Chapter 9. Biased Regression

9:1 Dealing with collinearity.

Collinearity in the predictors weakens the analysis by increasing the variance of the estimated parameters and by restricting the predictive power of the model to the reduced set of combinations of values for the predictor variables. Although no statistical methods will remove the collinearity from the data, there are techniques to reduce the impact of collinearity.

Variable selection, Ridge regression, and Principal components regression are techniques to reduce the impacts of collinearity in MLR.

9:2 Variable selection.

Variable selection, as performed by using the Cp criterion, is a useful means of eliminating excessive collinearity by leaving redundant variables out of the final model. Selection of variables must balance a reduction of collinearity versus an excessive biasing of the model achieved by removing variables. Keep in mind that the fact that a variable be a good candidate for removal from a model does not mean that the variable is not a part of the true model. The probability of making this type of error is not known, but should increase as more variables are deleted.

9:3 Biased regression.

Biased regression is a method to deal with multicollinearity that stabilizes partial regression coefficients by introducing bias. The bias is associated with a reduction in the variance of the estimated coefficients, so there is a gain that more than compensates for the increase in bias.

9:3.1 Concepts.

9:3.1.1 Mean squared error.

A measure of the average closeness of an estimator b of a parameter β is the Mean squared error (MSE) of the estimate. The key distinction here is that the MSE is measured as the expectation of deviations of b from β, whereas the variance of the estimator is measured as deviations from the expectation of the estimator E{b}. The difference between β and E{b} is the bias of the estimator b. OLS usually yield unbiased estimators, but modifications or other means of calculating the estimates may yield biased estimators. Of course, biased estimators have the disadvantage that they are biased; but they may have much lower variance and thus yield better confidence intervals for the true value of the parameter. Biased regression methods, such as RR and PCR use this advantage.

$$MSE(\hat{\beta}) = E\left\{(\hat{\beta} - \beta)^2\right\}$$

$$Var\{\hat{\beta}\} = E\left\{(\hat{\beta} - E\left\{\hat{\beta}\right\})^2\right\}$$

$$Bias\left(\hat{\beta}\right) = E\left\{\hat{\beta}\right\} - \beta$$

The total error in the estimate of a parameter can be expressed as the sum of the error due to the variance, plus the error due to the bias of the estimate:

Mean sq. error:

$$\begin{cases} E\{\hat{\beta} - \beta\}^2 = \Sigma \dfrac{(\hat{\beta}_i - \beta)^2}{n-1} = \\ = \Sigma \dfrac{(\hat{\beta}_i - E\{\hat{\beta}\} + E\{\hat{\beta}\} - \beta)^2}{n-1} = \\ = \Sigma \dfrac{(\hat{\beta}_i - E\{\hat{\beta}\})^2}{n-1} + 2\left(E\{\hat{\beta}\} - \beta\right)\Sigma \dfrac{(\hat{\beta}_i - E\{\hat{\beta}\})}{n-1} + E\{\hat{\beta}\} - \beta)^2 = \\ = \sigma^2\{\hat{\beta}\} + (E\{\hat{\beta}\} - \beta)^2 \end{cases}$$

If the estimator is not biased (as obtained by LS) then mean squared error = $\sigma^2\{\hat{\beta}\}$ meaning that the sample-based variance of the estimated parameter around its estimated value is also an estimate of the variance around the true parameter value.

However, a biased estimator of β can have a much smaller variance than an unbiased one.

Figure 1 illustrates the scope of biased regression, where the biased estimator is expected to have a much smaller variance, and thus greater probability of being close to the true parameter, than the unbiased LS estimator.

A simple example of the difference between a biased and an unbiased estimator can be found in any basic statistics textbook where the rationale for using a divisor equal to (n-1) instead of n for the estimated variance is explained. In that case, however, the biased estimator is not superior to the unbiased one.

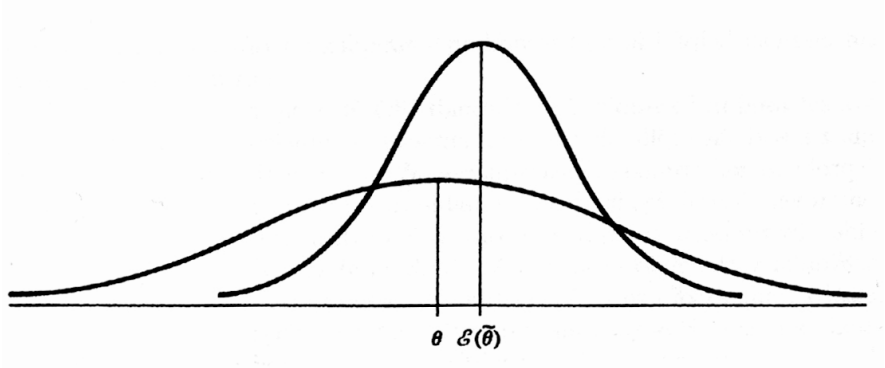


Figure 9-1. Relationship between the variance of an unbiased and a biased estimated parameters. Biased regression is based on the idea that the biased estimator will, on average be closer to the true parameter, although its mean will not be equal to the true parameter.

