

TESTES À IGUALDADE DOS PARÂMETROS DE UM MODELO DE REGRESSÃO: UMA APLICAÇÃO ESPECIAL DAS VARIÁVEIS BINÁRIAS (*DUMMY*)

¹Sérgio Ricardo Magalhães, ²Ednaldo Antônio de Andrade
serimag@oi.com.br; ednaldo@ufla.br

¹Professor da Universidade Vale do Rio Verde
²Professor da Universidade Federal do Mato Grosso

Resumo – O objetivo deste trabalho foi discutir e comparar duas metodologias que permitem testar se os parâmetros de um modelo de regressão são ou não diferentes entre duas ou mais sub-mostras: o teste tradicional de Chow e a técnica das variáveis *dummy*. Em particular, demonstrou-se analiticamente que a técnica das variáveis *dummy* substitui de forma muito eficaz o teste Chow, apresentando até algumas vantagens relativamente a este teste.

Palavras-chave – regressão linear, variáveis *dummy*, teste Chow

Abstract – The aim of this paper is to discuss and compare two methods for testing the parameters of a regression model are different or not between two or more sub-samples: the traditional test of Chow and the technique of dummy variables. In particular, we show analytically that the technique of dummy variables replaces very effectively the Chow test, with even some advantages to this test.

keywords – linear regression, dummy variables, Chow test

I. INTRODUÇÃO

A necessidade de se testar a hipótese de que parte ou a totalidade dos coeficientes de uma equação de regressão permanece estável para subconjuntos de observações de uma amostra, surge em estudos cuja causa é do tipo cronológico ou do tipo seccional.

Na primeira situação, os subconjuntos de observações dizem respeito a períodos de tempo diferentes e, no segundo caso, referem-se a grupos de indivíduos que diferem em alguma característica qualitativa (classe social, grupo etário, sexo, formação escolar, etc.).

O objetivo deste estudo foi sistematizar as contribuições de autores como Gujarati (1970), Valentine (1978) e Erlat (1985) no uso

de variáveis *dummy* na realização de testes à estabilidade dos parâmetros de um modelo de regressão e apresentar as demonstrações matemáticas que em geral foram muito abreviadas nos *papers* originais e que sustentam a equivalência entre este método e os clássicos testes de Chow.

Assim, pretendeu-se a partir desta análise comparativa, a apresentação de uma síntese das principais vantagens da abordagem segundo as variáveis *dummy* em relação à metodologia tradicional de Chow para testar variações nos parâmetros de um modelo.

II. DESENVOLVIMENTO

O estudo de situações, por meio da análise de regressão, em que se faz a comparação entre dois ou mais conjuntos de observações *n*-dimensionais, tem sido descrito na literatura por Gujarati (1970a), Scoloro (1997), Draper e Smith (1998), Regazzi (1999), entre outros.

Normalmente, preocupa-se primeiramente em estabelecer se os conjuntos de observações, representados por equações de regressão linear, diferem entre si. Se for notada a diferença entre as equações, pode ser interessante avaliar em que ponto diferem, ou seja, quais coeficientes diferem de uma equação para outra.

Em contrapartida, se for notado que as equações não diferem entre si, significa que uma única equação pode ser utilizada para representar todos os conjuntos de observações.

Em outras palavras, uma única equação pode ser estimada a partindo-se de todas as observações de todos os conjuntos envolvidos no estudo.

Deste modo, pode-se considerar que as diferentes situações em estudo comportam-se da mesma forma. Se isto for verdadeiro, ter-se-á uma equação estimada com melhor precisão e mais confiável, quando comparado à estimação de equações individuais.

Objetivando verificar a igualdade de duas regressões lineares, Chow (1960) sugeriu um teste geral, cujo algoritmo segue os seguintes passos:

1. Dadas as seguintes relações lineares:

$$\begin{aligned} y_{1i} &= a_1 + b_1 x_{1i} + e_{1i} & i &= 1, \dots, n_1 \\ y_{2i} &= a_2 + b_2 x_{2i} + e_{2i} & i &= 1, \dots, n_2 \end{aligned}$$

referentes a dois conjuntos de observações.

2. Combinam-se todas as $n_1 + n_2$ observações e calcula-se a estimativa de quadrados mínimos de a e b na regressão combinada $y = a + bx + e$. Desta equação obtém-se a soma de quadrados de resíduo (S_1) com grau de liberdade igual a $n_1 + n_2 - p$, em que p é o número de parâmetros a ser estimado. Neste caso, $p = 2$.
3. Obtém-se a soma de quadrados de resíduo para as duas equações, ou seja, S_2 e S_3 , com os graus de liberdade $n_1 - p$ e $n_2 - p$, respectivamente. Somam-se estas duas somas de quadrados de resíduo, isto é, $S_4 = S_2 + S_3$ e seus graus de liberdade $n_1 + n_2 - 2p$.
4. Obtém-se $S_5 = S_1 - S_4$.
5. Calcula-se a estatística F como:

$$F_c = \frac{S_5/p}{S_4/(n_1 + n_2 - 2p)}$$

com p e $n_1 + n_2 - 2p$ graus de liberdade.

Se $F_c > F$ tabelado, para um determinado nível de significância α , rejeita-se a hipótese de que os parâmetros a 's e b 's são os mesmos para os dois conjuntos de observações.

