Question 4

[Answer] 1. If k = 3 , the linear regression model y = X$\beta + \epsilon$ can be rewritten as below.

$$\mathrm{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

From the given information where an explanatory variable is multiplied by λ,

$$\mathrm{y} = \alpha x_1^* + \beta_2 x_2 + \beta_3 x_3 + \epsilon \text{ (where } \mathrm{x}_1^* = \lambda x_1)$$
$$\alpha x_1^* = \alpha\lambda x_1 = \frac{\beta_1}{\lambda}(\lambda x_1) = \beta_1 x_1$$
$$\alpha = \frac{\beta_1}{\lambda}$$

Applying the above information to the problem where x_1 is measured in thousands of kgs instead of millions of kgs:

$$\mathrm{x}_1 = 1000000\delta,\ \mathrm{x}_1^* = 1000\delta$$
$$\mathrm{x}_1^* = 1000\delta = \frac{1}{1000}1000000\delta = \frac{1}{1000}x_1$$
$$\lambda = \frac{1}{1000}$$
$$\alpha = \frac{\beta_1}{\lambda} = \frac{\beta_1}{\frac{1}{1000}} = 1000\beta_1$$

Hence, if an explanatory variable is measured $\frac{1}{1000}$ times to the original variable, its coefficient will turn out to be 1000 times bigger than the original coefficient.

2. If a constant λ is added to a particular explanatory variable in the given regression equation containing a constant term, the equation can be rewritten as below.

$$\begin{aligned} \mathrm{y} &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \\ \mathrm{y} &= \beta_0 + \beta_1(x_1 + \lambda) + \beta_2 x_2 + \beta_3 x_3 + \epsilon \\ &= \beta_0 + \beta_1 x_1 + \beta_1\lambda + \beta_2 x_2 + \beta_3 x_3 + \epsilon \\ &= \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_2 x_3 + \epsilon \end{aligned}$$

where $\alpha = \beta_0 + \beta_1\lambda$. Thus, if a constant is added to a particular explanatory variable, the intercept would be biased by the constant term multiplied by the coefficient of the explanatory variable.

Question 5

[Answer]

1. If the true relationship between X and Y is linear, the RSS for the linear regression will be lower than the one for the cubic regression. It's because all the non-linear curves of the cubic regression that is irrelevant to the true linear regression line will be regarded as residuals, consequently enlarging the RSS. On the contrary, the linear regression is likely to cover more data points of the true linear regression line compared to the former, hence resulting in a lower RSS.

2. There is not enough information to tell since there is too many possible forms of non-linearity that the true relationship between X and Y can take.