Properties of least squares estimates

If $Y = X\beta + \epsilon$, with $E(\epsilon) = 0$ and $Cov(\epsilon) = \sigma^2 I_n$. then $E(\hat{\beta}) = \beta$ since

$$\begin{aligned} E(\hat{\beta}) &= E((X'X)^{-1}X'Y) = (X'X)^{-1}X'E(Y) \\ &= (X'X)^{-1}X'X\beta = \beta, \text{ and} \\ Cov(\hat{\beta}) &= Cov((X'X)^{-1}X'Y) = (X'X)^{-1}X'Cov(Y)X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}. \end{aligned}$$

Theorem (Gauss-Markov). Consider the Gauss-Markov model, $Y = X\beta + \epsilon$, with $E(\epsilon) = 0$ and $Cov(\epsilon) = \sigma^2 I_n$. Let $\hat{\theta}$ be the least squares estimate of $\theta = X\beta$. Fix $c \in \mathcal{R}^p$ and consider estimating $c'\theta$. Then, in the class of all linear unbiased estimates of $c'\theta$, $c'\hat{\theta}$ is the unique estimate with minimum variance. (Thus $c'\hat{\theta}$ is BLUE of $c'\theta$.)

Proof. $\hat{\theta} = X\hat{\beta} = PY$, where P is the projection matrix onto $\mathcal{M}_C(X)$. In particular, $PX = X$. Therefore,

$$E(c'\hat{\theta}) = c'E(PY) = c'PE(Y) = c'PX\beta = c'X\beta = c'\theta,$$

so that $c'\hat{\theta} = PY$ is a linear unbiased estimate of $c'\theta$. Let $d'Y$ be any other linear unbiased estimate of $c'\theta$. Then $c'\theta = E(d'Y) = d'\theta$, or $(c-d)'\theta = 0$ for all $\theta \in \mathcal{M}_C(X)$. i.e., $(c-d)$ is orthogonal to $\mathcal{M}_C(X)$. Therefore $P(c-d) = 0$, and so $Pc = Pd$. Now,

$$\begin{aligned} Var(d'Y) - Var(c'\hat{\theta}) &= Var(d'Y) - Var(c'PY) \\ &= Var(d'Y) - Var(d'PY) \\ &= \sigma^2(d'd - d'P^2d) = \sigma^2(d'd - d'Pd) \\ &= \sigma^2 d'(I-P)d = \sigma^2 d'(I-P)(I-P)d \\ &\geq 0 \end{aligned}$$

with equality iff $(I-P)d = 0$ or $d = Pd = Pc$. i.e., $d'Y = c'PY = c'\hat{\theta}$.

Remark. Since we have assumed that X has full column rank, $P = X(X'X)^{-1}X'$ and so, if $\theta = X\beta$, then $X'\theta = X'X\beta$ or $\beta = (X'X)^{-1}X'\theta$. Therefore, for every $a \in \mathcal{R}^p$, $a'\beta = a'(X'X)^{-1}X'\theta = c'\theta$, where $c = X(X'X)^{-1}a$. i.e., every linear function of β is a linear function of θ. Therefore, for every $a \in \mathcal{R}^p$, we have that $a'\hat{\beta} = a'(X'X)^{-1}X'\hat{\theta} = c'\hat{\theta}$ is BLUE of $a'\beta$. Thus, when X has full column rank, all linear functions of β have BLUE, all components of β are estimable (BLUE exists). This will not be the case when X has less than full column rank.

Result. In the model, $Y = X\beta + \epsilon$, $E(\epsilon) = 0$, $Cov(\epsilon) = \sigma^2 I_n$ and X has full column rank (p), we have that

$$E(RSS) = E((Y - X\hat{\beta})'(Y - X\hat{\beta})) = (n-p)\sigma^2,$$

so that RSS$/(n-p)$ is an unbiased estimate of σ^2.

Proof. Note that $Y - X\hat{\beta} = Y - PY = (I-P)Y$. Therefore,

$$\text{RSS} = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'(I-P)^2 Y = Y'(I-P)Y,$$

where $I - P$ is symmetric idempotent with rank $n - p$.

$$\begin{aligned} E(RSS) &= E(Y'(I-P)Y) = \ \mathrm{tr}(\sigma^2(I-P)) + (X\beta)'(I-P)(X\beta) \\ &= \sigma^2(n-p) + \beta'X'(I-P)X\beta \\ &= (n-p)\sigma^2. \end{aligned}$$

For confidence statements and testing we need distribution theory.

Distribution Theory

Suppose ϵ_i are i.i.d. $N(0, \sigma^2)$. Then $\epsilon_{n\times 1} \sim N_n(0, \sigma^2 I_n)$ and so,
$Y \sim N_n(X\beta, \sigma^2 I_n)$.

Theorem. If $Y \sim N_n(X\beta, \sigma^2 I_n)$ and X has rank p, then
(i) $\hat{\beta} \sim N_p(\beta, \sigma^2(X'X)^{-1})$,
(ii) $(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)/\sigma^2 \sim \chi^2_p$,
(iii) $\hat{\beta}$ is independent of RSS $= (Y - X\hat{\beta})'(Y - X\hat{\beta})$,
(iv) RSS$/\sigma^2 \sim \chi^2_{n-p}$.

Proof. $Y \sim N_n(X\beta, \sigma^2 I_n)$, so (i)

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y \sim N_p((X'X)^{-1}X'X\beta, \sigma^2(X'X)^{-1}X'X(X'X)^{-1}) \\ &\quad = N(\beta, \sigma^2(X'X)^{-1}). \end{aligned}$$

(ii) Since $\hat{\beta} \sim N_p(\beta, \sigma^2(X'X)^{-1})$, note $(X'X)^{1/2}(\hat{\beta} - \beta) \sim N_p(0, \sigma^2 I_p)$, and hence
$(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)/\sigma^2 \sim \chi^2_p$.
(iii) $\hat{\beta} = (X'X)^{-1}X'Y = AY$ and RSS $= Y'(I-P)Y$. Since $Y \sim N_n(X\beta, \sigma^2 I_n)$, independence of $\hat{\beta}$ and $Y'(I-P)Y$ holds iff $A(I-P) = 0$. But $(I-P)A' = (I-P)X(X'X)^{-1} = 0$. Alternatively, $\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'P'Y = (X'X)^{-1}X'(PY)$, so that it is independent of $(I-P)Y$.
(iv) (a) RSS $= Y'(I-P)Y = (Y - X\beta)'(I-P)(Y - X\beta)$ since $(I-P)X = 0$. Note that since $Y - X\beta \sim N_n(0, \sigma^2 I_n)$, and $I - P$ is idempotent of rank $n-p$,

$(Y - X\beta)'(I - P)(Y - X\beta) \sim \chi^2_{n-p}$.
(b) Alternatively, note that $Q = (Y - X\beta)'(Y - X\beta) \sim \sigma^2\chi^2_n$. Now

$$\begin{aligned}
Q &= (Y - X\beta)'(Y - X\beta) \\
&= (Y - X\hat{\beta} + X\hat{\beta} - X\beta)'(Y - X\hat{\beta} + X\hat{\beta} - X\beta) \\
&= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \\
&= Q_1 + Q_2,
\end{aligned}$$

where $Q_2 \sim \sigma^2\chi^2_p$ and $Q_1 \geq 0$. Therefore, from a previous result, $Q_1 \sim \sigma^2\chi^2_{n-p}$ independent of Q_2.