

Stat 709: Mathematical Statistics

Lecture 36

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 36: The UMVUE and BLUE

Theorem 3.7.

Consider model

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

with assumption A1 (ε is distributed as $N_n(0, \sigma^2 I_n)$ with an unknown $\sigma^2 > 0$).

- (i) The LSE $I^r \hat{\beta}$ is the UMVUE of $I^r \beta$ for any estimable $I^r \beta$.
- (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 of (i)

Let $\hat{\beta}$ be an LSE of β . By $Z^r Z b = Z^r X$,

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

and, hence,

Lecture 36: The UMVUE and BLUE

Theorem 3.7.

Consider model

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

with assumption A1 (ε is distributed as $N_n(0, \sigma^2 I_n)$ with an unknown $\sigma^2 > 0$).

- (i) The LSE $I^r \hat{\beta}$ is the UMVUE of $I^r \beta$ for any estimable $I^r \beta$.
- (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 of (i)

Let $\hat{\beta}$ be an LSE of β . By $Z^r Zb = Z^r X$,

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

and, hence,

Proof of (i) continue

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

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 $Z\hat{\beta} = Z(Z^\tau Z)^{-1}Z^\tau X$ 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 $I^\tau \beta$ is estimable, then $I^\tau \hat{\beta}$ is unbiased for $I^\tau \beta$ (Theorem 3.6) and, hence, $I^\tau \hat{\beta}$ is the UMVUE of $I^\tau \beta$.

Proof of (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),

$$\begin{aligned} E\|X - Z\hat{\beta}\|^2 &= E(X - Z\beta)^\tau(X - Z\beta) - E(\beta - \hat{\beta})^\tau Z^\tau Z(\beta - \hat{\beta}) \\ &= \text{tr}(\text{Var}(X) - \text{Var}(Z\hat{\beta})) \\ &= \sigma^2[n - \text{tr}(Z(Z^\tau Z)^{-1}Z^\tau Z(Z^\tau Z)^{-1}Z^\tau)] \\ &= \sigma^2[n - \text{tr}((Z^\tau Z)^{-1}Z^\tau Z)]. \end{aligned}$$

Since each row of $Z \in \mathcal{R}(Z)$, $Z\hat{\beta}$ does not depend on the choice of $(Z^\tau Z)^{-1}$ in $\hat{\beta} = (Z^\tau Z)^{-1}Z^\tau X$ (Theorem 3.6).

- Hence, we can evaluate $\text{tr}((Z^\tau Z)^{-1}Z^\tau Z)$ using a particular $(Z^\tau Z)^{-1}$.

From the theory of linear algebra, there exists a $p \times p$ matrix C such that $CC^\tau = I_p$ and

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

Proof of (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),

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

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 $\text{tr}((Z^\tau Z)^- Z^\tau Z)$ 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

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

Proof of (ii) continue

Then, a particular choice of $(Z^T Z)^-$ is

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

and

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

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

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

Linear combination of LSEs

In general,

$$\text{Var}(I^r \hat{\beta}) = I^r (Z^T Z)^{-1} Z^T \text{Var}(\varepsilon) Z (Z^T Z)^{-1} I. \quad (3)$$

If $I \in \mathcal{R}(Z)$ and $\text{Var}(\varepsilon) = \sigma^2 I_n$ (assumption A2), then the use of the generalized inverse matrix in (2) leads to $\text{Var}(I^r \hat{\beta}) = \sigma^2 I^r (Z^T Z)^{-1} I$, which attains the Cramér-Rao lower bound under assumption A1 (Proposition 3.2).

Residual vector

- 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 $\text{SSR}/(n - r)$.
- Since $X - Z\hat{\beta} = [I_n - Z(Z^T Z)^{-1} Z^T]X$ and $I^r \hat{\beta} = I^r (Z^T Z)^{-1} Z^T X$ are linear in X , they are normally distributed under assumption A1.

Linear combination of LSEs

In general,

$$\text{Var}(I^r \hat{\beta}) = I^r (Z^T Z)^{-1} Z^T \text{Var}(\varepsilon) Z (Z^T Z)^{-1} I. \quad (3)$$

If $I \in \mathcal{R}(Z)$ and $\text{Var}(\varepsilon) = \sigma^2 I_n$ (assumption A2), then the use of the generalized inverse matrix in (2) leads to $\text{Var}(I^r \hat{\beta}) = \sigma^2 I^r (Z^T Z)^{-1} I$, which attains the Cramér-Rao lower bound under assumption A1 (Proposition 3.2).

Residual vector

- 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 $\text{SSR}/(n - r)$.
- Since $X - Z\hat{\beta} = [I_n - Z(Z^T Z)^{-1} Z^T]X$ and $I^r \hat{\beta} = I^r (Z^T Z)^{-1} Z^T X$ are linear in X , they are normally distributed under assumption A1.

Linear combination of LSEs

In general,

$$\text{Var}(I^r \hat{\beta}) = I^r (Z^T Z)^{-1} Z^T \text{Var}(\varepsilon) Z (Z^T Z)^{-1} I. \quad (3)$$

If $I \in \mathcal{R}(Z)$ and $\text{Var}(\varepsilon) = \sigma^2 I_n$ (assumption A2), then the use of the generalized inverse matrix in (2) leads to $\text{Var}(I^r \hat{\beta}) = \sigma^2 I^r (Z^T Z)^{-1} I$, which attains the Cramér-Rao lower bound under assumption A1 (Proposition 3.2).

