Econometria I Lista 8

Profa. Lorena Hakak

Entrega: 27/11/2022

Capítulo 6

Exercício 6

Para saber se vamos incluir no modelo ou não, vamos encontrar o R-quadrado ajustado:

$$\overline{R}^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

$$\overline{R}^2 = 1 - \frac{(1-0,232)(680-1)}{680-8-1}$$

$$\overline{R}^2 =,0223$$

Para testar ao nível de significância de 10%, faremos o teste F:

$$F = \frac{\frac{(R^2_{ir}-R^2_r)}{q}}{\frac{1-R^2_{ir}}{(n-k-1)}} = \frac{\frac{(0,232-0,229)}{673-671}}{\frac{1-0,232}{(671)}} = 1,31$$

Como $F > 1,18$ então rejeita-se a hipótese nula. Eu incluiria essas variáveis pelo teste estatístico aplicado, apesar de alterar o R-quadrado e o R-quadrado ajustado em valores muito pequenos. Há como defender a não inclusão das variáveis pelo critério da parcimônia, em escolher modelos mais simples.

Exercício 8

(i)

Sim, devemos inserir a variável $attend$ na regressão juntamente com $alcohol$, mas não na mesma variável. Isso porque cada uma dessas variáveis possuem impactos em direção contrárias: enquanto o aumento da frequência tende a aumentar as notas, o aumento do consumo de álcool tende a diminuir as notas.

(ii)

Essas variáveis são importantes de serem incluídas pois ajuda no controle do nível prévio dos alunos. Caso essas variáveis não sejam inseridas, o efeito do álcool nas notas de graduação podem estar superestimadas.

Capítulo 7

Exercício 2

(i)

Para cada cigarro a mais que a mãe fuma por dia durante a gravidez esperasse que se reduza 0,5% do peso da criança. Caso a mãe fume pelo menos 10 cigarros por dia, esperasse uma redução do peso da criança em 5%.

(ii)

Espera-se que uma criança branca pese aproximadamente 4,5% a mais que uma criança não branca. A diferença é estatisticamente significante, já que sua estatística t será de 3 (0,045/0,015), sendo significante a 1%.

(iii)

O que o efeito nos diz é que quanto maior os anos de escolaridade da mãe, menor é o peso da criança. Essa variável não se mostrou estatisticamente significante.

(iv)

Não será possível calcular a estatística F por conta de que os dois modelos usam observações diferentes, já que possui um número menor do que a regressão restrita (a primeira). Para poder dar certo, teríamos que utilizar a mesma base de dados para as duas equações.

Exercício 3

(i)

Sim, a variável $hsize^2$ deve ser incluída, já que sua estatística t é de 4,13, sendo estatisticamente significante a 1%. Para calcular o tamanho ótimo, temos que observar o efeito marginal do tamanho da sala:

$$\frac{\Delta sat}{\Delta hsize} = 19,30 - 4,38hsize = 0$$

$$hsize \approx 4,4$$

(ii)

A diferença passa a ser de aproximadamente 45,09 pontos no SAT. É estatisticamente significante a 1%, já que sua estatística t foi de 10,5.

(iii)

Sendo a regressão populacional escrita por $sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_4 black + \beta_5 female.black$, então o teste de hipótese seria:

$$H_0 : \beta_3 = 0 \text{ e } \beta_4 = 0$$
$$H_1 : \beta_3 = 0 \text{ e } \beta_4 \neq 0$$

Além disso, a diferença entre homens negros e não negros é de 169,81 pontos no SAT.

(iv)

A diferença de notas entre mulheres negras e não negras é de 107,05 pontos no SAT (62,31- 169,81). Para ver se é estatisticamente significante, eu precisaria fazer um teste t restringindo β_4 e β_5 do modelo populacional.

Exercício 6

Dado que a correlação entre a variável omitida (aptidão) e o treinamento é negativa, já que o treinamento ocorreu para aqueles com menor aptidão, e o coeficiente da variável omitida será positiva (quanto maior a aptidão, maior o salário esperado) então o viés é negativo, tendo um modelo

Exercício 7

(i)

Sendo a regressão populacional feita em (7.29):

$$inlf = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6$$

E agora utilizando como variável dependente $outlf$, temos que:

$$outlf = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6 + u$$

Como $outlf = 1 - inlf$, então:

$$1 - inlf = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6 + u$$

$$-inlf = -1 + \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6 + u$$

$$inlf = (1 - \beta_0) - \beta_1 nwifeinc - \beta_2 educ - \beta_3 exper - \beta_4 exper^2 - \beta_5 age - \beta_6 kidslt6 - \beta_7 kidsge6 - u$$

(ii)

Não acontece nada como erro padrão das variáveis, sendo que a mudança dos sinais dos betas irão mudar o sinal apenas da estatística t. O mesmo ocorre para o intercepto, já que:

$$Var(1 - \beta_0) = Var(-\beta_0) = Var(\beta_0)$$

(iii)

O R-quadrado não irá se alterar. Isso porque as mesmas variáveis estão sendo modeladas, alterando apenas a estrutura da regressão. Ao invés de ver como as variáveis influenciam a probabilidade de $inlf$ ser 1, no segundo modelo eles mostram a tendência da probabilidade de ser 0.