

# Principal Components Analysis (PCA) Evaluating Multicollinearity in Regression

SPSS output for the multicollinearity eigenvalues are incorrect when ran through the regular model because it includes the constant in the predictor matrix (one DV and 6 IVs will produce 7 eigenvalues incorrectly).

Incorrect eigenvalues:

**Collinearity Diagnostics<sup>a</sup>**

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions						
				(Constant)	x1	x2	x3	x4	x5	x6
1	1	2.633	1.000	.00	.00	.00	.00	.00	.02	.00
	2	1.821	1.203	.00	.00	.00	.00	.00	.05	.09
	3	1.033	1.596	.00	.00	.00	.00	.00	.03	.10
	4	.658	2.000	.00	.00	.00	.00	.00	.19	.40
	5	.606	2.085	.00	.00	.00	.00	.00	.00	.00
	6	.249	3.253	.00	.00	.00	.00	.00	.69	.40
	7	.000	92.253	1.00	1.00	.99	1.00	1.00	.02	.00

a. Dependent Variable: y

A way to remove the constant is to run the linear regression without the intercept in the equation. This will adjust the number of eigenvalues. However, these eigenvalues are still incorrect.

Incorrect eigenvalues from a regression with no intercept:

**Collinearity Diagnostics<sup>a,b</sup>**

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				x1	x2	x3	x4	x5	x6
1	1	1.990	1.000	.09	.10	.01	.00	.09	.01
	2	1.569	1.126	.03	.01	.16	.04	.00	.16
	3	.966	1.436	.02	.00	.01	.92	.01	.03
	4	.621	1.790	.17	.02	.18	.01	.20	.39
	5	.606	1.813	.47	.53	.04	.03	.00	.00
	6	.249	2.828	.23	.34	.60	.00	.70	.40

a. Dependent Variable: y

b. Linear Regression through the Origin

The technical reasons for this were best explained by Dr. Ming-mei Wang:

*SPSS 16.0 calculates the eigenvalues of the scaled and uncentered cross-product matrix  $X'X$  for collinearity diagnostics. For a model, it calculates the cross product matrix and then scales the matrix by its diagonals (as in calculating a correlation matrix). Explicitly, it obtains the scaled  $X'X$  as  $Dinv * X'X * Dinv$ , where  $Dinv$  is a diagonal matrix with elements equal to the reciprocal of the square roots of the diagonal elements of  $X'X$ . So the results depend on the model and differ for models with intercept and without. The scaled uncorrected (uncentered) cross product matrix (scaled  $X'X$ ) and the*

*eigenvalues obtained for this matrix using the Factor Analysis (FA) command of SPSS 16.0 would thus not be the same.*

The correct eigenvalues can be computed using the FA menu in SPSS:

Analyze → Data Reduction → Factor...

Input only your predictors as *Variables*. There is no need to request any rotations since we simply want to get a *Principal Component* solution with the correct correlation based eigenvalues.

Correct eigenvalues from PCA:

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.429	40.480	40.480	2.429	40.480	40.480
2	1.546	25.769	66.249	1.546	25.769	66.249
3	.922	15.368	81.617			
4	.794	13.233	94.850			
5	.208	5.132	99.982			
6	.001	.018	100.000			

Extraction Method: Principal Component Analysis.

Now kappa (the multicollinearity measure) can be computed using the correct eigenvalues. Compare these findings as opposed to the previous (regression with no intercept) kappa computations. It becomes clear that one must use the eigenvalues taken from the correlation matrix of the PCA.

$$Kappa = \kappa = \sqrt{\frac{Eigenvalue \max}{Eigenvalue \min}}$$

$$Incorrect \ Kappa = \sqrt{\frac{1.990}{.249}} = \sqrt{7.9919} = 2.8270$$

$$Correct \ Kappa = \sqrt{\frac{2.429}{.0019}} = \sqrt{2429} = 49.28$$

Comparing these kappa values you see that the high (49.28) kappa value, as opposed to the value computed from the incorrect eigenvalues, indicates serious issues of multicollinearity.