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

Statistical Inference for Several Coefficients

Often we are interested in testing the significance of a subset of regression coefficients. If we have k covariates in our model, suppose we wish to test if q of them are not important. Suppose they are the first q coefficients, so we wish to test $H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0$. We will highlight the theory for testing H_0 based on two approaches: i) a likelihood-based and ii) least-squares-based. We have already seen that the least-squares and maximum-likelihood estimators of $\boldsymbol{\beta}$ are identical for the general linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $E(\boldsymbol{\varepsilon}) = 0$ and $V(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_n$. It turns out that the estimators of the remaining parameters $\beta_{q+1}, \beta_{q+2}, \cdots, \beta_k$ when assuming that $\beta_1 = \beta_2 = \cdots = \beta_q = 0$ are also identical under both the likelihood and least squares approaches. When testing H_0 for either approach, we must estimate the remaining parameters $\beta_{q+1}, \beta_{q+2}, \cdots, \beta_k$ (we have called this the reduced model) under the constraint that $\beta_1 = \beta_2 = \cdots = \beta_q = 0$ is true, or more generally under the constraint $\boldsymbol{L\beta} = \mathbf{c}$, for some matrix $\boldsymbol{L}$ of dimension q x $k+1$ of full row rank q and a vector $\mathbf{c}$ of dimension q x 1. Taking the least-squares approach, to minimize the quantity $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ under the constraint $\boldsymbol{L\beta} = \mathbf{c}$, we can use the method of Lagrange multipliers, creating a function $S(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(\mathbf{c} - \boldsymbol{L\beta})$ and minimizing it with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. The partial derivative equations are:

$$\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}'\mathbf{y} - \boldsymbol{L}'\boldsymbol{\lambda},$$
$$\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \mathbf{c} - \boldsymbol{L\beta}$$

which we set to zero and solve. From these two equations we have, respectively,

$$\mathbf{b}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \frac{1}{2}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}'\,\boldsymbol{\lambda} = \mathbf{b} + \frac{1}{2}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}'\widehat{\boldsymbol{\lambda}},$$
$$\mathbf{c} = \boldsymbol{L}\mathbf{b}_0 = \boldsymbol{L}\mathbf{b} + \frac{1}{2}\boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}'\widehat{\boldsymbol{\lambda}}$$

Solving the second equation for $\widehat{\boldsymbol{\lambda}}$ we have:

$$\widehat{\boldsymbol{\lambda}} = 2[\boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}']^{-1}(\mathbf{c}-\boldsymbol{L}\mathbf{b}),$$

and substituting back in the first equation for $\mathbf{b}_0$ we have:

$$\mathbf{b}_0 = \mathbf{b}+(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}'[\boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}']^{-1}(\mathbf{c}-\boldsymbol{L}\mathbf{b}).$$

Thus $\mathbf{b}_0$ is both the least-squares and maximum-likelihood estimator of $\boldsymbol{\beta}$ under the constraint $\boldsymbol{L\beta} = \mathbf{c}$.

i) The likelihood-based approach

To test $H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0$ with this approach, we compare the maximized likelihood value under H_0 to its unrestricted maximum value. An examination of the likelihood function $L(\boldsymbol{\beta}, \sigma_\varepsilon^2)$ taking into account the MLE for σ_ε^2, $\widehat{\sigma}_\varepsilon^2 = \frac{(\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb})}{n}$, shows that $(\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb})$ cancels in the exponent, leaving the maximized value of $\max_* L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = (2\pi\widehat{\sigma}_*^2)^{-n/2}\exp(-n/2)$, with $\widehat{\sigma}_*^2 = (\mathbf{y}-\mathbf{Xb}_*)'(\mathbf{y}-\mathbf{Xb}_*)/n$ where $'*'$ indicates a particular model (either unrestricted or constrained under H_0). These maximized values are compared via the generalized likelihood ratio test statistic,

$$G_0^2 = -2\ln(L_0/L_1),$$

where L_0 is the maximized likelihood under H_0 and L_1 is the maximized likelihood for the unconstrained (complete) model. It can be shown that G_0^2 has an asymptotic chi-squared distribution with q degrees of freedom under H_0.

