1. [10 marks] 2 marks each for parts (a) and (d) and 3 marks each for parts (b) and (c).

   (a) The data from the file was put into a dataframe in *R* called `grade.df`. The following commands create the response vector **y**, the model matrix **X** and the projection matrix **H**.

```
y<-grade.df$GRADE
X<-cbind(1,grade.df$BOOKS,grade.df$ATTEND,grade.df$LATE)
H<-X%*%solve(t(X)%*%X)%*%t(X)
```

   (b) To find the fitted values:

```
> fvals<-H%*%y
> round(t(fvals),2)
      [,1]  [,2]  [,3]  [,4] [,5] [,6]  [,7]  [,8]  [,9] [,10] [,11] [,12]
[1,] 51.86 56.13 48.08 68.09 72.9 82.2 51.06 84.45 71.91 53.85 53.46 56.96
     [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
[1,] 71.81  52.6 48.91 48.11 63.63 60.01 67.93 69.82 62.38 58.92 67.38 71.91
     [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36]
[1,] 64.15 54.23 71.91 67.99 59.94 61.48 68.83  75.4 55.58 75.44 79.96 84.45
     [,37] [,38] [,39] [,40]
[1,] 56.44 51.19 68.09 52.57
```

   To find the residuals:

```
> res<-(diag(40)-H)%*%y
> round(t(res),2)
      [,1] [,2]  [,3]   [,4] [,5] [,6]  [,7] [,8]  [,9] [,10] [,11] [,12]
[1,] -6.86 0.87 -3.08 -17.09 -7.9  5.8 -7.06 2.55 17.09  5.15 12.54  8.04
      [,13] [,14] [,15] [,16] [,17]  [,18]  [,19]  [,20]  [,21] [,22] [,23]
[1,] -15.81  -5.6 17.09 -7.11 -7.63 -23.01 -22.93 -11.82 -15.38  5.08 29.62
      [,24]  [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32]  [,33] [,34] [,35]
[1,] -16.91 -13.15  6.77 -2.91 11.01 11.06  0.52 18.17 -21.4 -12.58 16.56  3.04
     [,36] [,37] [,38] [,39] [,40]
[1,]  9.55  3.56  4.81 19.91  9.43
```

   To estimate σ^2 calculate $\mathbf{r}^t\mathbf{r}/(40-4)$:

```
> sig2hat<-(t(res)%*%res)/(40-4)
> sig2hat
         [,1]
[1,] 184.9432
```

   (c) To find $\hat{\boldsymbol{\beta}}$ and its estimated covariance matrix:

```
> betahats<-solve(t(X)%*%X)%*%t(X)%*%y
> round(betahats,3)
       [,1]
[1,] 61.226
[2,]  2.276
[3,]  0.706
[4,] -2.245
>
> covmat<-solve(t(X)%*%X)*as.numeric(sig2hat)
> round(covmat,3)
```

```
         [,1]    [,2]   [,3]    [,4]
[1,] 218.404 -11.706 -7.624 -15.275
[2,] -11.706   3.763 -0.137   1.128
[3,]  -7.624  -0.137  0.418   0.370
[4,] -15.275   1.128  0.370   1.438
```

(d) Using `lm` in R we get:

```
> grade.lm<-lm(GRADE~.,data=grade.df)
> round(fitted.values(grade.lm),2)
    1     2     3     4     5     6     7     8     9    10    11    12    13
51.86 56.13 48.08 68.09 72.90 82.20 51.06 84.45 71.91 53.85 53.46 56.96 71.81
   14    15    16    17    18    19    20    21    22    23    24    25    26
52.60 48.91 48.11 63.63 60.01 67.93 69.82 62.38 58.92 67.38 71.91 64.15 54.23
   27    28    29    30    31    32    33    34    35    36    37    38    39
71.91 67.99 59.94 61.48 68.83 75.40 55.58 75.44 79.96 84.45 56.44 51.19 68.09
   40
52.57
> round(residuals(grade.lm),2)
     1      2      3      4      5      6      7      8      9     10     11
 -6.86   0.87  -3.08 -17.09  -7.90   5.80  -7.06   2.55  17.09   5.15  12.54
    12     13     14     15     16     17     18     19     20     21     22
  8.04 -15.81  -5.60  17.09  -7.11  -7.63 -23.01 -22.93 -11.82 -15.38   5.08
    23     24     25     26     27     28     29     30     31     32     33
 29.62 -16.91 -13.15   6.77  -2.91  11.01  11.06   0.52  18.17 -21.40 -12.58
    34     35     36     37     38     39     40
 16.56   3.04   9.55   3.56   4.81  19.91   9.43
> round(coefficients(grade.lm),3)
(Intercept)       BOOKS      ATTEND        LATE
     61.226       2.276       0.706      -2.245
```

