BLUE

Definition 1.4

An estimator of the form:

$$T = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n = \sum_{i=1}^{n} a_i Y_i \,,$$

for some constants $a_1, a_2, \dots, a_n$ is called a *linear estimator*.

e.g. $T = Y_1 + Y_2 + \dots + Y_n$ is a linear estimator.

$T = Y_1$ is also a linear estimator.

e.g. $T = Y_1^2$ is not a linear estimator.

Example 1.5

Is the sample mean $\bar{Y}$ a linear estimator?

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \boxed{\frac{1}{n}} Y_i$$

$a_i = \frac{1}{n}$

So $\bar{Y}$ is a linear estimator.

Lemma 1

Suppose $Y_1, Y_2, \dots, Y_n$ are independent random variables with

$$E[Y_i] = \mu_i \text{ and } \operatorname{var}(Y_i) = \sigma_i^2.$$

Let

$$T = \sum_{i=1}^{n} a_i Y_i ,$$

then

$$\boxed{E[T] = \sum_{i=1}^{n} a_i \mu_i} \text{ and } \boxed{\operatorname{var}(T) = \sum_{i=1}^{n} a_i^2 \sigma_i^2}.$$

Furthermore, if

$Y_i \sim N(\mu_i, \sigma_i^2)$ independently, then

$$\boxed{T \sim N\left(\sum_{i=1}^{n} a_i \mu_i \, , \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).}$$

Definition 1.5

The *best Linear Unbiased Estimator* (BLUE) for a parameter θ is the linear, unbiased estimator for θ that has minimum variance.

An estimator T is the BLUE if

① T is unbiased, i.e. $E(T) = \theta$

② T is linear, i.e. $T = \sum_{i=1}^{n} a_i Y_i$

③ T has minimum variance among all unbiased linear estimators of θ,
i.e. $var(T) \leq var(T')$ for all unbiased linear estimators T'

Lemma 2

Suppose $Y_1, Y_2, \ldots, Y_n$ are IID random variables with

$$E[Y_i] = \mu \text{ and } \mathrm{var}(Y_i) = \sigma^2.$$

The linear estimator

$$T = \sum_{i=1}^{n} a_i Y_i$$

is unbiased for μ if and only if

$$\boxed{\sum_{i=1}^{n} a_i = 1.}$$

Proof of Lemma 2

T is unbiased if $E[T] = \mu$.

$$E[T] = E\left[\sum_{i=1}^{n} a_i Y_i\right]$$

$$= \sum_{i=1}^{n} a_i E[Y_i]$$

$$= \sum_{i=1}^{n} a_i \mu$$

$$= \mu \sum_{i=1}^{n} a_i$$

We want $\mu = \mu \sum_{i=1}^{n} a_i$

$\Rightarrow \quad \underline{1} = \sum_{i=1}^{n} a_i$

Theorem 3

Suppose $Y_1, Y_2, \dots, Y_n$ are IID random variables with

$$E[Y_i] = \mu \text{ and } \operatorname{var}(Y_i) = \sigma^2,$$

then the BLUE for μ is given by

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i\,.$$

Proof of Theorem 3

Recall that $\text{var}(\bar{Y}) = \frac{\sigma^2}{n}$.

Let T be an unbiased linear estimator for μ.

So $E(T) = \mu$ and we can write $T = \sum_{i=1}^{n} a_i Y_i$.

$$
\begin{aligned}
\text{var}(T) &= \sum_{i=1}^{n} a_i^2 \sigma^2 \\
&= \sigma^2 \sum_{i=1}^{n} a_i^2 \\
&= \sigma^2 \sum_{i=1}^{n} \left(\underbrace{a_i - \tfrac{1}{n}} + \underbrace{\tfrac{1}{n}} \right)^2 \\
&= \sigma^2 \sum_{i=1}^{n} \left[\left(a_i - \tfrac{1}{n}\right)^2 + 2\tfrac{1}{n}\left(a_i - \tfrac{1}{n}\right) + \left(\tfrac{1}{n}\right)^2 \right] \\
&= \sigma^2 \left[\sum_{i=1}^{n} \left(a_i - \tfrac{1}{n}\right)^2 + \tfrac{2}{n} \underbrace{\sum_{i=1}^{n} \left(a_i - \tfrac{1}{n}\right)} + \tfrac{1}{n} \right]
\end{aligned}
$$

$$
\sum_{i=1}^{n} \left(a_i - \tfrac{1}{n}\right) = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n} \frac{1}{n} = \underbrace{\sum_{i=1}^{n} a_i}_{=1 \text{ by Lemma 2 since } T \text{ is unbiased}} - 1 = 0
$$

Proof of Theorem 3 (cont.)

$$\operatorname{var}(T) = \sigma^2 \left[\sum_{i=1}^{n} \left(a_i - \frac{1}{n}\right)^2 + \frac{1}{n} \right]$$

$$= \sigma^2 \sum_{i=1}^{n} \left(a_i - \frac{1}{n}\right)^2 + \frac{\sigma^2}{n}$$

$$= \underbrace{\sigma^2 \sum_{i=1}^{n} \left(a_i - \frac{1}{n}\right)^2}_{\geq 0} + \operatorname{var}(\bar{Y})$$

$$\geq \operatorname{var}(\bar{Y})$$

Any linear unbiased estimator T of μ has

$$\operatorname{var}(T) \geq \operatorname{var}(\bar{Y})$$

with equality if and only if $a_i = \frac{1}{n}$ for all i, that is, $T = \bar{Y}$.

Definition 1.6

If T is an unbiased estimator for θ, then the standard deviation of the estimator is called the *standard error*:

$$SE(T) = \sqrt{var(T)}.$$