**14.1** For the depression data set described in Appendix A, perform a principal components analysis on the last seven variables DRINK–CHRONILL (Table 3.3). Interpret the results.

**Solution**

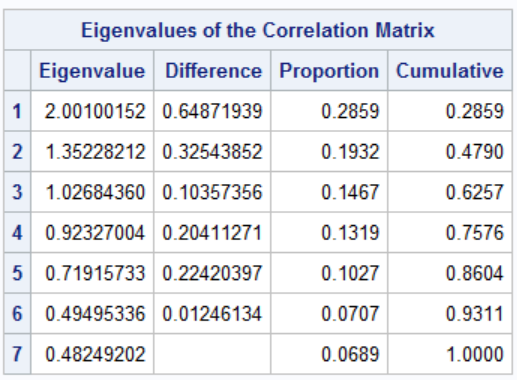
Title "PCA of last seven variables of Depression data set";

**proc** **princomp** data=depression out=depress\_pca;

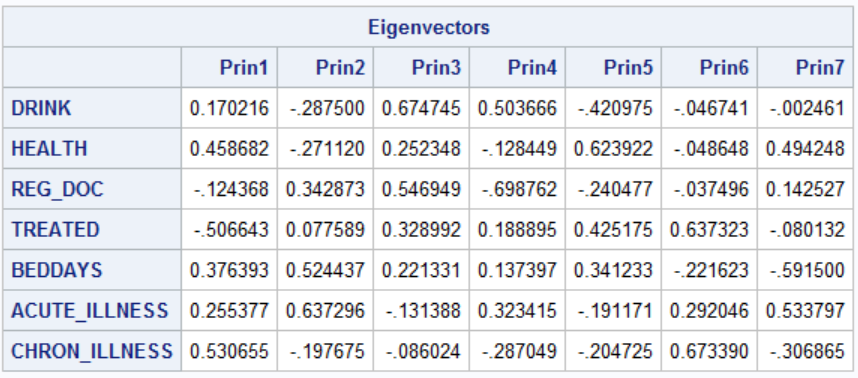
var DRINK HEALTH REG\_DOC TREATED BEDDAYS ACUTE\_ILLNESS CHRON\_ILLNESS;

**run**;

**quit**;



As can be seen from above result, Eigenvalues of first four components represents 75% of total variability.



Prin1 = 0.170216\*DRINK + 0.458682\*HEALTH - 0.124368\*REG\_DOC - 0.506643\*TREATED + 0.376393\*BEDDAYS + 0.255377\*ACUTE\_ILLNESS + 0.530655\*CHRON\_ILLNESS

Similarly, we can write equation for other principal components.

**14.6** Using the family lung function data described in Appendix A define a new variable RATIO = FEV1/FVC for the fathers. What is the correlation between RATIO and FEV1? Between RATIO and FVC? Perform a principal components analysis on FEV1 and FVC, plotting the results. Perform a principal components analysis on FEV1, FVC, and RATIO. Discuss the results.

**Solution**

**data** Lung;

set "F:\SAS\Lung";

**run**;

**data** Lung\_w\_ratio;

set Lung;

Ratio\_father = FEV1\_father / FVC\_father;

**run**;

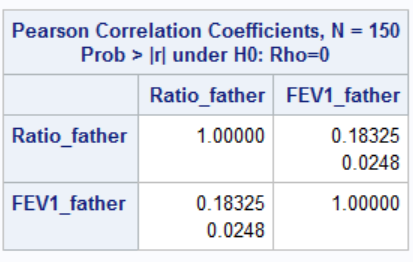
**Correlation between RATIO and FEV1**

**proc** **corr** data=Lung\_w\_ratio;

var RATIO\_father FEV1\_father;

**run**;

**quit**;



Correlation between Ratio\_father and FEV1\_father is 18%, which is significant. The ratio is positive which mean Ratio\_father and FEV1\_father is directly proportional to each other.

