SAS Code:

Step 1: In fist step we do plot all variables and select few significant principal components out

/\*Principal component analysis \*/

**proc** **princomp** data = sashelp.heart out = newdset;

var AgeAtStart Height Weight Diastolic Systolic MRW ;

**run**;

Step 2: In this step, we have plotted selected principal components to group relevant variables

/\* After preliminary analysis, 3 principal components have been selected

and plotting 3 components in order to picturize visually the relations

between all 6 variables \*/

**proc** **princomp** data = sashelp.heart n=**3** plots(ncomp=**3**)= pattern out = newdset;

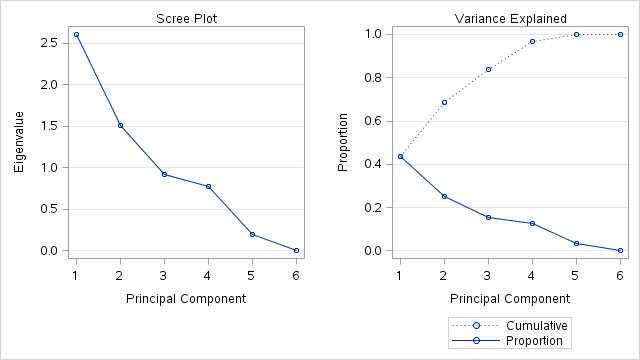
var AgeAtStart Height Weight Diastolic Systolic MRW ;

**run**;

Analysis of outputs from code:

The following Eigen values of the correlation matrix table from step 1 output (Complete SAS output is attached at the end), we have selected 3 principal components based on cumulative is 0.8383 (Cumulative > 80 %) and Eigen value >= 1 (here it is 0.92, but it is bit subjective though) and slope changed after 3rd principal component

| **Eigenvalues of the Correlation Matrix** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 2.60347936 | 1.09994784 | 0.4339 | 0.4339 |
| **2** | 1.50353152 | 0.58059105 | 0.2506 | 0.6845 |
| **3** | 0.92294047 | 0.15449450 | 0.1538 | 0.8383 |
| **4** | 0.76844597 | 0.57474136 | 0.1281 | 0.9664 |
| **5** | 0.19370461 | 0.18580654 | 0.0323 | 0.9987 |
| **6** | 0.00789807 |  | 0.0013 | 1.0000 |



Following plots are obtained for top 3 principal components and grouped manually by us based on nearby variables in all 3 diagrams (Diagram 1: Prin 1 Vs Prin2 , Diagrm2: Prin1 Vs Prin3 , Diagram 3: Prin2 Vs Prin 3)

Age at start, Diastolic and Systolic are under one group

Weight and MRW are under another group

Height under another group

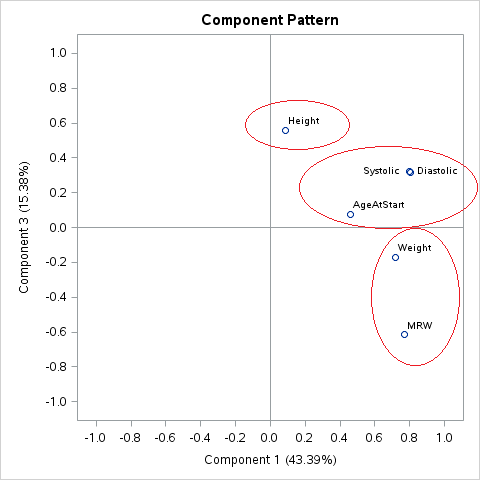
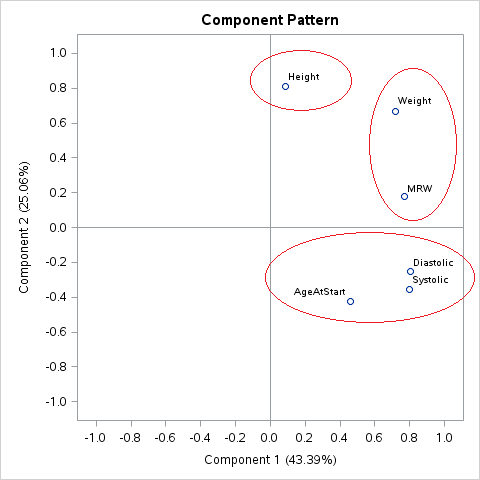


