ACH3657

Métodos Quantitativos para Avaliação de Políticas Públicas

Aula Teórica 10

Regressão Múltipla: Inferência

Alexandre Ribeiro Leichsenring
alexandre.leichsenring@usp.br

Organização

1. Regressão Múltipla: Inferência
   - Distribuição amostral dos estimadores MQO
   - Testando hipóteses sobre um parâmetro populacional

- Problema: testar hipóteses sobre parâmetros do modelo de regressão populacional
- Vamos descobrir a distribuição amostral dos estimadores MQO
- Adicionar hipótese: normalidade dos termos de erro

Distribuição amostral dos estimadores MQO

- Conhecer o valor esperado e a variância dos estimadores MQO é útil para descrever a precisão dos estimadores MQO
- Entretanto, para realizar inferência estatística, precisamos conhecer mais sobre a sua distribuição
- Precisamos conhecer a distribuição amostral inteira de $\hat{\beta}_j$.
- Sob as hipóteses de Gauss-Markov (RLM.1 a RLM.5), pode ter qualquer forma

Hipótese RLM.6 (Normalidade)

O termo de erro populacional u é independente das variáveis explicativas $x_1, x_2, \ldots, x_k$ e tem distribuição Normal com média 0 e variância σ^2:

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

- Suposição RLM.6 é bem mais forte do que qualquer outra das suposições anteriores
- De fato, sob RLM.6 u é independente, e portanto:

$$\begin{aligned} \mathbf{E}(u|x_1, x_2, \ldots, x_k) &= \mathbf{E}(u) = 0 \\ \text{Var}(u|x_1, x_2, \ldots, x_k) &= \text{Var}(u) = \sigma^2 \end{aligned}$$

- Logo, se assumimos RLM.6, estamos necessariamente assumindo RLM.3 e RLM.5.
- As 6 hipóteses RLM.1 a RLM.6 são chamadas de suposições do **modelo linear clássico (MLC)**

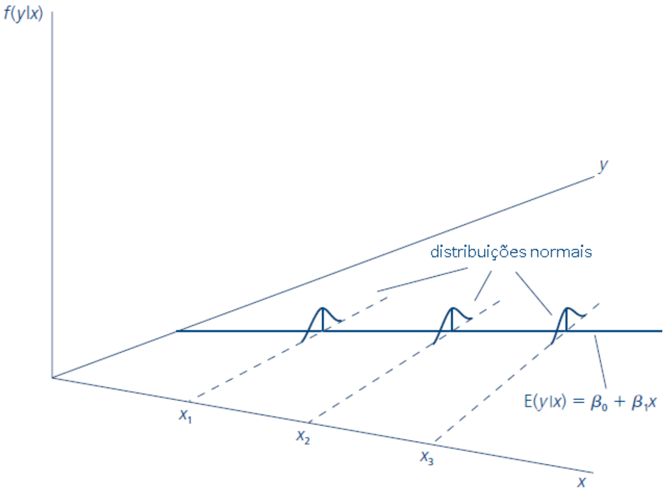
- Sob as suposições CLM, os estimadores MQO $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ têm propriedades de eficiência mais fortes do que sob as hipóteses de Gauss-Markov
- Pode-se mostrar que os estimadores MQO são os estimadores de variância mínima (MQO têm a menor variância entre todos os estimadores não viesados)
- Uma maneira sucinta de resumir as suposições MLC:

$$y|\boldsymbol{x} \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k, \sigma^2)$$

onde $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$

$\Rightarrow$ Condicional em $\boldsymbol{x}$, y tem distribuição normal com média linear em $x_1, x_2, \ldots, x_k$ e variância constante

A distribuição normal homoscedástica com uma variável explicativa



- Argumentos justificando normalidade vão na seguinte linha:
  - Sendo u a soma de vários fatores não observados diferentes afetando y, podemos invocar o teorema central do limite para concluir que u tem distribuição normal
- Tem seu mérito, mas também fragilidades:
  - Os fatores u podem ter distribuições muito diferentes na população (e.g: habilidade e qualidade da educação no termo de erro numa equação de salário)
  - Embora o teorema central do limite possa valer nesses casos, a aproximação normal pode ser pobre dependendo de quantos fatores aparecem em u e suas distribuições
  - Um problema mais sério com esse argumento é que ele assume que todos os fatores afetam y de uma maneira separada, aditiva (nada garante isso)
  - Se u é uma função complicada dos fatores não observados, o teorema central do limite não se aplica

- Em qualquer aplicação, a validade da suposição da normalidade é realmente um problema empírico. Não há, por exemplo, nenhum teorema que afirme que

$$salario|educ, exper, perm \sim \text{Normal}$$

- O simples raciocínio sugere que o oposto é verdade:
  - Como o salário não pode ser menor do que zero, não pode, estritamente falando, ter distribuição normal
  - Além disso, como há lei que rege salário mínimo, uma fração da população recebe salário mínimo, o que viola suposição de normalidade
- Para propósitos práticos, podemos perguntar se o salário condicional tem distribuição "aproximadamente" normal
- Evidencias passadas sugerem que normalidade não é uma boa suposição para salário
- Frequentemente, tomar uma transformação (log, especialmente) resulta numa distribuição que é próxima da normal

- Há exemplos em que RLM.6 é claramente falsa
- Quando y assume apenas alguns poucos valores, não pode ter distribuição normal
- A variável *narr86* (número de vezes que um jovem é preso) é um bom exmplo:
  - número inteiro
  - pequena amplitude
  - zero para a maior parte dos indivíduos