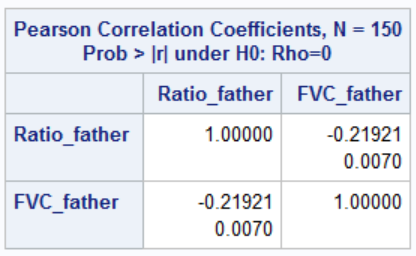
**Correlation between RATIO and FVC\_father**

**proc** **corr** data=Lung\_w\_ratio;

var RATIO\_father FVC\_father;

**run**;

**quit**;



Correlation between Ratio\_father and FVC\_father is -21%, which is also significant. The ratio is negative because Ratio\_father is inversely proportional to FVC\_father.

**Principal components analysis on FEV1 and FVC**

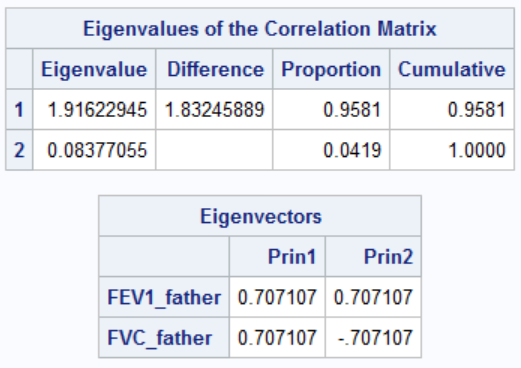
Title "PCA of FEV1\_father and FVC\_father";

**proc** **princomp** data=Lung\_w\_ratio out=Lung\_w\_pca;

var FEV1\_father FVC\_father;

**run**;

**quit**;



We can be see that 95 percent variability is achieved through first Principal Component.

Following is the computation of both Principal Component:

Prin1 = 0.707107\*FEV1\_father + 0.707107\*FVC\_father

Prin2 = 0.707107\*FEV1\_father – 0.707107\*FVC\_father

**Plotting the above result**

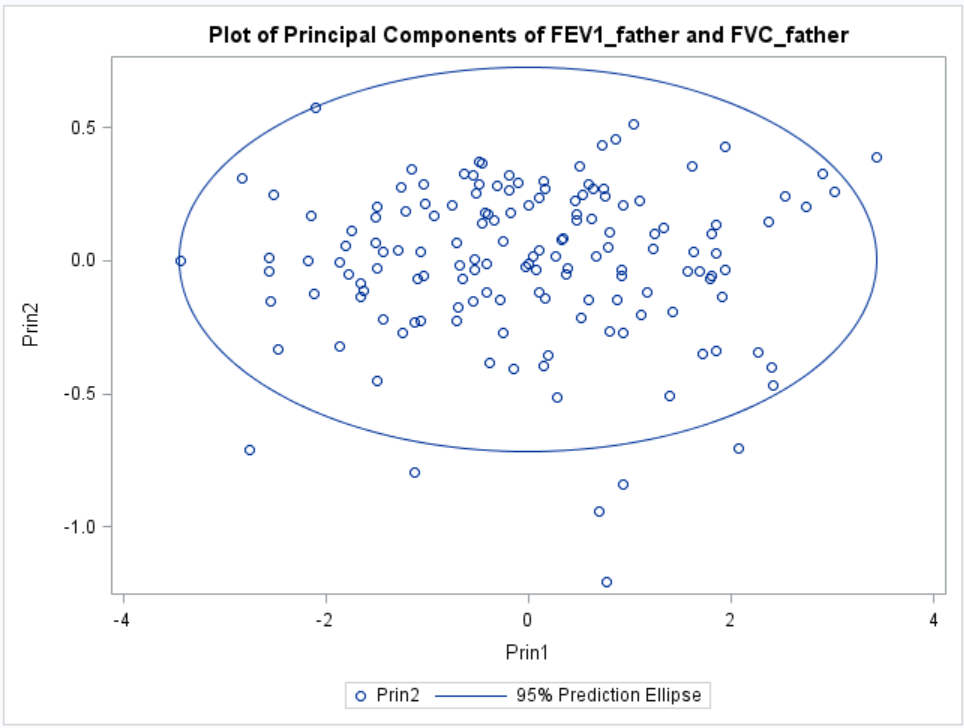
Title "Plot of Principal Components of FEV1\_father and FVC\_father";

