5.º teste (ímpares) – MRLM (previsão e forma funcional) – 4 de Dezembro de 2020

Questão aberta:

1. Num estudo para avaliar o consumo de cigarros estimou-se o seguinte modelo, usando o Método dos Mínimos Quadrados Ordinários:

$$\begin{aligned} cigs = {} & \underset{(63.2391)}{27.5445} - \underset{(1.5217)}{2.7011}\, preco + \underset{(0.0131)}{0.0229}\, preco^2 \\ & + \underset{(10.2022)}{10.1712} \ln(rend) - \underset{(0.5611)}{0.5497} \ln(rend)^2 + \underset{(0.1605)}{0.7361}\, idade - \underset{(0.0017)}{0.0085}\, idade^2 \end{aligned}$$

$$n = 807, \quad R^2 = 0.0385, \quad FStat = 5.3441, \quad \sigma = 13.5049$$

em que:

- $cigs$ é o número de cigarros fumados pelo indivíduo i;
- $preco$ é o preço médio do maço de cigarros, pago pelo indivíduo i;
- $rend$ é o rendimento do indivíduo i;
- $idade$ é a idade do indivíduo i.

Notas:

i. Se nada for dito em contrário assume-se que o modelo verifica as hipóteses 1 a 6 do MRLM;

ii. A função $\ln()$ corresponde ao logaritmo de base e;

iii. Considere uma **dimensão do ensaio de** 1%.

a) (10) Qual o valor previsto para o número de cigarros fumados, quando o preço médio do maço de cigarros é igual a 5€, o rendimento do indivíduo é 1000€ e a sua idade é 40 anos.
RESOLUÇÃO:

$$\begin{aligned} cigs = {} & 27.5445 - 2.7011 \times 5 + 0.0229 \times 5^2 \\ & + 10.1712 \ln(1000) - 0.5497 \ln(1000)^2 \\ & + 0.7361 \times 40 - 0.0085 \times 40^2 = 74.4857 \end{aligned}$$

b) (10) Obtenha o intervalo de previsão para a média do número de cigarros fumados, considerando o cenário da alínea anterior. Sabe-se ainda que o erro padrão da previsão da alínea anterior é de 36.7263.
RESOLUÇÃO:

- Variável fulcral:

$$T = \frac{\hat{\theta} - \theta}{\mathrm{se}\left(\hat{\theta}\right)} \sim \mathrm{t}_{(n-k-1)} \text{ OU } T = \frac{\hat{\theta} - \theta}{\mathrm{se}\left(\hat{\theta}\right)} \overset{a}{\sim} \mathrm{N}(0,1)$$

- Intervalo de Previsão para a média (teórico):

$$IP_{1-\alpha}^{\text{Média}} = \left(\hat{\theta} - t_{\alpha/2}\, \mathrm{se}\left(\hat{\theta}\right), \hat{\theta} + t_{\alpha/2}\, \mathrm{se}\left(\hat{\theta}\right)\right)$$

- Intervalo de Previsão para a média (observado):

  → $\mathrm{P}\left(-t_{\alpha/2} < t_{(807-6-1)} < t_{\alpha/2}\right) = 0.99$

  Pelo que $t_{0.005} = 2.582$ (exato) OU $t_{0.005} = 2.576$ (tabelas)

  → $IP_{99\%}^{\text{Média}} = 74.4857 \pm 2.582 \times 36.7263 = (-20.3416, 169.3130)$

  OU

  $IP_{99\%}^{\text{Média}} = 74.4857 \pm 2.576 \times 36.7263 = (-20.1213, 169.0926)$

c) (10) Obtenha o intervalo de previsão para a previsão pontual do número de cigarros fumados, considerando o cenário da alínea a). Compare-o com o intervalo de previsão obtido na alínea anterior.