Which are the same values as those found in (b) and (c).

2. [6 marks] The mean of the fitted values and the mean of the observed values are both equal to 63.55.

```
> mean(fvals)
[1] 63.55
> mean(y)
[1] 63.55
```

To show these means must be equal for all regression models (actually those with an intercept) it is sufficient to show that the sum of the fitted values equals the sum of the response which can be written as $\mathbf{1}^t\hat{\boldsymbol{\mu}}_{\mathbf{Y}} = \mathbf{1}^t\mathbf{y}$ where $\mathbf{1}$ is a vector of ones. Which can be done as follows:

$$
\begin{aligned}
\text{As } \mathbf{1} \in \text{colsp}(\mathbf{X})\text{:} \quad & \mathbf{1} = \mathbf{H}\mathbf{1} \\
\text{Taking the transpose gives:} \quad & \mathbf{1}^t = \mathbf{1}^t\mathbf{H}^t \\
\text{Since } \mathbf{H} \text{ is symmetric:} \quad & \mathbf{1}^t = \mathbf{1}^t\mathbf{H} \\
\text{Post multiply by } \mathbf{y}\text{:} \quad & \mathbf{1}^t\mathbf{y} = \mathbf{1}^t\mathbf{H}\mathbf{y} \\
\text{Which gives the result:} \quad & \mathbf{1}^t\mathbf{y} = \mathbf{1}^t\hat{\boldsymbol{\mu}}_{\mathbf{Y}}
\end{aligned}
$$

3. [24 marks] 2 marks for (a), 6 marks for (b), 2 marks for (c) and 14 marks for (d).

(a) To find the set of fitted values for the new model:

```
> W<-cbind(1,grade.df$BOOKS,grade.df$ATTEND-grade.df$LATE,grade.df$LATE)
> HW<-W%*%solve(t(W)%*%W)%*%t(W)
> fvalsW<-HW%*%y
> round(t(fvalsW),2)
      [,1]  [,2]  [,3]  [,4] [,5] [,6]  [,7]  [,8]  [,9] [,10] [,11] [,12]
[1,] 51.86 56.13 48.08 68.09 72.9 82.2 51.06 84.45 71.91 53.85 53.46 56.96
     [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
[1,] 71.81  52.6 48.91 48.11 63.63 60.01 67.93 69.82 62.38 58.92 67.38 71.91
     [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36]
[1,] 64.15 54.23 71.91 67.99 59.94 61.48 68.83  75.4 55.58 75.44 79.96 84.45
     [,37] [,38] [,39] [,40]
[1,] 56.44 51.19 68.09 52.57
```

These are the same as those in 1(b).

(b) The vector of estimated coefficients for the new model is given by $\hat{\boldsymbol{\beta}} = \left(\mathbf{W}^t\mathbf{W}\right)^{-1}\mathbf{W}^t\mathbf{y}$:

```
> betahatsW<-solve(t(W)%*%W)%*%t(W)%*%y
> round(betahatsW,3)
       [,1]
[1,] 61.226
[2,]  2.276
[3,]  0.706
[4,] -1.539
```

For this new model the coefficient for `LATE` has changed from -2.245 for the old model to -1.539. Where as the new variable `ONTIME` has the same coefficient (0.706) as `ATTEND` had in the old model.

