# SMIII lectures

Week 2

## Estimable functions and best linear unbiased estimates

Consider the linear model formulation

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{E}$$

- Theorems 1.2 and 1.3 (last lecture) describe the sampling distribution of the least squares estimates.
- The least squares estimate $\hat{\boldsymbol{\eta}} = X\hat{\boldsymbol{\beta}}$ is also the best linear unbiased estimate (BLUE) of $\boldsymbol{\eta} = E[\boldsymbol{Y}]$.

*Definition 1.4*

A quantity of the form $\boldsymbol{\lambda}^T\boldsymbol{\eta}$ where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$ is a fixed vector is said to be an **estimable function**.

□

## Examples

Consider the multiple regression model $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$ where the columns of $X$ are assumed to be linearly independent.

- Each individual regression coefficient $\beta_j$ is an estimable function. To see this, take
$$\boldsymbol{\lambda}^T = (0, \ldots, 0, 1, 0, \ldots, 0)(X^TX)^{-1}X^T.$$
- The point prediction, $\beta_0 + \beta_1 x_1 + \ldots + \beta_r x_r$ is an estimable function for any fixed values $x_1, x_2, \ldots, x_r$. To see this, take

$$\boldsymbol{\lambda}^T = (1, x_1, x_2, \ldots, x_r)(X^TX)^{-1}X^T.$$

## The Gauss-Markov Theorem

*Theorem 1.4*

Suppose $E(\boldsymbol{Y}) = \boldsymbol{\eta} = X\boldsymbol{\beta}$ and $\text{Var}(\boldsymbol{Y}) = \sigma^2 I$.

If $\boldsymbol{a}^T\boldsymbol{Y}$ is an unbiased linear estimator for $\boldsymbol{\lambda}^T\boldsymbol{\eta}$, then

$$\text{var}(\boldsymbol{a}^T\boldsymbol{Y}) \geq \text{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\eta}})$$

with equality if and only if

$$\boldsymbol{a} = X(X^T X)^{-1} X^T \boldsymbol{\lambda}.$$

$\square$

**The Gauss-Markov Theorem: proof**

Observe first that for $\boldsymbol{\eta} = X\boldsymbol{\beta}$,

$$E(\boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) = \boldsymbol{\lambda}^T \boldsymbol{\eta} \text{ and } \operatorname{var}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) = \sigma^2 \boldsymbol{\lambda}^T P \boldsymbol{\lambda}.$$

Next, observe that $\boldsymbol{a}^T \boldsymbol{Y}$ unbiased for $\boldsymbol{\lambda}^T \boldsymbol{\eta}$

$$\begin{aligned}
&\Leftrightarrow E(\boldsymbol{a}^T \boldsymbol{Y}) = \boldsymbol{\lambda}^T \boldsymbol{\eta} \\
&\Leftrightarrow (\boldsymbol{a} - \boldsymbol{\lambda})^T \boldsymbol{\eta} = 0 \text{ for all } \boldsymbol{\eta} = X\boldsymbol{\beta} \\
&\Leftrightarrow (\boldsymbol{a} - \boldsymbol{\lambda})^T X = \mathbf{0} \\
&\Leftrightarrow (\boldsymbol{a} - \boldsymbol{\lambda})^T P = \mathbf{0} \text{ where } P = X(X^T X)^{-1} X^T
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{var}(\boldsymbol{a}^T \boldsymbol{Y}) &= \operatorname{var}(\boldsymbol{a}^T \boldsymbol{Y} - \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}} + \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) \\
&= \operatorname{var}(\boldsymbol{a}^T \boldsymbol{Y} - \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) + \operatorname{var}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) \\
&\qquad + 2 \operatorname{cov}(\boldsymbol{a}^T \boldsymbol{Y} - \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}, \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}).
\end{aligned}$$

