Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ (i = 1, 2, ..., n)$. The model may be written in matrix notation as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

(a) Explain the terms $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$

(b) State the second-order distributional assumptions in matrix notation and then the normal theory assumptions using matrix notations

(c) Write the elements of $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$

(d) Show that the error sum of squares may be written as $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$

(e) The least squares estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ minimize $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$. State the normal equations (least squares equations) in matrix form.

(f)Assuming that $\mathbf{X}$ (and hence $\mathbf{X}'\mathbf{X}$) is of full rank, obtain $\hat{\boldsymbol{\beta}}$ using (e).

(g) Show that $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$.

(h) Derive $\mathbf{Cov}(\hat{\boldsymbol{\beta}})$, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$.

(i) What is the Maximum Likelihood Estimate of $\boldsymbol{\beta}$ when normality assumptions are made?

(j) Show that the vector of fitted values $\hat{\mathbf{Y}}$ may be written as

$\hat{\mathbf{Y}} = \mathbf{HY}$ where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

(k) Show that the matrix $\mathbf{H}$ defined above is a symmetric idempotent matrix.

(l) Show that $tr(\mathbf{H}) = 2$.

(m) The residual vector is $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$. Show that $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$.

(n) Show that the n x n matrix $(\mathbf{I} - \mathbf{H})$ is symmetric and idempotent with $tr(\mathbf{I} - \mathbf{H}) = n - 2$.

(o) Show that the *Residual Sum of Squares* SSE may be written as $SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$.

(p) The *Total (corrected) Sum of Squares* is defined as $SSTO = \sum_{i=1}^{n} Y_i^2 - [(\sum_{i=1}^{n} Y_i)^2 / n]$.

Show that $SSTO = \mathbf{Y}'[\mathbf{I} - (\frac{1}{n})\mathbf{J}]\mathbf{Y}$ where $\mathbf{J}$ is the matrix with all elements equal to 1.

(q) Show that the *Regression Sum of Squares* may be written as $SSR = \mathbf{Y}'[\mathbf{H} - (\frac{1}{n})\mathbf{J}]\mathbf{Y}$