Chapter 6

General Linear Model: Statistical Inference

6.1 Introduction

So far we have discussed formulation of linear models (Chapter 1), estimability of parameters in a linear model (Chapter 4), least square estimation (Chapter 4), and generalized least square estimation (Chapter 5). In discussing LS or GLS estimators, we have not made any probabilistic inference mainly because we have not assigned any probability distribution to our linear model structure. Statistical models very often demand more than just estimating the parameters. In particular, one is usually interested in putting a measure of uncertainty in terms of confidence levels or in testing whether some linear functions of the parameters such as difference between two treatment effects is significant or not.

As we know from our statistical methods courses that interval estimation or hypothesis testing almost always require a probability model and the inference depends on the particular model you chose. The most common probability model used in statistical inference is the normal model. We will start with the Gauss-Markov model, namely,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon, \tag{6.1.1}$$

where

Assumption I. $E(\epsilon) = 0$**,**

Assumption II. $cov(\epsilon) = \sigma^2\mathbf{I}$**,**

and introduce the assumption of normality to the error components. In other words,

Assumption IV. $\epsilon \sim N(0, \sigma^2\mathbf{I}_n)$**.**

As you can see, **Assumption IV** incorporates **Assumptions I and II**. The model (6.1.1) along with assumption IV will be referred to as normal theory linear model.

6.2 Maximum likelihood, sufficiency and UMVUE

Under the normal theory linear model described in previous section, the likelihood function for the parameter $\boldsymbol{\beta}$ and σ^2 can be written as

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y}, \mathbf{X}) &= \left\{\frac{1}{\sigma\sqrt{2\pi}}\right\}^n exp\left\{-\frac{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}{\sigma^2}\right\} \quad (6.2.1)\\ &= \left\{\frac{1}{\sigma\sqrt{2\pi}}\right\}^n exp\left\{-\frac{\mathbf{Y}^T\mathbf{Y}-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y}+\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})}{\sigma^2}\right\}, \quad (6.2.2) \end{aligned}$$

Since $\mathbf{X}$ is known, (6.2.2) implies that $L(\boldsymbol{\beta}, \sigma^2|\mathbf{Y}, \mathbf{X})$ belongs to an exponential family with a joint **complete sufficient statistic** $(\mathbf{Y}^T\mathbf{Y}, \mathbf{X}^T\mathbf{Y})$ for $(\boldsymbol{\beta}, \sigma^2)$. Also, (6.2.1) shows that the likelihood is maximized for $\boldsymbol{\beta}$ for given σ^2 when

$$\|\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \quad (6.2.3)$$

is minimized. But, when is (6.2.3) minimized? From chapter 4, we know that (6.2.3) is minimized when $\boldsymbol{\beta}$ is a solution to the normal equations

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{Y}. \quad (6.2.4)$$

Thus **maximum likelihood estimator (MLE)** of $\boldsymbol{\beta}$ also satisfies the normal equations. By the invariance property of maximum likelihood, MLE of an estimable function $\lambda^T\boldsymbol{\beta}$ is given by $\lambda^T\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is a solution to the normal equations (6.2.4). Using the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, we express the likelihood (6.2.1) as a function of σ^2 only, namely,

$$L(\sigma^2|\mathbf{Y},\mathbf{X}) = \left\{\frac{1}{\sigma\sqrt{2\pi}}\right\}^n exp\left\{-\frac{(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{\sigma^2}\right\} \tag{6.2.5}$$

with corresponding log-likelihood

$$lnL(\sigma^2|\mathbf{Y},\mathbf{X}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{\sigma^2}, \tag{6.2.6}$$

which is maximized for σ^2 when

$$\hat{\sigma}^2_{MLE} = \frac{(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{n}. \tag{6.2.7}$$

Note that while the MLE of $\boldsymbol{\beta}$ is identical to the LS estimator, $\hat{\sigma}^2_{MLE}$ is not the same as the LS estimator of σ^2.

Proposition 6.2.1. *MLE of σ^2 is biased.*

Proposition 6.2.2. *The MLE $\lambda^T\hat{\boldsymbol{\beta}}$ is the uniformly minimum variance unbiased estimator (UMVUE) of $\lambda^T\boldsymbol{\beta}$.*

Proposition 6.2.3. *The MLE $\lambda^T\hat{\boldsymbol{\beta}}$ and $n\hat{\sigma}^2_{MLE}/\sigma^2$ are independently distributed, respectively as $N(\lambda^T\boldsymbol{\beta}, \sigma^2\lambda^T(\mathbf{X}^T\mathbf{X})^g\lambda)$ and χ^2_{n-r}.*

6.3 Confidence interval for an estimable function

Proposition 6.3.1. *The quantity*

$$\frac{\lambda^T\hat{\boldsymbol{\beta}} - \lambda^T\boldsymbol{\beta}}{\sqrt{\hat{\sigma}^2_{LS}\lambda^T(\mathbf{X}^T\mathbf{X})^g\lambda}}$$

has a t distribution with $n-r$ df .

