STA 9705: HW6

Tanay Mukherjee

 # 12.6

The variance of y_1, y_2, x_1, x_2, and x_3 on diag(**S**)'= (0.016, 70.559, 1106.414, 2381.883, 2136.396).

Covariance Matrix					
	Y1	Y2	X1	X2	X3
Y1	0.016182	0.216029	0.787169	-0.213845	2.189072
Y2	0.216029	70.558937	26.228986	-23.956039	-20.841546
X1	0.787169	26.228986	1106.413527	396.732367	108.383575
X2	-0.213845	-23.956039	396.732367	2381.882609	1142.637681
X3	2.189072	-20.841546	108.383575	1142.637681	2136.396135

Eigen Values of S

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	3466.18182	2201.71128	0.6086	0.6086
2	1264.47054	369.20218	0.2220	0.8306
3	895.26836	825.93313	0.1572	0.9878
4	69.33524	69.32382	0.0122	1.0000
5	0.01142		0.0000	1.0000

Eigen Values of R

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	1.71724028	0.48339142	0.3434	0.3434
2	1.23384887	0.27364618	0.2468	0.5902
3	0.96020268	0.17357027	0.1920	0.7823
4	0.78663241	0.48455665	0.1573	0.9396
5	0.30207576		0.0604	1.0000

From the table above we can look at the proportion column and see for S, 2 principal components account for ~ 83% of the variances whereas for R, 3 principal components account for ~ 78% of the variances. The variances for S are significantly distanced as compared to R.

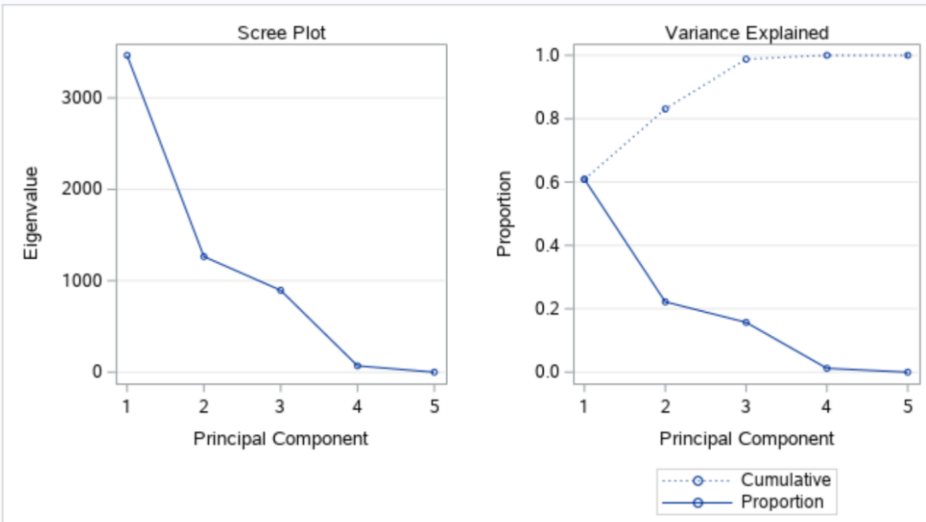
We look at the average eigen values of S and R and also the scree plot. Based on that we can say that based on the λ_{avg} for S and R is 1139.05 and 1 respectively. That makes λ_1 and λ_2 for S relevant and λ_1, λ_2 and λ_3 for R relevant. The values of which are highlighted above. Also, from the scree plot we look for the elbow point and that also gives us the same result, that is, 2 principal components for S and 3 for R.

The first principal component is largely a weighted average of the last two variables, x2 and x3, which have the largest variances. The second and third components represent contrasts in the last three variables and could be described as "shape" components. The fourth and fifth components are associated uniquely with y2 and y1, respectively. These components are "variable specific". As expected, the principal components of R show an entirely different pattern. All five variables contribute to the first three components of R, whereas in S, y1 and y2 have small variances and contribute almost nothing to the first three components.

Therefore, we conclude that we will go with R over S.

