AULA 6 - MQO Simples: Normalidade e Propriedades sob Normalidade

Econometria I

Área quantitativa - IE/UFRJ

Inferência em MQO simples

- A teoria clássica de inferência estatística consiste em dois ramos:
  - **Estimação**: já coberto nas aulas anteriores, no qual usando o método MQO estimamos α, β e σ^2. Sob hipóteses mostramos que estes estimadores são não viesados e com variância mínima.
  - **Teste de Hipótese**: além das estimativas, queremos fazer inferência sobre a função de regressão da população, tais como, o quão perto $\hat{\beta}$ está do verdadeiro β...
- Portanto, como α, β e σ^2 são variáveis aleatórias, precisamos descobrir suas distribuições de probabilidade, pois sem esse conhecimento não seremos capaz de relacioná-los aos seus verdadeiros valores.

A distribuição de probabilidade do erro u_i

- Considere

$$\hat{\beta} = \sum k_i y_i, \text{ onde } k_i = \frac{(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}$$

sendo $y_i = \alpha + \beta x_i + u_i$ escrevemos:

$$\hat{\beta} = \sum k_i(\alpha + \beta x_i + u_i)$$

Portanto, a distribuição de probabilidade de $\hat{\beta}$ dependerá da suposição feita sobre a distribuição de probabilidade de u_i.

- E como o conhecimento das distribuições de probabilidade dos estimadores MQO é necessário para inferir sobre seus valores populacionais, a natureza da distribuição de probabilidade de u_i assume um papel extremamente importante no teste de hipóteses.

A distribuição de probabilidade do erro u_i

- Como o método do MQO não faz nenhuma suposição sobre a natureza probabilística de u_i, é de pouca ajuda para o propósito de inferências.
- Esse vazio pode ser preenchido se estivermos dispostos a assumir que os u's seguem alguma distribuição de probabilidade.
- Por razões a serem explicadas em breve, no contexto de regressão, é geralmente assumido que os u's seguem a distribuição normal.
- Adicionando a suposição de normalidade para u_i às suposições do modelo de regressão linear clássico discutido anteriormente.

A hipótese da normalidade para u_i

- Assumimos que cada u_i é distribuído normalmente com:
  - Média: $E(u_i) = 0$
  - Variância: $E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$
  - Covariância: $E\{[(u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i u_j) = 0 \, i \neq j$

  De forma compacta escrevemos:

$$u_i \sim N(0, \sigma^2)$$

- Para duas variáveis normalmente distribuídas, covariância zero significa independência das duas variáveis (ver Apêndice A do Gujarati). Portanto, com a suposição de normalidade, significa que u_i e u_j são não correlacionados e também são independentemente distribuídos.

$$u_i \sim NID(0, \sigma^2)$$

  onde NID significa normalmente e independentemente distribuídos.

Por que a hipótese da normalidade

- Esperamos que a influência de variáveis omitidas no modelo seja pequena e aleatória. Pelo teorema do limite central (TCL), pode ser mostrado que se houver um grande número de variáveis aleatórias independentes e identicamente distribuídas, então, a sua soma tende a uma distribuição normal. (ver Apêndice A do Gujarati)
- Uma variante da TCL afirma que, mesmo que o número de variáveis não seja muito grande, sua soma pode ainda ser distribuído normalmente. (ver Apêndice A do Gujarati)
- Qualquer função linear de variáveis normalmente distribuídas é normalmente distribuída. Sendo os estimadores MQO funções lineares de u_i, temos que, se u_i for normalmente distribuído, α e β também são, o que torna a tarefa de testar hipóteses muito direta.

Por que a hipótese da normalidade

- A distribuição normal é simples envolvendo apenas dois parâmetros (média e variância); além disso, muitos fenômenos parecem seguir a distribuição normal.
- Se estamos lidando com um tamanho de amostra pequeno ou finito (ex menos de 100 observações), a suposição de normalidade assume um papel de nos ajudar a obter as distribuições de probabilidades exatas dos estimadores MQO e também nos permite usar os testes estatísticos t, F e χ^2 para modelos de regressão.

Propriedades do estimador MQO sob normalidade

- Pelo teorema de Markov mostramos que o estimador é não viesado e com mínima variância e agora temos também:
  - $\hat{\beta}$ (sendo uma função linear de u_i) é distribuido normalmente com:
    $$\text{Média:} \quad E(\hat{\beta}) = \beta$$
    $$var(\hat{\beta}): \quad \sigma^2_{\hat{\beta}} = \frac{\sigma^2}{\sum(x_i - \overline{x})^2}$$
    Então:
    $$Z = \frac{\hat{\beta} - \beta}{\sigma_{\hat{\beta}}}$$
    é uma distribuição normal padrão.
  - Para o intercepto temos também:
    $$Z = \frac{\hat{\alpha} - \alpha}{\sigma_{\hat{\alpha}}}$$

Propriedades do estimador MQO sob normalidade

- A hipótese da normalidade torna capaz derivarmos a distribuição de probabilidade de $\hat{\alpha}$ e $\hat{\beta}$ (ambas normal).
- Como veremos na próxima aula isso simplifica estabelecer o intervalo de confiança e teste de hipótese.
- Também temos que pela hipótese de normalidade de u_i, sendo y_i uma função linear de u_i, y_i é distribuida normalmente com média e variância dada por:

$$E(y_i) = \alpha + \beta x_i$$

$$var(y_i) = \sigma^2$$

De forma compacta podemos escrever:

$$y_i \sim N(\alpha + \beta x_i, \sigma^2)$$