1. Data on cotton aphid infestation rates, $Y$ (in aphids/100 leaves), average temperature, $x_1$ (in degrees Celsius), and mean relative humidity, $x_2$, are reported in the *Mesopotamia J. Agriculture* (1982). An analysis is based on the model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \quad i = 1, \ldots, 34,$$

where $\epsilon_i \sim N(0, \sigma^2)$. The data are available in the *data library* on R as *aphid*.

(a) Write down the fitted regression model.

(b) Construct a 95% confidence interval for $\beta_1$. What can you conclude from your interval?

(c) Calculate the multiple correlation coefficient.

(d) Test whether $Y$ depends on humidity (i.e. test $H_0 : \beta_2 = 0$). Give the $p$-value and explain how it is calculated.

(e) Fit the regression model using $x_2$ only. Determine a 95% confidence interval for the expected infestation rate if the mean humidity level is 60%.

2. The data below, from a study of computer assisted learning by 10 students, show the total number of correct and incorrect responses in completing a lesson ($X$) and the cost of computer time ($Y$ in cents) :

<table>
<thead>
<tr>
<th>$X$</th>
<th>16</th>
<th>14</th>
<th>22</th>
<th>10</th>
<th>14</th>
<th>17</th>
<th>10</th>
<th>13</th>
<th>19</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>77</td>
<td>70</td>
<td>85</td>
<td>50</td>
<td>62</td>
<td>70</td>
<td>52</td>
<td>63</td>
<td>88</td>
<td>57</td>
</tr>
</tbody>
</table>

(a) Fit a simple linear regression to the data and graph the residuals. Comment on any pattern. Estimate the standard error of the slope parameter.

(b) To account for the changes in variability we can fit the model

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

where $u_i \sim N(0, kx_i^2)$. Alternatively we can fit

$$y_i/x_i = \beta_0 (1/x_i) + \beta_1 + \epsilon_i$$

as this produces a model where the variance of the response variable is not increasing with $x_i$. Fit this regression model. The estimate the standard error of the estimator for $\beta_1$. Note that the estimated standard error of both parameter estimators decrease compared with the model in (i). Estimate the variance constant $k$. 
3. Consider the linear regression model \( Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \), \( i = 1, \ldots, n \) where the \( Y_i \) are independent. Write the least squares estimator for the intercept in the form \( \sum_{i=1}^{n} a_i Y_i \). Hence show that the least squares estimator is unbiased and prove that
\[
\text{Var}(\hat{\beta}_0) = \sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n S_{XX}}.
\]

4. (STAT 3912 Only)

Consider the linear model \( Y = X\beta + \epsilon \), where \( X \) is an \( n \times p \) matrix of known values of rank \( p \) and \( Y \) is a vector of independent random variables each with variance \( \sigma^2 \). Let \( \hat{\beta} \) be the least squares estimator for \( \beta \) and let \( R = Y - X\hat{\beta} \).

(a) Show that \( R = (I - H)\epsilon \), where \( H = X(X^TX)^{-1}X^T \).

(b) Show that the vector of fitted values can be written as
\[
\hat{Y} = X\beta + H\epsilon.
\]

(c) Prove that \( \text{Cov}(Y, \hat{Y}) = \text{Cov}(\hat{Y}, \hat{Y}) \).

(d) Prove that \( \sum_{i=1}^{n} \text{Var}(\hat{Y}_i) = p\sigma^2 \).