  $\Rightarrow$ Longe de ter distribuição normal!
- O que pode ser feito nesses casos?
- Para *amostras grandes*, esse desvio não é um grande problema (veremos adiante)

- Normalidade do termo de erro se traduz em distribuição amostral normal dos estimadores de MQO

Teorema 1 (Distribuição Amostral Normal)

Sob as suposições RLM.1 a RLM.6, condicional aos valores amostrais das variáveis independentes:

$$\hat{\beta}_j \sim \mathrm{N}\left(\beta_j, \mathrm{Var}(\hat{\beta}_j)\right)$$

Portanto:

$$\frac{(\hat{\beta}_j - \beta_j)}{\mathrm{dp}(\hat{\beta}_j)} \sim \mathrm{N}(0, 1)$$

Obs. Desvio padrão de $\hat{\beta}_j$:

$$\mathrm{dp}(\hat{\beta}_j) = \frac{\sigma}{\left[\mathrm{SQT}_j(1 - R_j^2)\right]^{1/2}}$$

Sendo σ desconhecido, o substituímos pelo seu estimador $\hat{\sigma}$. Isso nos dá o erro-padrão de $\hat{\beta}_j$:

$$\mathrm{ep}(\hat{\beta}_j) = \frac{\hat{\sigma}}{\left[\mathrm{SQT}_j(1 - R_j^2)\right]^{1/2}}$$

Teorema 2 (Distribuição dos estimadores padronizados)

Sob as suposições RLM.1 a RLM.6:

$$\frac{(\hat{\beta}_j - \beta_j)}{\text{ep}(\hat{\beta}_j)} \sim t_{n-k-1}$$

- $k+1$ é o número de parâmetros do modelo populacional

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- t_{n-k-1} é a distribuição t-student com $n-k-1$ graus de liberdade.

- A distribuição t no Teorema 2 vem do fato que a constante σ foi substituída pelo seu estimador $\hat{\sigma}$.

Testando hipóteses sobre um parâmetro populacional

- O modelo populacional pode ser escrito como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- Vamos assumir que o modelo satisfaz as suposições MLC
- Sabemos que os estimadores MQO produzem estimativas não viesadas dos β_j
- Vamos ver como testar hipóteses sobre um β_j particular
- Lembremos: β_j são desconhecidos e nunca conheceremos com exatidão
- Entretanto, podemos criar hipóteses sobre seu valor e usar a inferência estatística para testar essas hipóteses

- Na maioria das aplicações, nosso interesse se concentra em testar a hipótese nula:

$$H_0 : \beta_j = 0$$

- Como β_j mede o efeito parcial de x_j em y (controladas todas as outras variáveis independentes), H_0 significa que, controlados os efeitos de

$$x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$$

x_j não tem efeito nenhum no valor esperado de y

Exemplo

Considere a equação de salário:

$$\log(salario) = \beta_0 + \beta_1\, educ + \beta_2\, exper + \beta_3\, perm + u$$

- A hipótese nula $H_0 : \beta_2 = 0$ significa que, uma vez controlados educação e permanência no emprego, o número de anos de experiência não tem efeito no salário hora.
- É uma hipótese socialmente interessante: se for verdade, ela implica que o histórico da pessoa anterior ao atual trabalho não afeta o salário
- Se $\beta_2 > 0$, isso significa que a experiência de trabalho prévia contribui para a produtividade e portanto para o salário

Procedimento do teste de hipóteses

- O teorema 2 enuncia que, sob H_0 ($\beta_j = 0$):

$$\frac{(\hat{\beta}_j - \beta_j)}{\text{ep}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\text{ep}(\hat{\beta}_j)} \sim t_{n-k-1}$$

- Assim, se H_0 for verdadeira a distribuição da estatística

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\text{ep}(\hat{\beta}_j)}$$

  é t-student com $n - k - 1$ graus de liberdade
- $t_{\hat{\beta}_j}$ é simples de calcular dados $\hat{\beta}_j$ e o erro padrão ep$(\hat{\beta}_j)$ (computados pelos pacotes estatísticos)
- Em qualquer aplicação interessante, a estimativa pontual $\hat{\beta}_j$ nunca vai ser exatamente zero (independentemente da validade de H_0)
- A questão é: quão distante de zero?

- Um valor amostral de $\hat{\beta}_j$ distante de zero fornece evidência contra H_0
- Como reconhecemos que há erro amostral na nossa estimativa de $\hat{\beta}_j$, então é preciso confrontar seu erro amostral
- O erro padrão de $\hat{\beta}_j$ é uma estimativa do desvio padrão de $\hat{\beta}_j$, logo $t_{\hat{\beta}_j}$ mede quantos desvios padrão (estimados) $\hat{\beta}_j$ está de zero.
- Valores de $t_{\hat{\beta}_j}$ *suficientemente distantes* de zero resultam em rejeição de H_0
- A regra de decisão depende da definição da hipótese alternativa e do nível de significância adotado para o teste
- Determinar a regra de decisão para um determinado nível de significância – isto é, a probabilidade de rejeitar H_0 quando H_0 é verdadeira – requer conhecer a distribuição amostral de $t_{\hat{\beta}_j}$. Do Temorema 2:

$$t_{\hat{\beta}_j} \sim t_{n-k-1}$$

- Importante lembrar que estamos testando hipóteses sobre os parâmetros populacionais
- Não estamos interessados em testar estimativas de uma amostra particular