Exercise 1: The Mandel data set in the alr4 package contains eight artificial observations in a response y and two predictors x_1 and x_2. Use the data set to do the following:

a. Write the multiple linear regression model $y = X\beta + e$ in terms of the actual data. (In other words, write the model equation but substitute the response vector in place of y, the parameter vector in place of β, and the design matrix in place of X).

Solution:
From the following code we can put together the model equation from the Mandel data,
Code:

```
> df <- Mandel
> y = df$y
[1] 41.38 31.01 37.41 50.05 39.17 38.86 46.14 44.47

> X = cbind(rep(1, length(y)), df$x1, df$x2)
     [,1]  [,2]  [,3]
[1,]    1 16.85  1.46
[2,]    1 24.81 -4.61
[3,]    1 18.85 -0.21
[4,]    1 12.63  4.93
[5,]    1 21.38 -1.36
[6,]    1 18.78 -0.08
[7,]    1 15.58  2.98
[8,]    1 16.30  1.73
```

$$\begin{bmatrix} 41.38 \\ 31.01 \\ 37.41 \\ 50.05 \\ 39.17 \\ 38.86 \\ 46.14 \\ 44.47 \end{bmatrix} = \begin{bmatrix} 1 & 16.85 & 1.46 \\ 1 & 24.81 & -4.61 \\ 1 & 18.85 & -0.21 \\ 1 & 12.63 & 4.93 \\ 1 & 21.38 & -1.36 \\ 1 & 18.78 & -0.08 \\ 1 & 15.58 & 2.98 \\ 1 & 16.30 & 1.73 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{bmatrix}.$$

b. Fit the model by calculating the OLS estimators using $\hat{\beta} = (X^T X)^{-1} X^T y$.

Solution:
Using the following code we can solve for $\hat{\beta}$,
Code:

```
> y = df$y
> X = cbind(rep(1, length(y)), df$x1, df$x2)

> OLS_estimators = solve(t(X) %*% X, t(X) %*% y)

          [,1]
[1,] 0.7576208
[2,] 2.0664489
[3,] 4.6326417
```

c. Calculate the estimate of σ^2 using $RSS = (y - X\hat{\beta})^T(y - X\hat{\beta})$.

Solution:
First we need to compute the RSS then we get our estimator for σ^2 with the following formula,

$$\hat{\sigma}^2 = \frac{RSS}{n - (p + 1)}$$

Code:

```
> RSS = t(y - X %*% OLS_estimators)%*%(y - X %*% OLS_estimators)

    7.76326

SigmaSquared = RSS/(length(y) - (2 + 1))

    1.552652
```

d. Calculate the estimated variance-covariance matrix of the OLS estimators.

Solution:
We can view the estimated variance-covariance matrix with the following code. We could also calculate it, using the following matrix equation,

$$V(\hat{\beta}) = \hat{\sigma}^2(X^T X)^{-1}$$

Code:

```
> vcov(model <- lm(formula = y ~ ., data = df ))

              (Intercept)        x1         x2
(Intercept)     496.94874 -26.24768 -33.760218
x1              -26.24768   1.38691   1.783020
x2              -33.76022   1.78302   2.318787

> SigmaSquared*solve((t(X) %*% X))

           [,1]      [,2]       [,3]
[1,]  496.94876 -26.24768 -33.760219
[2,]  -26.24768   1.38691   1.783020
[3,]  -33.76022   1.78302   2.318787
```

Exercise 2: The wm2 data in the Alr4 package contains windspeed data for a location in South Dakota where developers were considering siting a wind turbine. Variables measured were: date, CSpd, RSpd, RDir, and Bin.

a. Fit the MLR model that uses CSps as response and RSpd and RDir as predictor. Report the estimated model.

Solution:
We can fit the model with the following code,
Code:

```
> df <- wm2

> MLR_Windspeed = lm(formula = CSpd ~ RSpd + RDir, data = df )
> summary(MLR_Windspeed)

Call:
lm(formula = CSpd ~ RSpd + RDir, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-7.7453 -1.6047 -0.1787  1.4355  9.2888

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.3985260  0.2111335   16.10   <2e-16 ***
RSpd         0.7645736  0.0200746   38.09   <2e-16 ***
RDir        -0.0015193  0.0007446   -2.04   0.0416 *
---


Residual standard error: 2.463 on 1113 degrees of freedom
Multiple R-squared:  0.5725,    Adjusted R-squared:  0.5718
F-statistic: 745.3 on 2 and 1113 DF,  p-value: < 2.2e-16
```

The final fitted model comes out to,

$$\hat{y}_{CSpd} = 0.7645736x_{RSpd} - 0.0015193x_{RDir} + 3.3985260.$$

b. Provide an interpretation of the regression coefficient you reported in the estimates mode, including the intercept.

Solution:
For $\hat{\beta}_2$(Ddir) when wind direction at the reference site increases by 1 degree, and windspeed at a reference site(RSpd) is held constant, we can expect windspeed at the South Dakota site(CSpd) to decrease by 0.0015193 knots.

For $\hat{\beta}_1$(RSpd), when wind speed at the reference site increases by 1 knot, and wind direction at the reference site(RDir) is held constant, we can expect the windspeed at the South Dakota site(CSpd) to increase by 0.7645736 knots.

$\hat{\beta}_0$(intercept) is the estimated mean windspeed at the South Dakota site when all other model parameters are set to zero.

c. Report the fitted model's estimate for σ^2 and use it to find the value of RSS

Solution:
Note that the lm() command reports the RSE value which is our estimator for σ. Squaring the reported value, then multiplying by the degrees of freedom we get our RSS value,
Code:

```
> SigmaSquared = 2.463^2
> RSS = SigmaSquared*1113
[1] 6751.869

> anova(MLR_Windspeed)

Analysis of Variance Table
Response: CSpd
            Df Sum Sq Mean Sq   F value   Pr(>F)
RSpd         1 9015.6  9015.6 1486.4820 < 2e-16 ***
RDir         1   25.2    25.2    4.1629 0.04156 *
Residuals 1113 6750.4     6.1
```

Double checking with the anova table and on further analysis the difference can be attributed to rounding error.

d. Obtain the effects plot for each predictor.

Solution:
Code:

```
> plot(Effect("RSpd", MLR_Windspeed),
       xlab = 'Reference Site Windspeed',
       ylab = 'South Dakota Site Windspeed',
       main = 'RSpd Effect')

> plot(Effect("RDir", MLR_Windspeed),
       xlab = 'Reference Site Wind Direction',
       ylab = 'South Dakota Site Windspeed',
       main = 'RDir Effect')
```