To see why this occurs simply use the fact that `ONTIME` = `ATTEND` − `LATE` to show that the two fitted models are equivalent.

Original model: $\hat{Y} = 61.226 + 2.276 \times \texttt{BOOKS} + 0.706 \times \texttt{ATTEND} - 2.245 \times \texttt{LATE}$

New model:

$$\hat{Y} = 61.226 + 2.276 \times \texttt{BOOKS} + 0.706 \times \texttt{ONTIME} - 1.539 \times \texttt{LATE}$$
$$\hat{Y} = 61.226 + 2.276 \times \texttt{BOOKS} + 0.706 \times (\texttt{ATTEND} - \texttt{LATE}) - 1.539 \times \texttt{LATE}$$
$$\hat{Y} = 61.226 + 2.276 \times \texttt{BOOKS} + 0.706 \times \texttt{ATTEND} + (-0.706 - 1.539) \times \texttt{LATE}$$
$$\hat{Y} = 61.226 + 2.276 \times \texttt{BOOKS} + 0.706 \times \texttt{ATTEND} - 2.245 \times \texttt{LATE}$$

(c) The matrix $\mathbf{A}$ is given by:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(d) i. To prove $\mathbf{H_W} = \mathbf{H_X}$ substitute $\mathbf{W} = \mathbf{XA}$ into the expression for $\mathbf{H_W}$ and use matrix algebra to show it is equivalent to $\mathbf{H_X}$.

$$
\begin{aligned}
\mathbf{H_W} &= \mathbf{W}\left(\mathbf{W}^t\mathbf{W}\right)^{-1}\mathbf{W}^t \\
&= \mathbf{XA}\left((\mathbf{XA})^t\,\mathbf{XA}\right)^{-1}(\mathbf{XA})^t \\
&= \mathbf{XA}\left(\mathbf{A}^t\mathbf{X}^t\mathbf{XA}\right)^{-1}\mathbf{A}^t\mathbf{X}^t \\
&= \mathbf{XA}\left(\mathbf{A}^t\left(\mathbf{X}^t\mathbf{X}\right)\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{X}^t \\
&= \mathbf{XA}\left(\mathbf{A}^{-1}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\left(\mathbf{A}^t\right)^{-1}\right)\mathbf{A}^t\mathbf{X}^t \\
&= \mathbf{X}\,\mathbf{AA}^{-1}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\left(\mathbf{A}^t\right)^{-1}\mathbf{A}^t\,\mathbf{X}^t \\
&= \mathbf{X}\left(\mathbf{AA}^{-1}\right)\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\left(\left(\mathbf{A}^t\right)^{-1}\mathbf{A}^t\right)\mathbf{X}^t \\
&= \mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t \\
&= \mathbf{H_X}
\end{aligned}
$$

ii. The expression is found by substituting $\mathbf{W} = \mathbf{XA}$ into the expression for $\hat{\boldsymbol{\beta}}_{\mathbf{W}}$:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathbf{W}} &= \left(\mathbf{W}^t\mathbf{W}\right)^{-1}\mathbf{W}^t\mathbf{y} \\
&= \left((\mathbf{XA})^t\,\mathbf{XA}\right)^{-1}(\mathbf{XA})^t\,\mathbf{y} \\
&= \left(\mathbf{A}^t\mathbf{X}^t\mathbf{XA}\right)^{-1}\mathbf{A}^t\mathbf{X}^t\mathbf{y} \\
&= \left(\mathbf{A}^{-1}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\left(\mathbf{A}^t\right)^{-1}\right)\mathbf{A}^t\mathbf{X}^t\mathbf{y} \\
&= \mathbf{A}^{-1}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\left(\left(\mathbf{A}^t\right)^{-1}\mathbf{A}^t\right)\mathbf{X}^t\mathbf{y} \\
&= \mathbf{A}^{-1}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{y} \\
&= \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}_{\mathbf{X}}
\end{aligned}
$$