- To be more general, we'll derive properties of a linear combination of $\widehat{\boldsymbol{\beta}}$ (rather than $\widehat{\boldsymbol{\beta}}$ itself)
- We can obtain a single element of $\boldsymbol{\beta}$ by appropriately setting the coefficients of the linear combination of $\boldsymbol{\beta}$
- Contrasts is a special type of linear combination of $\boldsymbol{\beta}$, where the coefficients of this linear combination sums to 0

Multiple linear regression: Statistical Properties of LSE

- Contrasts are useful for comparing different group means
- For example, we regress wine quality on 4 different types of wine
  - Two of the wines are red wine, while the other two are white wines
  - Using contrasts, we can setup a test compare within white wine group only
  - Similarly, we can compare within the red wine group only
  - We can also use contrasts to set up a test to test whether there is a difference between red and white wines

Lemma 7

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent with $E(Y_i) = \eta_i$ and $var(Y_i) = \sigma^2$. Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix}, \text{ and } \boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

and let $V = \boldsymbol{a}^\top \boldsymbol{Y}$. Then

$$E(V) = \boldsymbol{a}^\top \boldsymbol{\eta}, \quad = \sum_{i=1}^{n} a_i \eta_i$$

$$var(V) = \sigma^2 \boldsymbol{a}^\top \boldsymbol{a}. \quad = \sigma \sum_{i=1}^{n} a_i^2$$

If, furthermore, $Y_i \sim N(\eta_i, \sigma^2)$ independently, then

$$V \sim N(\boldsymbol{a}^\top \boldsymbol{\eta}, \sigma^2 \boldsymbol{a}^\top \boldsymbol{a})$$

Proof of Lemma 7

$$
\begin{aligned}
E[\boldsymbol{a}^T\boldsymbol{Y}] &= E\left[\sum_{i=1}^{n} a_i Y_i\right] \\
&= \sum_{i=1}^{n} a_i E[Y_i] \\
&= \sum_{i=1}^{n} a_i \eta_i \\
&= \boldsymbol{a}^T\boldsymbol{\eta}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(\boldsymbol{a}^T\boldsymbol{Y}) &= \mathrm{var}\left(\sum_{i=1}^{n} a_i Y_i\right) \\
&= \sum_{i=1}^{n} a_i^2 \mathrm{var}(Y_i) \qquad \text{by independence} \\
&= \sum_{i=1}^{n} a_i^2 \sigma^2 \\
&= \sigma^2 \sum_{i=1}^{n} a_i^2 \\
&= \sigma^2 \boldsymbol{a}^T\boldsymbol{a}.
\end{aligned}
$$

Theorem 11

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent with $E(Y_i) = \eta_i$ and $var(Y_i) = \sigma^2$, where

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} = \boldsymbol{X}\boldsymbol{\beta}.$$

and where $\boldsymbol{X}$ is an $x \times p$ matrix with linearly independently columns and let $\boldsymbol{\lambda}$ be a constant vector, then, consider $\lambda^T \hat{\beta}$, where $\hat{\beta} = (X^T X)^{-1} X^T Y$.

1. $E(\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}^\top \boldsymbol{\beta}$.
2. $var(\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}$.
3. $E(s_e^2) = \sigma^2$.
4. If, furthermore, $Y_i \sim N(\eta_i, \sigma^2)$, then

$$\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\lambda}^\top \boldsymbol{\beta}, \sigma^2 \boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}\right) \text{ and } \frac{(n-p)s_e^2}{\sigma^2} \sim \chi^2_{n-p}$$

independently.

Proof of Theorem 11

1 If $E(Y)=\eta$, then $E[a^T Y] = a^T\eta$ by Lemma 7.

$$E[\lambda^T\hat{\beta}] = E[\underbrace{\boxed{\lambda^T(X^TX)^{-1}X^T}}_{a^T}Y]$$

$$= E[a^TY]$$

$$= a^T\eta$$

$$= \lambda^T \cancel{(X^TX)^{-1}} \cancel{X^T} \overbrace{(X\beta)}^{E(Y)}$$

$$= \lambda^T\beta$$

2 If $var(Y)=\sigma^2 I$, then $var(a^TY) = \sigma^2 a^Ta$ by Lemma 7.

$$var(\lambda^T\hat{\beta}) = var(\underbrace{\boxed{\lambda^T(X^TX)^{-1}X^T}}_{a^T}Y)$$

$$= \sigma^2 a^Ta$$

$$= \sigma^2(\lambda^T \cancel{(X^TX)^{-1}X^T})(\cancel{X}(X^TX)^{-1}\lambda)$$

$$= \sigma^2\lambda^T(X^TX)^{-1}\lambda$$

Proof of Theorem 11

3 $\quad S_e^2 = \frac{1}{n-p} Y^T \underbrace{(I-H)} Y$ from Example 3.4

$E[x^TAx] = tr(A\Sigma) + \mu^T A \mu$ if $E(x) = \mu$ and $var(x) = \Sigma$

(We will prove this in the next tutorial.)

$$E[S_e^2] = \frac{1}{n-p} E[Y^T (I-H) Y] \quad (\text{Let } x = Y,\ A = I-H)$$

$$= \frac{1}{n-p}\left[tr\left((I-H)\sigma^2\right) + E(Y)^T (I-H) E(Y)\right]$$

$$= \frac{1}{n-p}\left[\sigma^2\, tr(I-H) + (X\beta)^T \underbrace{(I-H)(X\beta)}_{=0}\right] \quad \text{see proof of Theorem 10}$$

$$= \frac{\sigma^2}{n-p} tr(I-H)$$

$$= \frac{\sigma^2}{n-p}\left[tr(I) - tr(H)\right]$$

$$= \frac{\sigma^2}{\cancel{n-p}}\left[\cancel{n-p}\right]$$

$$= \sigma^2$$

$tr(A+B) = tr(A) + tr(B)$

$tr(AB) = tr(BA)$

$tr(H) = tr(X(X^TX)^{-1}X^T) = tr(\cancel{X^TX}\,\cancel{(X^TX)^{-1}}) = tr(I) = p$

BLUE for Multiple Linear Regression

Gauss-Markov Theorem

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent observations with $E(Y_i) = \eta_i$ and $var(Y_i) = \sigma^2$. Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \text{ and } \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix}$$

and suppose $\boldsymbol{\eta} = \boldsymbol{X\beta}$, where $\boldsymbol{X}$ is an $n \times p$ matrix whose columns are linearly independent.

Gauss-Markov Theorem (cont.)

If $\boldsymbol{a}^\top \boldsymbol{Y}$ is an unbiased linear estimator for $\boldsymbol{\lambda}^\top \boldsymbol{\beta}$ then

$$var(\boldsymbol{a}^\top \boldsymbol{Y}) \geq var\left(\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}}\right)$$

with equality if and only if

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

This means that $\lambda^T \hat{\beta}$ will have the smallest variance among all the unbiased linear estimator for $\lambda^T \beta$. So $\lambda^T \hat{\beta}$ is the BLUE for $\lambda^T \beta$.

Proof: Omitted here. It will be covered in Statistical Modelling III.