Para Gujarati (1970b), o teste Chow (1960) permite uma avaliação geral da equação, assegurando apenas se duas regressões lineares são iguais ou diferentes. Caso sejam diferentes, não especificam se a diferença é devida a interceptos ou inclinações.

Muitos autores priorizam a utilização de variáveis binárias, também mencionadas como variáveis *dummy*, indicadoras ou classificatórias, para testar a igualdade de equações ou coeficientes.

Gujarati (1970b) utilizou Variáveis *dummy*, que são definidas como aquelas que assumem somente dois valores 1 e 0, como uma alternativa para a análise padrão de métodos de análise de variância e do teste de Chow (1960).

O referido autor considerou a seguinte relação, referente a dois conjuntos de dados:

$$y_i = \alpha_0 + \alpha_1 D + \alpha_2 x_i + \alpha_3 (Dx_i) + e_i \quad i = 1, \dots, (n_1 + n_2)$$

em que:

$D = 1$ para observações do primeiro conjunto (n_1 observações);

$D = 0$ para observações do segundo conjunto (n_2 observações).

As variáveis binárias foram introduzidas na forma aditiva e multiplicativa. Os coeficientes α_1 e α_3 são diferenças de interceptos e inclinações, respectivamente.

Se $H_0 : \alpha_1 = 0$ é rejeitada, ou seja, α_1 é significativo, então, o valor do intercepto do

primeiro conjunto é obtido por $\alpha_1 + \alpha_0$. Neste caso, α_0 é o intercepto do segundo conjunto.

Se $H_0 : \alpha_1 = 0$ não é rejeitada, ou seja, α_1 é não significativo, então α_0 representa o intercepto comum para ambos os conjuntos.

Se $H_0 : \alpha_3 = 0$ é rejeitada, então o valor da inclinação do primeiro conjunto é obtido por $\alpha_2 + \alpha_3$. Neste caso, α_2 é a inclinação do segundo conjunto.

Se $H_0 : \alpha_3 = 0$ não é rejeitada, então α_2 representa a inclinação comum para ambos os conjuntos.

Logo, a inclusão de variáveis binárias aditivas ou multiplicativas permite verificar se duas equações lineares diferem em intercepto, em inclinação ou, ainda, em ambos.

Gujarati (1970b) notou que este método fornece resultados idênticos aos do teste de Chow (1960). Contudo, indica algumas vantagens para a técnica de variáveis binárias.

Esta técnica indica a(s) fonte(s) de diferença entre as regressões lineares, ou seja, se a diferença é devido a intercepto, ou inclinações, ou ambos. Em uma única regressão obtêm-se todas as informações necessárias, ao passo que o teste Chow é um procedimento de vários estágios.

III. CONSIDERAÇÕES FINAIS

Identificou-se algumas das vantagens que a técnica das variáveis *dummy* apresenta relativamente aos testes de Chow.

Em primeiro lugar, a aplicação do teste de Chow apenas permite testar se parte ou a totalidade dos parâmetros do modelo de regressão diferem ou não entre dois conjuntos de observações, mediante a não rejeição ou a rejeição da hipótese nula. Contudo, mediante a especificação de uma equação de regressão com variáveis *dummy* é sempre possível saber, através da realização de testes *t* ou *F*, qual é a origem da variabilidade.

Em segundo lugar, estimar uma única equação de regressão com todas as observações

constitui uma forma mais eficiente de testar a estabilidade dos parâmetros de uma equação de regressão entre várias sub-amostras do que recorrer ao teste de Chow.

Deste modo, em qualquer um dos contextos e independentemente do número de modelos de regressão cuja igualdade dos parâmetros se pretenda testar, através da técnica das variáveis *dummy*, apenas é necessário estimar um único modelo de regressão linear para obter toda a informação relevante.

REFERÊNCIAS

CHOW, G. C. Tests of equality between sets of coefficients in two linear regressions. **Econometrica**, Chicago, v. 28, p. 591-605, 1960.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 2. ed. New York: John Wiley & Sons, 1998. 709p.

ERLAT, H. Generalised Chow tests for structural change: a coordinate free approach. **International Economic Review**. v.23 , n.3, p565-575, 1978.

GRAYBILL, F. A. **Theory and application of the linear model**. Belmont: Duxbury Press, 1976. 704p.

GUJARATI, D. Use of dummy variables in testing for equality between sets of coefficients in linear regressions: a generalization. **The American Statistician**, Washington, v. 24, n. 5, p. 18-22, Dec. 1970a.

GUJARATI, D. Use of dummy variables in testing for equality between sets of coefficients in two linear regressions: a note. **The, n. American Statistician**, Washington, v. 24, n. 1, p. 50-52, Feb. 1970b.

HOFFMANN, R.; VIEIRA S. **Análise de regressão**: uma introdução à econometria. 3. ed. São Paulo: HUCITEC, 1998. 379p.

NETER, J.; WASSERMAN, W.; KUTNER, M. **Applied linear statistical models**. 3. ed. Burr Ridge, Illinois: Irwin, 1990. 1181p.

SEBER, G. A. F. **Linear regression analysis**. New York: John Wiley, 1977. 465p.

SWAMY, P. A. V. B.; MEHTA, J. S. Estimation of common coefficients in two regression evaluations. **Journal of Econometrics**, Lausanne, v. 10, p. 1-14, 1979.