The proposition 6.3.1 leads the way for us to construct a confidence interval (CI) for the estimable function $\lambda^T\boldsymbol{\beta}$. In fact, a $100(1-\alpha)\%$ CI for $\lambda^T\boldsymbol{\beta}$ is given as

$$\lambda^T\hat{\boldsymbol{\beta}} \pm t_{n-r,\alpha/2}\sqrt{\hat{\sigma}^2_{LS}\lambda^T(\mathbf{X}^T\mathbf{X})^g\lambda}. \tag{6.3.1}$$

This confidence interval is in the familiar form

$$Estimate \pm t_{\alpha/2} * SE.$$

Example 6.3.2. Consider the simple linear regression model considered in Example 1.1.3 given by Equation (1.1.7) where

$$Y_i = \beta_0 + \beta_1 w_i + \epsilon_i, \qquad (6.3.2)$$

where Y_i and w_i respectively represents the survival time and the age at prognosis for the ith patient. We have identified in Example 4.2.3 that a unique LS estimator for β_0 and β_1 is given by

$$\left.\begin{array}{r} \hat{\beta}_1 = \frac{\sum(w_i-\bar{w})(Y_i-\bar{Y})}{\sum(w_i-\bar{w})^2} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{w}, \end{array}\right\} \qquad (6.3.3)$$

provided $\sum(w_i - \bar{w})^2 > 0$. For a given patient who was diagnosed with leukemia at the age of w_0, one would be interested in predicting the length of survival. The LS estimator of the expected survival for this patient is given by

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 w_0. \qquad (6.3.4)$$

What is a 95% confidence interval for $\hat{Y}_0$?

Note that $\hat{Y}_0$ is in the form of $\lambda^T\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$, and $\lambda = (1, w_0)^T$. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by

$$\begin{aligned} cov(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2 \\ &= \frac{\sigma^2}{n\sum(w_i-\bar{w})^2}\begin{bmatrix} \sum w_i^2 & -\sum w_i \\ -\sum w_i & n \end{bmatrix}. \end{aligned} \qquad (6.3.5)$$

Thus the variance of the predictor $\hat{Y}_0$ is given by

$$\begin{aligned}
var(\hat{Y}_0) &= var(\lambda^T\hat{\boldsymbol{\beta}}) \\
&= \lambda^T cov(\hat{\boldsymbol{\beta}})\lambda \\
&= \frac{\sigma^2}{n\sum(w_i-\bar{w})^2}(1, w_0)\begin{bmatrix} \sum w_i^2 & -\sum w_i \\ -\sum w_i & n \end{bmatrix}\begin{pmatrix} 1 \\ w_0 \end{pmatrix} \\
&= \frac{\sigma^2}{n\sum(w_i-\bar{w})^2}(1, w_0)\begin{bmatrix} \sum w_i^2 - w_0\sum w_i \\ -\sum w_i + nw_0 \end{bmatrix} \\
&= \frac{\sigma^2}{n\sum(w_i-\bar{w})^2}\left[\sum w_i^2 - 2w_0\sum w_i + nw_0^2\right] \\
&= \frac{\sigma^2\sum(w_i-w_0)^2}{n\sum(w_i-\bar{w})^2}. \qquad (6.3.6)
\end{aligned}$$

Therefore, the standard error of $\hat{Y}_0$ is obtained as:

$$SE(\hat{Y}_0) = \hat{\sigma}_{LS}\sqrt{\frac{\sum(w_i-w_0)^2}{n\sum(w_i-\bar{w})^2}}, \qquad (6.3.7)$$

where $\hat{\sigma}_{LS}^2 = \sum_{i=1}^n(Y_i - \hat{\beta}_0 - \hat{\beta}_1 w_i)^2/(n-2)$. Using (6.3.1), a 95% CI for $\hat{Y}_0$ is then

$$\hat{Y}_0 \pm t_{n-2,0.025}\hat{\sigma}_{LS}\sqrt{\frac{\sum(w_i-w_0)^2}{n\sum(w_i-\bar{w})^2}}. \qquad (6.3.8)$$

6.4 Test of Hypothesis

Very often we are interested in testing hypothesis related to some linear function of the parameters in a linear model. From our discussions in Chapter 4, we have learned that not all linear functions of the parameter vector can be estimated. Similarly, not all hypothesis corresponding to linear functions of the parameter vector can be tested. We will know shortly, which hypothesis can be tested and which cannot. Let us first look at our favorite one-way-ANOVA model. Usually, the hypotheses of interest are:

1. Equality of a treatment effects:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_a. \tag{6.4.1}$$

2. Equality of two specific treatment effects, e.g.,

$$H_0 : \alpha_1 = \alpha_2, \tag{6.4.2}$$

3. A linear combination such as a contrast (to be discussed later) of treatment effects

$$H_0 : \sum c_i \alpha_i = 0, \tag{6.4.3}$$

   where c_i are known constants such that $\sum c_i = 0$.

Note that, if we consider the normal theory Gauss-Markov linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$, then all of the above hypotheses can be written as

$$\mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}, \tag{6.4.4}$$

where $\mathbf{A}$ is a $p \times q$ matrix with $rank(\mathbf{A}) = q$. If $\mathbf{A}$ is not of full column rank, we exclude the redundant columns from $\mathbf{A}$ to have a full column rank matrix. Now, how would you proceed to test the hypothesis (6.4.4)?

Let us deviate a little and refresh our memory about test of hypothesis. If $Y_1, Y_2, \ldots, Y_n$ are n iid observations from $N(\mu, \sigma^2)$ population, how do we construct a test statistic to test the hypothesis

$$H_0 : \mu = \mu_0? \tag{6.4.5}$$

Thus we took 3 major steps in constructing a test statistic:

- Estimated the parametric function (μ),
- Found the distribution of the estimate, and
- Eliminated the nuisance parameter.