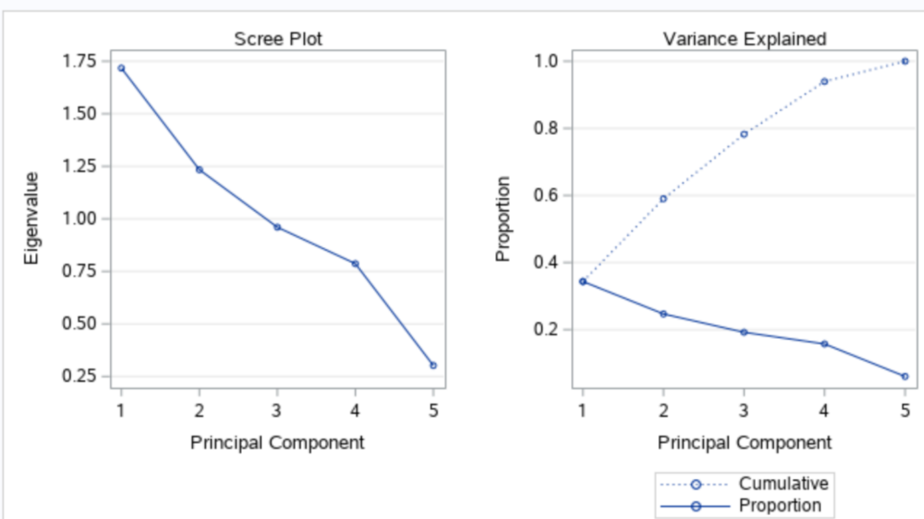
Eigen vectors for S and the respective scree plot is below:

Eigenvectors					
	Prin1	Prin2	Prin3	Prin4	Prin5
Y1	0.000400	-.000780	0.001791	0.002854	0.999994
Y2	-.008043	0.016602	0.028590	0.999417	-.002887
X1	0.154728	0.638220	0.753510	-.030915	-.000825
X2	0.742970	0.427940	-.514469	0.013591	0.000919
X3	0.651145	-.639739	0.408318	0.004182	-.001502



Eigen vectors for R and the respective scree plot is below:

Eigenvectors					
	Prin1	Prin2	Prin3	Prin4	Prin5
Y1	0.415530	0.529505	-.418127	-.399324	0.461160
Y2	0.073522	0.684079	0.155870	0.701202	-.103204
X1	0.363957	0.197448	0.762040	-.435645	-.240954
X2	0.542196	-.425870	0.249039	0.385717	0.560233
X3	0.628879	-.176947	-.397680	0.101449	-.636208



 # 12.9

The variance of y_1, y_2, y_3, y_4, y_5, y_6 on diag(**S**)'= (0.691, 5.401, 2006682.353, 90.290, 56.374,18.078).

Covariance Matrix						
	Y1	Y2	Y3	Y4	Y5	Y6
Y1	0.691	1.494	325.541	1.535	0.422	-0.268
Y2	1.494	5.401	1015.529	4.880	1.374	1.292
Y3	325.541	1015.529	2006682.353	10620.235	6465.529	406.706
Y4	1.535	4.880	10620.235	90.290	3.753	3.064
Y5	0.422	1.374	6465.529	3.753	56.374	0.579
Y6	-0.268	1.292	406.706	3.064	0.579	18.078

Eigen Values of S

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	2006760.04	2006694.71	1.0000	1.0000
2	65.33	47.28	0.0000	1.0000
3	18.06	11.47	0.0000	1.0000
4	6.59	3.66	0.0000	1.0000
5	2.93	2.70	0.0000	1.0000
6	0.23		0.0000	1.0000

Eigen Values of R

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.42446754	1.02153340	0.4041	0.4041
2	1.40293413	0.37415265	0.2338	0.6379
3	1.02878148	0.10851318	0.1715	0.8094
4	0.92026830	0.72077953	0.1534	0.9627
5	0.19948877	0.17542900	0.0332	0.9960
6	0.02405978		0.0040	1.0000

From the table above we can look at the proportion column and see for S, 1 principal components account for ~ 99.99% of the variances whereas for R, 3 principal components account for ~ 80% of the variances. We should definitely go for R in this case.