ii) The least-squares-based approach

To test $H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0$ with this approach, we can calculate the reduction in regression sums of squares under H_0, which is equivalent to calculating the increase in error sums of squares. Again using the notation that model 0 is constrained under H_0, and RSS and $\widehat{\mathbf{y}}_1$ are from the unconstrained model, we have:

$$RSS_0 - RSS = (\mathbf{y}-\widehat{\mathbf{y}}_0)'(\mathbf{y}-\widehat{\mathbf{y}}_0) - (\mathbf{y}-\widehat{\mathbf{y}}_1)'(\mathbf{y}-\widehat{\mathbf{y}}_1).$$

We will return to this equation later and show that the reduction in error sum of squares $RSS_0 - RSS$ is independent of RSS, and the ratio of these two quadratic forms divided by their degrees of freedom leads to:

$$F_0 = \frac{(RSS_0 - RSS)/q}{RSS/(n-k-1)},$$

which is distributed as an F statistic with q and $n-k-1$ degrees of freedom. As mentioned in the text, a test for a subset of regression coefficients can also be conducted using the subvector of $\mathbf{b}$ and its associated part of the covariance matrix $(\mathbf{X}'\mathbf{X})^{-1}$, which we denote by $\mathbf{V}_{11}$. Then to test $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)}$ we can use

$$F_0 = \frac{(\mathbf{b}_1 - \boldsymbol{\beta}_1^{(0)})'\mathbf{V}_{11}^{-1}((\mathbf{b}_1 - \boldsymbol{\beta}_1^{(0)}))}{qS_E^2},$$

which follows an $F_{q,n-k-1}$ distribution under H_0.

Finally the global test $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ is considered and they present a result on the expected value of the regression sum of squares for this test. It turns out that for the global test,

$$\begin{aligned}
E(\text{RegSS}) &= E[\mathbf{y}'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \frac{1}{n}\mathbf{J}_n)\mathbf{y}] \\
&= tr((\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \frac{1}{n}\mathbf{J}_n)\mathbf{I}_n\sigma_\varepsilon^2) + \mu'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \frac{1}{n}\mathbf{J}_n)\mu \\
&= k\sigma_\varepsilon^2 + (\mathbf{X}\boldsymbol{\beta}_\mathbf{1})'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \frac{1}{n}\mathbf{J}_n)(\mathbf{X}\boldsymbol{\beta}_\mathbf{1}) \\
&= k\sigma_\varepsilon^2 + \boldsymbol{\beta}_\mathbf{1}'(\mathbf{X}^{*\prime}\mathbf{X}^*)\boldsymbol{\beta}_\mathbf{1},
\end{aligned}$$

where as explained in the text, $\mathbf{X}^*$ is a matrix of mean-deviation regressors without an intercept column.

The General Linear Hypothesis

As was hinted in our discussion of inference for a subset of coefficients, a more general hypothesis than $H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0$ is $\boldsymbol{L\beta} = \mathbf{c}$, for some matrix $\boldsymbol{L}$ of dimension q x $k+1$ of full row rank q and a vector $\mathbf{c}$ of dimension q x 1. Since $\boldsymbol{L}\mathbf{b}$ is a function of the least-squares estimator, our previous results show that it is normally distributed, with mean $\boldsymbol{L\beta}$ and covariance matrix $V(\boldsymbol{L}\mathbf{b}) = V(\boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = \boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'V(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}' = \sigma_\varepsilon^2\boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}'$. Thus under $H_0 : \boldsymbol{L\beta} = \mathbf{c}$,

$$(\boldsymbol{L}\mathbf{b} - \mathbf{c})'[\boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}']^{-1}(\boldsymbol{L}\mathbf{b} - \mathbf{c})\,/\sigma_\varepsilon^2 \sim \chi_q^2$$

and

$$F_0 = \frac{(\boldsymbol{L}\mathbf{b} - \mathbf{c})'[\boldsymbol{L}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{L}']^{-1}(\boldsymbol{L}\mathbf{b} - \mathbf{c})}{qS_E^2}$$

follows an F distribution with q and $n-k-1$ degrees of freedom.