AULA 11 - Normalidade e Inferência em Regressão Múltipla - Parte 1

Econometria I

Área quantitativa - IE/UFRJ

Distribuições amostrais dos estimadores MQO

- Nas aulas passadas derivamos o valor esperado e variância dos estimadores
- Para a inferência estatísticas precisamos saber não só os dois primeiros momentos de $\hat{\beta}_j$, mas também a distribuição completa de $\hat{\beta}_j$.
- Pelo teorema de Markov que vimos na aula passada a distribuição de $\hat{\beta}_j$ poderia ser qualquer.
- Para tornar as distribuições amostrais de $\hat{\beta}_j$ passíveis de tratamento, vamos assumir que o erro não observado é normalmente distribuído na população (Hipótese da normalidade.

H6 **[Normalidade]**
O erro populacional u é independente das variáveis explicativas $x_1, x_2, ..., x_k$ e é normalmente distribuído com média zero e variância σ^2: $u \sim \text{Normal}(0, \sigma^2)$

Distribuições amostrais dos estimadores MQO

- Sendo y escrito como uma combinação linear de u, podemos escrever

$$y|\mathbf{x} \sim \text{Normal}(\alpha, \beta_1 x_1 + ... + \beta_k x_k, \sigma^2)$$

- Sob H1 a H6 condicional aos valores amostrais das variáveis independentes,

$$\hat{\beta}_j \sim \text{Normal}[\beta_j, Var(\hat{\beta}_j)]$$

e que $Var(\hat{\beta}_j) = (\hat{\beta}_j - \beta_j)/dp(\hat{\beta}_j) \sim \text{Normal}(0, 1)$
Isso é verdade pois cada $\hat{\beta}_j = \beta_j + \sum \omega_{ij} u_i$ onde ω_j são ponderadores formados por variáveis independentes (não aleatórios).
Assim $\hat{\beta}_j$ é uma combinação linear dos erros u_i na amostra.
Temos uma função linear de uma variável aleatória (va) possui a mesma distribuição da va.

Distribuições amostrais dos estimadores MQO

- Podemos dizer ainda mais: qualquer combinação linear dos $\hat{\alpha}$, $\hat{\beta}_1, ..., \hat{\beta}_k$ também é normalmente distribuída, e qualquer subconjunto dos $\hat{\beta}_j$ tem uma distribuição normal conjunta.
- A hipótese da normalidade dos erros, pode não ser realística para amostras pequenas, mas para amostras grandes a normalidade dos estimadores de MQO ainda é aproximadamente verdadeira, mesmo sem normalidade dos erros.

Teste de hipótese

- Como nunca saberemos quem são os β_j, a única coisa que podemos fazer são hipóteses sobre o valor de β_j e em seguida, utilizar inferência estatística para testar nossa hipótese.
- Vimos que a variância do estimador padronizado sob as hipóteses H1 a H6

  $$(\hat{\beta}_j - \beta_j)/dp(\hat{\beta}_j) \sim \text{Normal}(0,1)$$

  neste caso temos que o desvio padrão é função do desvio padrão do erro σ
- Entretanto, σ é um parâmetro populacional que não conhecemos
- Se considerarmos a estimativa do σ denotada por $\hat{\sigma}$ temos agora que $\hat{\sigma}$ é uma varável aleatória e não mais uma constante como o parâmetro populacional.
- Então

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

  agora passa a ter distribuição t

Teste de hipótese

- O resultado acima é importante porque permite testar hipóteses que envolvam os β_j.
- Na maioria das aplicações, nosso principal interesse é testar a hipótese nula
$$H_0 : \beta_j = 0$$
em que j corresponde a qualquer uma das k variáveis independentes.
- Interpretação correta: Uma vez que $x_1, x_2, ..., x_{j-1}, x_{j+1}, ..., x_k$ foram considerados, x_j não tem nenhum efeito sobre o valor esperado de y.

Teste de hipótese

Example

Equação do salário:

$$\log(\text{salárioh}) = \alpha + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{perm} + u$$

A $H_0 : \beta_2 = 0$ significa que, uma vez que a educação formal e a permanência foram consideradas, o número de anos no mercado de trabalho (exper) não tem nenhum efeito sobre o salário-hora. Se $\beta_2 > 0$ então a experiência prévia de trabalho contribui para a produtividade e, portanto, para o salário.

Teste de hipótese

- A estatística t de $\hat{\beta}_j$ é definida como

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

- Observe que como $ep(\hat{\beta}_j)$ é sempre positivo, $t_{\hat{\beta}_j}$ tem o mesmo sinal de $\hat{\beta}_j$
- Qualquer estimativa pontual de $\hat{\beta}_j$ nunca será exatamente zero, seja H_0 verdadeira ou não.
- A questão é: quão distante está $\hat{\beta}_j$ de zero?
- Um valor grande de $\hat{\beta}_j$ não garante isso, pois existe um erro amostral em nossa estimativa de $\hat{\beta}_j$, de modo que o tamanho de $\hat{\beta}_j$ deve ser ponderado pelo seu erro amostral.
- Como o erro-padrão de $\hat{\beta}_j$ é uma estimativa do desvio-padrão de $\hat{\beta}_j$, $t_{\hat{\beta}_j}$ mede quantos desvios-padrão estimados $\hat{\beta}_j$ estão afastados de zero.