We will basically follow the same procedure to test the hypothesis (6.4.4). First we need to estimate $\mathbf{A}^T\boldsymbol{\beta}$. That would mean that $\mathbf{A}^T\boldsymbol{\beta}$ must be estimable.

Definition 6.4.1. Testable hypothesis. A linear hypothesis $\mathbf{A}^T\boldsymbol{\beta}$ is testable if the rows of $\mathbf{A}^T\boldsymbol{\beta}$ are estimable. In other words (see chapter 4), there exists a matrix $\mathbf{C}$ such that

$$\mathbf{A} = \mathbf{X}^T\mathbf{C}. \tag{6.4.6}$$

The assumption that $\mathbf{A}$ has full column rank is just a matter of convenience. Given a set of equations $\mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$, one can easily eliminate redundant equations to transform it into a system of equations with a full column rank matrix.

Since $\mathbf{A}^T\boldsymbol{\beta}$ is estimable, the corresponding LS estimator of $\mathbf{A}^T\boldsymbol{\beta}$ is given by

$$\mathbf{A}^T\hat{\boldsymbol{\beta}} = \mathbf{A}^T(\mathbf{X}^T\mathbf{X})^g\mathbf{X}^T\mathbf{Y}, \tag{6.4.7}$$

which is a linear function of $\mathbf{Y}$. Under assumption IV, $\mathbf{A}^T\hat{\boldsymbol{\beta}}$ is distributed as

$$\mathbf{A}^T\hat{\boldsymbol{\beta}} \sim N_q(\mathbf{A}^T\boldsymbol{\beta}, \sigma^2\mathbf{A}^T(\mathbf{X}^T\mathbf{X})^g\mathbf{A} = \sigma^2\mathbf{B}), \tag{6.4.8}$$

where we introduced the notation $\mathbf{A}^T(\mathbf{X}^T\mathbf{X})^g\mathbf{A} = \mathbf{B}$. Therefore,

$$\frac{(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{A}^T\boldsymbol{\beta})^T\mathbf{B}^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{A}^T\boldsymbol{\beta})}{\sigma^2} \sim \chi^2_q, \tag{6.4.9}$$

or under the null hypothesis $H_0 : \mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$,

$$\frac{(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})^T\mathbf{B}^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})}{\sigma^2} \sim \chi^2_q. \tag{6.4.10}$$

Had we known σ^2 this test statistic could have been used to test the hypothesis $H_0 : \mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$. But since σ^2 is unknown, the left hand side of equation (6.4.10) cannot be used as a test statistic. In order to get rid of σ^2 from the test statistic, we note that

$$\frac{RSS}{\sigma^2} = \frac{\mathbf{Y}^T(\mathbf{I} - \mathbf{P})\mathbf{Y}}{\sigma^2} = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\sigma^2} \sim \chi^2_{n-r}. \tag{6.4.11}$$

If we could show that the statistics in (6.4.10) and (6.4.11) are independent, then we could construct a F-statistic as a ratio of two mean-chi-squares as follows:

$$F = \frac{\frac{(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})^T\mathbf{B}^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})}{\sigma^2} \Big/ q}{\frac{RSS}{\sigma^2} \Big/ (n-r)} \sim F_{q,n-r} \text{ under } H_0. \tag{6.4.12}$$

Notice that the nuisance parameter σ^2 is canceled out and we obtain the F-statistic

$$F = \frac{(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})^T B^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})}{q\hat{\sigma}^2_{LS}} \sim F_{q,n-r} \text{ under } H_0, \tag{6.4.13}$$

where $\hat{\sigma}^2_{LS}$, the residual mean square, is the LS estimator of σ^2. Let us denote the numerator of the F-statistic by Q.

Independence of Q and $\hat{\sigma}^2_{LS}$

First note that

$$\begin{aligned}
\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b} &= \mathbf{C}^T\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{b} \\
&= \mathbf{C}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^g\mathbf{X}^T\mathbf{Y} - \mathbf{b} \\
&= \mathbf{C}^T\mathbf{P}\mathbf{Y} - \mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{b} \\
&= \mathbf{C}^T\mathbf{P}\mathbf{Y} - \mathbf{C}^T\mathbf{X}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{b} \\
&= \mathbf{C}^T\mathbf{P}\mathbf{Y} - \mathbf{C}^T\mathbf{P}\mathbf{X}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{b} \\
&= \mathbf{C}^T\mathbf{P}(\mathbf{Y} - \mathbf{b}_*),
\end{aligned} \tag{6.4.14}$$

where $\mathbf{b}_* = \mathbf{X}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{b}$. Now,

$$\begin{aligned}
Q &= (\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})^T\mathbf{B}^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b}) \\
&= (\mathbf{Y} - \mathbf{b}_*)^T\mathbf{P}\mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T\mathbf{P}(\mathbf{Y} - \mathbf{b}_*) \\
&= (\mathbf{Y} - \mathbf{b}_*)^T\mathbf{A}_*(\mathbf{Y} - \mathbf{b}_*),
\end{aligned} \tag{6.4.15}$$

where $\mathbf{A}_* = \mathbf{P}\mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T\mathbf{P}$. On the other hand,

$$\begin{aligned} RSS &= (n-r)\hat{\sigma}^2_{LS} \\ &= \mathbf{Y}^T(\mathbf{I}-\mathbf{P})\mathbf{Y} \\ &= (\mathbf{Y}-\mathbf{b}_* + \mathbf{b}_*)^T(\mathbf{I}-\mathbf{P})(\mathbf{Y}-\mathbf{b}_* + \mathbf{b}_*) \\ &= (\mathbf{Y}-\mathbf{b}_*)^T(\mathbf{I}-\mathbf{P})(\mathbf{Y}-\mathbf{b}_*).(\text{Why?}) \end{aligned} \quad (6.4.16)$$

Thus, both Q and RSS are quadratic forms in $\mathbf{Y}-\mathbf{b}_*$, which under assumption IV is distributed as $N(\mathbf{X}\boldsymbol{\beta}-\mathbf{b}_*, \sigma^2\mathbf{I}_n)$. Therefore, SSE and Q are independently distributed if and only if $\mathbf{A}_*(\mathbf{I}-\mathbf{P}) = 0$, which follows immediately.

