STATS 500 - Practice Exam 1 Solutions

1. $n = 30$, $p = 4$ predictors (5 parameters) and $\hat{\sigma} = 31.5$.

2. $10 \times 28.92 = 289.2 Kcal/kg$ more.

3. This corresponds to $Water = 100$ and the rest=0 so the predicted content is $2245.09 - 100 \times 37.36 = -1490.9 Kcal/kg$ – this is clearly wrong (negative energy). This is called *extrapolation* (there are no real waste samples made up of water), so we could not expect a reliable result.

4. Noise in predictors generally worsens the fit, so without noise one would expect R^2 to be larger (assuming the linear model is true).

5. RSS will increase when we remove a variable.

6. $7.64 \pm 2 \times 2.31$ gives $(3.02, 12.26)$.

7. The F-test in this case is equivalent to the t-test for dropping `Garbage` , so $F = t^2 = 2.24^2 = 5.02$.

8. Not necessarily. $\hat{\sigma} = \sqrt{RSS/df}$ and removing a predictor will increase RSS but will also increase the df by one, so a decrease is possible.

9. The correlation between the predictors is likely to be negative (opposite of the correlation between coefficients).

10. Reject H_0, because the origin is outside the ellipse.

11. No: there is no non-linear trend or non-constant variance.

12. No: this is not big enough once you take into account the Bonferroni correction. 2-sided p-value is $0.021 \times 2 = 0.042$, which is bigger than $\alpha/n = 0.05/30$.

13. Yes: a reduction of 0.5 in the coefficient for `Garbage` might reduce the t-statistic below 2 which could lead to insignificance. Since the SE will also change when you drop a point, there is no way to tell for sure.

14.

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}((X^T X)^{-1} X^T y) \\ &= (X^T X)^{-1} X^T \text{Var}(y) X (X^T X)^{-1} \\ &= (X^T X)^{-1} X^T (\sigma^2 I) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} (X^T X) (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

15. Data broken up in two non-overlapping clusters in the partial residual plot indicates that the relationship between the response and predictor variables *may* be different in the two groups (clusters). To check this, one can fit the original model separately to the two subsets of the data.

16. Regression effect: points which are A standard units away from average x will get fitted values that are B standard units away from average y, where $B < A$. This is also called "regression to the mean" (everything is predicted to be closer to average). This effect is a feature of the linear model and does not indicate anything unusual or interesting about the data.

17. The estimates are typically biased due to results in class on errors in (observed) predictors – the underlying reason why the proposed argument is false is that the input variables are no longer deterministic, but are random and and typically have a correlation with errors, a violation of the standard linear model assumptions.