1) a) $\hat{\beta}_2$: quando área aumenta em 1 metro quadrado, há aumento esperado no preço de venda de 213 reais, ceteris paribus.

$\hat{\beta}_3$: quando número de quartos aumenta em 1 unidade, espera-se que o preço de venda caia, em média, em 3238 reais, ceteris paribus.

$\hat{\beta}_4$: quando número de banheiros aumenta em 1 unidade, espera-se que o preço de venda caia, em média, 1829 reais, tudo o mais constante.

O sinal associado à área está OK, uma vez que residências maiores tendem a ser mais caras. Explicar os outros sinais pode não ser tão simples, uma vez que se o tamanho da residência está fixo, aumentar o número de quartos e banheiros pode não fazer sentido

b) $H_0: \beta_3 = 2\beta_4$

$H_A: \beta_3 \neq 2\beta_4$

Para verificar tais hipóteses, devemos conduzir um teste F-parcial. Seria necessário estimar um modelo restrito e irrestrito.

Modelo irrestrito: proposto em (D)

Modelo restrito: $y_i = \beta_1 + \beta_2 X_{2i} + 2\beta_4 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$

Sob H_0, a estatística de teste é dada por

$$F = \frac{(R^2_{IR} - R^2_R)/q}{(1 - R^2_{IR})/(n-k-1)} \sim F[q;\, n-k-1]$$

onde $g = 1$ $\quad F_{critico}(1;10) = 4,26$

$n = 14$

$K = 3$

Região Crítica = $[4,26; +\infty)$

Se Fobs pertencer à RC, há evidências para rejeitar H_0 com 95% de confiança.

c) Para analisar a relevância de X_3 temos 2 opções:

- realizar um teste t no coeficiente β_3 no modelo (C)
- realizar um teste F-parcial considerando o modelo (C) como irrestrito e o modelo (B) como restrito.

Devido às informações disponíveis, faremos o teste F-parcial.

$H_0: \beta_3 = 0$

$H_A: \beta_3 \neq 0$

$$F = \frac{(R^2_{iR} - R^2_R)/g}{(1 - R^2_{iR})/(n-K-1)} \overset{\text{sob } H_0}{\sim} F_{[g;\, n-K-1]}$$

$$F_{obs} = \frac{0,835 - 0,821}{(1 - 0,835)/11} = 0,93$$

$RC = [4.84, +\infty)$ Como $F_{obs} \notin RC$, não há evidências para rejeitar H_0.

Portanto, o melhor modelo é o (B) pois não inclui uma variável irrelevante, o que poderia ocorrer perda de eficiência dos estimadores, mesmo que não cause viés.

2) a) $\beta_1 < 0$. Quanto maior a proporção alunos por professor, pior deve ser o aprendizado.

$\beta_2 < 0$. Alunos que não possuem inglês como língua materna devem ter mais dificuldades em um teste em inglês.

$\beta_3 < 0$. Se utilizarmos essa variável como um indicador de renda, espera-se que alunos com menor renda tenham um desempenho pior que alunos c/ renda mais elevada.

Interpretação: Todas as interpretações são feitas em função de modelo lin-lin. Assim, p/ variação unitária na variável X_j, é esperado que acarrete variação de β_j pontos no teste, tudo o mais constante.

b) Neste caso, estaríamos diante de um modelo log-lin, e portanto p/ variação unitária na variável X_j, esperaríamos variação de $(100 \cdot \beta_j)\%$ no teste, tudo o mais constante.

Se houvesse alguma escola com testscr = 0, não seria possível utilizar a transformação logarítmica p/ esse dado.

c) $\widehat{TestsCR_i} = 700.15 - 0,998 \cdot STR_i - 0,12\, El\_PCT_i - 0,547\, Meal\_PCT$

$\quad\quad\quad\quad (4,686) \quad (0,239) \quad\quad (0,032) \quad\quad (0,22)$

$R^2 = 0.7746 \qquad n = 420 \qquad \hat{\sigma} = 9,08$

d) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_A: \beta_1 \neq 0$ ou $\beta_2 \neq 0$ ou $\beta_3 \neq 0$