Now,

$$\begin{aligned}
\operatorname{cov}(\boldsymbol{a}^T \boldsymbol{Y} - \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}, \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) &= \operatorname{cov}\left((\boldsymbol{a}^T - \boldsymbol{\lambda}^T P)\boldsymbol{Y}, \boldsymbol{\lambda}^T P \boldsymbol{Y}\right) \\
&= (\boldsymbol{a}^T - \boldsymbol{\lambda}^T P)\sigma^2 I P \boldsymbol{\lambda} \\
&= \sigma^2 (\boldsymbol{a}^T - \boldsymbol{\lambda}^T) P \boldsymbol{\lambda} = 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
\operatorname{var}(\boldsymbol{a}^T \boldsymbol{Y}) &= \operatorname{var}(\boldsymbol{a}^T \boldsymbol{Y} - \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) + \operatorname{var}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) \\
&\geq \operatorname{var}(\boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}})
\end{aligned}$$

with equality if and only if

$$\operatorname{var}(\boldsymbol{a}^T \boldsymbol{Y} - \boldsymbol{\lambda}^T \hat{\boldsymbol{\eta}}) = 0.$$

That is, if and only if,

$$\boldsymbol{a} = P\boldsymbol{\lambda}.$$

## Inference for multiple regression: regression coefficients

- Theorem 1.3 (last lecture) allows for inference concerning individual regression coefficients.
- Suppose the multiple regression model

  $$M : \boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$$

  with

  $$\boldsymbol{\mathcal{E}} \sim N_n(\mathbf{0}, \sigma^2 I)$$

  holds.

## Pivotal quantity

A **pivotal quantity** for $\beta_j$ is given by

$$\frac{\hat{\beta}_j - \beta_j}{s_e\sqrt{c_{jj}}} \sim t_{n-p}$$

where $c_{jj}$ is the $j$th diagonal element of $(X^T X)^{-1}$.

## Confidence interval for $\beta_j$

The random interval

$$\hat{\beta}_j \pm t_{n-p}(\alpha/2) s_e \sqrt{c_{jj}}$$

is a $100(1-\alpha)\%$ confidence interval for $\beta_j$.

## Hypothesis testing of $\beta_j$

To test $H_0 : \beta_j = 0$ we use the test statistic

$$t = \frac{\hat{\beta}_j - 0}{s_e\sqrt{c_{jj}}}$$

and the rule,

$$\text{Reject } H_0 \text{ for } |t| \geq t_{n-p}(\alpha/2),$$

to define a two-sided test with significance level $\alpha$.

The $P$-value can be calculated in the usual way.

## Testing multiple coefficients

Hypotheses of the form

$$H_0 : \beta_p = \beta_{p-1} = \ldots = \beta_{p_0+1} = 0$$

may be tested using an F-test. The ANOVA table is

| Source | SS | DF | MS | F |
|---|---|---|---|---|
| $H_0$ vs $M$ | $Q(\hat{\boldsymbol{\beta}}_0) - Q(\hat{\boldsymbol{\beta}})$ | $p - p_0$ | $MS_H = \frac{Q(\hat{\boldsymbol{\beta}}_0) - Q(\hat{\boldsymbol{\beta}})}{p - p_0}$ | $F = \frac{MS_H}{MS_E}$ |
| Error | $Q(\hat{\boldsymbol{\beta}})$ | $n - p$ | $MS_E = \frac{Q(\hat{\boldsymbol{\beta}})}{n-p}$ | |
| Total | $Q(\hat{\boldsymbol{\beta}}_0)$ | $n - p_0$ | | |

- If $H_0$ is true then $F$ has the F-distribution, $F \sim F_{p-p_0,n-p}$ so we reject $H_0$, with significance level $\alpha$, if $F \geq F_{p-p_0,n-p}(\alpha)$.

