(a) If we believe that the case is an outlier because of a blunder, then we might delete the outlier and analyze the remaining cases without the suspected case. Otherwise, we can try to find out why a particular case is outlying.

(b) False.

(c) It is easy to interpret the transformation.

(d) The responses are constrained in [0, 1], and difficult to observe.

(e) Yes. All weights are equal to one.

(f) The plots will tell us more information about the fitting model, such as a few relatively large residuals and so on.

2. (a) nonlinearity trend and suspected outlier.

(b) non constant variance.

(c) suspected outlier and influential point.

(d) normality is in doubt and suspected outlier.

3. (a) $\hat{\beta}_1 = \frac{SXY}{SXX} = \frac{15924.51}{4168.62} \approx 3.8199$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx 10.2026$

(b) $E(Y|X=60) = 10.2026 + 3.8199 \times 60 = 239.3966$

$RSS = SYY - \frac{SXY^2}{SXX} \approx 118.2180$, $\hat{\sigma}^2 = \frac{RSS}{n-2} \approx 13.1353$

$\text{sefit}(y|x) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{11} + \frac{(60-\bar{x})^2}{SXX}\right)} \approx 1.4242$

$\therefore$ 95% confidence interval is : $(235.2412, 243.552)$ ~~[236.1779, 242.6153]~~

(c)

Source	df	SS	MS	F
Regression	1	60830.17	60830.17	4631.045
Residual	9	118.2180	13.1353	
Total	10	60948.39		

$\therefore$ Reject $H_0: E(Y|X=x) = \beta_0$

4\. (a) Because when $K \to \infty$, the regression model will become perfectly fitted.

Suggested method: choose some values of K, and use AIC, BIC and so on to do model comparison.

(b) $$\frac{\partial}{\partial b} \sum_{t=1}^{n} [Y_t - (3 + b\cos(4t))]^2 = -2 \sum_{t=1}^{n} [Y_t - 3 - b\cos(4t)] \cos(4t)$$

$$\therefore \hat{b} = \frac{\sum_{t=1}^{n} (Y_t - 3)\cos(4t)}{\sum_{t=1}^{n} \cos^2(4t)}$$

(c) $$\frac{\partial}{\partial w} \sum_{t=1}^{n} [Y_t - 3 - 2\cos(wt)]^2$$

$$= +2 \sum_{t=1}^{n} (Y_t - 3 - 2\cos(wt)) 2t \sin(wt)$$

$$= 4 \sum_{t=1}^{n} (Y_t - 3 - 2\cos(wt))\, t \cdot \sin(wt)$$

$\therefore$ The OLS estimator for w doesn't have a closed form. We should use some numerical methods, such as Newton's method and so on.

5. Define $U = (u_1, u_2, u_3, \cdots, u_n)$ where $u_4 = 1$ and $u_7 = 2$

others are all zero. We consider the following model.

$$E(Y \mid X = x) = \beta_0 + \beta_1 X + \delta U$$

and we use OLS to get the estimate of δ, denoted as $\hat{\delta}$.

and calculate the s.e. of $\hat{\delta}$, denoted as $se(\hat{\delta})$.

H_0: $\delta = 0$ v.s. H_1: $\delta \neq 0$

$|t| = \left| \dfrac{\hat{\delta}}{se(\hat{\delta})} \right|$ and we compare $|t|$ with $t_{0.025}(n-3)$ to

determine whether reject or accept H_0.

6. (a) $Var(\hat{\theta}) = Var(2\hat{\beta}_1 - \hat{\beta}_0) = Var(2\hat{\beta}_1) + Var(\hat{\beta}_0) + 2Cov(2\hat{\beta}_1, \hat{\beta}_0)$

$= 0.4766^2 \times [4 \times 0.13820855 + 0.045229535 + 4 \times 0.06215878]$

≈ 0.1923

$H_0: \theta = 0$ v.s. $H_1: \theta \neq 0$

$$t = \frac{2\hat{\beta}_1 - \hat{\beta}_0}{se(\hat{\theta})} = -0.1146963,$$

$\because |t| < 1.96$, $\therefore$ Don not reject H_0.

(b) $g = \frac{\hat{\beta}_1}{\hat{\beta}_2}$, $\therefore \left(\frac{\partial g}{\partial \underset{\sim}{\beta}}\right)' = \left(0, \frac{1}{\hat{\beta}_2}, -\frac{\hat{\beta}_1}{\hat{\beta}_2^2}\right)$

$$\therefore Var(g(\hat{\theta})) = \left(0 \quad \frac{1}{\hat{\beta}_2} \quad -\frac{\hat{\beta}_1}{\hat{\beta}_2^2}\right)\left[\hat{\sigma}^2(X'X)^{-1}\right]\begin{pmatrix} 0 \\ \frac{1}{\hat{\beta}_2} \\ \frac{-\hat{\beta}_1}{\hat{\beta}_2^2} \end{pmatrix} \approx 0.06021$$

(c) There is no problem in Figures 2(b) to 2(d)

(d) No. LOF test can tell us whether the model fit the data, but can not tell us whether the curvature exists.

(e) Because the p-value for x_i^2 is $0.02564 < 0.05$, $H_0: \beta_3 = 0$ is rejected, x_i^2 should be included in the model.