# Problem 1

(i)

$$X^{'}X = \begin{bmatrix} 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.50 & 2.00 \end{bmatrix} \begin{bmatrix} 1.00 & 1.00 \\ 1.00 & 1.00 \\ 1.00 & 0.50 \\ 1.00 & 2.00 \end{bmatrix} = \begin{bmatrix} 4.00 & 4.50 \\ 4.50 & 6.25 \end{bmatrix}$$

$$(X^{'}X)^{-1} = \begin{bmatrix} 4.00 & 4.50 \\ 4.50 & 6.25 \end{bmatrix}^{-1} = \frac{1}{4.75}\begin{bmatrix} 6.25 & -4.50 \\ -4.50 & 4 \end{bmatrix} = \begin{bmatrix} 1.315789 & -0.947368 \\ -0.947368 & 0.842105 \end{bmatrix}$$

(ii)

$$X^{'}Y = \begin{bmatrix} 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.50 & 2.00 \end{bmatrix} \begin{bmatrix} 0.50 \\ -0.50 \\ 0.30 \\ 1.20 \end{bmatrix} = \begin{bmatrix} 1.50 \\ 2.55 \end{bmatrix}$$

(iii)

$$\hat{\beta} = (X^{'}X)^{-1}X^{'}Y = \frac{1}{4.75}\begin{bmatrix} 6.25 & -4.50 \\ -4.50 & 4 \end{bmatrix} \begin{bmatrix} 1.50 \\ 2.55 \end{bmatrix} = \frac{1}{4.75}\begin{bmatrix} -2.1 \\ 3.45 \end{bmatrix} = \begin{bmatrix} -.4421 \\ .7263 \end{bmatrix}$$

(iv)

$$\hat{y} = X\hat{\beta} = \begin{bmatrix} 1.00 & 1.00 \\ 1.00 & 1.00 \\ 1.00 & 0.50 \\ 1.00 & 2.00 \end{bmatrix} \begin{bmatrix} -.4421 \\ .7263 \end{bmatrix} = \begin{bmatrix} 0.28421 \\ 0.28421 \\ -0.07895 \\ 1.01053 \end{bmatrix}$$

(v)

$$\hat{\epsilon} = y - \hat{y} = \begin{bmatrix} 0.50 \\ -0.50 \\ 0.30 \\ 1.20 \end{bmatrix} - \begin{bmatrix} 0.28421 \\ 0.28421 \\ -0.07895 \\ 1.01053 \end{bmatrix} = \begin{bmatrix} 0.21579 \\ -0.78421 \\ 0.37895 \\ 0.18947 \end{bmatrix}$$

# Problem 2

(a)

$$H_0 : \theta = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_2 - \beta_3 \\ \beta_3 - \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \theta_0$$

$$\theta_{3x1} = C_{3x5}\ \beta_{5x1}$$

$$\begin{bmatrix} \beta_1 - \beta_2 \\ \beta_2 - \beta_3 \\ \beta_3 - \beta_4 \end{bmatrix} = C_{3x5} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_2 - \beta_3 \\ \beta_3 - \beta_4 \end{bmatrix}$$

(b)

$$H_0 : \theta = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_3 - \beta_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \theta_0$$

$$\theta_{2x1} = C_{2x5}\ \beta_{5x1}$$

$$\begin{bmatrix} \beta_1 - \beta_2 \\ \beta_3 - \beta_4 \end{bmatrix} = C_{2x5} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_3 - \beta_4 \end{bmatrix}$$

(c)

$$H_0 : \theta = \begin{bmatrix} \beta_1 - 2\beta_2 - 4\beta_3 \\ \beta_1 + 2\beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -6 \end{bmatrix} = \theta_0$$

$$\theta_{2x1} = C_{2x5}\ \beta_{5x1}$$

$$\begin{bmatrix} \beta_1 - 2\beta_2 - 4\beta_3 \\ \beta_1 + 2\beta_2 \end{bmatrix} = C_{2x5} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 1 & -2 & -4 & 0 \\ 0 & 1 & 2 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & -2 & -4 & 0 \\ 0 & 1 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \beta_1 - 2\beta_2 - 4\beta_3 \\ \beta_1 + 2\beta_2 \end{bmatrix}$$

# Problem 3

(a)

Looking at the reduced echelon form of X we have:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The rank of X is 2 since we only have two non-zero rows.

Therefore X is not full rank and thus

$\beta$ is not estimable, but as many as 2 elements may be.

