Distributional Results

The random linear model:

option 1:

$$y_i = \sum_{j=1}^{p} z_{ij}\beta_j + \varepsilon_i \,, \quad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2) \quad i=1,\dots,n$$

option 2:

$$y = Z\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

$$y = (y_1, \dots, y_n)^T$$

$$Z = \begin{bmatrix} z_{11}, & \dots, & z_{1p} \\ \vdots & & \vdots \\ z_{n1} & \dots & z_{np} \end{bmatrix} \qquad \begin{aligned} \beta &= (\beta_1, \dots, \beta_p)^T \\ \varepsilon &= (\varepsilon_1, \dots, \varepsilon_n)^T \end{aligned}$$

Option 3: $\quad y \sim N(Z\beta, \sigma^2 I_n)$

Unbiasedness of $\hat{\beta}$

$$y = Z\beta + \varepsilon$$

ε_j are random

Z is full rank

we have:

$$\begin{aligned} \hat{\beta} &= (Z^T Z)^{-1} Z^T y \\ &= (Z^T Z)^{-1} Z^T (Z\beta + \varepsilon) \\ &= \beta + (Z^T Z)^{-1} Z^T \varepsilon \end{aligned}$$

$$\mathbb{E}[\hat{\beta}] = \beta + (Z^T Z)^{-1} Z^T \mathbb{E}[\varepsilon]$$

Conclusion: $\hat{\beta}$ is an <u>unbiased</u> estimator of β as long as $\mathbb{E}[\varepsilon]=0$, i.e. $\mathbb{E}[\hat{\beta}]=\beta$

<u>Variance $\hat{\beta}$</u> $\qquad Var(\varepsilon)=\sigma^2 I$

$$\hat{\beta}=\beta+(Z^TZ)^{-1}Z^T\varepsilon$$

$$\begin{aligned} Var[\hat{\beta}] &= Var\left((Z^TZ)^{-1}Z^T\varepsilon\right) \\ &= (Z^TZ)^{-1}Z^T Var(\varepsilon) Z (Z^TZ)^{-1} \\ &= (Z^TZ)^{-1}Z^T\sigma^2 I\, Z(Z^TZ)^{-1} \\ &= \sigma^2(Z^TZ)^{-1} \end{aligned}$$

Theorem.

Suppose that $y \sim N(Z\beta, \sigma^2 I)$ and that $(Z^TZ)^{-1}$ is invertible.

- $\hat{\beta} \sim N(\beta, \sigma^2(Z^TZ)^{-1})$
- $\hat{y}=Z\hat{\beta} \sim N(Z\beta, H\sigma^2)$
- $\hat{\varepsilon}=y-\hat{y} \sim N(0, (I-H)\sigma^2)$

Furthermore, $\hat{\varepsilon}$ is independent of $\hat{\beta}$ & $\hat{y}$

Proof.

$$Var(y) = Var(\varepsilon) = \sigma^2 I$$

$$H = Z\beta = Z(ZZ^T)^{-1}Z^T$$

$$cov(\hat{y}, \hat{\varepsilon}) = Cov(Hy, (I-H)y)$$

$$= H\, cov(y,y)\, (I-H)^T$$

$Var(y,y)$

$$= H \sigma^2 I (I-H) = \sigma^2 (H - H^2)$$

$$= 0 \qquad b.c.\ H^2 = H$$

uncorrelatedness $\Rightarrow$ independence

$$0 = cov(\hat{\varepsilon}, \hat{y}) = cov(\hat{\varepsilon}, Z\hat{\beta}) = cov(\hat{\varepsilon}, \hat{\beta}) Z^T$$

$$\Rightarrow \quad 0 = cov(\hat{\varepsilon}, \hat{\beta})\, Z^T \left(Z (Z^T Z)^{-1}\right)$$

$$\Rightarrow \quad 0 = cov(\hat{\varepsilon}, \hat{\beta}) \qquad \square$$

- The fact $cov(\hat{\varepsilon}, \hat{y}) = 0$ is sometimes referred to as the "orthogonality principle"
"the optimal linear estimator is uncorrelated with the residuals"
when things are Gaussian, we can replace uncorrelated with independent.