

Distributional Results

The random linear model:

option 1:

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

option 2:

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

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

$$Z = \begin{bmatrix} z_{11} & \dots & z_{1p} \\ \vdots & & \vdots \\ z_{n1} & \dots & z_{np} \end{bmatrix}$$

$$\beta = (\beta_1, \dots, \beta_p)^T$$
$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$$

option 3:

$$y \sim N(Z\beta, \sigma^2 I_n)$$

Unbiasedness of $\hat{\beta}$

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

ε_j are random
 Z is full rank

we have:

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

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

Conclusion: $\hat{\beta}$ is an unbiased estimator of β
as long as $E(\varepsilon) = 0$, i.e. $E(\hat{\beta}) = \beta$

Variance $\hat{\beta}$ $\text{Var}(\varepsilon) = \sigma^2 I$

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

$$\begin{aligned}\text{Var}(\hat{\beta}) &= \text{Var}((Z^T Z)^{-1} Z^T \varepsilon) \\ &= (Z^T Z)^{-1} Z^T \text{Var}(\varepsilon) Z (Z^T Z)^{-1} \\ &= (Z^T Z)^{-1} Z^T \sigma^2 I Z (Z^T Z)^{-1} \\ &= \sigma^2 (Z^T Z)^{-1}\end{aligned}$$

Theorem.

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

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

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

Proof.

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

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

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

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

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

$$= 0 \quad \text{b.c. } H^2 = H$$

Uncorrelatedness \Rightarrow independence

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

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

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

- The fact $\text{cov}(\hat{\varepsilon}, \hat{y}) = 0$ 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.