STAT 566 HW #4

Maggie Buffum

Problem 7.1

Prove Corollary 7.1.1.

Corollary 7.1.1. When $r(\mathbf{X}) = p$,

(1) $\hat{\beta} \sim N_p(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$

Our model is $y = \beta X + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I_n)$. Because the variability of y comes only from the error term, we have that $y \sim N(X\beta, \sigma^2 I_n)$. Let's consider the distribution of $\hat{\beta}$: $\hat{\beta} = (X'X)^{-1}X'y$, that is, $\hat{\beta}$ is a linear function of y. Let $A = (X'X)^{-1}X'$ such that $\hat{\beta} = Ay$. By Result 5.2.6 we have that

$$\hat{\beta} \sim N_p(A\mu_Y, A\Sigma_Y A')$$

We already know that $\mu_Y = X\beta$ and $\Sigma_Y = \sigma^2 I_n$, so we have

$$A\mu = (X'X)^{-1}X'X\beta = \beta$$

and

$$\begin{aligned}
A\Sigma_Y A' &= (X'X)^{-1}X'\sigma^2 I_n[(X'X)^{-1}X']' \\
&= \sigma^2(X'X)^{-1}X'[(X'X)^{-1}X']' \\
&= \sigma^2(X'X)^{-1}X'X[(X'X)^{-1}]' \\
&= \sigma^2[(X'X)^{-1}]' \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

since $(X'X)^{-1}$ is symmetric. Therefore, $\hat{\beta} \sim N_p(\beta, \sigma^2(X'X)^{-1})$.

(2) For any given vector $\mathbf{a} = (a_1, \ldots, a_p)'$, $\mathbf{a}'\hat{\beta} \sim N(\mathbf{a}'\beta, \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$.

From part (a) we know that $\hat{\beta} \sim N_p(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$. By Result 5.2.7 we know $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$ if and only if every linear combination $\mathbf{a}'\hat{\beta}$ has a univariate $N(\mathbf{a}'\beta, \mathbf{a}'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$ distribution, provided $\mathbf{a} \neq \mathbf{0}$. Thus, $\mathbf{a}'\hat{\beta} \sim N(\mathbf{a}'\beta, \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$ for all $\mathbf{a} \neq \mathbf{0}$.

(3) $\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}/\sigma^2 \sim \chi^2(p, \lambda)$, where $\lambda = \beta'\mathbf{X}'\mathbf{X}\beta/2\sigma^2$.

Consider rewriting such that we have

$$\hat{\beta}'\left(\frac{X'X}{\sigma^2}\right)\hat{\beta}$$

where, from part (a), we know that $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$. Note that

$$\left(\frac{X'X}{\sigma^2}\right)\sigma^2(X'X)^{-1} = X'X(X'X)^{-1} = I_p$$

which is idempotent, so $\frac{\hat{\beta}'X'X\hat{\beta}}{\sigma^2} \sim \chi^2_{(p,\lambda)}$ with $\lambda = \mu'A\mu/2 = \beta'\left(\frac{X'X}{\sigma^2}\right)\beta\frac{1}{2}$.

(4) $Cov(\hat{\beta}, \hat{\epsilon}) = \mathbf{0}$, so that $\hat{\beta}$ is independent of SSE.

Note that if we can show independence between $\hat{\beta}$ and $SSE = \hat{\epsilon}'\hat{\epsilon}$, then $Cov(\hat{\beta}, \hat{\epsilon}) = 0$. From Result 5.4.6 we know that if $y \sim N(X\beta, \sigma^2 I_n)$, then the linear form $B\mathbf{y}$ and the quadratic form $\mathbf{y}'A\mathbf{y}$ are independently distribution if and only if $B\sigma^2 I_n A = 0$. Consider that $\hat{\beta} = (X'X)^{-1}X\mathbf{y}$. If we let

$$B\mathbf{y} = \hat{\beta}\mathbf{y} = (X'X)^{-1}X\mathbf{y}$$

and

$$\mathbf{y}'A\mathbf{y} = SSE = \hat{\epsilon}'\hat{\epsilon} = \mathbf{y}'(I - P_y)\mathbf{y}$$

