1. Consider the general linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. Suppose that $\mathbf{X}$ is $n \times p$ with rank $r < p$. Let $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$ denote a least squares estimator of $\boldsymbol{\beta}$.
(a) Find $E(\widehat{\boldsymbol{\beta}})$ and $\text{cov}(\widehat{\boldsymbol{\beta}})$.
(b) Do your results in part (a) change when $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$, where $\mathbf{V} \neq \mathbf{I}$?
(c) Do your answers in part (a) change when $r = p$?

2. Consider the regression model $Y_i = \beta_0 + \beta_1 x_i + \beta_2(3x_i^2 - 2) + \epsilon_i$, for $i = 1, 2, 3$, where $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$.
(a) Put this model into $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ form.
(b) Find the least-squares estimates of β_0, β_1, and β_2.
(c) Show that the least-squares estimates of β_0 and β_1 are unchanged if $\beta_2 = 0$. Why do you think this happens?

3. Consider an experiment to study the effect of baking time, x, on the breaking strength of a ceramic, Y. The following eight data values were obtained:

x	2	6	8
y	15, 20, 25	21, 25, 29	33, 37

(a) Consider the cell means model $Y_{ij} = \mu_i + \epsilon_{ij}$, for $i = 1, 2, 3$ and $j = 1, 2, ..., n_i$, where $n_1 = n_2 = 3$ and $n_3 = 2$. Put this model into $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ form, where $\boldsymbol{\beta} = (\mu_1, \mu_2, \mu_3)'$.
(b) Consider the model $Y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_{ij}$, where $x_1 = 2$, $x_2 = 6$, and $x_3 = 8$. Write this model as $\mathbf{Y} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, where $\boldsymbol{\gamma} = (\beta_0, \beta_1, \beta_2)'$.
(c) Show that $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$. What does this imply about the fitted values and residuals for these two models?

4. Consider the general linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. Let $\widehat{\mathbf{Y}}$ denote the vector of least squares fitted values and $\widehat{\mathbf{e}}$ denote the vector of least squares residuals. Compute
(a) $E(\widehat{\mathbf{Y}})$
(b) $\text{cov}(\widehat{\mathbf{Y}})$
(c) $E(\widehat{\mathbf{e}})$
(d) $\text{cov}(\widehat{\mathbf{e}})$
(e) $\text{cov}(\widehat{\mathbf{Y}}, \widehat{\mathbf{e}})$.

5. The observed tension, Y, in a nonextensible string required to maintain a body of unknown weight, w, in equilibrium on a smooth inclined plane of angle θ, $0 < \theta < \pi/2$, is a random variable with mean $E(Y) = w \sin\theta$. For n known values $\theta_1, \theta_2, ..., \theta_n$, set by the experimenter and a given body, the observed data are $Y_1, Y_2, ..., Y_n$.
(a) Find $\widehat{w}$, the least squares estimator of w, the weight of this body.
(b) Compute $E(\widehat{w})$ and $\text{var}(\widehat{w})$. You may assume that $Y_1, Y_2, ..., Y_n$ are independent.

(c) Let $\widehat{Y}_1, \widehat{Y}_2, ..., \widehat{Y}_n$ denote the least squares fitted values. Is it necessarily true that $\sum_{i=1}^n (Y_i - \widehat{Y}_i) = 0$? Explain.

6. Define the matrices

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{W} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Take $\mathbf{Y} = (1, 0, 1, 2)'$.
(a) Show that $\mathcal{C}(\mathbf{W}) \subset \mathcal{C}(\mathbf{X})$.
(b) Express $\mathbf{Y}$ as the sum of two vectors: one in $\mathcal{C}(\mathbf{X})$ and one in $\mathcal{N}(\mathbf{X}')$.
(c) Compute the ppm onto $\mathcal{C}(\mathbf{W})^{\perp}_{\mathcal{C}(\mathbf{X})}$ and then project $\mathbf{Y}$ onto this space.

7. Consider the simple linear regression model in Section 2.3 (notes).
(a) Show algebraically that $\mathbf{P_X}$ and $\mathbf{P_W}$ are both equal to

$$\begin{pmatrix} \frac{1}{n} + \frac{(x_1-\overline{x})^2}{\sum_i (x_i-\overline{x})^2} & \frac{1}{n} + \frac{(x_1-\overline{x})(x_2-\overline{x})}{\sum_i (x_i-\overline{x})^2} & \cdots & \frac{1}{n} + \frac{(x_1-\overline{x})(x_n-\overline{x})}{\sum_i (x_i-\overline{x})^2} \\ \frac{1}{n} + \frac{(x_1-\overline{x})(x_2-\overline{x})}{\sum_i (x_i-\overline{x})^2} & \frac{1}{n} + \frac{(x_2-\overline{x})^2}{\sum_i (x_i-\overline{x})^2} & \cdots & \frac{1}{n} + \frac{(x_2-\overline{x})(x_n-\overline{x})}{\sum_i (x_i-\overline{x})^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} + \frac{(x_1-\overline{x})(x_n-\overline{x})}{\sum_i (x_i-\overline{x})^2} & \frac{1}{n} + \frac{(x_2-\overline{x})(x_n-\overline{x})}{\sum_i (x_i-\overline{x})^2} & \cdots & \frac{1}{n} + \frac{(x_n-\overline{x})^2}{\sum_i (x_i-\overline{x})^2} \end{pmatrix}.$$

