In all the questions that follow, X is an n by p design matrix with full column rank and P is the orthogonal projection onto the column space of X. Also, let X_0 be the matrix formed from the first $p_0 < p$ columns of X and let P_0 be the orthogonal projection onto the column space of X_0. The vector $Y \in \mathbb{R}^n$ will be a vector of responses and we will define $\hat{\beta} := (X^T X)^{-1} X^T Y$, $\hat{\beta}_0 := (X_0^T X_0)^{-1} X_0^T Y$ and $\tilde{\sigma}^2 := \|(I - P)Y\|^2/(n - p)$.

1. Consider the linear model
$$Y = X\beta + \varepsilon, \qquad \mathrm{Var}(\varepsilon) = \sigma^2 I. \tag{0.1}$$
Let $\tilde{\beta}$ be an unbiased linear estimator of β with $X\tilde{\beta} = AY$ for a matrix A.
   (a) Show that $AP = P$ so we may write $A = P + A(I - P)$.
   (b) By considering $\mathrm{Var}(AY) - \mathrm{Var}(PY)$ or otherwise, show that $\hat{\beta}$ is BLUE (the best linear unbiased estimator) for β in the linear model above (i.e. show that $\mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta})$ is positive semi-definite).
   (c) Conclude that for any $x^* \in \mathbb{R}^p$,
$$\mathbb{E}[\{x^{*T}(\hat{\beta} - \beta)\}^2] \leq \mathbb{E}[\{x^{*T}(\tilde{\beta} - \beta)\}^2].$$

2. Let W be an $n \times n$ positive definite matrix. Using the result from question 1 or otherwise, show that the best linear unbiased estimator for β in the model
$$Y = X\beta + \varepsilon, \qquad \mathrm{Var}(\varepsilon) = \sigma^2 W^{-1},$$
is
$$\hat{\beta}^{\mathrm{w}} := (X^T W X)^{-1} X^T W Y.$$
What are the variances of $\hat{\beta}^{\mathrm{w}}$ and $\hat{\beta} := (X^T X)^{-1} X^T Y$? If W is also diagonal, further show that
$$\hat{\beta}^{\mathrm{w}} = \underset{b \in \mathbb{R}^p}{\mathrm{argmin}} \left\{ \sum_{i=1}^{n} W_{ii}(Y_i - x_i^T b)^2 \right\},$$
where x_i^T is the i^{th} row of X.

3. Consider the model $Y = f + \varepsilon$ where $\mathbb{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I$ and $f \in \mathbb{R}^n$ is a non-random vector. Suppose we have have performed linear regression of Y on X so the fitted values are PY. Show that if $Y^* = f + \varepsilon^*$ where $\mathbb{E}(\varepsilon^*) = 0$, $\mathrm{Var}(\varepsilon^*) = \sigma^2 I$ and ε^* is independent of ε, then
$$\frac{1}{n}\mathbb{E}(\|PY - Y^*\|^2) = \frac{\sigma^2 p}{n} + \frac{1}{n}\|(I - P)f\|^2 + \sigma^2.$$
This is an instance of the (squared) bias-variance decomposition, which applies to any prediction $g(Y)$ in a regression problem.

4. Show that
$$\|(P - P_0)Y\|^2 = \|(I - P_0)Y\|^2 - \|(I - P)Y\|^2 = \|PY\|^2 - \|P_0 Y\|^2.$$

5. (a) Let V and W be linear subspaces of $\mathbb{R}^n$ with $V \subseteq W$. Let Π_V and Π_W denote orthogonal projections onto V and W respectively. Show that for all $v \in \mathbb{R}^n$, $\|v\|^2 \geq \|\Pi_W v\|^2 \geq \|\Pi_V v\|^2$.
   (b) Consider the linear model (0.1) but where only the first p_0 components of β are non-zero. Show that
$$\mathrm{Var}(\hat{\beta}_{0,j}) \leq \mathrm{Var}(\hat{\beta}_j) \qquad \text{for } j = 1, \ldots, p_0.$$
Here $\hat{\beta}_{0,j}$ denotes the j^{th} component of $\hat{\beta}_0$. *Hint: Use the alternative characterisation of $\hat{\beta}_j$ that if $X_j^{\perp}$ is the orthogonal projection of X_j (the j^{th} column of X) onto the orthogonal complement of the column space of X_{-j} (the matrix formed by removing the j^{th} column from X), then $\hat{\beta}_j = \frac{(X_j^{\perp})^T Y}{\|X_j^{\perp}\|^2}$.*

6. Consider a partitioning of X, β and $\hat{\beta}$ as

$$X = (X_0 \; X_1), \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \qquad \hat{\beta} = \begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{pmatrix},$$

where X_1 is $n \times p - p_0$, and correspondingly $\beta_0, \tilde{\beta}_0 \in \mathbb{R}^{p_0}$ and $\beta_1, \tilde{\beta}_1 \in \mathbb{R}^{p-p_0}$. Let $\tilde{X}_1 = (I - P_0)X_1$ and $\tilde{Y} = (I - P_0)Y$.