9:3.2 Biased estimated β's

Biased estimated β's can be obtained by using variables with the correlation transformation and adding a bias component to the correlation matrix. This is the procedure followed by ridge regression.

Correlation transformation: $$x' = \frac{1}{\sqrt{n-1}}\left(\frac{X - \bar{X}}{s_X}\right)$$

Note that the correlation transformation is equivalent to dividing the standardized variables by the square root of n-1. When variables are thus transformed, the X'X matrix becomes the correlation matrix and X'Y is the correlations of X's with Y.

$$\mathbf{r}_{xx}\,\mathbf{b} = \mathbf{r}_{xy}$$

The bias is introduced by a factor c and the identity matrix I

$$(\mathbf{r}_{xx} + c\mathbf{I})\,\mathbf{b}^R = \mathbf{r}_{xy}$$

Where b^R is the vector of biased standard partial regression coefficients.

$$\mathbf{b}^R = (\mathbf{r}_{xx} + c\mathbf{I})^{-1}\,\mathbf{r}_{xy}$$

9:3.3 How to choose c?

There is no formal objective rule to select the level of bias to be accepted. It is not possible to know the true optimum value of c. The best option is to look at the ridge trace and R^2 and select a level of c that produces acceptably stable coefficients without much loss of explanatory power. The ridge trace is a plot of the standardized partial regression coefficients (and R^2) as a function of bias.

As c increases from 0 to 1 the b^R's change at first rapidly and then slowly tend to zero. The goal is to pick a c that is in the range of stable b^R's but does not reduce R^2 excessively. One should consider the fact that it may not be possible to find such a value of c.

9:3.4 How to use the biased estimates?

Predictions with ridge estimated β's are more precise than LS β's when pattern of collinearity remains constant in the new sample of x values (and sometimes, even if it does not).

Ridge regression traces (the plot of the estimated coefficients against the bias) can be used to choose variables to drop from the final model:

1. drop X's with unstable trace that rapidly approach zero.
2. drop X's with stable but low trace.
3. drop X's with very unstable trace even if it does not approach 0.

9:3.5 Spartina example (SAS STAT, JMP does not do Ridge)

The use of ridge regression is illustrated with the Spartina data set. First, biomass is regressed against all variables. A RIDGE option is specified in the PROC REG statement. This option indicates that biases from 0 to 1 should be introduced in steps of 0.05. The OUTEST=s00.ridge indicates that the parameter values for each level of bias should be stored in file s00.ridge. I printed this file and transferred the data to Excel to standardize the coefficients and make the ridge plot. Because the ridge file has many variables and I wanted to print a complete observation in each line, I change my PAGESIZE option to about 240 columns.

```
proc reg data=s00.spartina ridge=0 to 1 by 0.05 outest=s00.ridge;
model bmss=h2s sal eh7 ph acid p k ca mg na mn zn cu nh4 / vif;
run;
quit;

proc print data=s00.ridge;
var  _ridge_ _rmse_ h2s sal eh7 ph acid p k ca mg na mn zn cu nh4 ;
run;
```

The contents of the file s00.ridge can be explored by opening the library and directory windows in SAS. This can be helpful in understanding the meaning of the variables. It is important to understand that the values listed under each variable name are the estimated regression coefficients for that variable for the different levels of bias.

Libref: S00
Dataset: RIDGE2

	Variable	Length	Label
_	_MODEL_	$8	Label of model
_	_TYPE_	$8	Type of statistics
_	_DEPVAR_	$8	Dependent variable
_	_RIDGE_	8	Ridge regression control value
_	_PCOMIT_	8	Number of principal components dropped
_	_RMSE_	8	Root mean squared error
_	INTERCEP	8	Intercept
_	H2S	8	
_	SAL	8	
_	EH7	8	
_	PH	8	
_	ACID	8	
_	P	8	
_	K	8	
_	CA	8	
_	MG	8	
_	NA	8	
_	MN	8	
_	ZN	8	
_	CU	8	
_	NH4	8	
_	BMSS	8	

Figure 9-2. Contents of the file created by the outest option when the ridge option is used in proc reg.

