The Matrix-Based Approach to the General Linear Model - continued

Properties of the Least-Squares Estimator

Using some of the matrix results that we have recently reviewed, it is simple to see that $E(\mathbf{b}) = \boldsymbol{\beta}$ so that the least-squares estimator is unbiased, and that the variance is $V(\mathbf{b}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$. Also, since $\mathbf{b}$ is a linear function of $\mathbf{y}$,it has a multivariate normal distribution, $\mathbf{b} \sim N_{k+1}(\boldsymbol{\beta},\sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1})$.

The Gauss-Markov Theorem

This theorem says that for the linear model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, if $E(\boldsymbol{\varepsilon}) = 0$ and $V(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2\mathbf{I}_n$, then the least-squares estimator has the lowest variance of any linear unbiased estimator, hence it is the BLUE (best linear unbiased estimator). To prove the theorem, we write the least-squares estimator as $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{My}$ and suppose that some other unbiased estimator $\widetilde{\mathbf{b}}$ has lower variance. Then we can write $\widetilde{\mathbf{b}}$ as $\widetilde{\mathbf{b}} = (\mathbf{M} + \mathbf{A})\mathbf{y}$ for some matrix $\mathbf{A}$. Since $\widetilde{\mathbf{b}}$ is also unbiased,

$$\begin{aligned} E(\widetilde{\mathbf{b}}) &= E((\mathbf{M} + \mathbf{A})\mathbf{y}) = E(\mathbf{My}) + E(\mathbf{Ay}) \\ &= E(\mathbf{b}) + \mathbf{A}E(\mathbf{y}) = \boldsymbol{\beta} + \mathbf{AX}\boldsymbol{\beta} \ \text{ must equal } \boldsymbol{\beta}, \end{aligned}$$

so $\mathbf{AX}\boldsymbol{\beta} = \mathbf{0}$ for all values of $\boldsymbol{\beta}$, hence the matrix $\mathbf{AX} = \mathbf{0}.$ Looking at the variance of $\widetilde{\mathbf{b}}$, we have:

$$\begin{aligned} V(\widetilde{\mathbf{b}}) &= V((\mathbf{M} + \mathbf{A})\mathbf{y}) = (\mathbf{M} + \mathbf{A})\sigma_\varepsilon^2\mathbf{I}_n(\mathbf{M} + \mathbf{A})' \\ &= \sigma_\varepsilon^2(\mathbf{MM}' + \mathbf{MA}' + \mathbf{AM}' + \mathbf{AA}'). \end{aligned}$$

However, both $\mathbf{MA}'$ and $\mathbf{AM}'$ must equal $\mathbf{0}$ because $\mathbf{AX} = \mathbf{0},$ so $V(\widetilde{\mathbf{b}}) = \sigma_\varepsilon^2(\mathbf{MM}' + \mathbf{AA}')$. Now if we look at an individual coefficient $\widetilde{B}_j$, it has two terms in its variance, one from each matrix:

$$V(\widetilde{B}_j) = \sigma_\varepsilon^2(\sum_{i=1}^{n} m_{ji}^2 + \sum_{i=1}^{n} a_{ji}^2).$$

Both terms are non-negative, so all the a_{ji} terms must be 0. We can repeat this argument for each coefficient $\widetilde{B}_j$, showing that the entire $\mathbf{A}$ matrix $= 0$. Thus this estimator is $\widetilde{\mathbf{b}} = (\mathbf{M} + \mathbf{0})\mathbf{y} = \mathbf{My} = \mathbf{b}$, so that the least-squares estimators is BLUE.

Maximum Likelihood Estimation

As we saw for simple linear regression, the maximum-likelihood estimator of $\boldsymbol{\beta}$ is the same as the least-squares estimator. The log-likelihood function for a sample is:

$$\ell(\boldsymbol{\beta}, \sigma_\varepsilon^2) = \log L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Taking partial derivatives we have:

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma_\varepsilon^2)}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma_\varepsilon^2}(2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}'\mathbf{y}),$$

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

which are set to zero and solved to yield:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

$$\widehat{\sigma}_\varepsilon^2 = \frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})}{n} = \frac{\mathbf{e}'\mathbf{e}}{n}$$

The MLE of $\boldsymbol{\beta}$ is thus the same as the least-squares estimator, and is clearly a maximum of the likelihood function. Once again, the MLE of σ_ε^2 is biased,so we will use the unbiased estimator $S_E^2 = \mathbf{e}'\mathbf{e}/(\mathbf{n} - \mathbf{k} - \mathbf{1})$.

Statistical Inference for Individual Coefficients

We can use our results about $\mathbf{b}$ to conduct inferences for individual coefficients B_j. Since an individual coefficient B_j can be considered a linear combination of the form $\mathbf{a}'\mathbf{b}$ where $\mathbf{a}$ is $\mathbf{0}$ except for a 1 corresponding to B_j, we can use the facts that $V(\mathbf{b}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$ and $V(\mathbf{a}'\mathbf{b}) = \mathbf{a}'V(\mathbf{b})\mathbf{a}$ to get $V(B_j) = V(\mathbf{a}'\mathbf{b}) = \sigma_\varepsilon^{\mathbf{2}}\,\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}$. It can be shown that $\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}$ is equal to the diagonal term of $(\mathbf{X}'\mathbf{X})^{-1}$ correponding to B_j (called v_{jj} in the text), so we have $B_j \sim N(\beta_j, \sigma_\varepsilon^2 v_{jj})$. Also, we can show that $\mathbf{b}$ and $\mathbf{e}$ are independent of each other (Result 4), so B_j and $\widehat{\sigma}_\varepsilon^2$ and independent. Since it can be shown that $(n-k-1)S_E^2/\sigma_\varepsilon^2 = \mathbf{e}'\mathbf{e}/\sigma_\varepsilon^2$ follows a chi-square distribution with $n-k-1$ degrees of freedom, the ratio

$$t = \frac{(B_j - \beta_j)/\sigma_\varepsilon\sqrt{v_{jj}}}{\sqrt{\frac{\mathbf{e}'\mathbf{e}/\sigma_\varepsilon^2}{n-k-1}}} = \frac{B_j - \beta_j}{S_E\sqrt{v_{jj}}}$$

follows a t-distribution with $n - k - 1$ degrees of freedom. Thus we can create test statistics and confidence intervals for β_j as shown in the text.