In summary, the linear testable hypothesis $H_0 : \mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$ can be tested by the F-statistic

$$F = \frac{(\mathbf{A}^T\hat{\boldsymbol{\beta}}-\mathbf{b})^T B^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}}-\mathbf{b})}{q\hat{\sigma}^2_{LS}} \quad (6.4.17)$$

by comparing it to the critical values from a $F_{q,n-r}$ distribution.

Example 6.4.1. Modeling weight loss as a function of initial weight. Suppose n individuals, m men and w women participated in a six-week weight-loss program. At the end of the program, investigators used a linear model with the reduction in weight as the outcome and the initial weight as the explanatory variable. They started with separate intercept and slopes but wanted to see if the rate of decline is similar between men and women.

Consider the linear model:

$$R_i = \begin{cases} \alpha_m + \beta_m W_{0i} + \epsilon_i, & \text{if the ith individual is a male} \\ \alpha_w + \beta_f W_{0i} + \epsilon_i, & \text{if the ith individual is a female} \end{cases} \tag{6.4.18}$$

The idea is to test the hypothesis

$$H_0 : \beta_m = \beta_f. \tag{6.4.19}$$

If we write $\boldsymbol{\beta} = (\alpha_m, \beta_m, \alpha_f, \beta_f)^T$, then H_0 can be written as

$$H_0 : \mathbf{A}^T \boldsymbol{\beta} = \mathbf{b}, \tag{6.4.20}$$

where

$$\mathbf{A} = (0\ 1\ 0\ -1),\ \mathbf{b} = 0. \tag{6.4.21}$$

Assuming (for simplicity) that the first m individuals are male ($i = 1, 2, \ldots, m$) and the rest ($i = m+1, m+2, \ldots, m+w = n$) are female, the linear model for this problem can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon,$$

where $\mathbf{Y} = (R_1, R_2, \ldots, R_n)$,and

$$\mathbf{X} = \begin{pmatrix} 1_m & \boldsymbol{W}_m & 0_m & 0_m \\ 0_f & 0_f & 1_f & \boldsymbol{W}_f \end{pmatrix}, \tag{6.4.22}$$

where $\boldsymbol{W}_m = (w_{01}, w_{02}, \ldots, w_{0m})^T$ and $\boldsymbol{W}_f = (w_{0(m+1)}, w_{0(m+2)}, \ldots, w_{0n})^T$. Obviously,

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} \mathbf{B}_m & 0_{2\times 2} \\ 0_{2\times 2} & \mathbf{B}_f \end{pmatrix}, \qquad (6.4.23)$$

where

$$\mathbf{B}_m = \begin{pmatrix} m & \sum_{i=1}^m w_{0i} \\ \sum_{i=1}^m w_{0i} & \sum_{i=1}^m w_{0i}^2 \end{pmatrix}, \qquad (6.4.24)$$

and

$$\mathbf{B}_f = \begin{pmatrix} m & \sum_{i=m+1}^n w_{0i} \\ \sum_{i=m+1}^n w_{0i} & \sum_{i=m+1}^n w_{0i}^2 \end{pmatrix}. \qquad (6.4.25)$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} \mathbf{B}_m^{-1} & 0_{2\times 2} \\ 0_{2\times 2} & \mathbf{B}_f^{-1} \end{pmatrix}, \qquad (6.4.26)$$

where

$$\mathbf{B}_m^{-1} = \frac{1}{mSS_m(W)} \begin{pmatrix} \sum_{i=1}^m w_{0i}^2 & -\sum_{i=1}^m w_{0i} \\ -\sum_{i=1}^m w_{0i} & m \end{pmatrix}, \qquad (6.4.27)$$

where $SS_m(W) = \sum_{i=1}^m (w_{0i} - \bar{W}_m)^2$ with $\bar{W}_m = \sum_{i=1}^m w_{0i}/m$. Similarly,

$$\mathbf{B}_f^{-1} = \frac{1}{mSS_f(W)} \begin{pmatrix} \sum_{i=m+1}^n w_{0i}^2 & -\sum_{i=m+1}^n w_{0i} \\ -\sum_{i=m+1}^n w_{0i} & m \end{pmatrix}, \qquad (6.4.28)$$

where $SS_f(W) = \sum_{i=m+1}^n (w_{0i} - \bar{W}_f)^2$ with $\bar{W}_f = \sum_{i=m+1}^n w_{0i}/w$. This leads to the usual LS estimator

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \bar{R}_m - \hat{\beta}_m \bar{W}_m \\ SP_m/SS_m(W) \\ \bar{R}_f - \hat{\beta}_f \bar{W}_f \\ SP_f/SS_f(W) \end{pmatrix} \tag{6.4.29}$$

where, $SP_m = \sum_{i=1}^{m}(R_i - \bar{R}_m)(w_{0i} - \bar{W}_m)$, and similarly, $SP_f = \sum_{i=m+1}^{n}(R_i - \bar{R}_f)(w_{0i} - \bar{W}_f)$.

