The Matrix-Based Approach to the General Linear Model - a little bit more

Random Regressors

All of the theory that we have been learning about for the general linear model so far has assumed that the matrix $\mathbf{X}$ is fixed, not random. In some applications this assumption is not true, so we will consider some situations when $\mathbf{X}$ is random. For our model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, when $\mathbf{X}$ was fixed we assumed that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2\mathbf{I}_n$. When $\mathbf{X}$ is random we will make three assumptions:

i) the distribution of $\boldsymbol{\varepsilon}$ is not affected by $\mathbf{X}$, so that $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are independent and thus the conditional distribution is $\boldsymbol{\varepsilon}|\mathbf{X}_0 \sim N_n(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_n)$ for any possible sample $\mathbf{X}_0$.

ii) the distribution of $\mathbf{X}$ does not depend on $\boldsymbol{\beta}$ or σ_ε^2.

iii) the covariance matrix of $\mathbf{X}$ is nonsingular.

When these assumptions hold, our results from this chapter are still valid. Here is an example of the reasoning: for a particular sample $\mathbf{X}_0$, the expected value of $\mathbf{b}$ is:

$$E(\mathbf{b}|\mathbf{X_0}) = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}|\mathbf{X_0}\right] = (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'E(\mathbf{y}|\mathbf{X}_0) = (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{X}_0\boldsymbol{\beta} = \boldsymbol{\beta}.$$

Since this conditional distribution of $\mathbf{b}|\mathbf{X_0}$ is the same for any value of $\mathbf{X_0}$, it is therefore the unconditional distribution of $\mathbf{b}$, so $\mathbf{b}$ is unconditionally unbiased. In a similar way, when conducting likelihood-based inference there is now a density for $\mathbf{X}$, $p(\mathbf{X}) = \prod_{i=1}^{n} p(\mathbf{x}_i')$, but as long as our assumptions above hold, it factors away from the part of the joint density $p(\mathbf{y}, \mathbf{X})$ involving $\mathbf{y}$, so our results (MLE, likelihood-ratio test, etc.) are unchanged.

Specification Error (consequences of underfitting)

One question at the end of this chapter is the effect on estimates of $\boldsymbol{\beta}$ when the model is underspecified, meaning that not all important covariates are included. Suppose the true model is

$$\mathbf{y}^* = \mathbf{X}^*\,\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1^*\boldsymbol{\beta}_1 + \mathbf{X_2^*}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

where $\mathbf{X}^*$ and $\boldsymbol{\beta}$ have been partitioned into a first part $\mathbf{X}_1^*$ based on variables in the model, and a second part $\mathbf{X_2^*}$ consisting of variables that should have been included but were not, and the corresponding partitions of $\boldsymbol{\beta}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. As we have used previously, the use of a $*$ with $\mathbf{X}$ and $\mathbf{y}$ indicates that they are expressed as deviations from their expectations ($\mathbf{y}^* = \mathbf{y} - E(\mathbf{y})$), so that the intercept term β_0 is eliminated, as illustrated in section 6.3. If we then fit a model just using the matrix $\mathbf{X}_1^*$, the model can be written as:

$$\mathbf{y}^* = \mathbf{X}_1^*\boldsymbol{\beta}_1 + \widetilde{\boldsymbol{\varepsilon}},$$

where $\widetilde{\boldsymbol{\varepsilon}} = \mathbf{X_2^*}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, so that the error term incorporates the missing part of the model. The least-squares estimator for this (incorrect) model can be written as:

$$\begin{aligned}
\mathbf{b}_1 &= (\mathbf{X}_1^{*\prime}\mathbf{X}_1^*)^{-1}\mathbf{X}_1^{*\prime}\mathbf{y}^* \\
&= (\frac{1}{n}\mathbf{X}_1^{*\prime}\mathbf{X}_1^*)^{-1}\frac{1}{n}\mathbf{X}_1^{*\prime}\mathbf{y}^* \\
&= (\frac{1}{n}\mathbf{X}_1^{*\prime}\mathbf{X}_1^*)^{-1}\frac{1}{n}\mathbf{X}_1^{*\prime}(\mathbf{X}_1^*\boldsymbol{\beta}_1 + \mathbf{X}_2^*\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta}_1 + (\frac{1}{n}\mathbf{X}_1^{*\prime}\mathbf{X}_1^*)^{-1}\frac{1}{n}\mathbf{X}_1^{*\prime}\mathbf{X}_2^*\boldsymbol{\beta}_2 + (\frac{1}{n}\mathbf{X}_1^{*\prime}\mathbf{X}_1^*)^{-1}\frac{1}{n}\mathbf{X}_1^{*\prime}\varepsilon.
\end{aligned}$$