$$\theta_1 = \beta_2 = C_{1x3} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$

$$\theta_i = C\beta = \beta_2$$

$$\theta_1 \text{ is estimable } \iff C\beta = t'_{1x5}\ E(y_{5x1})$$

$$E(y_{5x1}) = X\beta = \begin{bmatrix} \beta_0 + \beta_1 \\ \beta_0 + \beta_1 \\ \beta_0 - \beta_2 \\ \beta_0 - \beta_2 \\ \beta_0 + \beta_1 \end{bmatrix}$$

$$t'_{1x5}\ E(y_{5x1}) = \begin{bmatrix} t_1 & t_2 & t_3 & t_4 & t_5 \end{bmatrix} \begin{bmatrix} \beta_0 + \beta_1 \\ \beta_0 + \beta_1 \\ \beta_0 - \beta_2 \\ \beta_0 - \beta_2 \\ \beta_0 + \beta_1 \end{bmatrix}$$

$$= [(t_1 + t_2 + t_5)(\beta_0 + \beta_1) + (t_3 + t_4)(\beta_0 - \beta_2)]$$

Since we cannot obtain $\beta_2$ from the result, it is not estimable

(b)

$$\theta_2 = \begin{bmatrix} \beta_0 + \beta_1 \\ \beta_0 - \beta_2 \end{bmatrix}$$

$$t^{'}_{2x5}\ E(y_{5x1}) = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} & t_{15} \\ t_{21} & t_{22} & t_{23} & t_{24} & t_{25} \end{bmatrix} \begin{bmatrix} \beta_0 + \beta_1 \\ \beta_0 + \beta_1 \\ \beta_0 - \beta_2 \\ \beta_0 - \beta_2 \\ \beta_0 + \beta_1 \end{bmatrix} =$$

$$\begin{bmatrix} (t_{11} + t_{12} + t_{15})(\beta_0 + \beta_1) + (t_{13} + t_{14})(\beta_0 - \beta_2) \\ (t_{21} + t_{22} + t_{25})(\beta_0 + \beta_1) + (t_{23} + t_{24})(\beta_0 - \beta_2 \end{bmatrix}$$

$$t^{'}_{2x5} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \text{ is a solution}$$

Thus $\theta_2$ is estimable

$$C = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

$\theta_2$ testable since C is full rank and $\theta_2$ is estimable

# Problem 4

(a)

| Source | DF | Sum of Squares | Mean Square | F Value | $Pr > F$ |
|---|---|---|---|---|---|
| Model | 1 | 2624.670184 | 2624.670184 | 137.906162 | $\approx 0$ |
| Error | 96 | 1827.099916 | 19.0322908 | | |
| Corrected Total | 97 | 4451.7701 | | | |

$$MST = \frac{SSH}{DFH} \quad MSE = \frac{SSE}{DFE} \quad F = \frac{MSH}{MSE}$$

$$Pr > F = 1 - pf(137.906162, 1, 96) = 0 \quad \text{(using R)}$$

(b) The model assumptions are:
Homogeneity of variance- every element of $\epsilon$ (error terms) has the same variance
Independence- each element of $\epsilon$ is independent of all others
Linearity- expected values of WGHT are linear function of the parameters. $E(y) = X\beta$
Existence - $\epsilon_i$ has finite first and second moments.
Gaussian errors- error terms are normally distributed. $\epsilon_i \sim N(0, \sigma_i^2)$

(c) $H_0 : \beta_1 = 0$
Since the p-value is approximately 0, reject the null hypothesis and conclude that average daily exercise time is a significant predictor of weight loss.

# Problem 5

(a) $y = \beta_0 + \beta_1 X$
y=days x=index

```
mod=lm(days~index,data=hw2)
summary(mod)
```

```
##
## Call:
## lm(formula = days ~ index, data = hw2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -41.70 -21.54   2.12  18.56  36.42
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -192.984    163.503  -1.180    0.258
## index         15.296      9.421   1.624    0.127
##
## Residual standard error: 23.79 on 14 degrees of freedom
## Multiple R-squared:  0.1585, Adjusted R-squared:  0.09835
## F-statistic: 2.636 on 1 and 14 DF,  p-value: 0.1267
```

$\beta_0$: When index = 0 the model predicts -192.984 high ozone days with a standard error of 163.503.
$\beta_1$: with each increase of 1 in index, the model predicts an increase in 15.296 high ozone days with a standard error of 9.421.

(b) Since X is full rank, all of the $\beta$s are estimable

(c) $H_0 : \beta_1 = 0$
$\alpha = .05$

```r
anova(mod)
```

```
## Analysis of Variance Table
##
## Response: days
##           Df Sum Sq Mean Sq F value Pr(>F)
## index      1 1492.6  1492.6  2.6362 0.1267
## Residuals 14 7926.8   566.2
```

p-value=.1267
Conclusion: Fail to reject $H_0$ since p-value $> \alpha$
Thus we cannot conclude that the number high ozone days is associated with the meteorological index.

(d)

$$\alpha = .05$$

$$\text{t-statistic} = \frac{\hat{\beta}_1 - 12}{SE_{\hat{\beta}_1}} = (15.296 - 12)/9.421 = .35$$

$$\text{p-value} = (1 - pt(.35, 14)) * 2 = .7315$$

Conclusion: fail to reject $H_0$ since p-value $> \alpha$

We cannot conclude that a 1 degree increase in avg temp is associated with a 12 day increase in number of high ozone days

(e) 95% CI for expected number of high ozone days when index=16
$(21.76, 81, 76)$
95% Prediction Interval for expected number of high ozone days when index=16
$(-7.44, 110.96)$

```r
attach(hw2)
new = data.frame(index=16)
predict(mod, new, interval = "prediction")
```

```
##        fit      lwr      upr
## 1 51.75801 -7.44155 110.9576
```

```r
predict(mod, new, interval = "confidence")
```

```
##        fit      lwr      upr
## 1 51.75801 21.75791 81.75811
```