The interpretation of the ridge plot suggests which variables to remove from the model and how much bias to introduce. Once the level ob bias is selected, the OUTEST file can be inspected to get the values of the biased estimators. Further details about the biased estimators, and the final VIF's are not immediately available, but can be calculated with a bit of matrix algebra.

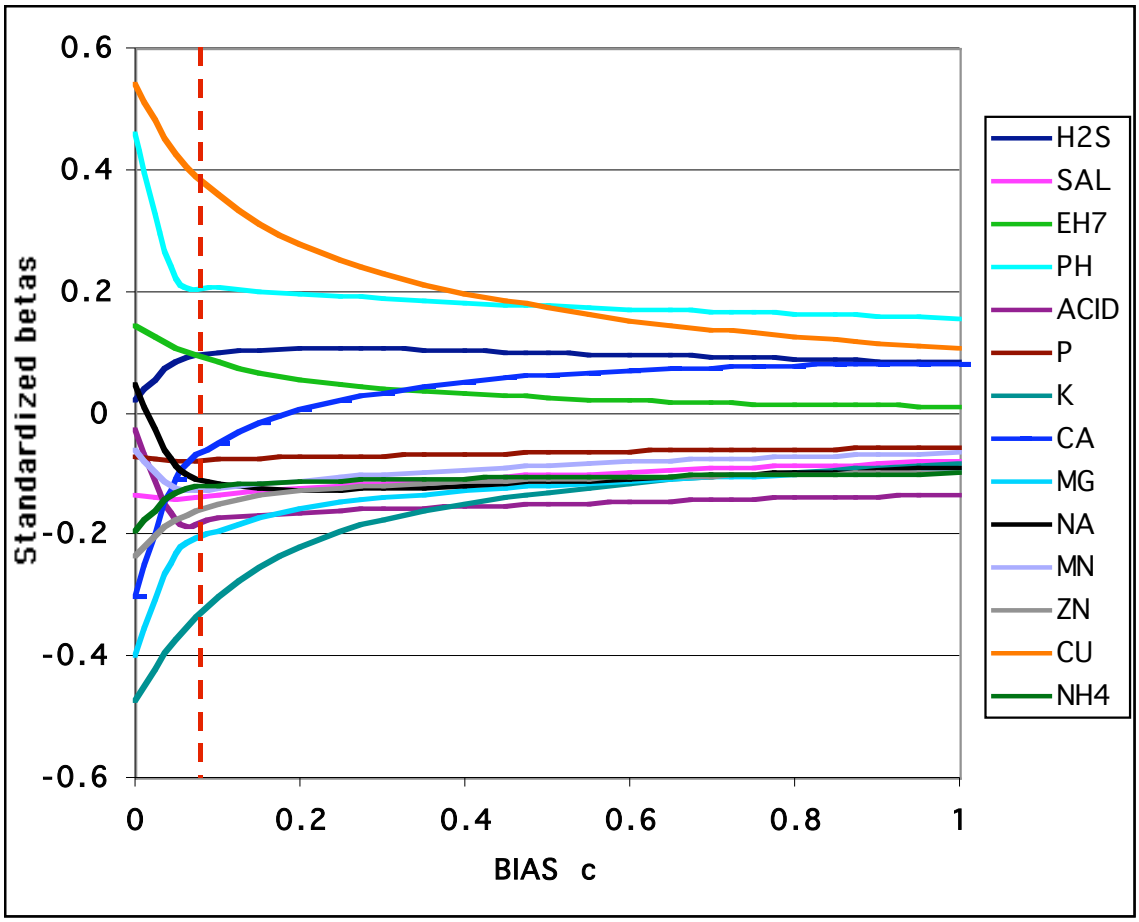


Figure 9-3. Ridge trace showing the changes in the values of estimated standardized partial regression coefficients as the bias is increased. The vertical red line indicated a good choice of bias to be selected for the final model. Alternatively, variables Cu, Eh7, K, and P could be eliminated from the model.

9:4 Principal Components Regression.

9:4.1 Purpose.