$$\begin{aligned} \mathbf{A}^T\hat{\boldsymbol{\beta}} &= (0\ 1\ 0\ -1) \begin{pmatrix} \bar{R}_m - \hat{\beta}_m \bar{W}_m \\ SP_m/SS_m(W) \\ \bar{R}_f - \hat{\beta}_f \bar{W}_f \\ SP_f/SS_f(W) \end{pmatrix} \\ &= SP_m/SS_m(W) - SP_f/SS_f(W). \end{aligned} \tag{6.4.30}$$

$$\begin{aligned} B &= \mathbf{A}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A} \\ &= (0\ 1\ 0\ -1) \begin{pmatrix} \mathbf{B}_m^{-1} & 0_{2\times 2} \\ 0_{2\times 2} & \mathbf{B}_f^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \\ &= \frac{SS_m(W) + SS_f(W)}{SS_m(W) * SS_f(W)}. \end{aligned} \tag{6.4.31}$$

Hence,

$$\begin{aligned} Q &= (\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})^T\mathbf{B}^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b}) \\ &= \frac{SS_m(W) * SS_f(W)}{SS_m(W) + SS_f(W)}(SP_m/SS_m(W) - SP_f/SS_f(W))^2 \quad (6.4.32) \end{aligned}$$

The residual sum of squares for this model is

$$\begin{aligned} RSS &= \sum_{i=1}^{m}(R_i - \hat{\alpha}_m - \hat{\beta_m}w_{0i})^2 + \sum_{i=m+1}^{n}(R_i - \hat{\alpha}_f - \hat{\beta}_f w_{0i})^2 \\ &= SS_m(R) - \hat{\beta}_m^2 SS_m(W) + SS_f(R) - \hat{\beta}_f^2 SS_f(W), \quad (6.4.33) \end{aligned}$$

where $SS_f(R) = \sum_{i=m+1}^{n}(R_i - \bar{R}_f)^2$ with $\bar{R}_f = \sum_{i=m+1}^{n} R_i/w$, and similarly for $SS_m(R)$. Corresponding F-statistic for testing $H_0 : \beta_m = \beta_f$ is then given by

$$F = \frac{Q}{RSS/(n-4)}. \quad (6.4.34)$$

6.4.1 Alternative motivation and derivation of F-test

Suppose we want to test the testable hypothesis $\mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$ in the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{6.4.35}$$

under assumption IV.

Under the null hypothesis, the model (6.4.35) can be treated as a restricted model with the jointly estimable restrictions imposed by the null hypothesis. Now whether this hypothesis is true can be tested by the extra error sum of squares due to the restriction (null hypothesis). By the theory developed in Chapter 4,

$$RSS_H = RSS + (\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})^T\mathbf{B}^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b}) = RSS + Q \tag{6.4.36}$$

leading to $RSS_H - RSS = Q$. If the null hypothesis is false, one would expect this difference to be large. But how large? We compare this extra error SS to the RSS from the original model. In other words, under the alternative, we expect Q/RSS to be large, or equivalently, $F = (n-r)Q/(q * RSS)$ to be large. The significance of this test can be calculated by noting the fact that under H_0, F follows a F-distribution with q and $n - r$ d.f.

Example 6.4.2. Example 6.4.1 continued. Under the null hypothesis, the model can be written as

$$R_i = \begin{cases} \alpha_m + \beta W_{0i} + \epsilon_i, & \text{if the ith individual is a male} \\ \alpha_w + \beta W_{0i} + \epsilon_i, & \text{if the ith individual is a female} \end{cases} \tag{6.4.37}$$

Notice the same slope for both subgroup. The parameter estimates for this problem are:

$$\begin{pmatrix} \hat{\alpha}_m \\ \hat{\alpha}_w \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \bar{R}_m - \hat{\beta}\bar{W}_m \\ \bar{R}_f - \hat{\beta}\bar{W}_f \\ \frac{SP_m + SP_f}{SS_m(W) + SS_f(W)} = \frac{\hat{\boldsymbol{\beta}}_m SS_m(W) + \hat{\boldsymbol{\beta}}_f SS_f(W)}{SS_m(W) + SS_f(W)} \end{pmatrix}. \tag{6.4.38}$$

It can be shown that the residual sum of squares for this reduced model can be written as

$$RSS_H = SS_m(R) + (\hat{\beta}^2 - 2\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}_m)SS_m(W) + SS_f(R) - (\hat{\beta}^2 - 2\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}_f)SS_f(W), \tag{6.4.39}$$

leading to

$$RSS_H - RSS = (\hat{\beta}_m - \hat{\beta})^2 SS_m(W) + (\hat{\beta}_f - \hat{\beta})^2 SS_f(W). \tag{6.4.40}$$

The test statistic F is then given by,

$$F = \frac{(n-4)\left\{(\hat{\beta}_m - \hat{\beta})^2 SS_m(W) + (\hat{\beta}_f - \hat{\beta})^2 SS_f(W)\right\}}{SS_m(R) - \hat{\beta}_m^2 SS_m(W) + SS_f(R) - \hat{\beta}_f^2 SS_f(W)}. \tag{6.4.41}$$

