1. Show that the sample mean of $\mathbf{y}$ is equal to the sample mean of $\hat{\mathbf{y}}$.

2. Let
$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$
and consider estimating $\boldsymbol{\beta}$ by minimizing the following criterion:
$$PLS(\boldsymbol{\beta}; \mathbf{y}) = \left[\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2\right] + \sum_{k=1}^{p}\lambda_k\beta_k^2$$
where $\lambda_1, \ldots, \lambda_p$ are known nonnegative real numbers.
   (a) Find the $\boldsymbol{\beta}$ that minimizes $PLS(\boldsymbol{\beta}; \mathbf{y})$. Hint: write $\sum_{k=1}^{p}\lambda_k\beta_k^2$ as a quadratic form.
   (b) What is the distribution of your estimate of $\boldsymbol{\beta}$ from (a)?
   (c) Under which conditions on $\{\lambda_k\}$ will your estimate from (a) be unbiased?

3. In the next two problems, you will derive the ML and REML estimates of σ^2. Under the linear model
$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$
the likelihood function is
$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) \propto |\sigma^2\mathbf{I}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$
where $|\cdot|$ is the determinant of $\cdot$.
Prove that the ML estimate of σ^2 is
$$\hat{\sigma^2}_{ML} = \frac{1}{n}\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}$$
by maximizing the likelihood function with respect to σ^2. You may use the fact that the MLE of $\boldsymbol{\beta}$ was obtained INDEPENDENTLY of σ^2, and thus you can freely substitute $\hat{\boldsymbol{\beta}}_{ML}$ for $\boldsymbol{\beta}$ in the likelihood when seeking the maximum.

4. It is common to estimate variance parameters like σ^2 using Restricted Maximum Likelihood, which involves maximizing a likelihood taken orthogonal to the fixed effects. In this case, that is done by considering the likelihood of the residuals
$$\hat{\boldsymbol{\epsilon}} = (\mathbf{I} - \mathbf{H})\mathbf{y} \sim N(\mathbf{0}, \sigma^2(\mathbf{I} - \mathbf{H}))$$
which is a *Singular Normal Distribution* because the covariance matrix $(\mathbf{I} - \mathbf{H})$ only has rank $n - p$ and is not invertible. The likelihood function for this Singular Normal Distribution is given by:
$$L(\sigma^2; \hat{\boldsymbol{\epsilon}}) \propto \left(\prod_{k=1}^{n-p}\frac{1}{\sqrt{\sigma^2\lambda_k}}\right)\exp\left\{-\frac{1}{2}\hat{\boldsymbol{\epsilon}}'\left[\sigma^2(\mathbf{I} - \mathbf{H})\right]^{-1}\hat{\boldsymbol{\epsilon}}\right\} \qquad (1)$$
where the determinant in the regular Multivariate Normal density is replaced by the product of the $n - p$ nonzero eigenvalues of $\sigma^2(\mathbf{I} - \mathbf{H})$, and the inverse covariance matrix in the exponent is now $\left[\sigma^2(\mathbf{I} - \mathbf{H})\right]^{-1}$, which is a generalized inverse, defined here as follows:
The product:
$$\mathbf{w} = (\mathbf{I} - \mathbf{H})^{-1}\hat{\boldsymbol{\epsilon}}$$
is any vector $\mathbf{w}$ that satisfies:
$$(\mathbf{I} - \mathbf{H})\mathbf{w} = \hat{\boldsymbol{\epsilon}}. \qquad (2)$$
   (a) Give two vectors: $\mathbf{w}_1$ and $\mathbf{w}_2$ that satisfy equation (2).
   (b) Prove that the REML estimate of σ^2 is
$$\sigma^2_{REML} = \frac{1}{n-p}\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}$$
   by maximizing the restricted likelihood (1).

5. Using the "cars" dataset, fit the following linear model

$$dist_i = \beta_0 + \beta_1 speed_i + \epsilon_i$$

in R using "lm". You do not need to consider any modeling assumptions, just fit the simple linear model.

Now use matrix operations in R to compute the following quantities. Turn in your R-code commented so that it is clear which sections of code correspond to the calculation of which quantities.

(a) $\hat{\boldsymbol{\beta}}$

(b) The residuals $\hat{\boldsymbol{\epsilon}}$ (Plot your residuals vs the residuals from lm).

(c) $\hat{\sigma}^2$

(d) $\hat{\mathbf{y}}$ (Plot yours vs the "fitted values" from lm)

(e) se_k, $k = 1, 2, \ldots, p$

(f) p-values for testing $H_0 : \beta_k = 0$ vs $H_1 : \beta_k \neq 0$ for each k.

(g) R^2 for the regression.