4. The mean and standard deviation will both be 1/10 of the mean and standard deviation measured on the original data in millimeters.

1.8 x3, y1, z2.

1.9

1. Median: 110, Mean: 111.67.
2. Range: 34, Q1: 105, Q3: 119, IQR: 14, standard deviation: 12.24, variance: 149.75.
3. Median: 110, Mean 115. The mean is changed but not the median.
4. Range: (94, 149). Standard deviation: 17.45, IQR 21.
5. First outliers should be checked that they are within a reasonable range of the measured variable. If they are not clearly in error the outliers should generally not be removed. Instead we need a statistical model that can accommodate the values.

1.10

1. To get information about the population that is analyzed.
2. The mean and standard deviation are only reasonable summary statistics if the variables that are presented are symmetric. Also, the general shape of the distribution should have just most of its mass in "the middle".
3. Can use the median and IQR

2.1

1. $\hat{\alpha} = 2.417,\ \hat{\beta} = 1.692$
2. Expecting 26.11 beats/minute on average.

2.3

4. Slope is 5.066, intercept is -36.943.
6. The correlation coefficient between volume and diameter is 0.9671, between volume and height is 0.5982, and between diameter and height is 0.5193.
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 if the constant is positive as it will be present in both the numerator and denominator and will "cancel out". If the constant is negative then the correlation coefficient will change sign but keep the same numeric value.

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 $\hat{c} = 0.4122$.

2.10

1. R^2 can attain values from 0 (no association or correlation) to 1 (perfect relation).
2. The correlation can be either positive or negative since we square the correlation to get the coefficient of determination. Thus, any signs are lost when squaring.
3. Generally a higher coefficient of determination is desirable (if we can see that the relationship is linear) since x better explains the y.
4. If x is constant then R^2 is not defined. When x and y are unrelated then R^2 should be close to zero.

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 $i = 1, \ldots, 9$; $g(i) = \mathrm{P_2O_7}$ for $i = 10, \ldots, 18$; $g(i) = $ HMP for $i = 19, \ldots, 26$;
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. $s = 0.368$.