss–Markov para o modelo generalizado:[1] em que sentido, e sob quais condições, o GLS é o estimador **linear não-viesado** de menor variância?

   (c) O FGLS substitui Ω por uma estimativa $\widehat{\Omega}$. Apresente razões pelas quais, na prática, o MQO com erros-padrão robustos pode ser preferível ao FGLS.

   (d) As n observações dividem-se em **dois grupos**, com $\mathrm{Var}[\varepsilon_i \mid \mathbf{x}_i] = \sigma_1^2$ para i no grupo 1 e σ_2^2 no grupo 2 (ambas conhecidas), e erros não-correlacionados entre si. Escreva Ω, obtenha a transformação GLS explicitamente e mostre que o GLS equivale a ponderar cada observação por $1/\sigma^2_{g(i)}$. No caso de um único regressor constante igual a 1 (estimação de uma média), obtenha o estimador GLS e mostre que ele dá **menos** peso ao grupo de maior variância.

5. **Heterocedasticidade: deteção e ponderação.**

   (a) Descreva os testes de **Breusch–Pagan** e de **White**: o que cada um regride, qual é a hipótese nula, e por que o de White detecta formas mais gerais de heterocedasticidade.

   (b) Argumente por que, no enquadramento moderno, um teste de heterocedasticidade **raramente** muda a conduta do pesquisador. O que, exatamente, mudou em relação à prática antiga de "testar e depois corrigir"?

   (c) Os mínimos quadrados ponderados (WLS) são um caso particular do GLS. Quando $\mathrm{Var}[\varepsilon_i \mid \mathbf{x}_i] = \sigma^2 h(\mathbf{x}_i)$ com h **conhecida**, qual é a transformação dos dados? Explique por que, se h for **mal especificada**, o WLS perde a eficiência prometida, enquanto o MQO com erros-padrão robustos permanece válido.

   (d) No modelo $y_i = \beta x_i + \varepsilon_i$ com $\mathbb{E}[\varepsilon_i \mid x_i] = 0$ e heterocedasticidade $\mathrm{Var}[\varepsilon_i \mid x_i] = \sigma^2 x_i^2$ (desvio-padrão proporcional a x_i, $x_i > 0$): **(i)** mostre que o MQO permanece não-viesado

[1] Este resultado é por vezes chamado *teorema de Aitken*.

e consistente para β; **(ii)** obtenha a transformação que torna os erros homocedásticos e identifique o estimador WLS resultante (divida a equação por x_i e interprete); **(iii)** escreva a variância verdadeira de $\hat{\beta}_{\text{MQO}}$ (a forma sanduíche) e explique por que a fórmula clássica $\sigma^2(\sum_i x_i^2)^{-1}$ é, em geral, **incorreta** — e o que o erro-padrão robusto corrige.

6. **Dependência intra-grupo e clusterização.**

   Seja $y_{ig} = \beta\, x_g + \varepsilon_{ig}$, com um regressor **invariante no cluster** x_g, G clusters de tamanho comum n, erros com $\text{Var}[\varepsilon_{ig}] = \sigma^2$ e correlação intra-cluster $\text{Corr}(\varepsilon_{ig}, \varepsilon_{jg}) = \rho$ para $i \neq j$ (e independência entre clusters).

   (a) Mostre que a variância verdadeira de $\hat{\beta}_{\text{MQO}}$ excede a fórmula convencional (que supõe erros i.i.d.) por um fator aproximado $1 + (n-1)\rho$ — o **fator de Moulton**. Interprete: por que regressores invariantes no cluster e $\rho > 0$ são a combinação perigosa?

   (b) O estimador de variância clusterizado tem forma de sanduíche com o **recheio** somando **sobre clusters**. Explique por que a escolha do **nível** de clusterização é uma decisão de **desenho amostral**, e não uma hipótese a ser testada nos dados.

   (c) Com **poucos** clusters, a inferência clusterizada falha. Explique por quê e cite duas correções possíveis (ajuste de graus de liberdade; *wild cluster bootstrap*).