Teste contra hipóteses alternativas unilaterais

- Precisamos decidir sobre a hipótese alternativa relevante.
- Considere primeiro a hipótese alternativa unilateral do tipo

$$H_1 : \beta_j > 0$$

  Esta hipótese pode ser baseada na teoria econômica .
- Temos que o valor esperado de $t_{\hat{\beta}_j}$ é positivo.
- Assim estamos procurando um valor positivo "suficientemente grande" de $t_{\hat{\beta}_j}$ a fim de rejeitar $H_0 : \beta_j = 0$ em favor de $H_1 : \beta_j > 0$
- A definição de suficientemente grande com por exemplo nível de significância de 5%, é o 95^o percentil de uma distribuição t com $n - k - 1$ graus de liberdade; denominaremos o ponto de c

Teste contra hipóteses alternativas unilaterais

- A regra de rejeição é que H_0 é rejeitada em favor de H_1, ao nível de significância de 5% se

$$t_{\hat{\beta}_j} > c$$

- Essa regra de rejeição é um exemplo de um teste monocaudal.
- Para obter c, necessitamos somente do nível de signifância e dos graus de liberdade.
- Ex: teste com nível de 5% e com $n - k - 1 = 28$ graus de liberdade o valor crítico é $c = 1,701$. Se $t_{\hat{\beta}_j} < 1,701$ não rejeitamos H_0, se for a 1% temos $c = 2,467$, ou seja, quando menor o nível, maior é o c de modo que para rejeitar H_0, exigimos um valor cada vez maior de $t_{\hat{\beta}_j}$.

Teste contra hipóteses alternativas unilaterais

- Se H_0 é rejeitada a 5% será automaticamente a 10%.
- A medida que aumentamos os graus de liberdade da distribuição t, ela se aproxima da distribuição normal padronizada (ex a partir de 120).
- Para o teste em que a hipótese alternativa unilateral cujo parâmetro é menor que zero, a regra de rejeição é:

$$t_{\hat{\beta}_j} < -c$$

Teste contra hipóteses alternativas bilaterais

- É comum testar $H_0 : \beta_j = 0$ contra uma hipótese alternativa bilateral

$$H_1 : \beta_j \neq 0$$

  ela é uma hipótese alternativa relevante quando o sinal de β_j não é bem determinada pela teoria.

- Quando a alternativa é bilateral, estamos interessados no valor absoluto da estatística t. A regra de rejeição é:

$$|t_{\hat{\beta}_j}| > c$$

  .

- Para achar c, vamos especificar novamente um nível de significância, por exemplo, de 5%. Assim, em cada cauda da distribuição t será igual a $2,5\%$.
- Quando a hipótese alternativa específica não é formulada, considera-se geralmente que é bilateral.

Teste de outras hipóteses sobre β_j

- Algumas vezes estamos interessados em testar se β_j é igual a alguma constante.
- Os exemplos mais comuns são $\beta_j = 1$ e $\beta_j = -1$.
- De forma geral temos:

$$H_0 : \beta_j = a_j$$

  em que a_j é o nosso valor hipotético de β_j.
- A estatística t apropriada é

$$t = (\hat{\beta}_j - a_j)/ep(\hat{\beta}_j)$$

Cálculos dos p-Valores dos testes t

- Podemos perguntar: dado o valor observado da estística t, qual é o menor nível de significância ao qual a hipótese nula seria rejeitada?
- Esse nível é conhecido como p-valor do teste.
- Obtemos o p-valor real ao calcular a probabilidade de que a variável aleatória t, com 40 gl seja maior que 1,85 em valor absoluto.
- O cálculo é muito trabalhoso, mas os programas de computador já fornecem esse resultado.
- O p-valor mais comum é:

$$P(|T| > |t|)$$

em que T representa a variável aleatória com distribuição t com $n - k - 1$ gl e t é o valor numérico da estatística do teste.

Significância econômica versus estatística

- Se a variável é estatisticamente signficante, discuta a magnitude do coeficiente para ter um idéia de sua importância prática ou econômica.
- Se a variável não é estatisticamente significante aos níveis usuais, pergunte se a variável tem o efeito esperado sobre y e se tal efeito é, na prática, grande. Se for grande calcule o p-valor.
- É comum encontrar variáveis com estatísticas t pequenas que têm o sinal " errado" (podemos apenar ignorar). Se a estatística for grande isso é um problema mais grave (em geral isso está relacionado com a omissão de uma variável relevante)

Intervalos de confiança

- Os intervalos de confiança (ou estimativas de intervalo), dão uma extensão dos valores prováveis do parâmetro populacional, e não somente uma estimativa pontual.
- Usando o fato que $(\hat{\beta}_j - \beta_j)/ep(\hat{\beta}_j) \sim t_{n-k-1}$ uma simples manipulação algébrica leva a um IC do β_j desconhecido.
- Um intervalo de confiança de 95% é dado por

$$\hat{\beta}_j \pm c \cdot ep(\hat{\beta}_j)$$

- Se as amostras aleatórias fossem obtidas repetidas vezes, então o valor populacional β_j estaria dentro do intervalo em 95% das amostras.
- Vale lembrar que o IC não será bom se temos MQO viesado, ou se heterocedasticidade está presente, então o erro-padrão não é válido.