Principal components regression (PCR) is a method for obtaining estimates of the parameters of a MLR with small mean squared error in the presence of collinearity. PCR is similar to Ridge regression (RR) in that it introduces a bias in the estimated parameters for the sake of reducing their variance. Whereas RR does this by introducing a bias c in the correlation matrix of the X variables, PCR achieves the same goal by performing a MLR of Y on a subset of the principal components of X. The results of RR and PCR are not numerically the same, but the address the same problem and should be similar.

A brief overview of PCR helps to understand the big picture before going into the details. PCR regresses Y on the PC's of the X's, so the total amount of explanatory power when all PC's are included is the same as with all the original X's. Then, the PC's that reflect the greatest amount of collinearity, as identified by their low eigenvalues and their high condition number (or index), and that are not significant in the regression, are dropped from the equation and new parameters are obtained for Y as a function of the remaining PC's. Because each PC is a linear combination of all the original variables, they can be expressed in terms of the original variables by using the eigenvectors. By substitution, the regression equation of Y on a subset of PC's is finally expressed as a function of all

of the original variables. The procedure focuses on determining how many of the least important PC's to eliminate.

From a matrix algebra conceptual standpoint, PCR amounts to partitioning all of the explanatory power in the X matrix into orthogonal, uncorrelated components. Then, only those components that explain a large amount of variation in Y are kept in the model. By reconstructing the original variables based only on a subset of the components, the corresponding partial regression coefficients become biased to respond to the "relevant" and less correlated part of each X.

The partition of an X matrix into additive components is called Singular Value Decomposition of a matrix, which is explained in pp. 60-68 in Rawlings et al., (1998). The matrix X is the sum of p matrices, each of rank 1. When the component matrices are ordered in decreasing order of eigenvalue, then the sum of the first k component matrices is the best (in the least squares sense) approximation of X of rank p. This concept is mentioned here because it is central to several methods for ordination of ecological community and ecosystem information. Note that p=number of variables in this context.

9:4.2 Model.

In the following, **Z** is the matrix of standardized observations for the X variables (it does not include a column of 1's), **W** is the matrix of principal components scores for all variables and all observations, **L** is a diagonal matrix that contains the eigenvalues in the corresponding ii locations in the main diagonal and 0's everywhere else, and **V** is the matrix of eigenvectors. This organization of names and matrices is the same that is shown in the matrix worksheet of xmpl_PCR.xls. The matrix algebra involved here is very, very simple (addition and multiplication) and it makes the equations much easier to see. For a numerical example with all details, refer to xmpl_PCR.xls.

Recall that

W=ZV

Where W is the matrix of PC scores (always based on correlation matrix **R** unless otherwise specified), **Z** is the matrix of standardized variables, and **V** is the matrix of eigenvectors.

$$\mathbf{Z} = \left\{ z_{ij} = \left(X_{ij} - \bar{X}_j \right) \frac{1}{S_j} \right\} = \left\{ \left(\frac{X_{ij}}{S_j} - \frac{\bar{X}_j}{S_j} \right) \right\},$$

where {} indicates a matrix, i refers to rows of the original data matrix and j refers to the columns that are the original variables. For example, in the Body Fat data X_{11} is triceps skin fold value for the first row of the data table (see file bodyfatPCR.jmp).

$$\mathbf{V} = \left\{ v_{jk} \right\},$$

where j refers to the original variables or columns of **X**, and k refers to the principal component. Thus, in the Body Fat example, v_{12} =0.0501056 is the coefficient for triceps skin fold in the second principal component. For this data set, i=1, ..., 20 (n=20); j=1, ..., 3 (p=3, number of variables); and k=1, ..., 3 (p=3, number of principal components).

Thus, the score for observation i in principal component k is:

$$w_{ik} = \sum_{j=1}^{p} z_{ij} v_{jk} = \sum_{j=1}^{p} \left(\frac{X_{ij}}{S_j} - \frac{\bar{X}_j}{S_j} \right) v_{jk}$$

As indicated in the chapter about PCA, all of the "explanatory power" of the X-matrix (matrix of predictors) is contained in the matrix W. Thus, a regression of Y (% body fat) on **W** is just a reparameterization of the regression on **X**.

$$\widehat{\mathbf{Y}} = \{Y_i\} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta} = \left\{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j\right\} = \mathbf{1}\gamma_0 + \mathbf{W}\boldsymbol{\gamma} = \left\{\gamma_0 + \sum_{k=1}^{p} w_{ik}\gamma_k\right\}$$