then

$$\begin{aligned} B\Sigma A &= (X'X)^{-1}X'\sigma^2 I_n(I_n - P_x) \\ &= (X'X)^{-1}X'\sigma^2 I_n - (X'X)^{-1}X'\sigma^2 P_x \\ &= \sigma^2(X'X)^{-1}X' - \sigma^2(X'X)^{-1}X'P_x \\ &= \sigma^2(X'X)^{-1}X' - \sigma^2(X'X)^{-1}X' \\ &= 0 \end{aligned}$$

so $B\mathbf{y} = \hat{\beta}$ and $\mathbf{y}'A\mathbf{y} = SSE = \hat{\epsilon}'\hat{\epsilon}$ are independently distributed, and therefore have a covariance of 0.

(5) $SSE/\sigma^2 = (N-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{N-p}$.

First, we can show that $SSE/\sigma^2 \sim \chi^2_{N-p}$ by showing that $(I - X(X'X)^{-1}X')\sigma^2$ is idempotent:

$$\begin{aligned} SSE &= y'(I - X(X'X)^{-1}X')y \\ \Sigma &= \sigma^2 I \\ &(I - X(X'X)^{-1}X')\sigma^2 I(I - X(X'X)^{-1}X')\sigma^2 I = \sigma^4(I - X(X'X)^{-1}X')(I - X(X'X)^{-1}X') \\ &= \sigma^2(I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X) \\ &= \sigma^4(I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X) \\ &= \sigma^4(I - X(X'X)^{-1}X') \\ &\implies \frac{SSE}{\sigma^2} \sim \frac{1}{\sigma^2}\chi^2_{N-p} \end{aligned}$$

Also, by equation 4.2.18, we have that

$$\hat{\sigma}^2 = \frac{SSE}{N-r} \implies SSE = (N-r)\hat{\sigma}^2 \implies \frac{SSE}{\sigma^2} = \frac{(N-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-p}$$

Problem 7.4

Consider the measurement error model in Exercise 4.3. Assuming that ϵ_is are iid $N(0, \sigma^2)$ variables, derive the distribution of the least squares estimator of the unknown force θ, and the level $(1-\alpha)$ confidence interval for θ.

From Exercise 4.3, we have that $Y_i = \frac{1}{2}t_i^2\theta = \epsilon_i$ for $i = 1, \ldots, N$. The least squares estimate of θ, $\hat{\theta}$, is $\hat{\theta} = (X'X)^{-1}X\mathbf{y}$ where $X = (\frac{1}{2}t_1^2, \ldots, \frac{1}{2}t_N^2)'$. Therefore,

$$X'X = \begin{bmatrix} \frac{1}{2}t_1^2 & \ldots & \frac{1}{2}t_N^2 \end{bmatrix} \begin{bmatrix} \frac{1}{2}t_1^2 \\ \vdots \\ \frac{1}{2}t_N^2 \end{bmatrix} = \frac{1}{4}t_1^4 + \cdots + \frac{1}{4}t_N^4 = \sum_{i=1}^{N} \frac{1}{4}t_i^4$$

and

$$X'\mathbf{y} = \begin{bmatrix} \frac{1}{2}t_1^2 & \dots & \frac{1}{2}t_N^2 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \frac{1}{2}t_1^2 y_1 + \dots + \frac{1}{2}t_N^2 y_N = \sum_{i=1}^{N} \frac{1}{2} t_i^2 y_i$$

Now,

$$\begin{aligned} \hat{\theta} &= \left(\frac{1}{4}t_i^4\right)^{-1} \sum_{i=1}^{N} \frac{1}{2} t_i^2 y_i \\ &= \frac{4\sum_{i=1}^{N} t_i^2 y_i}{2\sum_{i=1}^{N} t_i^4} \\ &= \frac{2\sum_{i=1}^{N} t_i^2 y_i}{\sum_{i=1}^{N} t_i^4} \end{aligned}$$