Modelo restrito é dado por $y_i = \beta_0 + \varepsilon_i$ $(R^2 = 0)$

$$F = \frac{R^2_{ir} / q}{(1 - R^2_{ir}) / n-k-1} \overset{\text{Sob } H_0}{\sim} F[q, n-k-1] \qquad q = 3 \quad n-k-1 = 416$$

$F_{obs} = 476,26$

$RC = [2,097; +\infty)$

$F_{obs} \in RC \rightarrow$ há evidências p/ rejeitar H_0

e) Modelo Restrito $y_i = \beta_0 + \beta_1 \cdot STR_i + \varepsilon_i$

$H_0: \beta_2 = \beta_3 = 0$

$H_A: \beta_2 \neq 0$ ou $\beta_3 \neq 0$

Após estimação do modelo restrito, temos que $R^2_R = 0,0512$

$$F_{obs} = \frac{(0,7745 - 0,0512) / 2}{(1 - 0,7746) / 416} = 667,18$$

$RC = [3,017; +\infty)$

$F_{obs} \in RC \rightarrow$ há evidências p/ rejeitar H_0.

3 a) Um modelo capaz de capturar o que os pesquisadores tem interesse é:

$$y_i = \beta_0 + \beta_1 h_i + \beta_2 \, dComp_i + \beta_3 \, h_i \cdot dComp_i + \varepsilon_i$$

em que y_i: nota do i-ésimo estudante

h_i: horas de estudo por semana

$dComp_i$: Dummy que assume o valor 1 quando o aluno i possui computador pessoal.

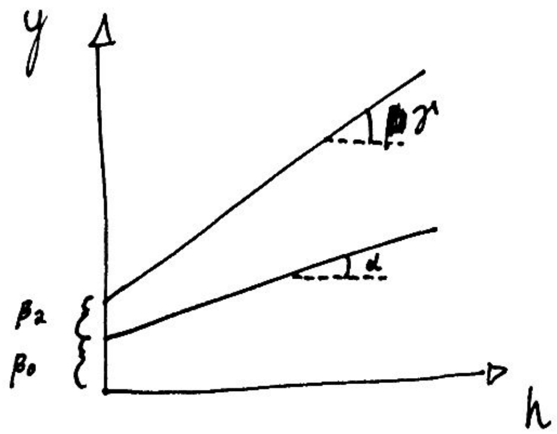
Temos que o efeito de uma hora a mais de estudo por semana sobre a nota média é:

$$\frac{\partial y_i}{\partial h_i} = \beta_1 + \beta_3 \, dComp_i$$

ou seja, quando o aluno tem um computador pessoal ($dComp_i = 1$), esse efeito é $\beta_1 + \beta_3$, e se o aluno não tem computador ($dComp_i = 0$), o efeito é β_1. Esses efeitos só serão diferentes se $\beta_3 \neq 0$. Então, os pesquisadores estariam interessados em testar

$$\begin{cases} H_0 : \beta_3 = 0 \\ H_A : \beta_3 \neq 0 \end{cases}$$

b) Assumindo que $\beta_0, \beta_1, \beta_2$ e β_3 são positivos, teríamos:



para $dComp_i = 0$

$$y_i = \beta_0 + \beta_1 h_i + \varepsilon_i$$

para $dComp_i = 1$

$$y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) h_i + \varepsilon_i$$

$tg\, \alpha = \beta_1$ $\qquad$ $tg\, \gamma = \beta_1 + \beta_3$

4) a) Um aumento de 1 cigarro na média diária de cigarros fumados durante a gravidez gera, em média, uma queda de 0,44% no peso da criança.

b) A criança branca tem, em média, um peso 5,5% maior que as não-brancas. $t_{obs} = 0{,}055/0{,}013 = 4{,}23 \rightarrow$ Significante

c) $t_{obs} = -0{,}003/0{,}003 = -1 \rightarrow$ o efeito é insignificante

d) note que as amostras são diferentes em cada modelo, o que gera um SST diferente para cada caso. Como $R^2 = \frac{SSE}{SST}$, se o SST de dois modelos são diferentes, os R^2 deles não são comparáveis, e por isso não podemos calcular a estatística-F:

$$F = \frac{(R^2_{IR} - R^2_R)/g}{(1 - R^2_{IR})/[n - (K+1)]}$$