7. After changing the largest volume value to 35, the correlation coefficient between volume and diameter is 0.9013, and between volume and height is 0.5298.
8. Slope is 4.021, intercept is -24.452.

2.4

1. No effect as the constant will "cancel out" since it will also be affect the mean of the x-values.
2. No effect as the constant will be present in both the numerator and denominator and will "cancel out".

2.5

2. The correlation coefficient between temperature and cookie score is -0.00715.
3. We see a strong association between temperature and cookie score but it is clearly quadratic and not linear. Despite the low correlation coefficient the temperature and score are clearly related.

2.6 The least square estimates are $\hat{\alpha} = \bar{y} - \bar{x}$ and $\hat{\alpha} = \bar{y} - 2 \cdot \bar{x}$, respectively.

2.7 Upper left: 0.70. Upper right: -0.25. Lower left: 0.17. Lower right: -0.86.

2.8

1. Correlation coefficient: 0.9340
2. Slope: $\hat{\beta} = 1.5737$. Intercept: $\hat{\alpha} = -0.1008$.
3. If we take logarithms on both sides the model is

$$\log(\mathrm{FW}) \approx c + \log(\mathrm{DM})$$

where $c = log(k)$. Thus we should fit a linear regression model where we force the slope to be equal to 1. Can use the result from the previous exercise and we get $\hat{c} = 0.4122$.

3.1

1. Bottom < upper left < upper right
2. Upper right < Upper left < bottom.
3. The important thing is how large $\mathrm{SS}_{\mathrm{grp}}$ is compared to SS_e.

3.2

1. $k = 3$, $n_{\mathrm{Cont}} = 9$, $n_{\mathrm{P_2O_7}} = 9$, $n_{\mathrm{HMP}} = 8$.

   $g(i) =$ Control for $i = 1, \dots, 9$; $g(i) = \mathrm{P_2O_7}$ for $i = 10, \dots, 18$; $g(i) =$ HMP for $i = 19, \dots, 26$;
2. The α's are interpreted as the expected values of the tartar index for a random dog that gets the corresponding treatments. The parameters are estimated by the sample means: 1.089, 0.747 and 0.438, respectively.
3. $s = 0.368$.

4. Use a command like `boxplot(index~treat)`.
5. Use a command like `lm(index~treat-1)`.
6. Use a command like `summary(lm(index~treat-1))`.

3.3

2. Only the j'term in the sum depends on α_j, so the minimum of Q is obtained by minimizing each of the terms.
3. For each j we get $f_j'(\alpha_j) = -2\sum_{i:g(i)=j}(y_i - \alpha_j) = -2\sum_{i:g(i)=j} y_i + 2n_j\alpha_j$ so $f_j'(\alpha_j) = 0$ if and only if $n_j\alpha_j = 2\sum_{i:g(i)=j} y_i$ if and only if $\alpha_j = \frac{1}{n_j}\sum_{i:g(i)=j} y_i = \bar{y}_j$.
4. There is only one solution to $f_j'(\alpha_j) = 0$, and $f_j''(\alpha_j) = \frac{\partial f_j^2}{\partial \alpha_j^2} = 2n_j > 0$.
5. The function Q is the smallest possible for $\alpha_j = \bar{y}_j$, hence $\hat{\alpha}_1, \ldots, \hat{\alpha}_k$ are the least squares estimates.

3.4 We have

$$
\begin{aligned}
s^2 &= \frac{1}{n-k}\sum_{i=1}^{n} r_i^2 = \frac{1}{n-k}\sum_{i=1}^{n}\left(y_i - \bar{y}_{g(i)}\right)^2 \\
&= \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i:g(i)=j}(y_i - \bar{y}_j)^2 = \frac{1}{n-k}\sum_{j=1}^{k} n_j\sigma_j^2
\end{aligned}
$$

as we should prove.

3.6

1. Two independent samples.
2. Two paired samples.
3. One-way ANOVA with four groups, or a two-way ANOVA since both sex and age are varied.

3.7

1. Use commands like

   ```
   > mean(parasites[stock=="atran"])
   > sd(parasites[stock=="atran"])
   ```

   and similarly for the Conon sample.
2. s^2 is the average of s_1^2 and s_2^2 so $s^2 = 43.397$ and $s = 6.59$.
3. Use the command `summary(lm(parasites~stock-1))`.

4.1

1. 10 and 1 (left), 50 and 15 (right).

4.2

1. $P(3000 < A < 4000) = P(8.006 < \log(A) < 8.294) = 0.084$.
2. 90% central area in distribution of $\log(A)$ is $7.48 \pm 1.645 \cdot 0.44 = (6.7562, 8.2038)$. Taking exponentials we get $(859, 3655)$.