Thus, each estimated Y can be expressed as a linear function of the original variables or as a linear function of the principal component scores. Because the principal component scores are themselves linear combinations of the original variables, it is possible to find equations for the relationship between the β's and the γ's. Based on the equality above, and setting p=3 as in the body fat example,

$$Y_i = \beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j = \gamma_0 + \sum_{k=1}^{p} w_{ik}\gamma_k = \gamma_0 + \sum_{k=1}^{p}\left(\sum_{j=1}^{p}\left(\frac{X_{ij}}{S_j} - \frac{\bar{X}_j}{S_j}\right)v_{jk}\right)\gamma_k =$$

$$= \gamma_0 + \sum_{k=1}^{p}\sum_{j=1}^{p}\gamma_k\left(\frac{X_{ij}}{S_j} - \frac{\bar{X}_j}{S_j}\right)v_{jk} = \gamma_0 + \sum_{k=1}^{p}\sum_{j=1}^{p}\frac{\gamma_k v_{jk}}{S_j}X_{ij} - \sum_{k=1}^{p}\sum_{j=1}^{p}\frac{\gamma_k v_{jk}}{S_j}\bar{X}_j =$$

$$\sum_{j=1}^{p} X_{ij}\left(\sum_{k=1}^{p}\frac{\gamma_k v_{jk}}{S_j}\right) + \gamma_0 - \sum_{k=1}^{p}\gamma_k\sum_{j=1}^{p}\frac{v_{jk}}{S_j}\bar{X}_j$$

The last expression is compared to the one based on the original variables to determine that the original β's are linear combinations of the γ's:

$$\beta_j = \sum_{k=1}^{p}\frac{\gamma_k v_{jk}}{S_j} \text{ and } \beta_0 = \gamma_0 - \sum_{k=1}^{p}\gamma_k\sum_{j=1}^{p}\frac{v_{jk}}{S_j}\bar{X}_j$$

In the body fat example, the set of linear combinations or customs tests necessary to recover the original partial regression coefficients (just to verify that we are applying the coefficients correctly) are given by the following table:

Predictor-coeff	β_0	β_1	β_2	β_3
Int. γ_0	1	0	0	0
PC1 γ_1	$\frac{v_{11}}{S_1}\bar{X}_1 + \frac{v_{21}}{S_2}\bar{X}_2 + \frac{v_{31}}{S_3}\bar{X}_3$	$\frac{v_{11}}{S_1}$	$\frac{v_{21}}{S_2}$	$\frac{v_{31}}{S_3}$
PC2 γ_2	$\frac{v_{12}}{S_1}\bar{X}_1 + \frac{v_{22}}{S_2}\bar{X}_2 + \frac{v_{33}}{S_3}\bar{X}_3$	$\frac{v_{12}}{S_1}$	$\frac{v_{22}}{S_2}$	$\frac{v_{32}}{S_3}$
PC3 β_3	$\frac{v_{13}}{S_1}\bar{X}_1 + \frac{v_{23}}{S_2}\bar{X}_2 + \frac{v_{33}}{S_3}\bar{X}_3$	$\frac{v_{13}}{S_1}$	$\frac{v_{23}}{S_2}$	$\frac{v_{33}}{S_3}$

The biased coefficients are obtained by dropping the rows of the table that correspond to non-significant principal components, after regressing body fat only on the significant ones. The custom tests will yield estimates of the coefficients AND of their variance, so we can assess the significance of, and construct confidence intervals for the biased coefficients.

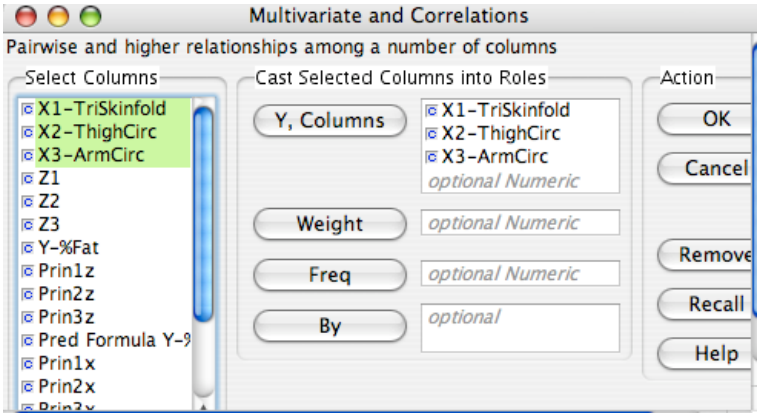
9:4.3 Assumptions.