Diagram 1: Prin 1 Vs Prin2 Diagram 2: Prin 1 Vs Prin3

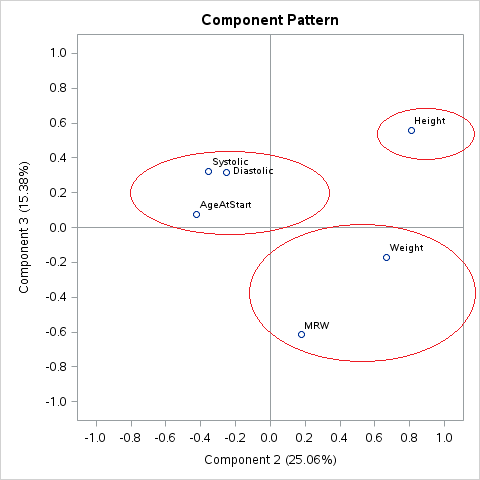


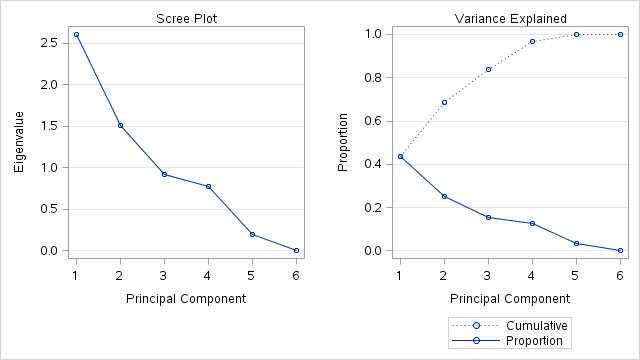
Diagram 3: Prin 2 Vs Prin 3

With this we have finished principal component analysis, Complete sas output is below

**Complete SAS Output from Step 1:**

|  |
| --- |
| **The PRINCOMP Procedure** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | **Observations** | 5199 | | --- | --- | | **Variables** | 6 | |
| | **Simple Statistics** | | | | | | | | --- | --- | --- | --- | --- | --- | --- | |  | **AgeAtStart** | **Height** | **Weight** | **Diastolic** | **Systolic** | **MRW** | | **Mean** | 44.06732064 | 64.81405078 | 153.0859781 | 85.37411041 | 136.9282554 | 119.9540296 | | **StD** | 8.57127775 | 3.58322011 | 28.9248558 | 12.97347870 | 23.7507831 | 19.9889832 | |
| | **Correlation Matrix** | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | --- | |  |  | **AgeAtStart** | **Height** | **Weight** | **Diastolic** | **Systolic** | **MRW** | | **AgeAtStart** | Age at Start | 1.0000 | -.1325 | 0.0936 | 0.2751 | 0.3798 | 0.2045 | | **Height** |  | -.1325 | 1.0000 | 0.5174 | -.0149 | -.0714 | -.1363 | | **Weight** |  | 0.0936 | 0.5174 | 1.0000 | 0.3277 | 0.2639 | 0.7673 | | **Diastolic** |  | 0.2751 | -.0149 | 0.3277 | 1.0000 | 0.7962 | 0.3852 | | **Systolic** |  | 0.3798 | -.0714 | 0.2639 | 0.7962 | 1.0000 | 0.3627 | | **MRW** | Metropolitan Relative Weight | 0.2045 | -.1363 | 0.7673 | 0.3852 | 0.3627 | 1.0000 | |
| | **Eigenvalues of the Correlation Matrix** | | | | | | --- | --- | --- | --- | --- | |  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** | | **1** | 2.60347936 | 1.09994784 | 0.4339 | 0.4339 | | **2** | 1.50353152 | 0.58059105 | 0.2506 | 0.6845 | | **3** | 0.92294047 | 0.15449450 | 0.1538 | 0.8383 | | **4** | 0.76844597 | 0.57474136 | 0.1281 | 0.9664 | | **5** | 0.19370461 | 0.18580654 | 0.0323 | 0.9987 | | **6** | 0.00789807 |  | 0.0013 | 1.0000 | |
| | **Eigenvectors** | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | --- | |  |  | **Prin1** | **Prin2** | **Prin3** | **Prin4** | **Prin5** | **Prin6** | | **AgeAtStart** | Age at Start | 0.284692 | -.346542 | 0.079844 | 0.883904 | 0.105765 | 0.003232 | | **Height** |  | 0.052866 | 0.659624 | 0.578329 | 0.191418 | -.030664 | 0.435953 | | **Weight** |  | 0.444517 | 0.544805 | -.178226 | 0.090392 | -.011488 | -.682294 | | **Diastolic** |  | 0.498559 | -.205878 | 0.331138 | -.353652 | 0.688543 | 0.015477 | | **Systolic** |  | 0.493448 | -.291685 | 0.333109 | -.217632 | -.716151 | -.015212 | | **MRW** | Metropolitan Relative Weight | 0.475928 | 0.143268 | -.637794 | -.038314 | -.027901 | 0.586463 | |