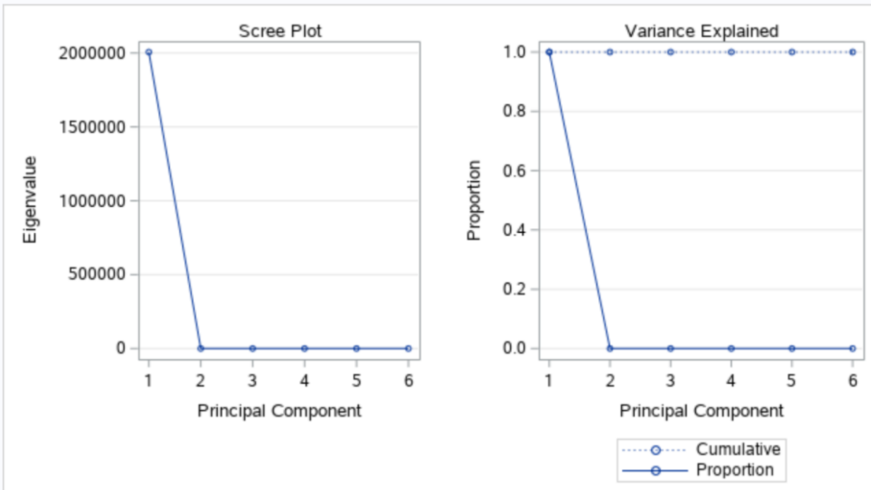
We look at the average eigen values of S and R and also the scree plot. Based on that we can say that based on the λ_{avg} for S and R is 334475.53 and 1 respectively. That makes λ_1 only for S relevant and λ_1, λ_2 and λ_3 for R relevant. The values of which are highlighted above. Also, from the scree plot we look for the elbow point and that also gives us the same result, that is, 2 principal components for S and 3 for R.

The first principal component is largely a weighted average of the y3. The second principal component contrasts with y4 and y5 and third component represent contrasts for y6. Thus, S accounts for most of the variance and y3 would essentially constitute that single principal component. As expected, the principal components of R show an entirely different pattern. The principal components from R, on the other hand, are not dominated by y3.

Therefore, we conclude that we will go with R over S.

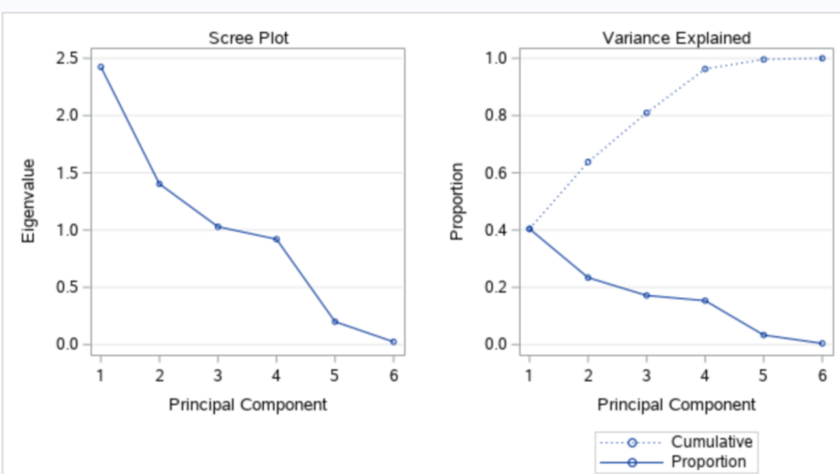
Eigen vectors for S and the respective scree plot is below:

Eigenvectors						
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
Y1	0.000162	0.005133	-.013633	0.229244	0.156517	0.960592
Y2	0.000506	0.017309	0.078667	0.748028	0.598603	-.275027
Y3	0.999981	-.001403	-.000210	0.003358	-.005040	-.000145
Y4	0.005292	0.697921	-.017368	-.452870	0.554342	0.013776
Y5	0.003222	-.715514	0.019480	-.423930	0.554718	0.014885
Y6	0.000203	0.024856	0.996466	-.055523	-.046299	0.034803



Eigen vectors for R and the respective scree plot is below:

Eigenvectors						
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
Y1	0.424249	-.560579	-.149785	0.068485	0.691325	0.026619
Y2	0.446357	-.527912	0.087368	0.148804	-.699772	0.051179
Y3	0.563341	0.387433	-.050540	-.071750	-.007498	-.724420
Y4	0.454229	0.267193	0.165534	-.629081	0.015044	0.546730
Y5	0.302916	0.425266	-.295999	0.683156	0.011125	0.415873
Y6	0.073105	0.068656	0.923233	0.324922	0.178841	-.004875