In regression analysis, this matrix is called the **hat matrix**.
(b) Compute the trace of this matrix.
(c) Use the ceramic data from Problem 3 and fit both the centered and uncentered simple linear regression models. Report least squares estimates for both models. Also, show that the fitted values and residuals are the same for both fits.

8. Consider two linear models for the same data:

$$\begin{aligned} &\text{Model 1:} \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &\text{Model 2:} \quad \mathbf{Y} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon}. \end{aligned}$$

Here, $\mathbf{X}$ is $n \times p$, $\mathbf{W}$ is $n \times q$, $\boldsymbol{\beta}$ is $p \times 1$, and $\boldsymbol{\gamma}$ is $q \times 1$. Suppose that $\mathcal{C}(\mathbf{W}) \subset \mathcal{C}(\mathbf{X})$. For Model 1, let $\widehat{\mathbf{Y}}_X$, $\widehat{\mathbf{e}}_X$, and $\mathbf{P_X}$ denote the vector of (least-squares) fitted values, the vector of (least-squares) residuals, and the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$. The quantities $\widehat{\mathbf{Y}}_W$, $\widehat{\mathbf{e}}_W$, and $\mathbf{P_W}$ are defined analogously.
(a) Show that $\mathbf{W} = \mathbf{XC}$, for some $p \times q$ matrix $\mathbf{C}$.
(b) Show that $\mathbf{P_X P_W} = \mathbf{P_W}$.
(c) Show that $(\widehat{\mathbf{Y}}_X - \widehat{\mathbf{Y}}_W)'\widehat{\mathbf{Y}}_W = 0$.
(d) Show that $\mathbf{Y}'\mathbf{Y} = \widehat{\mathbf{Y}}_W'\widehat{\mathbf{Y}}_W + (\widehat{\mathbf{Y}}_X - \widehat{\mathbf{Y}}_W)'(\widehat{\mathbf{Y}}_X - \widehat{\mathbf{Y}}_W) + \widehat{\mathbf{e}}_X'\widehat{\mathbf{e}}_X$.

9. The effectiveness of three skin creams was studied in an experiment on s subjects. On the forearm of each subject, three locations were specified. The three creams were randomly allocated to locations on each subject; that is, each subject received a complete set of three treatments. It is assumed that the three observations on the same individual are correlated and that observations on different subjects are uncorrelated. A statistical model for this experiment is

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where Y_{ij} denotes the ith measurement on subject j and μ_i denotes the mean response for the ith skin cream. Assume that ϵ_{ij}, for $i = 1, 2, 3$ and $j = 1, 2, ..., s$, are random variables with $E(\epsilon_{ij}) = 0$, $\text{var}(\epsilon_{ij}) = \sigma^2$, and $\text{corr}(\epsilon_{ij}, \epsilon_{i'j}) = \rho$, for $i \neq i'$. Note that $\text{corr}(\epsilon_{ij}, \epsilon_{ij'}) = 0$ when $j \neq j'$ (regardless of i) because ϵ_{ij} and $\epsilon_{ij'}$ correspond to different subjects.
(a) Assuming that μ_i is fixed (not random), express this model as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Define all vectors and matrices. Your design matrix $\mathbf{X}$ should be full rank.
(b) Compute $\text{cov}(\mathbf{Y})$.
(c) Find $\widehat{\boldsymbol{\beta}}$, the least-squares estimator of $\boldsymbol{\beta}$. Your answer should be a vector (I don't want to see matrices in your final answer).
(d) Compute $\text{cov}(\widehat{\boldsymbol{\beta}})$.

10. Let $\mathbf{P_X}$ denote the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$.
(a) Give a detailed argument showing that $\mathbf{I} - \mathbf{P_X}$ is the perpendicular projection matrix onto $\mathcal{N}(\mathbf{X}')$.
(b) Let

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Compute $\mathbf{P_X}$ and $\mathbf{I} - \mathbf{P_X}$.
(c) Express

$$\mathbf{Y} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$$

as the sum of two vectors, one in $\mathcal{C}(\mathbf{X})$ and one in $\mathcal{N}(\mathbf{X}')$.
(d) For the $\mathbf{X}$ matrix in part (b), describe in words what $\mathcal{C}(\mathbf{X})$ and $\mathcal{N}(\mathbf{X}')$ are.