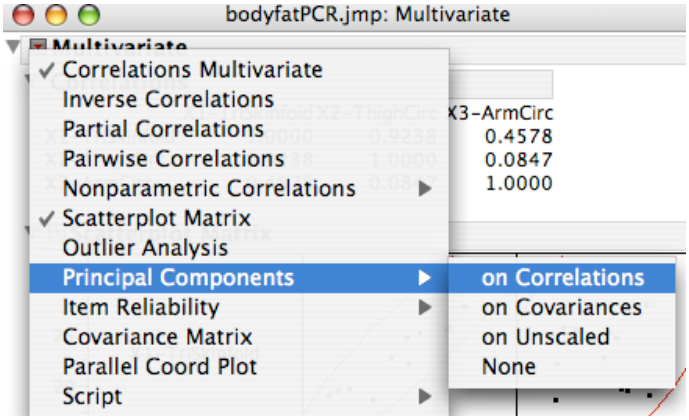
The assumptions for biased regression are the same as for MLR, and should be checked accordingly.

9:4.4 Procedure in JMP

There are three main steps to do PCR in JMP. First, obtain the Principal Components and save all the scores for all components. Use the PC's based on the correlations. Second, regress the response variable (body fat %) on all PC's and determine which PC's will be dropped. Finally, regress the response variable on the PC's that were not dropped and obtain the estimated parameters for the original variables by using the Custom Tests.

Step 1

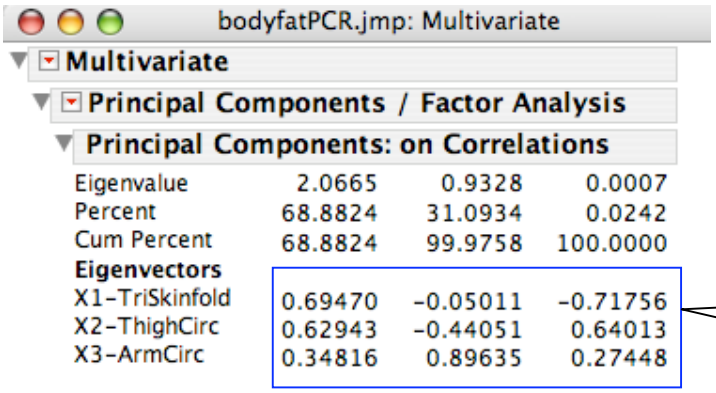




First, perform a PCA using the three predictors.

We use the principal components based on the correlation matrix.

Keep in mind that we are using the body fat example because of simplicity and because we are familiar with it. However, this is not a case where we would normally want to get biased parameter estimates, because the main goal of the real analysis is to make predictions.

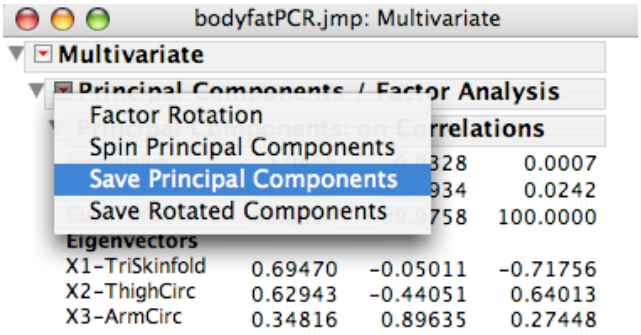


There is a very high degree of collinearity in the predictors, as indicated by the very small value for the third eigenvalue.

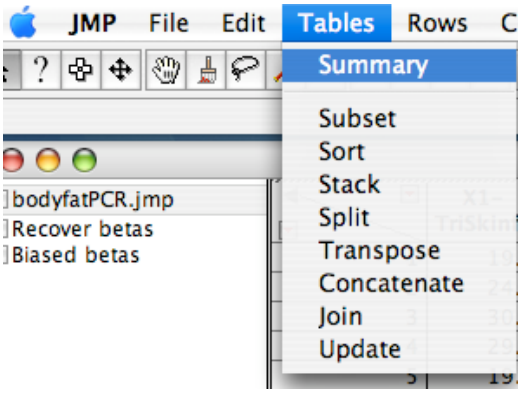
This output was copied on an Excel spreadsheet and transposed to facilitate the calculations of the coefficients for the custom tests.