 # 13.8 (a)

The output for each rotation is given below as obtained from SAS:

Principal Component Loadings

Factor Pattern			
		Factor1	Factor2
Y1	8 yr	0.94913	-0.29528
Y2	8 1/2 yr	0.97386	-0.19285
Y3	9 yr	0.97822	0.17114
Y4	9 1/2 yr	0.94292	0.31886

Variance Explained by Each Factor	
Factor1	Factor2
3.6952471	0.2553458

Varimax Rotated Laodings

Rotated Factor Pattern			
		Factor1	Factor2
Y1	8 yr	0.88388	0.45475
Y2	8 1/2 yr	0.82971	0.54514
Y3	9 yr	0.57766	0.80778
Y4	9 1/2 yr	0.44893	0.88838

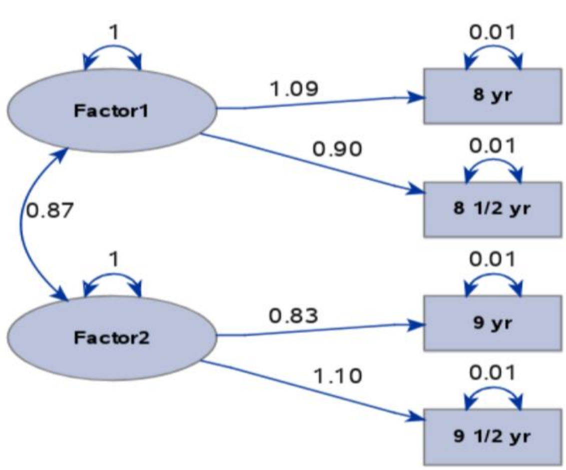
Variance Explained by Each Factor	
Factor1	Factor2
2.0048990	1.9456939

The Harris–Kaiser orthoblique rotation produced loadings for which the variables have a complexity of 1. These oblique loadings provide a much cleaner simple structure than that given by the varimax loadings. For interpretation, we see that one factor represents variables 1 and 2, and the other factor represents variables 3 and 4. This same clustering of variables can be deduced from the varimax loadings if we simply use the larger of the two loadings for each variable.

Orthoblique Pattern Loadings

Rotated Factor Pattern (Standardized Regression Coefficients)			
		Factor1	Factor2
Y1	8 yr	1.08683	-0.10827
Y2	8 1/2 yr	0.89922	0.10589
Y3	9 yr	0.18786	0.82522
Y4	9 1/2 yr	-0.12067	1.09865

Variance Explained by Each Factor Eliminating Other Factors	
Factor1	Factor2
0.49376765	0.46261349



13.12

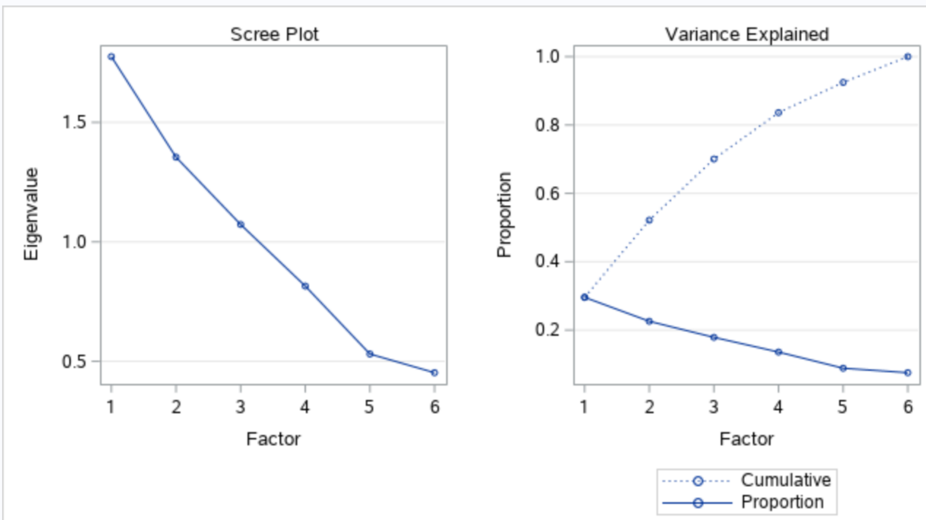
(a)

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Community Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 6 Average = 1				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	1.77512769	0.42071182	0.2959	0.2959
2	1.35441587	0.28176535	0.2257	0.5216
3	1.07265052	0.25785471	0.1788	0.7004
4	0.81479582	0.28418301	0.1358	0.8362
5	0.53061281	0.07821551	0.0884	0.9246
6	0.45239730		0.0754	1.0000

3 factors will be retained by the NFACTOR criterion.



