Chapter 20
Analysis of Regression Models

C. R. Henderson

1984 - Guelph

A regression model is one in which $\mathbf{Zu}$ does not exist, the first column of $\mathbf{X}$ is a vector of 1's, and all other elements of $\mathbf{X}$ are general (not 0's and 1's) as in the classification model. The elements of $\mathbf{X}$ other than the first column are commonly called covariates or independent variables. The latter is not a desirable description since they are not variables but rather are constants. In hypothetical repeated sampling the value of $\mathbf{X}$ remains constant. In contrast $\mathbf{e}$ is a sample from a multivariate population with mean $= \mathbf{0}$ and variance $= \mathbf{R}$, often $\mathbf{I}\sigma_e^2$. Accordingly $\mathbf{e}$ varies from one hypothetical sample to the next. It is usually assumed that the columns of $\mathbf{X}$ are linearly independent, that is, $\mathbf{X}$ has full column rank. This should not be taken for granted in all situations, for it could happen that linear dependencies exist. A more common problem is that near but not complete dependencies exist. In that case, $(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}$ can be quite inaccurate, and the variance of some or all of the elements of $\hat{\boldsymbol{\beta}}$ can be extremely large. Methods for dealing with this problem are discussed in Section 20.2.

1 Simple Regression Model

The most simple regression model is

$$y_i = \mu + w_i\gamma + e_i,$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & w_1 \\ 1 & w_2 \\ \vdots & \vdots \\ 1 & w_n \end{pmatrix}.$$

The most simple form of $Var(\mathbf{e}) = \mathbf{R}$ is $\mathbf{I}\sigma_e^2$. Then the BLUE equations are

$$\begin{pmatrix} n & w. \\ w. & \sum w_i^2 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} y. \\ \sum w_i y_i \end{pmatrix}. \quad (1)$$

To illustrate suppose n=5,

$$\mathbf{w}' = (6, 5, 3, 4, 2), \;\; \mathbf{y}' = (8, 6, 5, 6, 5).$$

The BLUE equations are

$$\frac{1}{\sigma_e^2}\begin{pmatrix} 5 & 20 \\ 20 & 90 \end{pmatrix}\begin{pmatrix} \hat{\mu} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 30 \\ 127 \end{pmatrix} /\sigma_e^2.$$

The inverse of the coefficient matrix is

$$\begin{pmatrix} 1.8 & -.4 \\ -.4 & .1 \end{pmatrix} \sigma_e^2.$$

The solution is (3.2, .7).

$$Var(\hat{\mu}) = 1.8\ \sigma_e^2,\ Var(\hat{\gamma}) = .1\ \sigma_e^2, Cov(\hat{\mu}, \hat{\gamma}) = -.4\ \sigma_e^2.$$

Some text books describe the model above as

$$y_i \ = \ \alpha + (w - \bar{w}_.)\gamma + e_i.$$

The BLUE equations in this case are

$$\begin{pmatrix} n & 0 \\ 0 & \sum(w_i - \bar{w}_.)^2 \end{pmatrix}\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} y_. \\ \sum(w_i - w_.)\bar{y}_i \end{pmatrix}. \tag{2}$$

This gives the same solution to $\hat{\gamma}$ as (19.1) but $\hat{\mu} \neq \hat{\alpha}$ except when $\bar{w}_. \ = \ 0$. The equations of (20.2) in our example are

$$\frac{1}{\sigma_e^2}\begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}\begin{pmatrix} \hat{\mu} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 30 \\ 7 \end{pmatrix} /\sigma_e^2.$$

$$\hat{\alpha} \ = \ 6,\ \hat{\gamma} \ = \ .7.$$

$$Var(\hat{\alpha}) = .2\sigma_e^2,\ Var(\hat{\gamma}) \ = \ .1\sigma_e^2,\ Cov(\hat{\alpha}, \hat{\gamma}) = 0.$$

It is easy to verify that $\hat{\mu} \ = \ \hat{\alpha} - \bar{w}_.\hat{\gamma}_.$. These two alternative models meet the requirements of linear equivalence, Section 1.5.

BLUP of a future y say y_0 with $w_i \ = \ w_0$ is

$$\hat{\mu} + w_0\hat{\gamma} + \hat{e}_0 \text{ or } \hat{\alpha} + (w_0 - \bar{w})\hat{\gamma} + \hat{e}_0,$$

where $\hat{e}_0$ is BLUP of $e_0 \ = \ 0$, with prediction error variance, σ_e^2. If $w_0 \ = \ 3, \ \ y_0$ would be 5.3 in our example. This result assumes that future μ or α) have the same value as in the population from which the original sample was taken. The prediction error variance is

$$(1 \ \ 3)\begin{pmatrix} 1.8 & -.4 \\ -.4 & .1 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix} \sigma_e^2 \ + \ \sigma_e^2 \ = \ 1.3\ \sigma_e^2.$$

Also using the second model it is

$$(1 \ \ -1)\begin{pmatrix} .2 & 0 \\ 0 & .1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \sigma_e^2 \ + \ \sigma_e^2 \ = \ 1.3\ \sigma_e^2$$

as in the equivalent model.

