Lecture 36: The UMVUE and BLUE

Theorem 3.7. Consider model

$$X = Z\beta + \varepsilon \qquad (1)$$

with assumption A1 (ε is distributed as $N_n(0, \sigma^2 I_n)$ with an unknown $\sigma^2 > 0$).
(i) The LSE $l^\tau \hat{\beta}$ is the UMVUE of $l^\tau \beta$ for any estimable $l^\tau \beta$.
(ii) The UMVUE of σ^2 is $\hat{\sigma}^2 = (n-r)^{-1}\|X - Z\hat{\beta}\|^2$, where r is the rank of Z.
Proof. (i) Let $\hat{\beta}$ be an LSE of β. By $Z^\tau Z b = Z^\tau X$,

$$(X - Z\hat{\beta})^\tau Z(\hat{\beta} - \beta) = (X^\tau Z - X^\tau Z)(\hat{\beta} - \beta) = 0$$

and, hence,

$$\begin{aligned} \|X - Z\beta\|^2 &= \|X - Z\hat{\beta} + Z\hat{\beta} - Z\beta\|^2 \\ &= \|X - Z\hat{\beta}\|^2 + \|Z\hat{\beta} - Z\beta\|^2 \\ &= \|X - Z\hat{\beta}\|^2 - 2\beta^\tau Z^\tau X + \|Z\beta\|^2 + \|Z\hat{\beta}\|^2. \end{aligned}$$

Using this result and assumption A1, we obtain the following joint Lebesgue p.d.f. of X:

$$(2\pi\sigma^2)^{-n/2} \exp\left\{ \frac{\beta^\tau Z^\tau x}{\sigma^2} - \frac{\|x - Z\hat{\beta}\|^2 + \|Z\hat{\beta}\|^2}{2\sigma^2} - \frac{\|Z\beta\|^2}{2\sigma^2} \right\}.$$

By Proposition 2.1 and the fact that $Z\hat{\beta} = Z(Z^\tau Z)^- Z^\tau X$ is a function of $Z^\tau X$, the statistic $(Z^\tau X, \|X - Z\hat{\beta}\|^2)$ is complete and sufficient for $\theta = (\beta, \sigma^2)$.
Note that $\hat{\beta}$ is a function of $Z^\tau X$ and, hence, a function of the complete sufficient statistic.
If $l^\tau \beta$ is estimable, then $l^\tau \hat{\beta}$ is unbiased for $l^\tau \beta$ (Theorem 3.6) and, hence, $l^\tau \hat{\beta}$ is the UMVUE of $l^\tau \beta$.
(ii) From $\|X - Z\beta\|^2 = \|X - Z\hat{\beta}\|^2 + \|Z\hat{\beta} - Z\beta\|^2$ and $E(Z\hat{\beta}) = Z\beta$ (Theorem 3.6),

$$\begin{aligned} E\|X - Z\hat{\beta}\|^2 &= E(X - Z\beta)^\tau (X - Z\beta) - E(\beta - \hat{\beta})^\tau Z^\tau Z(\beta - \hat{\beta}) \\ &= \mathrm{tr}\left(\mathrm{Var}(X) - \mathrm{Var}(Z\hat{\beta})\right) \\ &= \sigma^2[n - \mathrm{tr}\left(Z(Z^\tau Z)^- Z^\tau Z (Z^\tau Z)^- Z^\tau\right)] \\ &= \sigma^2[n - \mathrm{tr}\left((Z^\tau Z)^- Z^\tau Z\right)]. \end{aligned}$$

