Biost/Stat 533
Spring 2008

Homework #3

Problem 1

Let $\mathbf{Y} \sim N(\mu, \sigma^2\mathbf{I})$. Define matrices $\mathbf{A}_1 = \frac{1}{3}\mathbf{J}_3\mathbf{J}_3^T$, $\mathbf{A}_2 = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and $\mathbf{A}_3 == \frac{1}{6}\begin{bmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ -2 & -2 & 4 \end{bmatrix}$.

Define $Q_i = \mathbf{Y}^T\mathbf{A}_i\mathbf{Y}$.
(a) Find distributions of Q_1, Q_2, Q_3.
(b) Prove or disprove: the Q_i are pairwise independent.

Solution: (a) Let $\mu = (\mu_1, \mu_2, \mu_3)^T$. It is easy to check that A_i satisfy that

$$\begin{aligned} A_i^T &= A_i, \\ A_i^2 &= A_i, \\ rank[A_i] &= 1. \end{aligned}$$

Thus we can apply Theorem in page 6 of Lecture 6 to obtain

$$\frac{\mathbf{Y}^T\mathbf{A}_i\mathbf{Y}}{\sigma^2} \sim \chi_1^2\left(\frac{\mu^t\mathbf{A}_i\mu}{2\sigma^2}\right).$$

Because

$$\begin{aligned} \frac{\mu^t\mathbf{A}_1\mu}{2\sigma^2} &= \frac{(\sum_{i=1}^3 \mu_i)^2}{6\sigma^2} \\ \frac{\mu^t\mathbf{A}_2\mu}{2\sigma^2} &= \frac{(\mu_1 - \mu_2)^2}{4\sigma^2} \\ \frac{\mu^t\mathbf{A}_3\mu}{2\sigma^2} &= \frac{(\mu_1 + \mu_2 - 2\mu_3)^2}{12\sigma^2}, \end{aligned}$$

we have

$$\begin{aligned} \mathbf{Y}^T\mathbf{A}_1\mathbf{Y} &\sim \sigma^2\chi_1^2\left(\frac{(\sum_{i=1}^3 \mu_i)^2}{6\sigma^2}\right) \\ \mathbf{Y}^T\mathbf{A}_2\mathbf{Y} &\sim \sigma^2\chi_1^2\left(\frac{(\mu_1 - \mu_2)^2}{4\sigma^2}\right) \\ \mathbf{Y}^T\mathbf{A}_3\mathbf{Y} &\sim \sigma^2\chi_1^2\left(\frac{(\mu_1 + \mu_2 - 2\mu_3)^2}{12\sigma^2}\right) \end{aligned}$$

(b) It is easy to see that

$$\mathbf{A}_1\sigma^2\mathbf{I}\mathbf{A}_2^T = 0, \quad \mathbf{A}_2\sigma^2\mathbf{I}\mathbf{A}_3^T = 0, \quad \mathbf{A}_3\sigma^2\mathbf{I}\mathbf{A}_1^T = 0$$

Then by Theorem 2.5 of Seber & Lee, $\mathbf{A}_i\mathbf{Y}$ and $\mathbf{A}_j\mathbf{Y}, i \neq j, i, j = 1, 2, 3$ are independent. This implies that $(\mathbf{A}_i\mathbf{Y})^T(\mathbf{A}_i\mathbf{Y})$ and $(\mathbf{A}_j\mathbf{Y})^T(\mathbf{A}_j\mathbf{Y}), i \neq j, i, j = 1, 2, 3$ are independent but

$$\begin{aligned}(\mathbf{A}_i\mathbf{Y})^T(\mathbf{A}_i\mathbf{Y}) &= \mathbf{Y}^T\mathbf{A}_i^T\mathbf{A}_i\mathbf{Y} \\ &= \mathbf{Y}^T\mathbf{A}_i\mathbf{A}_i\mathbf{Y} \\ &= \mathbf{Y}^T\mathbf{A}_i\mathbf{Y} \\ &= Q_i.\end{aligned}$$

Thus the Q_i are pairwise independent.