Clearly X is full rank, so we can use Result 7.1.1 (1) which states that

$$\begin{aligned} \theta^0 &\sim SN(GX'X\theta, \sigma^2 GX'XG') \\ &\sim SN((X'X)^{-1}X'X\theta, \sigma^2 (X'X)^{-1}X'X[(X'X)^{-1}]') \\ &\sim SN(\theta, \sigma^2 (X'X)^{-1}) \\ &\sim SN\left(\theta, \sigma^2 \frac{4}{\sum_{i=1}^{N} t_i^4}\right) \end{aligned}$$

Finally, the $(1-\alpha)$ confidence interval for θ^0 is

$$\begin{aligned} &\theta^0 \pm S \times t_{N-1,\alpha/2} \\ &\frac{2\sum_{i=1}^{N} t_i^2 y_i}{\sum_{i=1}^{N} t_i^4} \pm \sigma^2 \frac{4}{\sum_{i=1}^{N} t_i^4} \times t_{N-1,\alpha/2} \end{aligned}$$

Problem 7.6

In an experiment where several treatments are compared with a control, it may be desirable to replicate the control more than the experimental treatments, since the control enters into every difference investigated. Suppose each of the m experimental treatments is replicated t times while the control is replicated c times. Let Y_{ij} denote the jth observation on the ith experimental treatment, $j = 1, \dots, t$, $i = 1, \dots, m$, and let Y_{0j} denote the jth observation on the control, $j = 1, \dots, c$. Assume that $Y_{ij} = \tau_i + \epsilon_{ij}$, $i = 0, \dots, m$, where ϵ_{ij} are iid $N(0, \sigma^2)$ variables. Find the distribution of the least squares estimate of $\theta_i = \tau_i - \tau_0$, $i = 1, \dots, m$.

The model in matrix form is

$$Y = X\beta + \epsilon = \begin{bmatrix} Y_{01} \\ \vdots \\ Y_{0c} \\ \hline Y_{11} \\ \vdots \\ Y_{1t} \\ \hline \vdots \\ \hline Y_{m1} \\ \vdots \\ Y_{mt} \end{bmatrix}_{(mt+c)\times 1} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \hline 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}_{(mt+c)\times(m+1)} \begin{bmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_m \end{bmatrix} + \epsilon$$

We are trying to estimate

$$\theta = \begin{bmatrix} \tau_1 - \tau_0 \\ \tau_2 - \tau_0 \\ \vdots \\ \tau_m - \tau_0 \end{bmatrix}_{m\times 1} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix}_{m\times(m+1)} \begin{bmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_m \end{bmatrix}_{(m+1)\times 1} = \mathbf{C}'\beta^0$$

If $\theta = \mathbf{C}'\beta^0$ is estimable, and $\mathbf{C}'\mathbf{H} = \mathbf{C}'$, then from Result 7.5.1 we have that

$$\mathbf{C}'\beta^0 \sim N(\mathbf{C}'\beta, \sigma^2\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C})$$

To show estimability, there must exist a matrix of constants $\mathbf{T}$ such that $\mathbf{C}'\beta^0 = \mathbf{C}'(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}\mathbf{y} = \mathbf{T}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y}$, or simply that $\mathbf{C}' = \mathbf{T}'\mathbf{X}$. In matrix form,

$$\mathbf{C}' = \mathbf{T}'\mathbf{X} \implies \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix}_{m\times(m+1)} = [\mathbf{T}]'_{m\times(mt+c)+} \left[\begin{array}{cccc} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \hline 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{array}\right]_{(mt+c)\times(m+1)}$$

By observation we see that $[\mathbf{T}]'$ must be

$$[\mathbf{T}]' = \left[\begin{array}{ccc|ccc|ccc|c|ccc} -1/c & \dots & -1/c & 1/m & \dots & 1/m & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ -1/c & \dots & -1/c & 0 & \dots & 0 & 1/m & \dots & 1/m & \dots & 0 & \dots & 0 \\ \dots & \ddots & \dots & \dots & \ddots & \dots & \dots & \ddots & \dots & \dots & \dots & \ddots & \dots \\ -1/c & \dots & -1/c & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1/m & \dots & 1/m \end{array}\right]$$

and so $\mathbf{C}'\beta^0$ is estimable. Therefore, by Result 7.1.2 we have that

$$\mathbf{C}'\beta^0 \sim N(\mathbf{C}'\beta, \sigma^2\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C})$$