Three of the eigen values are greater than one but it takes 4 eigen values to constitute more than 80% in proportion. Also, from the scree plot we can see the elbow point says the value of ‘m’ should be 4.

(b) The value of loadings is below:

1. Principal Component Loadings

Factor Pattern			
	Factor1	Factor2	Factor3
X1	0.53612	0.46140	0.47826
X2	-0.12945	0.86962	-0.18155
X3	0.51350	-0.25389	-0.44841
X4	0.72391	-0.36597	-0.11031
X5	-0.41554	-0.41424	0.64920
X6	0.71452	0.12367	0.41982

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
1.7751277	1.3544159	1.0726505

Final Community Estimates: Total = 4.202194					
X1	X2	X3	X4	X5	X6
0.72905685	0.80595890	0.52921335	0.67014233	0.76573408	0.70208857

2. Varimax Rotated Loadings

Orthogonal Transformation Matrix			
	1	2	3
1	0.72707	0.68234	0.07603
2	-0.41710	0.35103	0.83833
3	-0.54534	0.64124	-0.53983

Rotated Factor Pattern			
	Factor1	Factor2	Factor3
X1	-0.06347	0.83447	0.16939
X2	-0.35783	0.10052	0.81720
X3	0.72378	-0.02628	0.06826
X4	0.73913	0.29475	-0.19222
X5	-0.48338	-0.01266	-0.72932
X6	0.23898	0.80017	-0.06863

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
1.4930209	1.4344375	1.2747357

Final Community Estimates: Total = 4.202194					
X1	X2	X3	X4	X5	X6
0.72905685	0.80595890	0.52921335	0.67014233	0.76573408	0.70208857

(c) The value of loadings is below:

1. Principal Component Loadings

Factor Pattern			
	Factor1	Factor2	Factor3
X1	0.40292	0.31201	0.22658
X2	-0.10624	0.56866	-0.10018
X3	0.34267	-0.13858	-0.19721
X4	0.55861	-0.24673	-0.08970
X5	-0.28641	-0.24640	0.32767
X6	0.55571	0.08897	0.19721

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
0.99395196	0.56942354	0.25456943

Final Communality Estimates: Total = 1.817945					
X1	X2	X3	X4	X5	X6
0.31103433	0.34469386	0.17551746	0.38096675	0.25010590	0.35562663

2. Varimax Rotated Loadings

Orthogonal Transformation Matrix			
	1	2	3
1	0.72184	0.68616	0.09013
2	-0.46399	0.38321	0.79867
3	-0.51348	0.61833	-0.59499

Rotated Factor Pattern			
	Factor1	Factor2	Factor3
X1	0.02973	0.53614	0.15069
X2	-0.28910	0.08307	0.50419
X3	0.41291	0.06009	0.03754
X4	0.56377	0.23329	-0.09333
X5	-0.26067	-0.08834	-0.41756
X6	0.25859	0.53735	0.00381

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
0.70761307	0.64892094	0.46141092