Problem 2

Recall our definition of $\hat{\beta}$: $\hat{\mathbf{Y}}$ is the projection of $\mathbf{Y}$ onto the column space of $\mathbf{X}$ and $\hat{\beta}$ is a vector such that $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$. Show that if $\mathbf{X}$ has a full rank, then

$$(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) = (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta)^T\mathbf{X}^T\mathbf{X}(\hat{\beta} - \beta).$$

and hence deduce that the left side is minimized when $\beta = \hat{\beta}$.

Solution: Because

$$\mathbf{X}^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) = 0,$$

we have

$$\begin{aligned}(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) &= (\mathbf{Y} - \mathbf{X}\hat{\beta} + \mathbf{X}\hat{\beta} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\hat{\beta} + \mathbf{X}\hat{\beta} - \mathbf{X}\beta) \\ &= ((\mathbf{Y} - \mathbf{X}\hat{\beta}) + \mathbf{X}(\hat{\beta} - \beta))^T((\mathbf{Y} - \mathbf{X}\hat{\beta}) + \mathbf{X}(\hat{\beta} - \beta)) \\ &= (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\mathbf{Y} - \mathbf{X}\hat{\beta})^T\mathbf{X}(\hat{\beta} - \beta) \\ &\quad +(\mathbf{X}(\hat{\beta} - \beta))^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\mathbf{X}(\hat{\beta} - \beta))^T(\mathbf{X}(\hat{\beta} - \beta)) \\ &= (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\mathbf{X}^T(\mathbf{Y} - \mathbf{X}\hat{\beta}))^T(\hat{\beta} - \beta) \\ &\quad +(\hat{\beta} - \beta))^T\mathbf{X}^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta))^T\mathbf{X}^T(\mathbf{X}(\hat{\beta} - \beta)) \\ &= (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta))^T\mathbf{X}^T\mathbf{X}(\hat{\beta} - \beta)).\end{aligned}$$

We can write

$$\begin{aligned}(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) &= (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta)^T\mathbf{X}^T\mathbf{X}(\hat{\beta} - \beta) \\ &= \|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2 + \|\mathbf{X}(\hat{\beta} - \beta)\|^2.\end{aligned}$$

Note that $\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2$ does not depend on β and that $\|\mathbf{X}(\hat{\beta} - \beta)\|^2 \geq 0$. Because $\|\mathbf{X}(\hat{\beta} - \beta)\|^2 = 0$ when $\beta = \hat{\beta}$, the claim follows.

Problem 3

Suppose that $\hat{\beta}_1 \neq \hat{\beta}_2$ are two different least squares estimate of β. Show that there are infinitely many least squares estimate of β.

Solution: Let $c \in (0,1)$. Then $c\hat{\beta}_1 + (1-c)\hat{\beta}_2$ is also a least squares estimate of β different from $\hat{\beta}_1$ and $\hat{\beta}_2$ because this quantity satisfies the normal equation:

$$\begin{aligned}
\mathbf{X}^T(\mathbf{Y} - \mathbf{X}(c\hat{\beta}_1 + (1-c)\hat{\beta}_2) &= \mathbf{X}^T\mathbf{Y} - \mathbf{X}^T\mathbf{X}c\hat{\beta}_1 + \mathbf{X}^T\mathbf{X}(1-c)\hat{\beta}_2 \\
&= (c+1-c)\mathbf{X}^T\mathbf{Y} - c\mathbf{X}^T\mathbf{X}\hat{\beta}_1 + (1-c)\mathbf{X}^T\mathbf{X}\hat{\beta}_2 \\
&= c(\mathbf{X}^T\mathbf{Y} - \mathbf{X}^T\mathbf{X}\hat{\beta}_1) + (1-c)(\mathbf{X}^T\mathbf{Y} - \mathbf{X}^T\mathbf{X}\hat{\beta}_2) \\
&= 0.
\end{aligned}$$

Since the choice of c is infinitely many, the claim follows.

