Teste RESET

Erros de especificação na forma funcional

A omisso de variveis explicativas no o único modo de um modelo ser mal especificado.

Por exemplo, se o verdadeiro processo gerador do dados for

$$\log y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k + \varepsilon$$

e utilizarmos y, no lugar de $\log y$, como variável resposta, então obteremos estimadores viesados e inconsistentes para parâmetros do modelo de interesse.

Teste RESET - Ramsey (1969)

Se o modelo original é da forma

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k + \varepsilon$$

satisfazendo à suposição de que

$$E(\varepsilon | x_2, x_3, \ldots, x_k) = E(\varepsilon) = 0$$

então, funções não-lineares de variáveis independentes não devem ser relevantes quando acrescentadas acima.

1. Assim, para implementarmos o teste RESET, teremos que decidir sobre quais funções não-lineares das variáveis explicativas deveremos incluir na equação a ser expandida;

2. Entretanto, não existe uma resposta direta para esta pergunta.

3. Todavia, segundo Ramsey (1969), a inclusão de termos quadráticos e cúbicos se mostrou bastante adequado em várias aplicações.

Seja $\hat{y}_i$ os valores estimados, segundo o modelo proposto.

Segundo Ramsey (1969), a partir da equação expandida, dada por:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + \varepsilon$$

poderemos testar se existem problemas de especificação na forma funcional (ausências de não-linearidades importantes):

$$H_0 : \delta_1 = \delta_2 = 0,$$

versus

$$H_a : H_0 \text{ é falsa.}$$

Sob H_0, o modelo foi especificado corretamente.

Então, sob H_0, o teste RESET nada mais é do que implementação de um teste de restrição nos coeficientes, que usa a estatística F para testar $H_0 : \delta_1 = \delta_2 = 0$, no modelo expandido.

A significância desta estatística sugere algum problema com não-linearidades no modelo original.

A distribuição da estatística F, sob H_0 e admitindo a validade das suposições de 1 a 5, é aproximadamente $F_{2,n-p-2}$, para amostras grandes (isto é, n grande).

Exemplo `tyler.txt`

Utilizando o arquivo de dados `tyler.txt`, estime os parâmetros do seguinte modelo:

$$\texttt{vendas} = \beta_1 + \beta_2\texttt{gastopub} + \beta_3\texttt{preço} + \varepsilon$$

onde

- `preço`: 1, 2 ou 3 dólares,
- `gastopub`: gastos com publicidade (50/100 mil dólares),
- `vendas`: vendas do produto (em milhares de unidades).

Conduza um teste RESET para verificar se há problemas de especificação na forma funcional.

tyler.txt

obs	gastopub	preco	vendas
1	50	1	15.050
2	50	1	15.322
3	50	2	11.199
4	50	2	9.508
5	50	3	6.248
6	50	3	5.355
7	100	1	24.277
8	100	1	24.813
9	100	2	15.316
10	100	2	15.667
11	100	3	5.837
12	100	3	6.271
13	50	1	14.561
14	50	1	15.838
15	50	2	11.067
16	50	2	10.083
17	50	3	5.457
18	50	3	5.155
19	100	1	25.011
20	100	1	23.941
21	100	2	15.398
22	100	2	15.417
23	100	3	5.815
24	100	3	5.657

Modelo estimado:

$$\texttt{vendas} = \beta_1 + \beta_2\texttt{gastopub} + \beta_3\texttt{preço} + \varepsilon$$

tal que valores ajustados de vendas sejam:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2\texttt{gastopub} + \hat{\beta}_3\texttt{preço}$$

Modelo expandido:

$$\texttt{vendas} = \beta_1 + \beta_2\texttt{gastopub} + \beta_3\texttt{preço} + \delta_1\hat{y}^2 + \delta_2\hat{y}^3 + \varepsilon$$

Hipóteses de interesse:

$$H_0 : (\delta_1, \delta_2) = (0, 0) \quad \text{versus} \quad H_a : (\delta_1, \delta_2) \neq (0, 0)$$

```
data = read.table("tyler.txt",header=TRUE)
attach(data)
n=nrow(data)

# Modelo estimado
reg = lm(vendas~gastopub+preco)
summary(reg)
SQR = sum(reg$res^2)
yhat = reg$fit
yhat2 = yhat^2
yhat3 = yhat^3

# Modelo expandido
reg1 = lm(vendas~gastopub+preco+yhat2+yhat3)
summary(reg1)
SQR1 = sum(reg1$res^2)

# F test
num   = (SQR-SQR1)/2
den   = SQR1/(n-5)
Ftest = num/den

c(Ftest,qf(0.95,2,n-5))
[1] 10.179698  3.521893
```