11. Consider the general linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $E(\boldsymbol{\epsilon}) = \mathbf{0}$. Let $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ denote the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$ and denote by $\widehat{\mathbf{e}}$ the vector of residuals obtained from the least squares fit. Prove that $\widehat{\boldsymbol{\beta}}$ is a least squares estimate of $\boldsymbol{\beta}$ if and only if $\widehat{\mathbf{e}} \perp \mathcal{C}(\mathbf{X})$.

12. Consider the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$, where $p = k+1$, and k is the number of independent variables in the model (the model also includes an intercept term). Assume that $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. Let $\mathbf{M}$ denote the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$, let $\mathbf{J}$ denote an $n \times n$ matrix of ones, and let $\widehat{\boldsymbol{\beta}}$ denote the (unique) OLS estimator.
(a) Show that $\mathbf{Y}'(\mathbf{M} - n^{-1}\mathbf{J})\mathbf{Y} = \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2$, where $\overline{Y}$ is the sample mean of $Y_1, Y_2, ..., Y_n$.
(b) Prove that $r(\mathbf{M} - n^{-1}\mathbf{J}) = k$.
(c) Let $\widehat{\mathbf{e}}$ denote the vector of residuals from the OLS fit. Find $E(\widehat{\mathbf{e}})$ and $\text{cov}(\widehat{\mathbf{e}})$. Do the least squares residuals have constant variance?

TERMINOLOGY: The **Gram-Schmidt** procedure is a method for orthonormalizing a set of basis vectors. Let $\mathcal{V}$ be a vector space with basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r\}$. For $s = 1, 2, ..., r$, define inductively

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{u}_1/||\mathbf{u}_1|| \\ \mathbf{w}_s &= \mathbf{u}_s - \sum_{i=1}^{s-1}(\mathbf{u}_s'\mathbf{v}_i)\mathbf{v}_i \\ \mathbf{v}_s &= \mathbf{w}_s/||\mathbf{w}_s||. \end{aligned}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ is an orthonormal basis for $\mathcal{V}$, where $\mathbf{v}_s \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_s\}$.

13. Consider the vector space $\mathcal{V} = \mathcal{R}^3$ and the vectors $\mathbf{u}_1 = (1,1,1)'$, $\mathbf{u}_2 = (0,1,1)'$, and $\mathbf{u}_3 = (0,0,1)'$. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a basis for $\mathcal{V}$. Then use the Gram-Schmidt procedure to orthonormalize the basis.

14. Let $\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_r$ be an orthonormal basis for $\mathcal{C}(\mathbf{X})$ and $\mathbf{O} = (\mathbf{o}_1\ \mathbf{o}_2\ \cdots\ \mathbf{o}_r)$. Prove that $\mathbf{O}\mathbf{O}' = \sum_{i=1}^{r}\mathbf{o}_i\mathbf{o}_i'$ is the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$.

15. Consider the one-way ANOVA model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, for $i = 1, 2, ..., a$ and $j = 1, 2, ..., n_i$, so that the design matrix is

$$\mathbf{X}_{n\times p} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_a} & \mathbf{0}_{n_a} & \mathbf{0}_{n_a} & \cdots & \mathbf{1}_{n_a} \end{pmatrix},$$

where $p = a+1$ and $n = \sum_i n_i$.
(a) Show that the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$ is given by the $n \times n$ matrix

$$\mathbf{P}_{\mathbf{X}} = \text{Blk Diag}(n_i^{-1}\mathbf{J}_{n_i\times n_i}),$$

where $\mathbf{J}_{n_i \times n_i}$ is the $n_i \times n_i$ matrix of ones and "Blk Diag" stands for "block diagonal." For example, if $a = 3$, $n_1 = n_2 = 2$, and $n_3 = 3$, then $n = 7$ and

$$\mathbf{P_X} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}_{7\times 7}.$$

(b) We have seen that $\mathbf{P_1} = n^{-1}\mathbf{J}_{n\times n}$ is the perpendicular projection matrix responsible for removing the effects of the intercept term; in this model, the intercept term is μ. We have also seen that $\mathbf{P_X} - \mathbf{P_1}$ is the perpendicular projection matrix which projects $\mathbf{Y}$ onto the orthogonal complement of $\mathcal{C}(\mathbf{1})$ with respect to $\mathcal{C}(\mathbf{X})$, a subspace of dimension $r(\mathbf{P_X} - \mathbf{P_1}) = a - 1$. Show that

$$\mathbf{Y}'(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y} = \sum_{i=1}^{a} n_i(\overline{Y}_{i+} - \overline{Y}_{++})^2.$$