Problem 4

Let $\mathbf{P}$ be the projection operator onto $\mathcal{R}(\mathbf{X})$. For least squares estimation, recall that $\hat{\epsilon} = (\mathbf{I} - \mathbf{P})\mathbf{Y}$. Derive
(a) $E[\hat{\epsilon}]$
(b) $\text{cov}(\hat{\epsilon})$
(c) $\text{cov}(\hat{\epsilon}, \mathbf{PY})$
(d) $E[RSS]$.

Solution: Note that $(\mathbf{I} - \mathbf{P})\mathbf{P} = 0$ and that $(\mathbf{I} - \mathbf{P})\mathbf{X} = 0$.
(a) We have

$$\begin{aligned}
E[\hat{\epsilon}] &= E[(\mathbf{I} - \mathbf{P})\mathbf{Y}] \\
&= (\mathbf{I} - \mathbf{P})E[\mathbf{Y}] \\
&= (\mathbf{I} - \mathbf{P})E[\mathbf{X}\beta + \epsilon] \\
&= (\mathbf{I} - \mathbf{P})\mathbf{X}\beta \\
&= 0.
\end{aligned}$$

(b) Noting that $\mathbf{I} - \mathbf{P}$ is idempotent, we have

$$\begin{aligned}
\text{cov}(\hat{\epsilon}) &= \text{cov}((\mathbf{I} - \mathbf{P})\mathbf{Y}) \\
&= (\mathbf{I} - \mathbf{P})\text{cov}(\mathbf{Y})(\mathbf{I} - \mathbf{P})^T \\
&= (\mathbf{I} - \mathbf{P})\sigma^2\mathbf{I})(\mathbf{I} - \mathbf{P}) \\
&= \sigma^2(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) \\
&= \sigma^2(\mathbf{I} - \mathbf{P}).
\end{aligned}$$

(c) We have

$$\begin{aligned}
\text{cov}(\hat{\epsilon}, \mathbf{PY}) &= \text{cov}((\mathbf{I} - \mathbf{P})\mathbf{Y}, \mathbf{PY}) \\
&= (\mathbf{I} - \mathbf{P})\text{cov}(\mathbf{Y}, \mathbf{Y})\mathbf{P}^T \\
&= (\mathbf{I} - \mathbf{P})\sigma^2\mathbf{IP} \\
&= \sigma^2(\mathbf{I} - \mathbf{P})\mathbf{P} \\
&= 0.
\end{aligned}$$

(d) Let $\mathbf{Y} \in \mathbb{R}^n$ and $rank(\mathbf{P}) = p$. Then $rank(\mathbf{I} - \mathbf{P}) = n - p$ and furthermore $\text{tr}(\mathbf{I} - \mathbf{P}) = n - p$. Thus, we have

$$\begin{aligned}
E[RSS] &= E[\hat{\epsilon}^T \hat{\epsilon}] \\
&= E[(\mathbf{I} - \mathbf{P})\mathbf{Y})^T (\mathbf{I} - \mathbf{P})\mathbf{Y})] \\
&= E[\mathbf{Y}^T (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})\mathbf{Y})] \\
&= E[\mathbf{Y}^T (\mathbf{I} - \mathbf{P})\mathbf{Y})] \\
&= \text{tr}((\mathbf{I} - \mathbf{P})\sigma^2 \mathbf{I}) + E[\mathbf{Y}]^T (\mathbf{I} - \mathbf{P}) E[\mathbf{Y}] \\
&= \sigma^2 \text{tr}((\mathbf{I} - \mathbf{P})) + (\mathbf{X}\beta)^T (\mathbf{I} - \mathbf{P})(\mathbf{X}\beta) \\
&= \sigma^2 (n - p).
\end{aligned}$$