Matrices for MLR

$$Y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \varepsilon_1$$
$$Y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \varepsilon_2$$
$$\vdots \qquad \vdots$$
$$Y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \varepsilon_n$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\underset{\sim}{Y} = \underset{\sim}{X} \; \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

- <u>Least Squares Estimates:</u> $\hat{\underset{\sim}{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)'$ chosen to minimize SSE
  - $\hat{\underset{\sim}{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)' = \left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1} \underset{\sim}{X}'\underset{\sim}{Y}$
  - $\text{var}(\hat{\underset{\sim}{\beta}}) = \left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1} \sigma^2$ is the Variance-Covariance (VCV) matrix for the $\hat{\beta}_i$
    - The VCV matrix has $\widehat{\text{var}}(\hat{\beta}_i)$ on the diagonal and $\widehat{\text{cov}}(\hat{\beta}_i, \hat{\beta}_j)$ off the diagonal

9.6.5 Inference Methods for Linear Functions of Regression Coefficients

It is often of interest to consider statistical inferences about a *linear sum*

$L = c_1\beta_1 + c_2\beta_2 + \cdots + c_k\beta_k$

of regression coefficients, where $c_1, c_2, \ldots, c_k$ are specified constants.

Consider a study where varying doses of two drugs A and B in milligrams (mg) were prescribed to lower systolic blood pressure (SBP). Consider the following situations:

1. Suppose that it is conjectured that the amount of drug A given is twice as effective as the amount of drug B with regard to lowering SBP. This conjecture can be addressed by considering the null hypothesis H_0: $\beta_A = 2\beta_B$.

2. We may be interested in estimating *how much* SBP decreases, on average, as a result of 1 mg (or larger) doses of both drugs. In the 1 mg case, we would want to estimate the quantity $(\beta_A + \beta_B)$ and construct a confidence interval for this unknown parameter.

$\hat{L} = \hat{\beta}_A - 2\hat{\beta}_A$. Thus the c_i values corresponding to β_A and β_B would be 1 and −2, respectively, and are equal to 0 for any remaining *regression* coefficients considered in the model. More generally, when $(c_1 + c_2 + \cdots + c_k) = 0$, the linear function is called a *linear contrast*.

Hypothesis tests about, and confidence intervals for, L utilize the estimated value $\hat{L}$ and its estimated standard error $S_{\hat{L}}$. Hypothesis tests use the test statistic $\frac{\hat{L}}{S_{\hat{L}}}$, which follows the t_{n-k-1} distribution under the null hypothesis that $L = 0$. Confidence intervals are constructed using the formula $\hat{L} \pm t_{n-k-1, 1-\frac{\alpha}{2}}(S_{\hat{L}})$.

A key difference from earlier is that the standard error of $\hat{L}$, $S_{\hat{L}}$—or, equivalently, its variance, $S_{\hat{L}}^2$—is not readily obtained from standard output and requires extra calculations and/or programming. The formula for $S_{\hat{L}}^2 = \widehat{\text{Var}}(\hat{L})$ is

$$S_{\hat{L}}^2 = \sum_{i=1}^{k} c_i^2 S_{\hat{\beta}_i}^2 + 2\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} c_i c_j \,\widehat{\text{cov}}\,(\hat{\beta}_i, \hat{\beta}_j) \qquad (9.11)$$

Note that this complicated expression involves the k estimated variances of the estimated regression coefficients and the estimated covariances[6] for each of the $k(k-1)/2$ pairs of estimated regression coefficients. The individual variance estimates may be found by squaring the estimated standard errors of the estimated regression coefficients that are given in standard computer output, but the estimated covariances need to be obtained by further requesting the estimated variance–covariance matrix.[7]