4. (1) $y = \beta_0 \mathbb{1} + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k + \varepsilon$ where x_j is measured in millions of kgs

If we convert the unit of x_j in millions of kgs to in thousands of kgs, we would multiply 1000 to x_j

Then, $y = \beta_0 \mathbb{1} + \beta_1 x_1 + \dots + \frac{\beta_j}{1000}(1000 x_j) + \dots + \beta_k x_k + \varepsilon$

Therefore, the corresponding regression coefficient would be multiplied by $\frac{1}{1000}$. If we define $\beta_j{}^* = \frac{\hat{\beta}_j}{1000}$, the variance would also be changed.

$Var(\hat{\beta}_j{}^*) = \frac{1}{10^6} Var(\hat{\beta}_j)$. However, t-statistic would be same since

$$t_{\beta_j^*} = \frac{\hat{\beta}_j^*}{s.e(\hat{\beta}_j^*)} = \frac{10^{-3}\hat{\beta}_j}{10^{-3}s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} = t_{\beta_j}$$

(2) $y = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k + \varepsilon$ add λ to x_j (particular explanatory variable)

$y = (\beta_0 - \beta_j \lambda \cdot \mathbb{1}) + \beta_1 x_1 + \dots + \beta_j (x_j + \lambda \cdot \mathbb{1}) + \dots + \beta_k x_k + \varepsilon$

Therefore, the corresponding regression coefficient β_j is unaffected, while the intercept is shifted by $-\beta_j \lambda \cdot \mathbb{1}$

Let's use this result to logarithmic form.

$y = \beta_0 + \beta_1 \log(x_1) + \dots + \beta_j \log(x_j) + \dots + \beta_k \log(x_k) + \varepsilon$

If we multiply λ to a particular explanatory variable x_j,

$y = \beta_0 + \beta_1 \log(x_1) + \dots + \beta_j \log(\lambda x_j \frac{1}{\lambda}) + \dots + \beta_k \log(x_k) + \varepsilon$

Since $\log(ab) = \log(a) + \log(b)$, the above equation will become $y = \beta_0 + \beta_1 \log(x_1) + \dots + \beta_j(\log(\lambda x_j) - \log \lambda) + \dots + \beta_k \log(x_k) + \varepsilon$

$= (\beta_0 - \beta_j \log \lambda) + \beta_1 \log(x_1) + \dots + \beta_j \log(\lambda x_j) + \dots + \beta_k \log(x_k) + \varepsilon$

Therefore, the intercept would be shifted by $-\beta_j \log \lambda$, while the slope coefficient would be unaffected.

In conclusion, the corresponding coefficient is independent of the units in which the variable is measured

5. (1) I think training RSS will be lower in cubic regression than in linear regression even if the true relationship between X and Y is linear. Cubic regression has a higher flexiblity than the linear one, which leads to lower RSS. Training RSS is regardless of the true relationship
In contrast, test RSS would be higher in cubic regression than linear regression since linear regression has lower bias and variance than cubic regression if the true relationship is linear

(2) If the true relationship between X and Y is not linear, the cubic regression would still be better than the linear regression since the increased flexiblity of the cubic regression will make lower RSS by closely following points.
Moreover, we cannot compare test RSS since cubic regression has less bias than linear one if the true relationship is not linear. But cubic regression has larger variance. Therefore, we cannot conclude which regression has lower test RSS.