**Required Part:**

**0.** Read §4f A General Model for Two-way Data and Variance Components, §4g The Theory and Application of statistical Regression.

**1.** Consider $Y \sim N_n(X\beta, \sigma^2 I_n)$, where $X$ is $n \times m$ with rank $r \le m < n$. Assume $\sigma^2 > 0$ is known.

(a) Let $c'\beta$ be estimable and $b'Y$ be an unbiased estimator of $c'\beta$, not necessarily the BLUE. Find a $(1-\alpha)\%$ confidence interval $b'Y \pm \frac{1}{2}l(b, Y)$ for $c'\beta$, that is, specify $l(b, Y)$.

(b) Given $c'\beta$, find the unbiased estimator $b_0'Y$ of $c'\beta$ with minimum expected length $E(l(b, Y))$.
*Hint: See "Extremum of a Quadratic Form" on page 229.*

(c) Consider the class of all normed estimable functions

$$S = \{c'\beta \mid c \in \mathcal{M}(X'), c'c = 1\}.$$

Find the function $c_0'\beta$ in $S$ which has the smallest minimum expected length confidence interval. That is,

$$\min_{b:E(b'Y)=c_0'\beta} E(l(b, Y)) = \min_{c'\beta \in S} \min_{b:E(b'Y)=c'\beta} E(l(b, Y)).$$

*Hint: Suppose* $X'X = \lambda_1 P_1 P_1' + \cdots + \lambda_r P_r P_r'$*, then* $c \in \mathcal{M}(X'X)$ *if and only if* $c = a_1 P_1 + \cdots + a_r P_r$.

**2.** (**One-way ANOVA**) Suppose the linear regression model is given by

$$Y_{ij} = \mu_i + \epsilon_{ij},\ i = 1, \ldots, k;\ j = 1, \ldots, b,$$

where $\epsilon_{ij}$'s are iid $\sim N(0, \sigma^2)$. In its matrix form, we write

$$Y = (Y_{11}, \ldots, Y_{1b}, Y_{21}, \ldots, Y_{2b}, \ldots, Y_{kb})'$$

as an $n \times 1$ vector of responses, $\beta = (\mu_1, \ldots, \mu_k)'$ as a $k \times 1$ vector of parameters, and

$$\epsilon = (\epsilon_{11}, \ldots, \epsilon_{1b}, \epsilon_{21}, \ldots, \epsilon_{2b}, \ldots, \epsilon_{kb})'$$

as an $n \times 1$ vector of noises, where $n = kb$.

(a) Find the matrix $X$ in the form of $Y = X\beta + \epsilon$. Determine its rank.

(b) Determine $X'X$ and derive the least square estimator $\hat{\beta}$ for $\beta$.

(c) Determine $R_0^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})$ and derive an unbiased estimator $s^2$ for $\sigma^2$ based on $R_0^2$.

(d) Consider the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$. Under $H_0$, rewrite the design matrix $X$ into $X_0$. Determine $R_1^2 = \min_{\mu \in \mathbb{R}} (Y - X_0\mu)'(Y - X_0\mu)$.

(e) Show that

$$R_1^2 - R_0^2 = b\sum_{i=1}^{k} \left(\bar{Y}_i - \bar{Y}\right)^2 = b\sum_{i=1}^{k} \bar{Y}_i^2 - n\bar{Y}^2,$$

where $\bar{Y}_i = \frac{1}{b}\sum_{j=1}^{b} Y_{ij}, i = 1, \ldots, k$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{b} Y_{ij}$ .

(f) Rewrite the null hypothesis $H_0$ in (d) into $H'\beta = 0$, where $H$ is a $k \times (k-1)$ matrix such that $\mathcal{M}(H) \subset \mathcal{M}(X')$. Let $Z = H'\hat{\beta}$ with the same $\hat{\beta}$ as in (b). Determine the matrix $D$ such that the dispersion matrix of $Z$ is $\sigma^2 D$. Show that

$$Z'D^{-1}Z = b\sum_{i=1}^{k} \left(\bar{Y}_i - \bar{Y}\right)^2 = b\sum_{i=1}^{k} \bar{Y}_i^2 - n\bar{Y}^2.$$

Note that it is equal to $R_1^2 - R_0^2$ as in (e). Determine a test statistic $F$ to test $H_0$ based on your results.
*Hint: You may choose $H$ with columns $(1, -1, 0, 0, \ldots, 0)'$, $(1, 0, -1, 0, \ldots, 0)'$, $\ldots$, $(1, 0, 0, 0, \ldots, 0, -1)'$. Then use the formulas and test statistic derived in §4b.2 . You may also need the results of Problem 2 in Hw3.*