Projeto 3 – Parte Teórica

Filipe Borba e Michel Becker

a) Obter B_0 e B_1 a partir do MMQ • $\hat{B}_0$ e $\hat{B}_1$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (\underbrace{y_i - \hat{y}_i}_{[y_i - (B_0 + B_1 x_i)]})^2 = (y_i - B_0 - B_1 x_i)^2$$

$$\frac{\partial \varepsilon}{\partial \hat{B}_0}\bigg|_{\hat{B}_0,\hat{B}_1} = 2(y_1 - \hat{B}_0 - \hat{B}_1 x_1)^{2-1}(-1) + 2(y_2 - \hat{B}_0 - \hat{B}_1 x_2)^{2-1}(-1) + \dots + 2(y_n - \hat{B}_0 - \hat{B}_1 x_n)^{2-1}(-1)$$

$$\frac{\partial \varepsilon}{\partial \hat{B}_0}\bigg|_{\hat{B}_0,\hat{B}_1} = (-2)\sum_{i=1}^{n}(y_i - \hat{B}_0 - \hat{B}_1 x_i) = 0 \Rightarrow \boxed{n\hat{B}_0 + \hat{B}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i} \quad (1)$$

SIMPLIFICANDO

$$\frac{\partial \varepsilon}{\partial \hat{B}_1}\bigg|_{\hat{B}_0,\hat{B}_1} = (-2)\sum_{i=1}^{n}(y_i - \hat{B}_0 - \hat{B}_1 x_i)x_i = 0 \Rightarrow \boxed{\hat{B}_0 \sum x_i + \hat{B}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i} \quad (2)$$

Dividindo (1) por n: $\hat{B}_0 + \hat{B}_1 \bar{x} = \bar{y} \to \boxed{\hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x}}$ (3)

Substituindo (3) em (2) e dividindo por n:

$$\frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i}{n} - \hat{B}_1 \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} + \hat{B}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i$$

$$\hat{B}_1 \left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right) = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right) \cdot \left(\sum_{i=1}^{n} x_i\right)}{n}$$

$$\boxed{\hat{B}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right) \cdot \left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}} \quad (4)$$

b) Na regressão linear, podemos assumir que os erros possuem distribuição Normal, média (µ) ou valor esperado (E(ε_i)) igual a zero, variância (Var(ε_i)) constante (homoscedasticidade), além de não haver correlação ($Corr(\varepsilon_i, \varepsilon_j) = 0$) entre eles. A adequação dessas suposições pode ser checada, na prática, através da análise de curvas de probabilidade cumulativa dos resíduos e da distribuição normal, verificando se as curvas são parecidas para identificar se o erro é ou não uma distribuição normal; construindo um intervalo de confiança para a média como sendo nula, a fim de verificar a suposição da média, além de verificar graficamente se isso se cofirma, entre outros recursos.

c) Os testes de hipótese na regressão simples servem para verificar quão boa é a regressão para explicar a variável resposta. Para tanto, seria interessante a realização de um teste t-Student para verificar se $\beta_1 = 0$, sendo essa a hipótese nula. A rejeição de H_0 indicaria então que não há relação linear entre a variável explicativa (x) e a variável resposta (y), ao passo que a não-rejeição dela indicaria alguma relação.

d) Sim, é possível fazer uma regressão múltipla, mudando apenas a equação inicial e o teste de hipóteses, com as suposições do modelo permanecendo iguais. Na equação, devem ser adicionados mais termos para que as novas variáveis explicativas sejam levadas em conta, de modo que a nova equação seja da forma $y = B_0 + B_1 x_1 + B_2 x_2 + \cdots + B_p x_p + \varepsilon_i$. Sendo p a quantidade de variáveis. No caso do teste de hipótese, ele teria de ser feito para cada uma das variáveis explicativas, logo, se existem n variáveis explicativas, devem existir n testes de hipótese. (OBS: Poderia ser feito um teste-F para avaliar todas as variáveis de uma vez).