**proc** **sgplot** data=Lung\_w\_pca;

scatter x = Prin1 y= Prin2 ;

ellipse x = Prin1 y= Prin2;

**quit**;



We can see from the above SG plot that a more variability is provided by Prin1 and less variability by Prin2.

**Principal components analysis on FEV1, FVC, and RATIO**

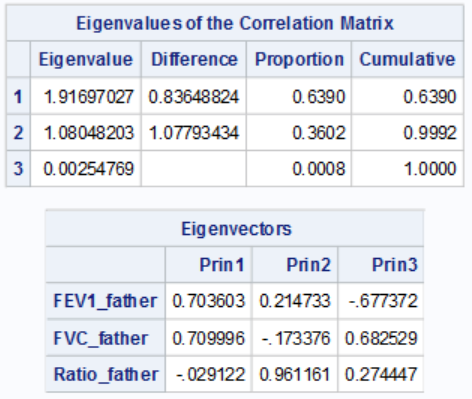
Title "PCA of FEV1\_father, FVC\_father and Ratio\_father";

**proc** **princomp** data=Lung\_w\_ratio out=Lung\_w\_ratio\_pca;

var FEV1\_father FVC\_father Ratio\_father;

**run**;

**quit**;



We can see above, 99% variability is provided by Principal Component 1 and 2.

Following is the computation of all the Principal Components:

Prin1 = 0.703603\*FEV1\_father + 0.709996\*FVC\_father – 0.029122\*Ratio\_father

Prin2 = 0.214733\*FEV1\_father – 0.173376\*FVC\_father + 0.961161\*Ratio\_father

Prin3 = -0.677372\*FEV1\_father + 0.682529\*FVC\_father + 0.274447\*Ratio\_father

**14.7** Using the family lung function data, perform a principal components analysis

on age, height, and weight for the oldest child.

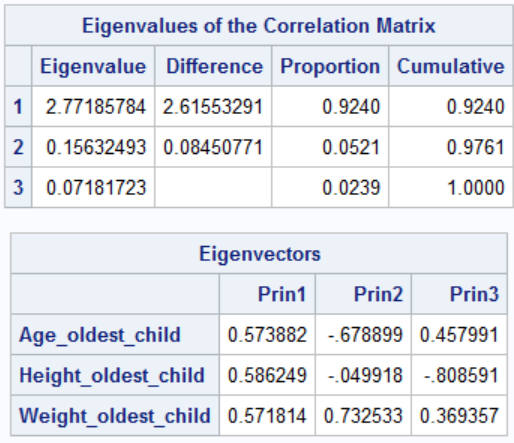
**Solution**

Title "PCA of Age, Height and Weight of Oldest Child";

**proc** **princomp** data=Lung out=Lung\_oldest\_child\_pca;

var Age\_oldest\_child Height\_oldest\_child Weight\_oldest\_child;

**run**;



We can see from above result that 92% variability is achieved through Principal component 1 itself.

Following is the computation of all the Principal Components:

Prin1 = 0.583882\*Age\_oldest\_child + 0.586249\*Height\_oldest\_child + 0.571814\*Weight\_oldest\_child

Prin2 = -0.678899\*Age\_oldest\_child - 0.049918\*Height\_oldest\_child + 0.732533\*Weight\_oldest\_child

Prin3 = 0.457991\*Age\_oldest\_child - 0.808591\*Height\_oldest\_child + 0.369357\*Weight\_oldest\_child