This is the V matrix:

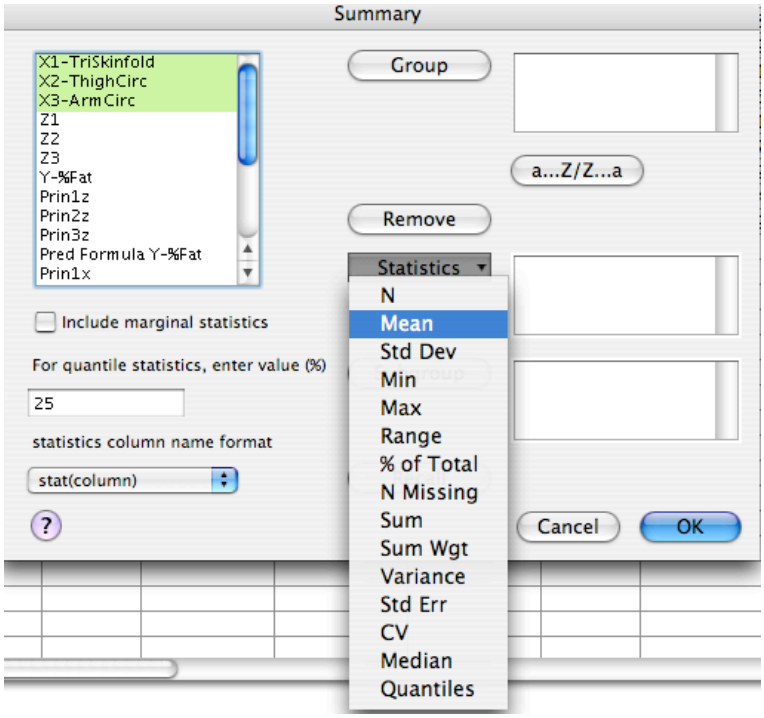
$$\begin{matrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{matrix}$$



Save all (3) principal components. That adds columns and rows to the data table. It adds the matrix W to the data table.



Use the Table->Summary command to obtain the averages and standard deviations of the predictors. These were added to the spreadsheet where the calculations to obtain the elements of the L-vectors are performed.



bodyfatPCA.xls

	A	B	C	D	E	F	G	H	I	J	K
1								Custom test Coefficients			
2	Label	Eigenvalue	%	Cum %	X1-TriSkinfo	X2-ThighCirc	X3-ArmCirc	Beta1	Beta2	Beta3	Beta0
3	PC1	2.06647	68.882	68.88	0.6946957	0.6294279	0.3481645	0.1382958	0.1202435	0.0954621	-12.2891
4	PC2	0.9328	31.093	99.98	-0.050106	-0.440509	0.8963488	-0.009975	-0.084153	0.2457671	-2.22956
5	PC3	0.00073	0.0242	100	-0.717557	0.6401347	0.2744818	-0.142847	0.1222889	0.0752593	-4.72144
6											
7				stdev	5.0232591	5.2346115	3.6471474	%fat			
8				means	25.305	51.17	27.62	20.195			

Use the spreadsheet to implement the calculations based on the formulas given above. The file bodyfatPCA.xls is available for you to double check the calculation. That file uses named ranges to facilitate the writing and reading of equations. The columns labeled Beta1, Beta2, etc., contain the elements of the L-vectors that will be applied in Custom tests to eventually obtain the biased betas.

Response Y-%Fat

Summary of Fit

RSquare	0.801359
RSquare Adj	0.764113
Root Mean Square Error	2.479981
Mean of Response	20.195
Observations (or Sum Wgts)	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	396.98461	132.328	21.5157
Error	16	98.40489	6.150	Prob > F
C. Total	19	495.38950		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	117.08469	99.7824	1.17	0.2578
X1-TriSkinfold	4.334092	3.015511	1.44	0.1699
X2-ThighCirc	-2.856848	2.582015	-1.11	0.2849
X3-ArmCirc	-2.18606	1.595499	-1.37	0.1896

Effect Tests

Effect Details

First, regress the response variable on all the predictors to get the unbiased betas and their standard deviations.

These values will be used first to make sure we apply the L-vectors correctly, and then, to compare them with the corresponding biased partial regression coefficients.

Step 2

In step 2 we regress the response variable on all PC's and determine which ones are significant and which ones will not be included in the model. By removing PC's, we remove collinearity without eliminating any of the original variables.