We already know that $\theta = \mathbf{C}'\beta = \tau_i - \tau_0$. To find the variance, let's first find $\mathbf{X}'\mathbf{X}$:

$$\mathbf{X}'\mathbf{X} = \left[\begin{array}{ccc|ccc|c|ccc} 1 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{array}\right] \left[\begin{array}{cccc} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \hline 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{array}\right]$$

$$= \left[\begin{array}{c|ccc} c & 0 & \dots & 0 \\ \hline 0 & t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t \end{array}\right] = \left[\begin{array}{c|c} c & \mathbf{0} \\ \hline \mathbf{0} & t\mathbf{I}_m \end{array}\right]$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} = \left[\begin{array}{c|c} 1/c & \mathbf{0} \\ \hline \mathbf{0} & (1/t)\mathbf{I}_m \end{array}\right]$$

Now we can calculate the variance as

$$\sigma^2\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C} = \sigma^2 \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix} \left[\begin{array}{c|c} 1/c & \mathbf{0} \\ \hline \mathbf{0} & (1/t)\mathbf{I}_m \end{array}\right] \begin{bmatrix} -1 & -1 & \dots & -1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} -(1/c) & 1/t & 0 & \dots & 0 \\ -(1/c) & 0 & 1/t & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -(1/c) & 0 & 0 & \dots & 1/t \end{bmatrix} \begin{bmatrix} -1 & -1 & \dots & -1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1/c + 1/t & 1/c & \dots & 1/c \\ 1/c & 1/c + 1/t & \dots & 1/c \\ \vdots & \vdots & \ddots & \vdots \\ 1/c & 1/c & \dots & 1/c + 1/t \end{bmatrix}$$

$$= \sigma^2(\frac{1}{t}\mathbf{I} + \frac{1}{c}\mathbf{J})$$

such that

$$\theta = \tau_i - \tau_0 \sim N(\tau_i - \tau_0, \sigma^2(\frac{1}{t}\mathbf{I} + \frac{1}{c}\mathbf{J}))$$

Problem 7.7

Suppose $\epsilon_1, \dots, \epsilon_N$ are iid random variables following a normal distribution with mean 0 and variance 1. Suppose $Y_0 = 0$ and $Y_i = \theta Y_{i-1} + \epsilon_i$, $i = 1, \dots, N$, and $|\theta| < 1$. Find the maximum likelihood estimate of θ.

Result 7.5.1 states that for a linear model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ with normally distributed errors, the MLE of β, $\theta = \mathbf{C}'\beta$, is the same as the least squares estimate of β, and in the full-rank case, setting $\mathbf{C} = \mathbf{I}_N$ we have that the MLE of β is $\hat{\beta}_{ML} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Thus, the MLE of θ is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$:

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} Y_0 & Y_1 & \dots & Y_{N-1} \end{bmatrix} \begin{bmatrix} Y_0 \\ Y_1 \\ \dots \\ Y_{N-1} \end{bmatrix} = Y_0^2 + Y_1^2 + \dots + Y_{N-1}^2 = \sum_{i=1}^{N} Y_{i-1}^2$$

and

$$(\mathbf{X}'\mathbf{y}) = \begin{bmatrix} Y_0 & Y_1 & \dots & Y_{N-1} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_N \end{bmatrix} = Y_0 Y_1 + Y_1 Y_2 + \dots + Y_{N-1} Y_N = \sum_{i=1}^{N} Y_{i-1} Y_N$$

Therefore, we have

$$\hat{\theta} = \left(\sum_{i=1}^{N} Y_{i-1}^2 \right)^{-1} \sum_{i=1}^{N} Y_{i-1} Y_N = \frac{\sum_{i=1}^{N} Y_{i-1} Y_N}{\sum_{i=1}^{N} Y_{i-1}^2}$$