6.5 Non-testable Hypothesis

In the previous section, we only considered hypothesis which are testable. What happens if you start with a non-testable hypothesis? A non-testable hypothesis $H_0 : \mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$ is one for which every row of the linear function $\mathbf{A}^T\boldsymbol{\beta}$ are non-estimable and no linear combinations of them are estimable. From our discussion in Chapter 4, we know that when we impose non-estimable restrictions on the linear model, the restricted solution becomes yet another solution to the linear model and hence the residual sum of squares does not change. Thus, in this case, extra error sum of squares due to the non-testable hypothesis is identically equal to zero. However, one may still calculate a F-statistic using the formula (6.4.13)

$$F = \frac{(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})^T\mathbf{B}^{-1}(\mathbf{A}^T\hat{\boldsymbol{\beta}} - \mathbf{b})}{q\hat{\sigma}^2_{LS}}, \tag{6.5.1}$$

as long as $\mathbf{B}$ is invertible. What would this statistic test if $H_0 : \mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$ is non-testable? In fact, when $\mathbf{A}^T\boldsymbol{\beta}$ is not estimable, the test statistic F would actually test the hypothesis $\mathbf{A}^T(\mathbf{X}^T\mathbf{X})^g\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{b}$. [See Problem 6 in this chapter.]

Partially Testable Hypothesis

Suppose you want to test the hypothesis

$$\mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}, \quad \mathbf{A} = (\mathbf{A}_1^{p\times q_1}\ \mathbf{A}_2^{p\times q_2});\ q_1 + q_2 = q,\ \text{and } \mathbf{b} = (\mathbf{b}_1^T\ \mathbf{b}_2^T)^T \tag{6.5.2}$$

such that $\mathbf{A}$ has rank q and $\mathbf{A}_1^T\boldsymbol{\beta}$ is estimable while $\mathbf{A}_2^T\boldsymbol{\beta}$ is not. From our discussion on the testable and non-testable hypotheses, we can easily recognize that if we use the F-statistic based on $RSS_H - RSS$, we will end up with a test statistic which will test the hypothesis $H_0 : \mathbf{A}_1^T\boldsymbol{\beta} = \mathbf{b}_1$. However, if we use the test statistic (6.4.13), we will be testing the hypothesis

$$\mathbf{A}_*^T\boldsymbol{\beta} = \mathbf{b}, \tag{6.5.3}$$

where $\mathbf{A}_* = (\mathbf{A}_1 \quad \mathbf{H}^T\mathbf{A}_2)$ with $\mathbf{H} = (\mathbf{X}^T\mathbf{X})^g\mathbf{X}^T\mathbf{X}$.

6.6 Problems

1. Consider the simple regression model:

$$Y_i = \mu + \beta x_i + \epsilon_i,\ i = 1, 2, \ldots, n, \tag{6.6.1}$$

where $\epsilon_i, i = 1, 2, \ldots, n$, are iid normal random variables with mean zero and variance σ^2.

   (a) For a given value x_0 of x, obtain the least square estimator of $\theta(x_0) = \mu + \beta x_0$.

   (b) Derive a test statistic for testing the hypothesis $H_0 : \theta(x_0) = \theta_0$, where θ_0 is a known constant.

2. Suppose $\theta = a^T\beta$ is estimable in the model $Y = X\beta + \epsilon$, where a and β are $p \times 1$ vectors, Y is an $n \times 1$ random vector, X is an $n \times p$ matrix with $rank(X) = r \leq p$, and $\epsilon \sim N(0, \sigma^2 I_n)$. Assume $\hat{\beta}$ and S^2 to be the least squares estimators of β and σ^2 respectively.

   (a) What is the least squares estimator for $a^T\beta$? Call it $\hat{\theta}$.

   (b) What is the distribution of $\hat{\theta}$?

   (c) What is the distribution of $(n - r)S^2/\sigma^2$?

   (d) Show that $\hat{\theta}$ and $(n - r)S^2$ are independently distributed.

   (e) Show how one can construct a 95% confidence interval for θ.

(While answering parts (b) and (c), do not forget to mention the specific parameters of the corresponding distributions.)

3. Suppose $Y_{11}, Y_{12}, \ldots, Y_{1n_i}$ be n_1 independent observations from a $N(\mu + \alpha_1, \sigma^2)$ distribution and $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ be n_2 independent observations from a $N(\mu - \alpha_2, \sigma^2)$ distribution. Notice that the two populations have different means but the same standard deviation. Assume that Y_{1j} and Y_{2j} are independent for all j. Define $n_1 + n_2 = n$, and $Y = (Y_{11}, Y_{12}, \ldots, Y_{1n_1}, Y_{21}, Y_{22}, \ldots, Y_{2n_2})^T$ as the $n \times 1$ vector consisting of all n observations. We write Y as

$$Y = X\beta + \epsilon, \tag{6.6.2}$$

where $\beta = (\mu, \alpha_1, \alpha_2)^T$, $X = \begin{bmatrix} 1_{n_1} & 1_{n_1} & 0 \\ 1_{n_2} & 0 & -1_{n_2} \end{bmatrix}$ with 1_{n_i} representing a $n_i \times 1$ vector of 1's, and ϵ is an $n \times 1$ vector distributed as $N(0, \sigma^2 I_n)$.

(a) A particular solution to the normal equations for estimating β is given by $\hat{\beta} = GX^TY$, where $G = diag(0, 1/n_1, 1/n_2)$. Show that the error sums of squares for this model is given by $SSE = (Y - X\hat{\beta})^T(Y - X\hat{\beta}) = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$.

