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

Questão aberta:

1. Num estudo para avaliar o impato do modelo de ensino na nota final da disciplina de Estatística II, foi estimado o seguinte modelo, usando o Método dos Mínimos Quadrados Ordinários:

$$\begin{aligned} nfinal = & \underset{(2.6185)}{27.9545} - \underset{(0.0596)}{0.0003}\, assid - \underset{(0.0004)}{0.00004}\, assid^2 \\ & + \underset{(0.0496)}{0.0094}\, tpc + \underset{(0.0004)}{0.00005}\, tpc^2 - \underset{(3.9167)}{15.9292} \ln(med) + \underset{(2.1955)}{13.2691} \left[\ln(med)\right]^2 \\ & n = 674, \quad R^2 = 0.1642, \quad FStat = 21.8333, \quad \sigma = 4.3391 \end{aligned}$$

em que:

- $nfinal$ é a nota final (na escala de 0 a 100 pontos) do aluno $i$, à disciplina de Estatística II;
- $assid$ é percentagem de aulas de Estatística II assistidas pelo aluno $i$ (em relação ao número total de aulas);
- $tpc$ é a percentagem de trabalhos de casa da disciplina de Estatística II que foram entregues pelo aluno $i$;
- $med$ é a média (ponderada pelo número de créditos e medida na escala de 0 a 20 valores) do aluno $i$, antes do exame à disciplina de Estatística II.

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 a nota final prevista para um aluno que frequentou 50% das aulas, fez 25% dos trabalhos de casa e tinha uma média (até ao momento anterior ao exame) de 10 valores.
**RESOLUÇÃO:**

$$\begin{aligned} nfinal = & 27.9545 - 0.0003 \times 50 + 0.00004 \times 50^2 \\ & + 0.0094 \times 25 + 0.00005 \times 25^2 \\ & - 15.9292 \ln(10) + 13.2691 \left[\ln(10)\right]^2 = 61.6288 \end{aligned}$$

b) (10) Sabe-se que o intervalo de previsão para a média da nota final à disciplina de Estatística II, considerando o cenário da alínea anterior, é igual a $(53.9971, 69.2005)$. Qual o grau de confiança associado a este intervalo, sabendo que o erro padrão da previsão para a média é igual a 4.621?
**RESOLUÇÃO:**

- Variável fulcral:

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

- Intervalo de Previsão para a média (teórico):

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

com $\mathrm{P}\left(-t_{\alpha/2} < t_{(n-k-1)} < t_{\alpha/2}\right) = 1-\alpha$, logo $\mathrm{P}\left(t_{(n-k-1)} > t_{\alpha/2}\right) = \alpha/2$

- Amplitude do intervalo de Previsão para a média (observado):

$$2\,t_{\alpha/2}\,\mathrm{se}\left(\hat{\theta}\right) = 69.2005 - 53.9971 \Leftrightarrow t_{\alpha/2} = \frac{15.2034}{2 \times 4.621} \Leftrightarrow t_{\alpha/2} = 1.645$$

Logo $\mathrm{P}\left(t_{(674-6-1)} > 1.645\right) = 0.05$ e $\alpha = 0.1$. O grau de confiança associado ao intervalo de previsão para a média é de 90%.

c) (10) Obtenha o intervalo de previsão para a previsão pontual da nota final à disciplina de Estatística II, considerando o cenário da alínea a). É possível que o aluno chumbe? Justifique.

**RESOLUÇÃO:**

- Variável fulcral:

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

- 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 $t_{0.005} = 2.583$ (exato) OU $t_{0.005} = 2.576$ (tabelas)
  - $$\begin{aligned} IP_{99\%}^{\text{Pontual}} &= 61.6288 \pm 2.583 \times \sqrt{4.621^2 + 4.3391^2} \\ &= (45.2555, 78.0022) \\ &\text{OU} \\ IP_{99\%}^{\text{Pontual}} &= 61.6288 \pm 2.576 \times \sqrt{4.621^2 + 4.3391^2} \\ &= (45.2999, 77.9578) \end{aligned}$$

- Interpretação:

  Uma vez que valores abaixo de 50 pontos estão contidos no intervalo de previsão acima, é possível afirmar, com 99% de confiança, que o aluno poderá chumbar.

d) (10) Determine o efeito parcial da percentagem de trabalhos de casa da disciplina de Estatística II que foram entregues sobre a nota final à disciplina. É possível que ocorra mudança do sinal deste efeito, para o conjunto de valores admissíveis da percentagem de trabalhos de casa da disciplina de Estatística II que foram entregues? Justifique.
**RESOLUÇÃO:**

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

$$\frac{\partial \widehat{nfinal}}{\partial tpc} = \hat{\beta}_3 + 2\hat{\beta}_4\, tpc = 0.0094 + 0.00005\, tpc$$

- Interpretação: Uma vez que os valores estimados para $\beta_3$ e $\beta_4$ são positivos, o efeito da percentagem de trabalhos de casa da disciplina de Estatística II que foram entregues sobre a nota final à disciplina é estritamente positivo e crescente. Como tal, nunca irá existir mudança do sinal deste efeito parcial, *ceteris paribus.*

e) (10) Como procederia para testar a correta especificação do modelo? **Nota: não tem de fazer cálculos. Não se esqueça de enunciar o nome do teste para o efeito.**
**RESOLUÇÃO:**

- Passo 1: Obter os valores previstos para $nfinal$;
- Passo 2: Estimar uma regressão auxiliar que adiciona $\widehat{nfinal}^2$, $\widehat{nfinal}^3$ e eventualmente termos de ordem superior, à regressão original:

$$\begin{aligned} nfinal = &\beta_0 + \beta_1 assid + \beta_2 assid^2 + \beta_3 tpc + \beta_4 tpc^2 \\ &+ \beta_5 \ln(med) + \beta_6 \left[\ln(med)\right]^2 + \beta_7 \widehat{nfinal}^2 + \beta_8 \widehat{nfinal}^3 \end{aligned}$$

- Passo 3: Utilizar o teste RESET

$$\begin{cases} H_0 : \beta_7 = \beta_8 = 0 \\ H_1 : \exists \beta_j \neq 0,\, j = 7, 8 \end{cases}$$

  - Estatística de teste:

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

  - Região de rejeição:

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

  Pelo que $f_{0.01} = 4.637$ (exato) OU $f_{0.01} = 4.61$ (tabelas).