Response Y-%Fat

Summary of Fit

RSquare	0.801359
RSquare Adj	0.764113
Root Mean Square Error	2.479981
Mean of Response	20.195
Observations (or Sum Wgts)	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	396.98461	132.328	21.5157
Error	16	98.40489	6.150	Prob > F
C. Total	19	495.38950		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	20.195	0.554541	36.42	<.0001
Prin1z	2.935758	0.395783	7.42	<.0001
Prin2z	-1.649761	0.589084	-2.80	0.0128
Prin3z	-27.38341	21.10659	-1.30	0.2129

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Prin1z	1	1	338.39487	55.0208	<.0001
Prin2z	1	1	48.23747	7.8431	0.0128
Prin3z	1	1	10.35227	1.6832	0.2129

Effect Details

Custom Test

Parameter			
Intercept	0	0	0
Prin1z	0.13829581	0.12024347	0.09546213
Prin2z	-0.0099747	-0.0841531	0.2457671
Prin3z	-0.1428468	0.12228886	0.07525932
=	0	0	0
Value	4.3340920076	-2.856847936	-2.186060262
Std Error	3.0155113636	2.58201527	1.5954990064
t Ratio	1.4372660172	-1.106441147	-1.370142039
Prob>|t|	0.1699110654	0.2848943702	0.1895628469
SS	12.70489277	7.529277877	11.545902155

Sum of Squares	396.98461183
Numerator DF	3
F Ratio	21.515712304
Prob > F	0.0000073433

Custom Test

Parameter	
Intercept	1
Prin1z	-12.289098
Prin2z	-2.2295604
Prin3z	-4.7214447
=	0
Value	117.08469453
Std Error	99.782402994
t Ratio	1.1734002291
Prob>|t|	0.2578077977
SS	8.4681594477

Now, regress the response variable on all the principal components saved in one of the steps above.

Note that the R-square and the RMSE are the same as above.

The values listed in the Parameter estimates table are the gamma's. Note that gamma0 is always equal to the mean of the response when you use the PC's based on the correlation matrix.

The custom tests show the location of the coefficients to recover the original betas.

The tests show that the gamma for PC3 is not significant, so PC3 will be dropped to obtain the biased betas.

These values are the same as in the Parameter Estimates table of the previous analysis, which confirms that we calculated and applied the coefficients correctly.

Response Y-%Fat

Summary of Fit

RSquare	0.780461
RSquare Adj	0.754633
Root Mean Square Error	2.529324
Mean of Response	20.195
Observations (or Sum Wgts)	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	386.63234	193.316	30.2175
Error	17	108.75716	6.397	Prob > F
C. Total	19	495.38950		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	20.195	0.565574	35.71	<.0001
Prin1z	2.935758	0.403657	7.27	<.0001
Prin2z	-1.649761	0.600805	-2.75	0.0138

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Prin1z	1	1	338.39487	52.8950	<.0001
Prin2z	1	1	48.23747	7.5401	0.0138

Effect Details

Custom Test

Parameter	β^b_2	β^b_3
Intercept	0	0
Prin1z	0.12024347	0.09546213
Prin2z	-0.0841531	0.2457671
=	0	0
Value	0.4918383028	-0.125203254
Std Error	0.0700865774	0.1526032534
t Ratio	7.0175819842	-0.820449441
Prob>|t|	0.0000020705	0.4233155919
SS	315.05322727	4.3063823797

Sum of Squares	386.63233857
Numerator DF	2
F Ratio	30.217549212
Prob > F	0.0000025284

Custom Test

Parameter	β^b_0	β^b_1
Intercept	1	0
Prin1z	-12.289098	0.13829581
Prin2z	-2.2295604	-0.0099747
=	0	0
Value	-12.20457574	0.4224589352
Std Error	5.1692951733	0.0561448688
t Ratio	-2.36097482	7.5244442534
Prob>|t|	0.0304307475	8.32146e-7
SS	35.66084692	362.20780174

Sum of Squares	1322.3018981
Numerator DF	2
F Ratio	103.34552673
Prob > F	3.0676e-10

Step 3

Note that the R^2 has declined slightly. Because the PC's are orthogonal, their partial regression coefficients and SS do not change when PC's are dropped.

The biased coefficients are very different from the original ones, and they have much lower variances. When there is a large change in the partial regression coefficients like here, it is necessary to compare the coefficient of variation. For example, the CV for β^b_1 is 0.05614/0.42246=0.133, whereas for β_1 it is 3.0155/4.3341=0.696

The coefficient for X_3 is not significant. In this specific example, X_3 is almost perfectly collinear with the other two. Thus, it is likely that the elimination of X3 will work better than PCR. As an exercise, compare the results of PCR with those of eliminating X_3.

9:4.5 How to eliminate PC's.

Eliminate PC's that:

1. Have small eigenvalues and thus are causing variance inflation on the parameter estimates.
2. Do not have a significant effect on the response variable (regression coefficient is not significantly different from 0 at α=0.10).