By taking probability limits, it can be shown that

$$\text{plim } \mathbf{b}_1 = \boldsymbol{\beta}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\beta_2,$$

showing that $\mathbf{b}_1$ is inconsistent unless the two sets of covariates are uncorrelated ($\boldsymbol{\Sigma}_{12} = 0$) or the excluded covariates have slopes of $\beta_2 = \mathbf{0}$. It is also shown that the asymptotic covariance between $\mathbf{X}_1^*$ and $\widetilde{\boldsymbol{\varepsilon}}$ is not generally 0.

Confidence Intervals and Prediction Intervals for a particular covariate vector

This topic is presented as a problem in our text (Exercise 9.14), and is important to understand. Often, once we have a fitted model $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ we wish to predict either $E(y)$ or y for a particular covariate vector value $\mathbf{x}_0' = [1, x_{01}, \ldots, x_{0k}]$. The way to predict $E(y)$ or y is straightforward, we use $\widehat{E(y_0)} = \widehat{y}_0 = \mathbf{x}_0'\mathbf{b}$. However, we also typically want an interval estimate of this value. If we are predicting $E(y)$, we call an interval estimate a confidence interval, our error in estimation is $\delta = \widehat{y}_0 - E(y_0)$, and we can see that:

$$\begin{aligned}
E(\delta) &= E(\widehat{y}_0 - E(y_0)) = E(\widehat{y}_0) - \mathbf{x}_0'\boldsymbol{\beta} = \mathbf{x}_0'E(\mathbf{b}) - \mathbf{x}_0'\,\boldsymbol{\beta} = 0 \text{ and} \\
V(\delta) &= V(\widehat{y}_0 - E(y_0)) = V(\mathbf{x}_0'\mathbf{b} - \mathbf{x}_0'\boldsymbol{\beta}) = V(\mathbf{x}_0'\mathbf{b}) \\
&= \mathbf{x}_0'V(\mathbf{b})\mathbf{x_0} = \mathbf{x}_0'(\sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1})\mathbf{x_0} = \sigma_\varepsilon^{\mathbf{2}}\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_0}.
\end{aligned}$$

However, if we are predicting an actual value $\widehat{y}_0$ then we call an interval estimate a prediction interval, and we need to account for the variation of an observation about its mean. We can reflect this by writing the true value y_0 as $y_0 = \mathbf{x}_0'\,\boldsymbol{\beta} + \varepsilon_0$, then we can define the error as $D \equiv \widehat{y}_0 - y_0 = \mathbf{x}_0'\mathbf{b} - (\mathbf{x}_\mathbf{0}'\boldsymbol{\beta} + \varepsilon_\mathbf{0}) = \mathbf{x}_0'(\mathbf{b} - \boldsymbol{\beta}) - \varepsilon_\mathbf{0}$. Then we see that:

$$\begin{aligned}
E(D) &= E(\mathbf{x}_0'(\mathbf{b} - \boldsymbol{\beta}) - \varepsilon_\mathbf{0}) = \mathbf{x}_0'E(\mathbf{b} - \boldsymbol{\beta}) - E(\varepsilon_\mathbf{0}) = 0 \text{ and} \\
V(D) &= V(\mathbf{x}_0'(\mathbf{b} - \boldsymbol{\beta}) - \varepsilon_\mathbf{0}) = \mathbf{x}_0'V(\mathbf{b} - \boldsymbol{\beta})\mathbf{x}_0 + V(\varepsilon_\mathbf{0}) \\
&= \mathbf{x}_0'V(\mathbf{b})\mathbf{x}_0 + \sigma_\varepsilon^2 = \sigma_\varepsilon^{\mathbf{2}}\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_0} + \sigma_\varepsilon^2 \\
&= \sigma_\varepsilon^{\mathbf{2}}(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_0}).
\end{aligned}$$

Clearly we are less certain in our forecast of a particular value than in our forecast of a mean value. Since we do not know the value of $\sigma_\varepsilon^{\mathbf{2}}$, we use S_E^2 as an estimate and use the t distribution for constructing confidence and prediction intervals.