The data used in this assignment comes from *Building Models with Regression and Correlation* by Jeremy Miles. The response is the grade (`GRADE`) received by 40 students in a particular course. There are three explanatory variables: `BOOKS` represents the number of statistics books read by the student, `ATTEND` represents the number of lectures the student attended and `LATE` represents the number of times they were late for a lecture (given they attended). The data is in a file called **Grade Data** in "Data Sets" on the STATS 330 webpage.

1. [10 marks] Use R to do the following:
   (a) Create the response vector $\mathbf{y}$, the model matrix $\mathbf{X}$ (include the intercept column and all three regressors) and the projection matrix $\mathbf{H}$. Note: you don't need to print out these matrices but you do need to give the R commands you used to create them.
   (b) Find the fitted values and the residuals by projecting $\mathbf{y}$ onto the "model space" and onto the "error space". Use the vector of residuals to estimate the error variance σ^2.
   (c) Find (and print out) the vector of estimated coefficients $\hat{\boldsymbol{\beta}}$ and the estimated covariance matrix for $\hat{\boldsymbol{\beta}}$.
   (d) Use the `lm` command to fit this regression model and confirm that you get the same fitted values, residuals and estimated coefficients as above.
2. [6 marks] Find the mean of the fitted values and the mean of the observed values for the response for the fitted model. Prove that for any regression model the mean of $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$ is equal to the mean of $\mathbf{y}$.
3. [24 marks] Create a new explanatory variable `ONTIME` where `ONTIME` = `ATTEND` − `LATE`. Note that `ONTIME` represents the number of lectures that a student arrives for "on time." Create a new model matrix, call this one $\mathbf{W}$, that includes columns for `BOOKS`, `ONTIME` and `LATE` in addition to the intercept column.
   (a) Find the projection matrix $\mathbf{H}_{\mathbf{W}} = \mathbf{W}\left(\mathbf{W}^t\mathbf{W}\right)^{-1}\mathbf{W}^t$ and show that it produces the same set of fitted values that you found in 1(b).
   (b) Find the vector of estimated coefficients $\hat{\boldsymbol{\beta}}$ for this new model. Compare this to estimated coefficients you found in 1(c). Note that the estimated coefficient for `ONTIME` in the new model is the same as that for `ATTEND` in the old model but the coefficient for `LATE` has changed. Explain why this has occurred (i.e. show that the two models are equivalent).
   (c) Find a 4×4 matrix $\mathbf{A}$ such that $\mathbf{W} = \mathbf{X}\mathbf{A}$.
   (d) In general, consider $\mathbf{W} = \mathbf{X}\mathbf{A}$ for a $p \times (k+1)$ model matrix $\mathbf{X}$ and a non-singular $(k+1) \times (k+1)$ matrix $\mathbf{A}$.
      i. Prove that $\mathbf{H}_{\mathbf{W}} = \mathbf{W}\left(\mathbf{W}^t\mathbf{W}\right)^{-1}\mathbf{W}^t$ is the same as $\mathbf{H}_{\mathbf{X}} = \mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t$.
      ii. Find an expression that relates the vector of estimated coefficients for $\mathbf{W}$ (call this $\hat{\boldsymbol{\beta}}_{\mathbf{W}}$), to that for $\mathbf{X}$ (call this $\hat{\boldsymbol{\beta}}_{\mathbf{X}}$).