|  |
| --- |
| **Complete SAS Output from Step 2:**  **The PRINCOMP Procedure** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| | **Observations** | 5199 | | --- | --- | | **Variables** | 6 | |
| | **Simple Statistics** | | | | | | | | --- | --- | --- | --- | --- | --- | --- | |  | **AgeAtStart** | **Height** | **Weight** | **Diastolic** | **Systolic** | **MRW** | | **Mean** | 44.06732064 | 64.81405078 | 153.0859781 | 85.37411041 | 136.9282554 | 119.9540296 | | **StD** | 8.57127775 | 3.58322011 | 28.9248558 | 12.97347870 | 23.7507831 | 19.9889832 | |
| | **Correlation Matrix** | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | --- | |  |  | **AgeAtStart** | **Height** | **Weight** | **Diastolic** | **Systolic** | **MRW** | | **AgeAtStart** | Age at Start | 1.0000 | -.1325 | 0.0936 | 0.2751 | 0.3798 | 0.2045 | | **Height** |  | -.1325 | 1.0000 | 0.5174 | -.0149 | -.0714 | -.1363 | | **Weight** |  | 0.0936 | 0.5174 | 1.0000 | 0.3277 | 0.2639 | 0.7673 | | **Diastolic** |  | 0.2751 | -.0149 | 0.3277 | 1.0000 | 0.7962 | 0.3852 | | **Systolic** |  | 0.3798 | -.0714 | 0.2639 | 0.7962 | 1.0000 | 0.3627 | | **MRW** | Metropolitan Relative Weight | 0.2045 | -.1363 | 0.7673 | 0.3852 | 0.3627 | 1.0000 | |
| | **Eigenvalues of the Correlation Matrix** | | | | | | --- | --- | --- | --- | --- | |  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** | | **1** | 2.60347936 | 1.09994784 | 0.4339 | 0.4339 | | **2** | 1.50353152 | 0.58059105 | 0.2506 | 0.6845 | | **3** | 0.92294047 |  | 0.1538 | 0.8383 | |
| | **Eigenvectors** | | | | | | --- | --- | --- | --- | --- | |  |  | **Prin1** | **Prin2** | **Prin3** | | **AgeAtStart** | Age at Start | 0.284692 | -.346542 | 0.079844 | | **Height** |  | 0.052866 | 0.659624 | 0.578329 | | **Weight** |  | 0.444517 | 0.544805 | -.178226 | | **Diastolic** |  | 0.498559 | -.205878 | 0.331138 | | **Systolic** |  | 0.493448 | -.291685 | 0.333109 | | **MRW** | Metropolitan Relative Weight | 0.475928 | 0.143268 | -.637794 | |