- $\hat{\boldsymbol{\beta}}_0$ is obtained by applying least squares calculation to the reduced regression model

$$\boldsymbol{Y} = X_0\boldsymbol{\beta}_0 + \boldsymbol{\mathcal{E}}$$

where $X_0$ is the $n \times p_0$ matrix comprising the first $p_0$ columns of $X$. That is, we have $\hat{\boldsymbol{\beta}}_0 = (X_0^T X_0)^{-1} X_0^T \boldsymbol{Y}$. - It can be proved that the F-statistic is given equivalently by

$$F = \frac{\hat{\boldsymbol{\beta}}_1^T C_{11}^{-1} \hat{\boldsymbol{\beta}}_1}{(p - p_0)s_e^2}$$

where $\hat{\boldsymbol{\beta}}_1^T = (\hat{\beta}_{p_0+1}, \ldots, \hat{\beta}_p)$ and $C_{11}$ is the lower right $(p-p_0)\times(p-p_0)$ block of $(X^TX)^{-1}$. - This form of the F-statistic reduces to $F = t^2$ when the test concerns a single parameter, $H_0 : \beta_p = 0$.

## Prediction in multiple regression

Consider the multiple regression model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\mathcal{E}} \text{ with } \boldsymbol{\mathcal{E}} \sim N_n(\boldsymbol{0}, \sigma^2 I).$$

- Suppose we have a new observation $(Y_0, \boldsymbol{x}_0)$
- with $Y_0 \sim N(\boldsymbol{x}_0^T\boldsymbol{\beta}, \sigma^2)$ (independently from $\boldsymbol{Y}$).

## A $100(1-\alpha)\%$ confidence interval for $\eta_0 = \boldsymbol{x}_0^T\boldsymbol{\beta}$

The $100(1-\alpha)\%$ confidence interval for $\eta_0$ is therefore

$$\hat{\eta}_0 \pm t_{n-p}(\alpha/2)s_e\sqrt{\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0}$$

It follows from the Gauss-Markov theorem that $\hat{\eta}_0 = \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}}$ is the BLUE for $\eta_0$. Moreover, since

$$\hat{\boldsymbol{\beta}} \sim N_p\left(\boldsymbol{\beta}, \sigma^2(X^TX)^{-1}\right),$$

it can be shown that

$$\hat{\eta}_0 \sim N(\eta_0, \sigma^2\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0)$$

independently of $s_e^2$. The pivotal quantity is

$$\frac{\hat{\eta}_0 - \eta_0}{s_e\sqrt{\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0}} \sim t_{n-p}.$$

## A $100(1-\alpha)\%$ prediction interval for $Y_0$

The $100(1-\alpha)\%$ prediction interval for $Y_0$ is therefore

$$\hat{\eta}_0 \pm t_{n-p}(\alpha/2)s_e\sqrt{1+\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0}$$

Since $Y_0 \sim N\left(\eta_0, \sigma^2\right)$ independently of $\boldsymbol{Y}$ it follows that $Y_0$ is independent of $\hat{\eta}_0$ so that

$$Y_0 - \hat{\eta}_0 \sim N\left(0, \sigma^2(1+\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0\right)$$

and hence

$$\frac{Y_0 - \hat{\eta}_0}{s_e\sqrt{1+\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0}} \sim t_{n-p}.$$

**Remarks**

The confidence interval is simply a confidence interval for the estimable function

$$\eta_0 = \boldsymbol{x}_0^T\boldsymbol{\beta}.$$

- The prediction interval is different in nature in the sense that it is predicting a likely range of values for the random variable $Y_0$.

If the parameters $\boldsymbol{\beta}$ and $\sigma^2$ were known without error, it would still be relevant to consider the question of prediction and resulting interval would simply be

$$\eta_0 \pm z(\alpha/2)\sigma.$$

The interval

$$\hat{\eta}_0 \pm t_{n-p}(\alpha/2)s_e\sqrt{1+\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0}$$

differs from this simple form

$$\eta_0 \pm z(\alpha/2)\sigma$$

because the parameters have been replaced by estimates and adjustments to allow for this fact have been made.

In particular, we use

- $t_{n-p}(\alpha/2)$ instead of $z(\alpha/2)$
- to allow for the fact that we are using $s_e$ in place of $\sigma$.
- The term $\boldsymbol{x}_0^T(X^TX)^{-1}\boldsymbol{x}_0$ is introduced
- to allow for errors in the estimation of $\eta_0$ by $\hat{\eta}_0$.

## Case study

Rubber example