Multicolinearity Diagnostics :

Some of the diagnostics we have just discussed are sensitive to multicolinearity. For example, we know that with multicolinearity, additions and deletions of data cause shifts in the regression coefficients, this will be picked up by Cook's D or DFBETAS.

Regression coefficients may even have a different sign from the expected.

Other effects of multicolinearity are

1) EXTRA SS results which appear nonsignificant when they should be significant,

2) large confidence intervals on the regression coefficients.

Detecting multicolinearity : Variance Inflation Factor (VIF)

This provides a measure of how much the variance of the estimated regression coefficient is inflated as compared to when the variables are not linearly related.

First we know that the variance of the regression coefficients is given by

$$\sigma^2_{b_i} = \sigma^2(X'X)^{-1}$$

The comparison is made to the variance calculated on the standardized variables,

$$\sigma^2_{b'_i} = (\sigma')^2\ r^{-1}_{XX}$$

The elements of the diagonal of the r^{-1}_{XX} matrix are the VIF's. These are

$$VIF_k = (1-R^2_k)^{-1}$$

This value is 1 when $R^2_k = 0$ (ie. when the variable X_k is not correlated to other independent variables).

Otherwise, the value will be greater than one.

Another value, called the tolerance, is the reciprocal of the VIF

$$\text{Tolerance}_k = 1\text{-}R_k^2$$

where R^2 is the squared correlation of the regressor of X_k on the other regressors

SAS provides both of these values

The uses of VIF

1) The largest value of VIF is an indicator of multicolinearity.

This value should not exceed 10.

2) The mean of the VIF is related to the severity of multicolinearity and to how far the estimated standardized regression coefficients are from the true standardized values.

$$\text{Mean VIF} = \frac{\sum_{k=1}^{p-1} \text{VIF}_k}{p-1}$$

If this value is much greater than 1, serious problems are indicated.

Another measure is provided by the Condition number (a multivariate evaluation). Eigen values are extracted from the regressors.

1) These are variances of linear combinations of the regressors, and go from larger to smaller. If one or more, at the end, are zero then the matrix is not full rank.

2) These sum to p, and if the X_k are independent, each would equal 1.

3) The condition number is the square root of the ratio of the largest (always the first) to each of the others. If this value exceeds 30 then multicolinearity may be a problem.

What to do about Multicolinearity

1) Ignore the problem. If you do not need the reg. coeff, just predicted values, then the results may be adequate.

2) In polynomial regressions, go deviations or orthogonal

3) Get more data, especially if the data set is small. This may reduce the correlation between the variables

4) Delete some variables, fit a simpler model (variable selection discussed soon). It may also be possible to use some substitute variable or some surrogate variable.

5) Use Ridge regression.

What is Ridge regression?

Ridge regression is BIASED.

However, as we have seen, variance is inflated when multicolinearity is present.
Therefore, we have a trade-off; we introduce a “small" bias, and hope for a big reduction in the variance.

Overall, we hope for a better estimate.

How can introducing bias be advantageous?
very high variability may make an estimate worthless,
while a slight bias with low variability may be close enough

see sketch,

Calculations for Ridge regression.

$$X'X\underline{b} = X'\underline{Y}$$

the standardized equivalent

$$r_{XX}b = r_{XY}$$

the standardized calculations for Ridge Regression are

$$(r_{XX}+cI)b^{R} = r_{XY}$$

then

$$b^{R} = (r_{XX}+cI)^{-1}r_{XY}$$

The value c is the biasing constant. The smaller c, the smaller the bias.

The presence of colinearity does not affect the unbiased estimates, but the regression coefficients tend to be too large in magnitude.

Adding a small constant to the diagonal of the r_{XX} drives the regression coefficients toward zero; if the value of c is large enough, eventually the values of the regression coefficients will equal 0.

However, they do not approach 0 at the same rate, and presumably at some point the values variances are sufficiently reduced to provide improved estimates.

So we start by adding small values of c for effect. Increasingly larger values are tried until the desired point is reached.

See handout

How large a value should we add?

Usually the plotted lines of successive b_k^R change quite quickly for k very near 0

At some point the estimates stabilize

At the same time we can calculate the VIF, which will also fall quickly initially and stabilize

There is no objective criteria, one simply changes the value of k by small increments, and plots the regression coefficients on the value of k.

Judgment is used to find the smallest k (to get smallest bias) which achieves stability.

The calculations can be done with the standardized values. Once the best values are chosen, the regression coefficients can be transformed back to the original units using the same equation we originally used to calculate the standardized regression coefficient.

$$\beta_k = \beta'_k \left(\frac{s_Y}{s_X} \right)$$