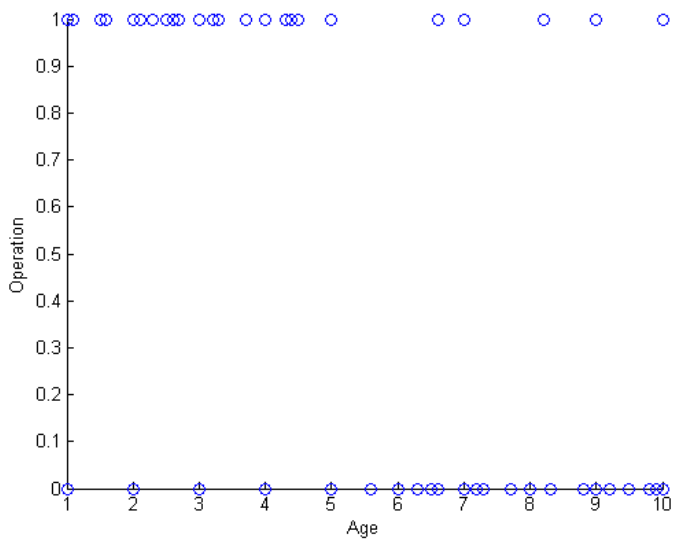
1.

Solution:
Use $\sum_{i=1}^{8} \hat{e}_i = 0$ to get $\hat{e}_8$,
the 8^{th} residual of the model fitting Y versus $X_1, X_2, X_3 = -0.34918$,
the 8^{th} residual of the model fitting X_4 versus $X_1, X_2, X_3 = 0.417$,
$\therefore \hat{\beta}_4 = \frac{\sum \hat{e}_{Y|X_1,X_2,X_3} \hat{e}_{X_4|X_1,X_2,X_3}}{\sum \hat{e}^2_{X_4|X_1,X_2,X_3}} = \frac{-0.8728}{2.5873} = -0.3373.$

2.



Solution:
I will not use simple linear regression model to study the data.
If $y = \beta_0 + \beta_1 X + e$, we cannot guarantee the right hand side (fitted value/prediction) can't take integer value (0 or 1).

3. It is assumed that there is an intercept.

(a)

Solution:
The number of dummy variables without interaction,= (3-1)+(2-1)=3.

(b)

Solution:
The number of dummy variables with interaction,= 2×3 -1=5.

4.

Solution: Please refer to lecture slide chapter 7.

5. (a)

Solution: $RSS = \hat{\sigma}^2(n-p-1) = 0.6635^2 \times 3 = 1.3207$.

(b)

Solution:
$R^2 = 1 - \frac{RSS}{TSS}$
$TSS = \frac{RSS}{1-R^2} = \frac{1.3207}{1-0.9323} = 19.5081$.

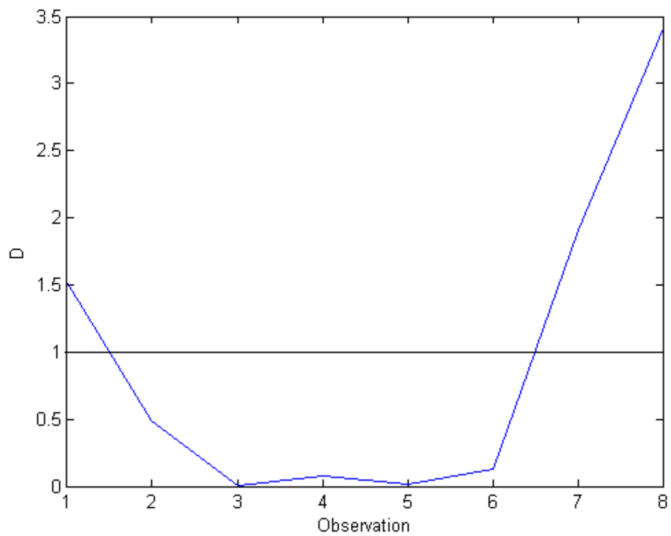
(c)

Solution:
The degree of freedom of the F-statistic is 4 and 3.

