STA502 Math and Stat for MBA

Assignment 2.

Problem Set 4 & 5

Taylor H. Moon

21/09/13

---

4. multiple Regression $y = X\beta + \varepsilon$ with k explanatory variables.

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_K X_K + \varepsilon$

4-1) $y = \beta_0 + \beta_1 X_1 + \cdots + \frac{\beta_i}{\lambda}(\lambda X_i) + \cdots + \beta_K X_K + \varepsilon$

we have to standardize our data set with same scale,

we have to devide corresponding regression coefficient by

It is important that our data have standardized scale of unit. Therefore in the case where one particular explanatory variable has millions of kgs when others have thousands of kgs, we have to devide the corresponding regression coefficient by 1000 to have uniform scale.

4-2) $y = \beta_0 + \beta_1 X_1 + \cdots + \beta_i (X_i + \lambda) + \cdots + \beta_K X_K + \varepsilon$

$y = (\beta_0 + \lambda\beta_i) + \beta_1 X_1 + \cdots + \beta_i X_i + \cdots + \beta_K X_K + \varepsilon$

When a constant λ is added to all observations of one particular variable, X_i, constant term of regression is increase by corresponding regression coefficient, β_i, multiplied by λ. All other coefficients remain the same.

From the perspective for logarithmic transformation,

Linear model $Y_i = \alpha + \beta X_i + \varepsilon$ can be transformed into $Y_i = \alpha + \beta \ln X_i + \varepsilon_i$.

Since $\log X + 1 = \log X + \log e$, $e^Y = e^{\beta_0} \cdot X_1^{\beta_1} \cdot X_2^{\beta_2} \cdot X_3^{\beta_3} \cdot \cdots \cdot e^{\beta_i \lambda} \cdot X_i^{\beta_i} \cdot \cdots \cdot X_K^{\beta_K} \cdot e^{\varepsilon}$

Therefore, corresponding coefficient is independent from λ that was added to explanatory variable, X_i..

5. $n=100$ $Y=\beta_0+\beta_1 X+\beta_2 X^2+\beta_3 X^3+\varepsilon$

5-1) true relationship between X & Y is Linear $Y=\beta_0+\beta_1 X+\varepsilon$

if true relationship to X and Y is linear, it is highly likely that predicted model, $\hat{y}=\hat{\beta}_0+\hat{\beta}_1 X+\hat{\varepsilon}$, will explain the data better than cubic regression model assuming her linear model was the most optimized function. Therefore, I would like to conclude that training RSS of linear model has lower value than training RSS for cubic model.

5-2) Since I do not have enough information on given data, I wouldnot be able to make a judgement. However, the complex the polynomial regression gets, the more I can explain the given data since the regression model fits better. Given the cubic regression model was the most optimized model, cubic regression model has higher chance of getting lower value for RSS than linear regression model.