(b) Suppose one wants to test the hypothesis $H_0 : \alpha_1 = \alpha_2$. The restricted model under the null hypothesis is given by

$$Y = X_H\beta_H + \epsilon, \tag{6.6.3}$$

where $X_H = \begin{bmatrix} 1_{n_1} & 1_{n_1} \\ 1_{n_2} & -1_{n_2} \end{bmatrix}$, and $\beta_H = (\mu, \alpha_1)^T$. The least squares estimator of β_H is given by $\hat{\beta}_H = \frac{1}{2}\begin{bmatrix} \bar{Y}_{1.} + \bar{Y}_{2.} \\ \bar{Y}_{1.} - \bar{Y}_{2.} \end{bmatrix}$. Find the error sum of squares for the restricted model. Call it SSE_H. Compare SSE and SSE_H. Are they equal? Why or why not?

(c) Is the hypothesis $H_1 : \alpha_1 + \alpha_2 = 0$ testable? If so, give a F-statistic to test this hypothesis. If not, state the reason why it is not testable.

4. Suppose two objects, say A_1 and A_2 with unknown weights β_1 and β_2 respectively are measured on a balance using the following scheme, all of these actions being repeated twice:

- both objects on the balance, resulting in weights Y_{11} and Y_{12},
- Only A_1 on the balance, resulting in weights Y_{21} and Y_{22}, and
- Only A_2 on the balance, resulting in weights Y_{31} and Y_{32}.

Assume $Y_{ij}'s$ are independent, normally distributed variables with common variance σ^2. Also assume that the balance has an unknown systematic (fixed) error β_0. The model can then be written as:

$$Y_{11} = Y_{12} = \beta_0 + \beta_1 + \beta_2 + \epsilon_1$$
$$Y_{21} = Y_{22} = \beta_0 + \beta_1 + \epsilon_2$$
$$Y_{31} = Y_{32} = \beta_0 + \beta_2 + \epsilon_3,$$

where $\epsilon_i's$ are i.i.d. random variables. Let $\bar{Y}_{i.} = (Y_{i1} + Y_{i2})/2$.

(a) Express the least square estimates $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ in terms of $\bar{Y}_{i.}, i = 1, 2, 3$.

(b) Find $Var(\hat{\boldsymbol{\beta}})$. Argue that $\hat{\boldsymbol{\beta}}$ has a normal distribution.

(c) Suppose $\hat{\sigma}^2 = \sum_{i=1}^{3}\sum_{j=1}^{2}(Y_{ij} - \bar{Y}_{i.})^2/3$. What is the distribution of $\hat{\sigma}^2$?.

(d) Write the hypothesis that both objects have the same weight in the form $H_0 : \mathbf{A}^T\boldsymbol{\beta} = \mathbf{b}$.

(e) Propose an F-statistic for testing the above hypothesis.

5. A clinical study recruited 150 untreated hepatitis C patients along with 75 hepatitis C patients who have been treated previously with Interferon-α monotherapy. There are two new drugs under study, pegintron and peginterferon. Investigators randomly allocated the two treatments to all 225 patients. The goal is to compare the effect of Pegintron and Peginterferon in reducing the hepatitis C. The response variable Y is the difference in viral levels between baseline and 24 weeks post treatment. Denote by α_P and α_{PI} the effect of pegintron and peginterferon respectively.

(a) Write this problem as a general linear model assuming that the new treatments may interact with whether the patient was previously treated or not. **Use your own notation for effects, subscripts and error terms not defined here**.

(b) What is the design matrix for the model you specified in part (a)?

(c) Is $\alpha_{PI} - \alpha_P$ estimable in the linear model you have proposed in part (a)?

i. If yes, give its BLUE and a $100(1-\alpha)\%$ confidence interval for the difference between α_{PI} and α_P.

ii. If not, give an estimable function that would represent the difference between the two new treatments and give a $100(1-\alpha)\%$ confidence interval for this estimable function.

6. Consider the linear model

$$Y = X\beta + \epsilon,$$

where $\epsilon\ N_n(0, \sigma^2 I_n)$, X is a $n \times p$ matrix of rank $r(< p)$ and β is of order $p \times 1$.

(a) Let $H = (X^T X)^g X^T X$. Show that $A^T H\beta$ is estimable for any $p \times q$ matrix A.

(b) For a matrix A of order $p \times q$ of rank q, the ***testable*** hypothesis $H_0 : A^T\beta = b$ can be tested using the test statistic $F = \frac{(n-r)}{q}\frac{Q}{SSE}$, where

$$Q = (A^T\hat{\beta} - b)^T[A^T(X^TX)^gA]^{-1}(A^T\hat{\beta} - b) \tag{6.6.4}$$

which is identical to $SSE_H - SSE$. Here, $\hat{\beta} = (X^TX)^gX^Ty$ is a solution to the normal equations $X^TX\beta = X^TY$. When $A^T\beta = b$ is ***non-testable***, $SSE_H - SSE$ is identically equal to zero. However, it is still possible to calculate Q using expression (6.6.4) if $A^T(X^TX)^gA$ is invertible.

Now suppose that $A^T\beta = b$ is ***non-testable*** but $A^T(X^TX)^gA$ is invertible. Derive Q for testing the hypothesis $H_0 : A^TH\beta = b$ and show that it is algebraically identical to the test statistic Q that would be calculated for testing the non-testable hypothesis $H_0 : A^T\beta = b$ using (6.6.4).