Since each row of $Z \in \mathcal{R}(Z)$, $Z\hat{\beta}$ does not depend on the choice of $(Z^\tau Z)^-$ in $\hat{\beta} = (Z^\tau Z)^- Z^\tau X$ (Theorem 3.6).
Hence, we can evaluate $\mathrm{tr}((Z^\tau Z)^- Z^\tau Z)$ using a particular $(Z^\tau Z)^-$.
From the theory of linear algebra, there exists a $p \times p$ matrix C such that $CC^\tau = I_p$ and

$$C^\tau (Z^\tau Z) C = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix},$$

where Λ is an $r \times r$ diagonal matrix whose diagonal elements are positive.
Then, a particular choice of $(Z^\tau Z)^-$ is

$$(Z^\tau Z)^- = C \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{pmatrix} C^\tau \tag{2}$$

and

$$(Z^\tau Z)^- Z^\tau Z = C \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} C^\tau$$

whose trace is r.
Hence $\hat\sigma^2$ is the UMVUE of σ^2, since it is a function of the complete sufficient statistic and

$$E\hat\sigma^2 = (n-r)^{-1} E\|X - Z\hat\beta\|^2 = \sigma^2.$$

In general,

$$\mathrm{Var}(l^\tau \hat\beta) = l^\tau (Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) Z (Z^\tau Z)^- l. \tag{3}$$

If $l \in \mathcal{R}(Z)$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$ (assumption A2), then the use of the generalized inverse matrix in (2) leads to $\mathrm{Var}(l^\tau\hat\beta) = \sigma^2 l^\tau (Z^\tau Z)^- l$, which attains the Cramér-Rao lower bound under assumption A1 (Proposition 3.2).

The vector $X - Z\hat\beta$ is called the *residual vector* and $\|X - Z\hat\beta\|^2$ is called the *sum of squared residuals* and is denoted by SSR.
The estimator $\hat\sigma^2$ is then equal to $SSR/(n-r)$.

Since $X - Z\hat\beta = [I_n - Z(Z^\tau Z)^- Z^\tau]X$ and $l^\tau\hat\beta = l^\tau (Z^\tau Z)^- Z^\tau X$ are linear in X, they are normally distributed under assumption A1.
Also, using the generalized inverse matrix in (2), we obtain that

$$[I_n - Z(Z^\tau Z)^- Z^\tau] Z (Z^\tau Z)^- = Z(Z^\tau Z)^- - Z(Z^\tau Z)^- Z^\tau Z (Z^\tau Z)^- = 0,$$

which implies that $\hat\sigma^2$ and $l^\tau\hat\beta$ are independent (Exercise 58 in §1.6) for any estimable $l^\tau\beta$.
Furthermore,

$$[Z(Z^\tau Z)^- Z^\tau]^2 = Z(Z^\tau Z)^- Z^\tau$$

(i.e., $Z(Z^\tau Z)^- Z^\tau$ is a projection matrix) and

$$SSR = X^\tau [I_n - Z(Z^\tau Z)^- Z^\tau] X.$$

The rank of $Z(Z^\tau Z)^- Z^\tau$ is $\mathrm{tr}(Z(Z^\tau Z)^- Z^\tau) = r$.
Similarly, the rank of the projection matrix $I_n - Z(Z^\tau Z)^- Z^\tau$ is $n - r$.
From

$$X^\tau X = X^\tau [Z(Z^\tau Z)^- Z^\tau] X + X^\tau [I_n - Z(Z^\tau Z)^- Z^\tau] X$$

and Theorem 1.5 (Cochran's theorem), SSR/σ^2 has the chi-square distribution $\chi^2_{n-r}(\delta)$ with

$$\delta = \sigma^{-2} \beta^\tau Z^\tau [I_n - Z(Z^\tau Z)^- Z^\tau] Z\beta = 0.$$

Thus, we have proved the following result.

Theorem 3.8. Consider model (1) with assumption A1. For any estimable parameter $l^\tau\beta$, the UMVUE's $l^\tau\hat{\beta}$ and $\hat{\sigma}^2$ are independent; the distribution of $l^\tau\hat{\beta}$ is $N(l^\tau\beta, \sigma^2 l^\tau(Z^\tau Z)^- l)$; and $(n-r)\hat{\sigma}^2/\sigma^2$ has the chi-square distribution χ^2_{n-r}.