RESOLUÇÃO:

- Variável fulcral:

$$T = \frac{\hat{y}^0 - y^0}{\sqrt{\left[\mathrm{se}\left(\hat{\theta}\right)\right]^2 + \hat{\sigma}^2}} \sim \mathrm{t}_{(n-k-1)} \text{ OU } T = \frac{\hat{y}^0 - y^0}{\sqrt{\left[\mathrm{se}\left(\hat{\theta}\right)\right]^2 + \hat{\sigma}^2}} \overset{a}{\sim} \mathrm{N}(0,1)$$

- Intervalo de Previsão (teórico):

$$IP_{1-\alpha}^{\text{Pontual}} = \left(\hat{y}^0 - t_{\alpha/2}\sqrt{\left[\mathrm{se}\left(\hat{\theta}\right)\right]^2 + \hat{\sigma}^2}, \hat{y}^0 + t_{\alpha/2}\sqrt{\left[\mathrm{se}\left(\hat{\theta}\right)\right]^2 + \hat{\sigma}^2}\right)$$

- Intervalo de Previsão (observado):
  - $\mathrm{P}\left(-t_{\alpha/2} < t_{(807-6-1)} < t_{\alpha/2}\right) = 0.99$

    Pelo que $t_{0.005} = 2.582$ (exato) OU $t_{0.005} = 2.576$ (tabelas)
  - $$\begin{aligned} IP_{99\%}^{\text{Pontual}} &= 74.4857 \pm 2.582 \times \sqrt{36.7263^2 + 13.5049^2} \\ &= (-26.5495, 175.5209) \end{aligned}$$

    OU

    $$\begin{aligned} IP_{99\%}^{\text{Pontual}} &= 74.4857 \pm 2.576 \times \sqrt{36.7263^2 + 13.5049^2} \\ &= (-26.3147, 175.2861) \end{aligned}$$

- Comparação:

  O intervalo de previsão para y^0 é cerca de 30% mais amplo que o intervalo de previsão para θ_0, conforme era esperado. Tal fenómeno deve-se à incerteza relativa aos fatores omitidos no nosso modelo e que permitem explicar o consumo de cigarros. Isto também permite explicar os valores negativos dos intervalos de previsão.

d) (10) Calcule e interprete o ponto de viragem da idade sobre o número de cigarros fumados.

RESOLUÇÃO:

- O efeito parcial da idade sobre o número de cigarros fumados:

$$\frac{\partial \widehat{cigs}}{\partial idade} = \hat{\beta}_5 + 2\hat{\beta}_6\, idade = 0.7361 - 0.017\, idade$$

- Ponto de viragem:

$$\frac{\partial \widehat{cigs}}{\partial idade} = 0 \Leftrightarrow 0.7361 - 0.017\, idade = 0$$

$$\Leftrightarrow idade^* = |0.7361/0.017| \Leftrightarrow idade^* = 43.3$$

- Interpretação: A partir dos 43.3 anos, um ano adicional de Vida provoca uma redução, em média, do número de cigarros fumados, *ceteris paribus*.

e) (10) Procedeu-se a nova estimação do modelo, desta vez sem os termos quadráticos. O R^2 desse modelo é igual a 0.0061. O que pode concluir acerca da especificação do modelo inicial?
RESOLUÇÃO:

- Hipóteses:

$$\begin{cases} H_0 : \beta_2 = \beta_4 = \beta_6 = 0 \\ H_1 : \exists \beta_j \neq 0,\ j = 2, 4, 6 \end{cases}$$

- Estatística de teste:

$$F = \frac{R^2_{UR} - R^2_R}{1 - R^2_{UR}} \times \frac{n-k-1}{q} \sim F_{(q,n-k-1)}$$

- Valor observado da Estatística de teste:

$$f_{obs} = \frac{0.0385 - 0.0061}{1 - 0.0385} \times \frac{807 - 6 - 1}{3} = 8.986$$

- Região de rejeição:

$$W = \{f : f > f_\alpha\} \text{ com } \mathrm{P}\left(F_{(q,n-k-1)} > f_\alpha\right) = \alpha$$

Pelo que $f_{0.01} = 3.806$ (exato) OU $f_{0.01} = 3.78$ (tabelas).

OU

Valor-p:

$$\mathrm{P}\left(F_{(3,800)} > 8.986\right) = 0.00001$$

- Decisão: rejeitar H_0. É possível concluir que os termos quadráticos são conjuntamente significativos. Assim sendo, existem indícios estatísticos que nos permitem concluir que o modelo se encontra bem especificado.