(d)

Solution:
$\hat{y}|_{x_6} = x_6\hat{\beta} = 5.4566$
$\hat{Var}(\hat{y}|_{x_6}) = \hat{\sigma}^2 x_6 (X'X)^{-1} x_6' = \hat{\sigma}^2 h_{66} = 0.6635^2 \times 0.378 = 0.1664.$
$t_{0.005,3} = 5.8409,$
The required C.I. is:
$\hat{y}|_{x_6} \pm t_{0.005,3}\sqrt{\hat{Var}(\hat{y}|_{x_6})} = 5.4566 \pm 5.8409\sqrt{0.1664} = [3.0740, 7.8392].$

(e)



Solution: $D_i = (\frac{\hat{e}_i}{1-h_{ii}})^2 \frac{h_{ii}}{(p+1)\hat{\sigma}^2} = (\frac{\hat{e}_i}{1-h_{ii}})^2 \frac{h_{ii}}{5\times 0.6635^2}$
$D_1 = 1.5212, D_2 = 0.4837, D_3 = 0.0039, D_4 = 0.0721,$
$D_5 = 0.0183, D_6 = 0.1310, D_7 = 1.9112, D_8 = 3.4174,$
Since $D_1, D_7 \& D_8 > 1$, observations 1,7,8 are potential outliers.

(f)

Solution: $t_i = r_i\sqrt{\frac{n-p-2}{n-p-1-r_i^2}}$
$r_4 = \frac{\hat{e}_4}{\hat{\sigma}\sqrt{1-h_{44}}} = \frac{0.1155}{0.6635\sqrt{1-0.749}} = 0.3475$
$t_4 = 0.2770 < t_{0.025,2} = 4.303$
$\therefore$ Observation 4 is not an outlier.

6. (a)

Solution: Let $\bar{Y} = \mathbb{1}\bar{y} = \frac{1}{n}\mathbb{1}\mathbb{1}'Y = \frac{1}{n}JY$,
$Cov(\bar{Y}, \hat{y}) = Cov(\frac{1}{n}JY, HY) = \frac{1}{n}JCov(Y,Y)H = \frac{1}{n}\sigma^2 JH = \frac{\sigma^2}{n}J.$
$\therefore Cov(\bar{y}, \hat{y}) = \frac{\sigma^2}{n}\mathbb{1}.$

(b)

Solution: $E[\sum(\hat{y}_i - \bar{y})^2] = E[Y'(H - J/n)Y]$
$= tr\{E[Y'(H - J/n)Y]\}$
$= E\{tr[Y'(H - J/n)Y]\}$
$= E\{tr[(H - J/n)YY']\}$
$= tr[(H - J/n)E(YY')]$
$= tr[(H - J/n)(\sigma^2 I + X\beta\beta'X')]$
$= \sigma^2 tr(H - J/n) + tr\{(H - J/n)X\beta\beta'X'\}$
$= \sigma^2(p + 1 - 1) + tr\{\beta'X'(H - J/n)X\beta\}$
$= p\sigma^2 + tr\{\beta'X'(I - J/n)X\beta\} \qquad \because X'HX = X'X$
$= p\sigma^2 + \beta'X'(I - J/n)X\beta.$

7. (a)

Solution: Problems: Curve relationship/non-linear mean function.
Solutions: Transformation/ polynomial regression.

(b)

Solution: Problems: Non-constant variance.
Solutions: Weighted least squares.

(c)

Solution: Problems: Potential outliers and influential points.
Solutions: Do outlier detection test. Remove it if it is an outlier.

8.

Solution: $\hat{e} = Y - \hat{Y} = Y - HY = (I-H)Y = (I-H)(X\beta + e)$
$= (I-H)X\beta + (I-H)e = (X\beta - X\beta) + (I-H)e = (I-H)e$
Since $(I-H)X = 0$, $I-H$ is not invertible. $\therefore (I-H)^{-1}$ doesn't exist and we don't have $e = (I-H)^{-1}\hat{e}$.

9.

Solution: The model is:
$$E(y|Temp) = \begin{cases} c_1 & \text{if} \quad Temp < K \\ c_2 & \text{if} \quad Temp \geq K \end{cases}$$
Therefore, $y_i = c_1 \mathbb{1}_{\{Temp_i < K\}} + c_2 \mathbb{1}_{\{Temp_i \geq K\}} + e_i$.
For a fixed value of K, we can do OLS for model fitting.
If K is unknown, we should fit the model with different values of K and use the one with the minimum RSS.
Therefore, we can help the scientists that the temperature should be kept below $\hat{K}$ so as to reduce the emission rate.