2 Multiple Regression Model

In the multiple regression model the first column of $\mathbf{X}$ is a vector of 1's, and there are 2 or more additional columns of covariates. For example, the second column could represent age in days and the third column could represent initial weight, while $\mathbf{y}$ represents final weight. Note that in this model the regression on age is asserted to be the same for every initial weight. Is this a reasonable assumption? Probably it is not. A possible modification of the model to account for effect of initial weight upon the regression of final weight on age and for effect of age upon the regression of final weight on initial weight is

$$\begin{aligned} y_i &= \mu + \gamma_1 w_1 + \gamma_2 w_2 + \gamma_3 w_3 + e_i, \\ \text{where} \quad w_3 &= w_1 w_2. \end{aligned}$$

This model implies that the regression coefficient for $\mathbf{y}$ on w_1 is a simple linear function of w_2 and the regression coefficient for $\mathbf{y}$ on w_2 is a simple linear function of w_1. A model like this sometimes gives trouble because of the relationship between columns 2 and 3 with column 4 of $\mathbf{X}$. We illustrate with

$$\mathbf{X} = \begin{pmatrix} 1 & 6 & 8 & 48 \\ 1 & 5 & 9 & 45 \\ 1 & 5 & 8 & 40 \\ 1 & 6 & 7 & 42 \\ 1 & 7 & 9 & 63 \end{pmatrix}.$$

The elements of column 4 are the products of the corresponding elements of columns 2 and 3. The coefficient matrix is

$$\begin{pmatrix} 5 & 29 & 41 & 238 \\ & 171 & 238 & 1406 \\ & & 339 & 1970 \\ & & & 11662 \end{pmatrix}. \qquad (3)$$

The inverse of this is

$$\begin{pmatrix} 4780.27 & -801.54 & -548.45 & 91.73 \\ & 135.09 & 91.91 & -15.45 \\ & & 63.10 & -10.55 \\ & & & 1.773 \end{pmatrix}. \qquad (4)$$

Suppose that we wish to predict y for $w_1 = \bar{w}_{1.} = 5.8$, $\bar{w}_2 = 8.2$, $w_3 = 47.56 = (5.8)(8.2)$. The variance of the error of prediction is

$$(1 \;\; 5.8 \;\; 8.2 \;\; 47.56)(\text{matrix } 20.4) \begin{pmatrix} 1 \\ 5.8 \\ 8.2 \\ 47.56 \end{pmatrix} \sigma_e^2 \;+\; \sigma_e^2 \;=\; 1.203\; \sigma_e^2$$

Suppose we predict y for $w_1 = 3$, $w_2 = 5$, $w_3 = 15$. Then the variance of the error of prediction is 215.77 σ_e^2, a substantial increase. The variance of the prediction error is extremely vulnerable to departures of w_i from $\bar{w}_i$.

Suppose we had not included w_3 in the model. Then the inverse of the coefficient matrix is

$$\begin{pmatrix} 33.974 & -1.872 & -2.795 \\ & .359 & -.026 \\ & & .359 \end{pmatrix}.$$

The variances of the errors of prediction of the two predictors above would then be 1.20 and 7.23, the second of which is much smaller than when w_3 is included. But if $w_3 \neq 0$, the predictor is biased when w_3 is not included.

Let us look at the solution when w_3 is included and $\mathbf{y}' = (6, 4, 8, 7, 5)$. The solution is

$$(157.82, \quad -23.64, \quad -17.36, \quad 2.68).$$

This is a strange solution that is the consequence of the large elements in $(\mathbf{X}'\mathbf{X})^{-1}$. A better solution might result if a prior is placed on w_3. When the prior is 1, we add 1 to the lower diagonal element of the coefficient matrix. The resulting solution is

$$(69.10, \quad -8.69, \quad -7.16, \quad .967).$$

This type of solution is similar to ridge regression, Hoerl and Kennard (1970). There is an extensive statistics literature on the problem of ill-behaved $\mathbf{X}'\mathbf{X}$. Most solutions to this problem that have been proposed are (1) biased (shrunken estimation) or (2) dropping one or more elements of $\boldsymbol{\beta}$ from the model with either backward or forward type of elimination, Draper and Smith (1966). See for example a paper by Dempster et al. (1977) with an extensive list of references. Also Hocking (1976) has many references.

Another type of covariate is involved in fitting polynomials, for example

$$y_i = \mu + x_i\gamma_1 + x_i^2\gamma_2 + x_i^3\gamma_3 + x_i^4\gamma_4 + e_i.$$

As in the case when covariates involve products, the sampling variances of predictors are large when x_i departs far from $\bar{x}$. The numerator mean square with 1 d.f. can be computed easily. For the i^{th} γ_i it is

$$\hat{\gamma}_i^2 / c^{i+1},$$

where c^{i+1} is the i+1 diagonal of the inverse of the coefficient matrix. The numerator can also be computed by reduction under the full model minus the reduction when γ_i is dropped from the solution.