HW # 9. 10.8, 10.12, 10.18, 10.24

10.8 (See MTB File & Word Doc.) a.) Added Variable Plots show X_1, X_2, & X_4 provide coefficients of partial determination that are indications those variables would be added to a model containing the other 3: $R_{Y1|234} = 36.8\%$, $R_{Y2|134} = 20.8\%$; $R_{Y3|124} = 0.4\%$; $R_{Y4|123} = 30.1\%$. [Note that best subsets found X_1, X_2, X_4 is the best model for first order additive.] b. The plots don't show much evidence of higher order curvature effects. The p-value for adding X_3 to the model given $X_1 X_2 X_4 = 0.568$.

10.12 a) Bonferroni threshold significance level $\alpha^* = \frac{\alpha}{n} = \frac{.01}{81} = 0.000123$. Bonferroni critical point is $t_{.000123/2,\,81-5-1} = t_{.0000617\,28,\,75} = \underline{4.05}\ (\alpha = .01)$ or $\underline{3.35}\ (\alpha = .1)$. Largest $t_i = \underline{-3.0721}$.

(SEE .MTB and .xlsx files)

b) $2 \cdot \bar{h}_{ii} = 2\left(\frac{p}{n}\right) = 2\left(\frac{5}{81}\right) = 0.1235$. $h_{ii} \geq 0.1235$ are #3, 8, 53, 61, 65 (See plot in MTB).

c) $X'_{new} = [1\ 10\ 12\ .05\ 350{,}000]$ $h_{new,new} = X'_{new}(X'X^{-1})X_{new} = \underline{0.0529}$. Not a hidden extrapolation.

d) Compare Cook's D_i to $F_{.50,5,76} = \underline{.878075}$. Column F(5,76) are D_i percentiles. None are too large. DFFITS comparison $= 2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{5}{81}} = .497 \Rightarrow$ 6, 53, 61, 62 all influential. DFBETAs on transformed data $\Rightarrow$ #6 & #62 are most influential, #61, 3, 53 affect one parameter each; #8 DFBETAs are smaller than $2\sqrt{\frac{1}{n}}$.

e) AVG % DIFF between $\hat{Y}$ and $\hat{Y}_{(i)}$ are 3: .192%; 6: $\underline{.556\%}$; 8: .054%; 53: .235% 61: .360%; 62: .417%. All are less than 1%.

f) Cook's D_i plot in MTB. All are < .2, max is $\underline{.137}$. (#6).

10.18 a) Scatter Plot Matrix and Correlations show significant correlations. The Smoother & Regression Lines show possible higher order effects

b) VIF for transformed data: X_1^*: $\underline{1.240}$; $X_2^* = \underline{1.648}$; $X_3^* = \underline{1.324}$; $X_4^* = 1.413$ $\overline{VIF} = 1.406$, none above 10 $\Rightarrow$ no multicollinearity problems.

10.24 $(AB)^{-1} = B^{-1}A^{-1}$ from (5.34). If $n = p$ $X_{p\times p} = X_{n\times n}$ and $H_{p\times p} = H_{n\times n}$.

$H = X(X'X)^{-1}X' = X[X^{-1}(X')^{-1}]X' = [X \cdot X^{-1}][(X')^{-1} \cdot X'] = I \cdot I = I_{n\times n} = I_{p\times p}$

Since $\hat{Y} = HY = I_n \cdot Y = Y$, all points fall on regression function. This means $\hat{\sigma}^2 = 0$ ... an overspecified model.