7. Consider the one-way-ANOVA model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}; \quad i = 1, 2, 3; j = 1, 2, \ldots n_i.$$

   (a) Develop a test procedure for testing the hypotheseis $H_0 : A^T\beta = 0$ where

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(b) Develop a test procedure for testing the hypothesis $H_0 : A^T\beta = 0$ where

$$A = \begin{pmatrix} 0 & 0 \\ 1 & -1 \\ -1 & -1 \\ 0 & 2 \end{pmatrix}.$$

(c) Compare the test statistics from part (a) and (b). Are they same? If so, why?

(d) If possible, derive a test procedure for testing the hypothesis $H_0 : \alpha_1 = \alpha_2 = 5$.

8. For testing the hypothesis $H_0 : A^T\beta = b$, the test statistic derived in class was

$$F = \frac{Q/q}{SSE/(n-r)} = \frac{n-r}{q}\frac{Q}{SSE}.$$

(a) If $U \sim \chi^2_n(\lambda)$ distribution, what is $E(U)$?

(b) If $V \sim \chi^2_n$ distribution, what is $E(1/V)$?

(c) Use the independence of Q and SSE, and your results from parts (a) and (b) to find $E(F)$.

(d) Show that $E(F) \geq 1$.

9. Silymarin is a milk thistle product frequently used by non-alcoholic

steatohepatitis (NASH) patients. A clinical trial sponsored by National Center for Alternative Medicine investigated the use of 3 doses of silymarin (140mg, 280mg, and 420mg) in relation to a placebo dose. The outcome was measured by the reduction in NASH Activity Score (NAS, a continuous random variable). There were 6 patients in each treatment group. In addition to the doses, analysis of NAS score included two other additional variables (X and Z below) as covariates. Let

Y_{ij} denote the reduction in NAS score for the jth patient in the ith dose group, $i = 0, 1, 2, 3;\ j = 1, .., 6.$

D_i be the silymarin dose for the ith group ($D_0 = 0$ for placebo, $D_1 = 140, \ldots,$ and so on)

α_i be the effect of the ith dose

X_{ij} denote the baseline ALT levels, a continuous variable indicating severity of liver damage

Z_{ij} be the baseline measure of insulin sensitivity (a continuous measure).

Consider the following two linear models:

Model A (treat dose as categorical):

$$Y_{ij} = \mu + \alpha_i + \theta X_{ij} + \gamma Z_{ij} + \epsilon_{ij} \tag{6.6.5}$$

Table 6.1: Results from the analysis of silymarin data

Model A				Model B		
Source	df	Type I SS		Source	df	Type I SS
Intercept	1	107.48		Intercept	1	107.48
Dose	3	30.34		Dose	1	25.09
X	1	3.27		X	1	2.54
Z	1	0.20		Z	1	0.27
Error	18	9.62		Error	20	15.53

Model B (treat dose as continuous):

$$Y_{ij} = \mu + \delta D_i + \theta X_{ij} + \gamma Z_{ij} + \epsilon_{ij} \tag{6.6.6}$$

with standard assumptions of normality and independence for the error terms.

(a) Write the null hypothesis $H_0 : (\alpha_1 - \alpha_0)/D_1 = (\alpha_2 - \alpha_0)/2D_1 = (\alpha_3 - \alpha_0)/3D_1$ as $\mathbf{A}^T\boldsymbol{\beta} = b$.

(b) Show algebraically that under H_0, Model A reduces to Model B.

(c) Suppose you have analyzed the data using SAS and obtained the output from proc GLM (given in Table 6.1). Test the above hypothesis from your output.

10. Suppose $(Y_i, X_i), i = 1, 2, \ldots, n$ be a pair of n independent observations such that the conditional distribution of $Y_i|X_i$ is normal with mean $\mu_1 + \frac{\rho\sigma_1}{\sigma_2(1-\rho)}(X_i - \mu_2)$ and variance σ_1^2 $(\sigma_1^2, \sigma_2^2 > 0, |\rho| < 1)$.

(a) Can you write the problem as a linear model problem? [Hint: Use transformed parameters, for example, define $\beta_0 = \mu_1 - \frac{\rho\sigma_1\mu_2}{\sigma_2(1-\rho)}$].

(b) Is the hypothesis $\mu_1 = 0$ testable?

(c) Is the hypothesis $\rho = 0$ linear in terms of your transformed parameters? If so, is it testable?

(d) For part (b) and (c), if possible, provide the F-statistic for testing the above hypotheses.

11. Consider the simple linear model

$$Y_i = \mu + \alpha\,(-1)^i,\; i = 1, 2, \ldots, 2n-1, 2n, \tag{6.6.7}$$

where Y_i's are independent normal random variables with common variance σ^2.

(a) Show that $U = (Y_2 + Y_1)/2$ and $V = (Y_2 - Y_1)/2$ are unbiased estimators of μ and α, respectively. Can you find separate 95% confidence intervals for μ and α based on U and V?

(b) Find 95% confidence intervals for μ and α based on the least square estimators of μ and α.

(c) Compare the intervals from parts (a) and (b) and comment.

12. Consider the multiple regression model

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \epsilon_i,\; i = 1, 2, \ldots, n.$$

with $E(\epsilon_i) = 0$ and $var(\epsilon_i) = \sigma^2$. Assume ϵ_i's are independent.

(a) Derive a test statistic for testing $H_0 : \beta^{(1)} = 0$, where $\beta^{(1)}$ includes only the first $q(< p)$ components of β.

(b) Discuss two different ways to test the hypothesis $\beta_1 = \beta_0$, a given value of β_1.