Example 3.15. In Examples 3.12-3.14, UMVUE's of estimable $l^\tau\beta$ are the LSE's $l^\tau\hat{\beta}$, under assumption A1. In Example 3.13,

$$SSR = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2;$$

in Example 3.14, if $c > 1$,

$$SSR = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(X_{ijk} - \bar{X}_{ij\cdot})^2.$$

We now study properties of $l^\tau\hat{\beta}$ and $\hat{\sigma}^2$ under assumption A2, i.e., without the normality assumption on ε.
From Theorem 3.6 and the proof of Theorem 3.7(ii), $l^\tau\hat{\beta}$ (with an $l \in \mathcal{R}(Z)$) and $\hat{\sigma}^2$ are still unbiased without the normality assumption.
In what sense are $l^\tau\hat{\beta}$ and $\hat{\sigma}^2$ optimal beyond being unbiased?
We have the following result for the LSE $l^\tau\hat{\beta}$.
Some discussion about $\hat{\sigma}^2$ can be found, for example, in Rao (1973, p. 228).

Theorem 3.9. Consider model (1) with assumption A2.
(i) A necessary and sufficient condition for the existence of a linear unbiased estimator of $l^\tau\beta$ (i.e., an unbiased estimator that is linear in X) is $l \in \mathcal{R}(Z)$.
(ii) (Gauss-Markov theorem). If $l \in \mathcal{R}(Z)$, then the LSE $l^\tau\hat{\beta}$ is the *best linear unbiased estimator* (BLUE) of $l^\tau\beta$ in the sense that it has the minimum variance in the class of linear unbiased estimators of $l^\tau\beta$.
Proof. (i) The sufficiency has been established in Theorem 3.6.
Suppose now a linear function of X, $c^\tau X$ with $c \in \mathcal{R}^n$, is unbiased for $l^\tau\beta$. Then

$$l^\tau\beta = E(c^\tau X) = c^\tau EX = c^\tau Z\beta.$$

Since this equality holds for all β, $l = Z^\tau c$, i.e., $l \in \mathcal{R}(Z)$.
(ii) Let $l \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$.
Then $l = (Z^\tau Z)\zeta$ for some ζ and $l^\tau\hat{\beta} = \zeta^\tau(Z^\tau Z)\hat{\beta} = \zeta^\tau Z^\tau X$ by $Z^\tau Zb = Z^\tau X$.
Let $c^\tau X$ be any linear unbiased estimator of $l^\tau\beta$. From the proof of (i), $Z^\tau c = l$. Then

$$\begin{aligned}
\mathrm{Cov}(\zeta^\tau Z^\tau X, c^\tau X - \zeta^\tau Z^\tau X) &= E(X^\tau Z\zeta c^\tau X) - E(X^\tau Z\zeta\zeta^\tau Z^\tau X) \\
&= \sigma^2\mathrm{tr}(Z\zeta c^\tau) + \beta^\tau Z^\tau Z\zeta c^\tau Z\beta \\
&\quad - \sigma^2\mathrm{tr}(Z\zeta\zeta^\tau Z^\tau) - \beta^\tau Z^\tau Z\zeta\zeta^\tau Z^\tau Z\beta \\
&= \sigma^2\zeta^\tau l + (l^\tau\beta)^2 - \sigma^2\zeta^\tau l - (l^\tau\beta)^2 \\
&= 0.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{Var}(c^\tau X) &= \mathrm{Var}(c^\tau X - \zeta^\tau Z^\tau X + \zeta^\tau Z^\tau X) \\
&= \mathrm{Var}(c^\tau X - \zeta^\tau Z^\tau X) + \mathrm{Var}(\zeta^\tau Z^\tau X) \\
&\quad + 2\mathrm{Cov}(\zeta^\tau Z^\tau X, c^\tau X - \zeta^\tau Z^\tau X) \\
&= \mathrm{Var}(c^\tau X - \zeta^\tau Z^\tau X) + \mathrm{Var}(l^\tau \hat{\beta}) \\
&\geq \mathrm{Var}(l^\tau \hat{\beta}).
\end{aligned}$$