6 Properties of Least Squares Estimates

6.1 Note: The basic distributional assumptions of the linear model are

(a) The errors are unbiased: $E[\boldsymbol{\varepsilon}] = \mathbf{0}$.

(b) The errors are uncorrelated with common variance: $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$.

These assumptions imply that $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ and $\text{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$.

6.2 Theorem: If $\mathbf{X}$ is of full rank, then

(a) The least squares estimate is unbiased: $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$.

(b) The covariance matrix of the least squares estimate is $\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

6.3 Theorem: Let $\text{rank}(\mathbf{X}) = r < p$ and $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$, where $(\mathbf{X}'\mathbf{X})^{-}$ is a generalized inverse of $\mathbf{X}'\mathbf{X}$.

(a) $\mathbf{P}$ and $\mathbf{I} - \mathbf{P}$ are projection matrices.

(b) $\text{rank}(\mathbf{I} - \mathbf{P}) = \text{tr}(\mathbf{I} - \mathbf{P}) = n - r$.

(c) $\mathbf{X}'(\mathbf{I} - \mathbf{P}) = \mathbf{0}$.

6.4 Note: In general, $\hat{\boldsymbol{\beta}}$ is not unique so we consider the properties of $\hat{\boldsymbol{\mu}}$, which is unique. It is an unbiased estimate of the mean vector $\boldsymbol{\mu} = E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$:

$$E[\hat{\boldsymbol{\mu}}] = E[\mathbf{P}\mathbf{Y}] = \mathbf{P}E[\mathbf{Y}] = \mathbf{P}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\mu},$$

since $\mathbf{P}\mathbf{X} = \mathbf{X}$ by Theorem 6.3 (c).

6.5 Theorem: Let $\hat{\boldsymbol{\mu}}$ be the least-squares estimate. For any linear combination $\mathbf{c}'\boldsymbol{\mu}$, $\mathbf{c}'\hat{\boldsymbol{\mu}}$ is the unique estimate with minimum variance among all linear unbiased estimates.

6.6 Note: The above shows that $\hat{\boldsymbol{\mu}}$ is optimal in the sense of having minimum variance among all linear estimators. This result is the basis of the Gauss-Markov theorem on the estimation of estimable functions in ANOVA models, which we will study in a later lecture.

6.7 Note: We call $\mathbf{c}'\hat{\boldsymbol{\mu}}$ the Best Linear Unbiased Estimate (BLUE) of $\mathbf{c}'\boldsymbol{\mu}$.

6.8 Theorem: If rank$(\mathbf{X}_{n\times p}) = p$, then $\mathbf{a}'\hat{\boldsymbol{\beta}}$ is the BLUE of $\mathbf{a}'\boldsymbol{\beta}$ for any $\mathbf{a}$.

6.9 Note: The Gauss-Markov theorem will generalize the above to the less than full rank case, for the set of estimable linear combinations $\mathbf{a}'\boldsymbol{\beta}$.

6.10 Definition: Let rank$(\mathbf{X}) = r$. Define

$$S^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-r) = RSS/(n-r).$$

This is a generalization of the sample variance.

6.11 Theorem: S^2 is an unbiased estimate of σ^2.

6.12 Note: If we assume that $\boldsymbol{\varepsilon}$ has a multivariate normal distribution in addition to the assumptions $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, i. e. if we assume $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$, we have $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.

6.13 Theorem: Let $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where rank$(\mathbf{X}_{n\times p}) = p$. Then

(a) $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$,

(b) $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma^2 \sim \chi^2_p$,

(c) $\hat{\boldsymbol{\beta}}$ is independent of S^2,

(d) $RSS/\sigma^2 = (n-p)S^2/\sigma^2 \sim \chi^2_{n-p}$.