Residual vector

- 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 $\text{SSR}/(n - r)$.
- Since $X - Z\hat{\beta} = [I_n - Z(Z^T Z)^{-1} Z^T]X$ and $I^r \hat{\beta} = I^r (Z^T Z)^{-1} Z^T X$ are linear in X , they are normally distributed under assumption A1.

- Also, using the generalized inverse matrix in (2), we obtain that

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

which implies that $\hat{\sigma}^2$ and $I^T \hat{\beta}$ are independent (Exercise 58 in §1.6) for any estimable $I^T \beta$.

- Furthermore,

$$[Z(Z^T Z)^- Z^T]^2 = Z(Z^T Z)^- Z^T$$

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

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

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

From

$$X^{\tau}X = X^{\tau}[Z(Z^{\tau}Z)^{-1}Z^{\tau}]X + X^{\tau}[I_n - Z(Z^{\tau}Z)^{-1}Z^{\tau}]X$$

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

$$\delta = \sigma^{-2}\beta^{\tau}Z^{\tau}[I_n - Z(Z^{\tau}Z)^{-1}Z^{\tau}]Z\beta = 0.$$

Thus, we have proved the following result.

Theorem 3.8.

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

From

$$X^{\tau}X = X^{\tau}[Z(Z^{\tau}Z)^{-1}Z^{\tau}]X + X^{\tau}[I_n - Z(Z^{\tau}Z)^{-1}Z^{\tau}]X$$

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

$$\delta = \sigma^{-2}\beta^{\tau}Z^{\tau}[I_n - Z(Z^{\tau}Z)^{-1}Z^{\tau}]Z\beta = 0.$$

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 $N(l^{\tau}\beta, \sigma^2 l^{\tau}(Z^{\tau}Z)^{-1}l)$; and $(n-r)\hat{\sigma}^2/\sigma^2$ has the chi-square distribution χ_{n-r}^2 .

Example 3.15.

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

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

in Example 3.14, if $c > 1$,

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

Non normal assumptions

We now study properties of $I^r \hat{\beta}$ 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), $I^r \hat{\beta}$ (with an $I \in \mathcal{R}(Z)$) and $\hat{\sigma}^2$ are still unbiased without the normality assumption.
- In what sense are $I^r \hat{\beta}$ and $\hat{\sigma}^2$ optimal beyond being unbiased?
- We have the following result for the LSE $I^r \hat{\beta}$.
- 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 $I^T\beta$ (i.e., an unbiased estimator that is linear in X) is $I \in \mathcal{R}(Z)$.
- (ii) (Gauss-Markov theorem). If $I \in \mathcal{R}(Z)$, then the LSE $I^T\hat{\beta}$ is the *best linear unbiased estimator* (BLUE) of $I^T\beta$ in the sense that it has the minimum variance in the class of linear unbiased estimators of $I^T\beta$.

Proof of (i)

The sufficiency has been established in Theorem 3.6.

- Suppose now a linear function of X , $c^T X$ with $c \in \mathcal{R}^n$, is unbiased for $I^T\beta$. Then

$$I^T\beta = E(c^T X) = c^T EX = c^T Z\beta.$$

Since this equality holds for all β , $I = Z^T c$, i.e., $I \in \mathcal{R}(Z)$.

Theorem 3.9

Consider model (1) with assumption A2.

- (i) A necessary and sufficient condition for the existence of a linear unbiased estimator of $I^T\beta$ (i.e., an unbiased estimator that is linear in X) is $I \in \mathcal{R}(Z)$.
- (ii) (Gauss-Markov theorem). If $I \in \mathcal{R}(Z)$, then the LSE $I^T\hat{\beta}$ is the *best linear unbiased estimator* (BLUE) of $I^T\beta$ in the sense that it has the minimum variance in the class of linear unbiased estimators of $I^T\beta$.

Proof of (i)

The sufficiency has been established in Theorem 3.6.

- Suppose now a linear function of X , $c^T X$ with $c \in \mathcal{R}^n$, is unbiased for $I^T\beta$. Then

$$I^T\beta = E(c^T X) = c^T EX = c^T Z\beta.$$

Since this equality holds for all β , $I = Z^T c$, i.e., $I \in \mathcal{R}(Z)$.

Proof of (ii)

Let $l \in \mathcal{R}(Z) = \mathcal{R}(Z^T Z)$.

- Then $l = (Z^T Z)\zeta$ for some ζ and $l^T \hat{\beta} = \zeta^T (Z^T Z)\hat{\beta} = \zeta^T Z^T X$ by $Z^T Zb = Z^T X$.
- Let $c^T X$ be any linear unbiased estimator of $l^T \beta$. From the proof of (i), $Z^T c = l$. Then

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

Proof of (ii) continue

- Hence

$$\begin{aligned}\text{Var}(\mathbf{c}^T \mathbf{X}) &= \text{Var}(\mathbf{c}^T \mathbf{X} - \zeta^T \mathbf{Z}^T \mathbf{X} + \zeta^T \mathbf{Z}^T \mathbf{X}) \\ &= \text{Var}(\mathbf{c}^T \mathbf{X} - \zeta^T \mathbf{Z}^T \mathbf{X}) + \text{Var}(\zeta^T \mathbf{Z}^T \mathbf{X}) \\ &\quad + 2 \text{Cov}(\zeta^T \mathbf{Z}^T \mathbf{X}, \mathbf{c}^T \mathbf{X} - \zeta^T \mathbf{Z}^T \mathbf{X}) \\ &= \text{Var}(\mathbf{c}^T \mathbf{X} - \zeta^T \mathbf{Z}^T \mathbf{X}) + \text{Var}(I^T \hat{\beta}) \\ &\geq \text{Var}(I^T \hat{\beta}).\end{aligned}$$