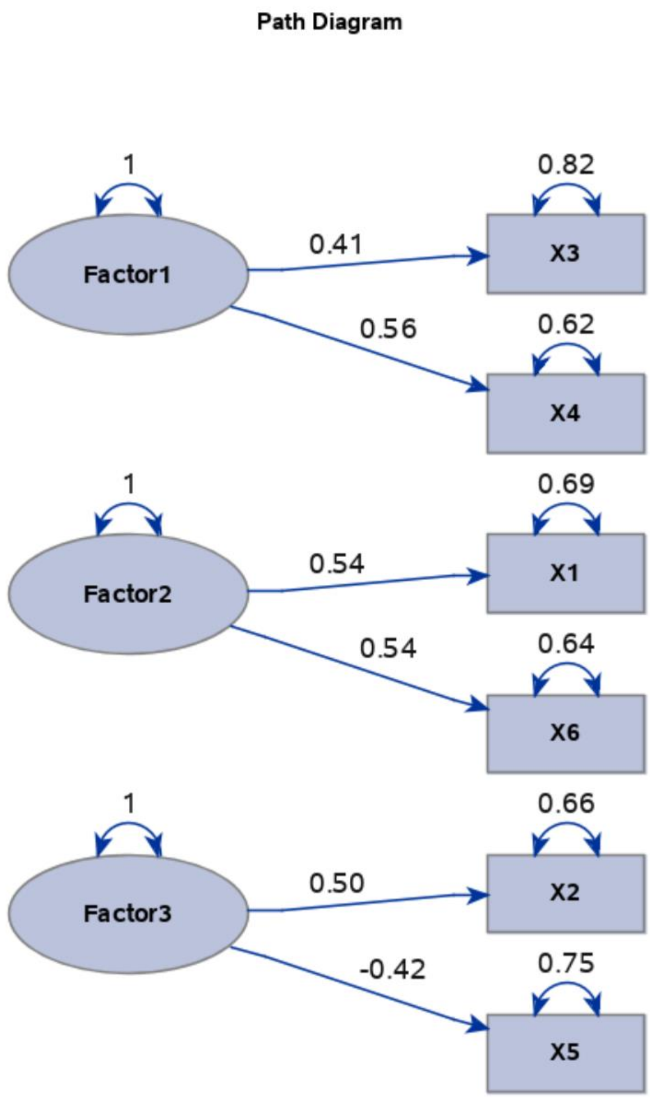
Final Communality Estimates: Total = 1.817945					
X1	X2	X3	X4	X5	X6
0.31103433	0.34469386	0.17551746	0.38096675	0.25010590	0.35562663

(d) The pattern of loadings is similar in parts (b) and (c), and the interpretation of the three factors would be the same. This means, the interpretation of the three factors in terms of variables are like:

factor 1 represents x3 and x4,

factor 2 represents x1 and x6, and

factor 3 represents x2 and x5.



APPENDIX:

This section will have the entire SAS code.

 # 12.6

```
DATA work.DIABETES;

  INFILE "/folders/myfolders/data/T3_4_DIABETES.dat";
  INPUT OBS Y1 Y2 X1 X2 X3;

TITLE "HW6 Q-12.6";

PROC PRINCOMP COV;
VAR Y1 Y2 X1 X2 X3;
RUN;

PROC PRINCOMP;
VAR Y1 Y2 X1 X2 X3;
RUN;
```

 # 12.9

```
DATA work.HEMATOL;

  INFILE "/folders/myfolders/data/T4_3_HEMATOL.dat";
  INPUT Y1 Y2 Y3 Y4 Y5 Y6;

TITLE "HW6 Q-12.9";

PROC PRINCOMP COV;
VAR Y1 Y2 Y3 Y4 Y5 Y6;
RUN;

PROC PRINCOMP;
VAR Y1 Y2 Y3 Y4 Y5 Y6;
RUN;
```

 # 13.8 (a)

```
DATA work.BONE;

  INFILE "/folders/myfolders/data/T3_6_BONE.dat";
  INPUT OBS Y1 Y2 Y3 Y4;

TITLE "HW6 Q-13.8 (a)";



LABEL Y1='8 yr' Y2='8 1/2 yr' Y3='9 yr' Y4='9 1/2 yr';

PROC FACTOR METHOD=PRIN NFACT=2 ROTATE=VARIMAX PLOT = ALL;
VAR Y1-Y4;
RUN;

PROC FACTOR METHOD=PRIN NFACT=2 ROTATE=HK PLOT = ALL;
VAR Y1-Y4;
RUN;
```

 # 13.12

```
DATA work.PILOT;

  INFILE "/folders/myfolders/data/T5_6_PILOT.dat";
  INPUT OBS X1 X2 X3 X4 X5 X6;

TITLE "HW6 Q-13.12";



PROC FACTOR METHOD=PRIN NFACT=3 ROTATE=VARIMAX PLOT = ALL;
VAR X1-X6;

RUN;

PROC FACTOR METHOD=PRIN PRIORS=SMC NFACT=3 ROTATE=VARIMAX PLOT = ALL;
VAR X1-X6;
RUN;RUN;
```