REGRESSION ANALYSIS - HOMEWORK 3 SOLUTIONS

ALAITTIN KIRTISOGLU

1) We use the residual plots to check if the LINE condition is met in SLR or MLR.

Consider the sample correlation coefficients:

$$cor(e_i, \hat{y}_i) = \frac{\sum(e_i - \bar{e})(\hat{y}_i - \bar{y})}{\sqrt{\sum(e_i - \bar{e})^2 \sum(\hat{y}_i - \bar{y})^2}} = 0 \text{ when the LINE condition is met.}$$

$$cor(e_i, y_i) = \frac{\sum(e_i - \bar{e})(y_i - \bar{y})}{\sqrt{\sum(e_i - \bar{e})^2 \sum(y_i - \bar{y})^2}} \text{ may not be zero even if the LINE condition is met.}$$

Note that $\bar{\hat{y}} = \bar{y}$.

This makes the residual plot e_i against $\hat{y}_i$ more meaningful.

2.1) $Var(\hat{y_0}) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0) = Var(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_0) = Var(\bar{y} + \hat{\beta}_1(x_0 - \bar{x}))$

. $= Var(\bar{y}) + Var(\hat{\beta}_1)(x_0 - \bar{x})^2 + 2Cov(\bar{y}, \hat{\beta}_1(x_0 - \bar{x})) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)$. To conclude this,

. you need to show that $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ and $Var(\bar{y}) = \frac{\sigma^2}{n}$ and $Cov(\bar{y}, \hat{\beta}_1) = 0$.

2.2) $Var(\hat{\beta}) = Var((X^T X)^{-1} X^T Y) = ((X^T X)^{-1} X^T) Var(Y) ((X^T X)^{-1} X^T)^T = \sigma^2 (X^T X)^{-1}$. Hence,
. $Var(\hat{y_0}) = Var(x_0^T \hat{\beta}) = x_0^T Var(\hat{\beta}) x_0 = \sigma^2 x_0^T (X^T X)^{-1} x_0$.

2.3) $Var(y_0 - \hat{y_0}) = Var(y_0) + Var(\hat{y_0}) + 2Cov(y_0, \hat{y_0}) = \sigma^2 + \sigma^2 x_0^T (X^T X)^{-1} x_0$, since $Cov(y_0, \hat{y_0}) = 0$.

3) The full model is given.

General F-statistic:

$F^\star = \frac{SSE(R) - SSE(F)}{df_R - df_F} \Big/ \frac{SSE(F)}{df_F} \sim F(p-q, n-p-1)$, $\quad df_F = n - p - 1$, $\quad df_R = n - q - 1$, and $p = 4$.

3.1) Reduced model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$, $\quad q = 3$ and $F^\star = \frac{SSE(R) - SSE(F)}{1} \Big/ \frac{SSE(F)}{n-5} \sim F(1, n-5)$.

3.2) Reduced model $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \epsilon$, $\quad q = 2$, $\quad F^\star = \frac{SSE(R) - SSE(F)}{2} \Big/ \frac{SSE(F)}{n-5} \sim F(2, n-5)$.

3.3) Reduced model $Y = \beta_0 + \epsilon$, $q = 0$, $SSE(R) - SSE(F) = SST - SSE(F) = SSR$. $F^\star = \frac{SSR}{4} \Big/ \frac{SSE}{n-5} \sim F(4, n-5)$.

Date: October 24, 2023.

4) $\sum_{i=1}^{n} Var(\hat{y}_i) = \sum_{i=1}^{n} Var\left(\sum_{j=1}^{n} h_{ij}y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}^2 Var(y_j) = \sigma^2 \sum_{i=1}^{n} h_{ii} = \sigma^2(p+1)$ by 5.2.

Note that $\sum_{j=1}^{n} h_{ij}^2 = h_{ii}$ due to $H = HH^T$.

5.1) $h_{ii} = \mathbf{h_i}^T\mathbf{h_i} = \sum_{j=1}^{n} h_{ij}^2 = h_{ii}^2 + \sum_{j\neq i} h_{ij}^2$. Hence, we have $h_{ii} \geq h_{ii}^2$, which implies that $0 \leq h_{ii} \leq 1$.

5.2) $\sum_{i=1}^{n} h_{ii} = trace(H) = trace(X(X^TX)^{-1}X^T) = trace((X^TX)^{-1}X^TX) = trace(I_{p+1}) = p+1$

5.3) `The answer is in the following page.`

Illinois Institute of Technology, Applied Mathematics, Chicago, IL, USA

Email address: `akirtisoglu@hawk.iit.edu`