   (a) Show that $P - P_0$ is the orthogonal projection onto the column space of $\tilde{X}_1$.

   (b) Show that $\tilde{X}_1$ has full rank, and show that $\tilde{\beta}_1$ is equal to the OLS solution of regressing $\tilde{Y}$ on $\tilde{X}_1$, that is, $\tilde{\beta}_1 = (\tilde{X}_1^T \tilde{X}_1)^{-1} \tilde{X}_1^T \tilde{Y}$.

   (c) Find a condition on X_0 and X_1 such that $\tilde{\beta}_0 = \hat{\beta}_0$.

7. Show that the maximum likelihood estimator of σ^2 in the normal linear model ($Y = X\beta + \varepsilon$ with $\varepsilon \sim N_n(0, \sigma^2 I)$) is $\hat{\sigma}^2 = \|(I - P)Y\|^2/n$.

8. Let the cuboid C be defined as $C := \prod_{j=1}^{p} C_j(\alpha/p)$, where

$$C_j(\alpha) = \left[\hat{\beta}_j - \sqrt{\tilde{\sigma}^2 (X^T X)^{-1}_{jj}}\, t_{n-p}(\alpha/2),\ \hat{\beta}_j + \sqrt{\tilde{\sigma}^2 (X^T X)^{-1}_{jj}}\, t_{n-p}(\alpha/2) \right].$$

Assuming the normal linear model, show that $\mathbb{P}(\beta \in C) \geq 1 - \alpha$.

9. Data are available on weights of two groups of three rats at the beginning of a fortnight and at its end. During the fortnight, one group was fed normally, and the other was given a growth inhibitor. The weights of the k^{th} rat in the j^{th} group before and after the fortnight are X_{jk} and y_{jk} respectively. The y_{jk} are taken as realisations of random variables Y_{jk} that follow the model $Y_{jk} = \alpha_j + \beta_j X_{jk} + \varepsilon_{jk}$.

   (a) Let W be the vector of responses, so $W = (Y_{11}, Y_{12}, Y_{13}, Y_{21}, Y_{22}, Y_{23})^T$, and similarly let δ be the vector of random errors. Write down the model above in the form $W = A\theta + \delta$, giving the design matrix A explicitly and expressing the vector of parameters θ in terms of the α_j and β_j.

   (b) The model is to be reparametrised in such a way that it can be specialised to (i) two parallel lines for the two groups, (ii) two lines with the same intercept, (iii) one common line for both groups, just by setting parameters to zero. Give one design matrix that can be made to correspond to (i), (ii) and (iii), just by dropping columns, specifying which columns are to be dropped for which cases.

10. Using a suitable pivotal quantity, construct a $(1 - \alpha)$-confidence interval of σ^2.

11. An α-level test is called F-test (t-test), if the test statistic follows an F-distribution (t-distribution) under the null hypothesis.

   (a) Fix $\beta_0 \in \mathbb{R}^p$. Find an α-level F-test for testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.

   (b) Fix $\beta_0, v \in \mathbb{R}^p$. Find a two-sided α-level t-test for testing $H_0 : \beta^T v = \beta_0^T v$ against $H_1 : \beta^T v \neq \beta_0^T v$.

12. Prove that in the linear model $Y = X\beta + \varepsilon$ with $\varepsilon \sim N_n(0, \sigma^2 I)$, the F-test and t-test for testing the significance of a single predictor are equivalent. That is, prove that, taking $p_0 = p - 1$, so P_0 is the orthogonal projection onto the first $p - 1$ columns of X,

$$\frac{\hat{\beta}_p^2}{\{(X^T X)^{-1}\}_{pp} \tilde{\sigma}^2} = \frac{\|(P - P_0)Y\|^2}{\frac{1}{n-p} \|(I - P)Y\|^2}.$$

Hint: Look at the hint to question 5(b).