This is the called the corrected treatment (model) sum of squares.
(c) The quantity $\mathbf{Y}'(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y}$ is useful. An uninteresting use of this quantity involves testing the hypothesis that all the α_i's are equal; i.e., testing $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a$. Informally, this is done by comparing the size of $\mathbf{Y}'(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y}$ to the size of $\mathbf{Y}'(\mathbf{I} - \mathbf{P_X})\mathbf{Y}$, the residual sum of squares, while adjusting for the ranks of $\mathbf{P_X} - \mathbf{P_1}$ and $\mathbf{I} - \mathbf{P_X}$; i.e., $a - 1$ and $n - a$. It is more useful (and more interesting) to break this quantity up into smaller pieces and test more refined hypotheses that correspond to the pieces. One way to do this is to break up $\mathbf{Y}'(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y}$ into $a - 1$ components

$$\mathbf{Y}'(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y} = \mathbf{Y}'\mathbf{M}_1\mathbf{Y} + \mathbf{Y}'\mathbf{M}_2\mathbf{Y} + \cdots + \mathbf{Y}'\mathbf{M}_{a-1}\mathbf{Y},$$

where $\mathbf{M}_i\mathbf{M}_j = \mathbf{0}$, for all $i \neq j$, and $\mathbf{M}_1, \mathbf{M}_2, ..., \mathbf{M}_{a-1}$ are perpendicular projection matrices onto $a - 1$ orthogonal subspaces of $\mathcal{C}(\mathbf{P_X} - \mathbf{P_1})$. The sums of squares $\mathbf{Y}'\mathbf{M}_i\mathbf{Y}$, $i = 1, 2, ..., a - 1$ each have 1 degree of freedom and can be used to test **orthogonal contrasts**. Breaking up $\mathcal{C}(\mathbf{P_X} - \mathbf{P_1})$ in this fashion can be done using the Gram-Schmidt orthonormalization procedure. Set $\mathbf{M}_* = \mathbf{P_X} - \mathbf{P_1}$. We now break up $\mathcal{C}(\mathbf{M}_*)$ into $a - 1$ orthogonal subspaces. Let $\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_{a-1}$ be an orthonormal basis for $\mathcal{C}(\mathbf{M}_*)$. Note that, using Gram-Schmidt, $\mathbf{o}_1$ can be any normalized vector in $\mathcal{C}(\mathbf{M}_*)$, $\mathbf{o}_2$ can be any normalized vector in $\mathcal{C}(\mathbf{M}_*)$ orthogonal to $\mathbf{o}_1$, and so on. Set $\mathbf{O} = (\mathbf{o}_1\ \mathbf{o}_2\ \cdots\ \mathbf{o}_{a-1})$. From Problem 14, we have

$$\mathbf{M}_* = \mathbf{O}\mathbf{O}' = \sum_{i=1}^{a-1} \mathbf{o}_i\mathbf{o}_i'.$$

Take $\mathbf{M}_i = \mathbf{o}_i\mathbf{o}_i'$. Then, $\mathbf{M}_i$ is a perpendicular projection matrix in its own right and $\mathbf{M}_i\mathbf{M}_j = \mathbf{0}$, for $i \neq j$, because of orthogonality. Finally, note that

$$\mathbf{Y}'\mathbf{M}_*\mathbf{Y} = \mathbf{Y}'(\mathbf{M}_1 + \mathbf{M}_2 + \cdots + \mathbf{M}_{a-1})\mathbf{Y} = \mathbf{Y}'\mathbf{M}_1\mathbf{Y} + \mathbf{Y}'\mathbf{M}_2\mathbf{Y} + \cdots + \mathbf{Y}'\mathbf{M}_{a-1}\mathbf{Y}.$$

This demonstrates that the corrected model sum of squares $\mathbf{Y}'\mathbf{M}_*\mathbf{Y}$ can be written as the sum of $a-1$ pieces, as claimed. Now, for specificity, take $a=3$ and $n_1=n_2=n_3=2$. Break up $\mathcal{C}(\mathbf{P_X}-\mathbf{P_1})$ into $a-1=2$ orthogonal subspaces. With your orthogonal subspaces (and their associated ppms, say, $\mathbf{M}_1$ and $\mathbf{M}_2$), verify that

$$\mathbf{Y}'(\mathbf{P_X}-\mathbf{P_1})\mathbf{Y}=\mathbf{Y}'\mathbf{M}_1\mathbf{Y}+\mathbf{Y}'\mathbf{M}_2\mathbf{Y},$$

using the observed data $\mathbf{Y}=(1,0,2,1,3,4)'$.