Nithin Das, 10422784, Date: 12/02/2019

Assignment 6

QUESTION 1

Q) For the depression data set described in Appendix A, perform a principal components analysis on the last seven variables DRINK–CHRONILL (Table3.3).Interpret the results.

Solution:

The required columns on which we perform PCA are:

"DRINK","HEALTH","REGDOC","TREAT","BEDDAYS","ACUTEILL", "CHRONILL"

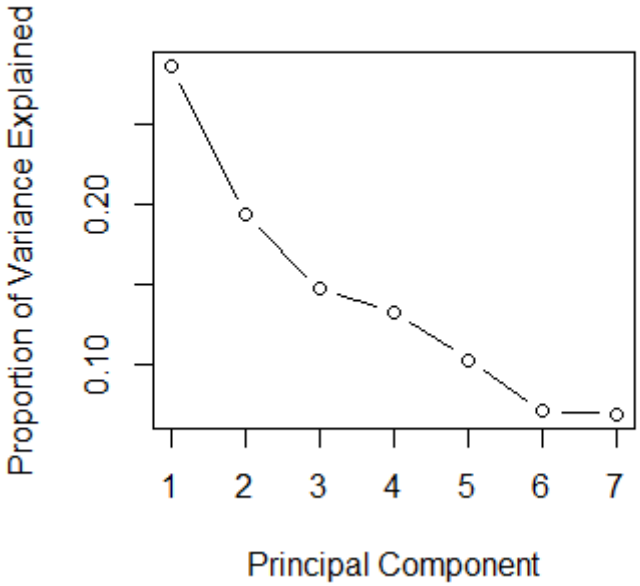
```
Importance of components:
                          PC1    PC2    PC3    PC4    PC5     PC6     PC7
Standard deviation     1.4146 1.1629 1.0133 0.9609 0.8480 0.70353 0.69462
Proportion of Variance 0.2859 0.1932 0.1467 0.1319 0.1027 0.07071 0.06893
Cumulative Proportion  0.2859 0.4790 0.6257 0.7576 0.8604 0.93107 1.00000
```

The proportion of variance after applying PCA is given below:

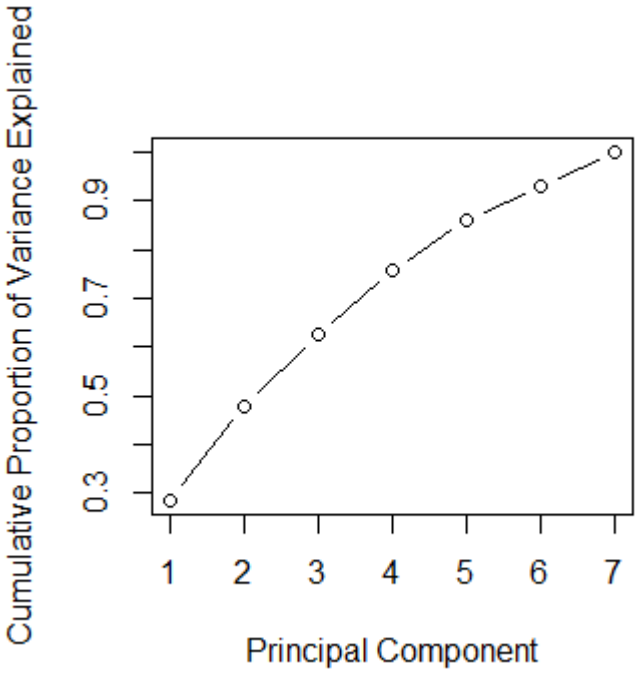
"DRINK", "HEALTH", "REGDOC", "TREAT", "BEDDAYS", "ACUTEILL", "CHRONILL"

0.28585736, 0.19318316, 0.14669194, 0.13189572, 0.10273676, 0.07070762, 0.06892743

This means first component, ‘’DRINK” explains 28.5% variance, “HEALTH” explains 19.3% variance and



so on.



The cumulative proportion chart shows that First 6 variables explain almost 94% of total variance.

Therefore, we can remove “CHRONILL” without compromising on the explained variance

Question 2

Using the family lung functiondata described in AppendixA 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 componentsanalysisonFEV1andFVC,plottingtheresults.PerformaprincipalcomponentsanalysisonFEV1,FVC,andRATIO.Discusstheresults

Solution:

Correlation between "RATIO" and "FFEV1" = 0.1832515 or 18.32%

Correlation between "RATIO" and "FFVC" = 0.2192122 or 21.92%

PCA on FEV1 and FFVC

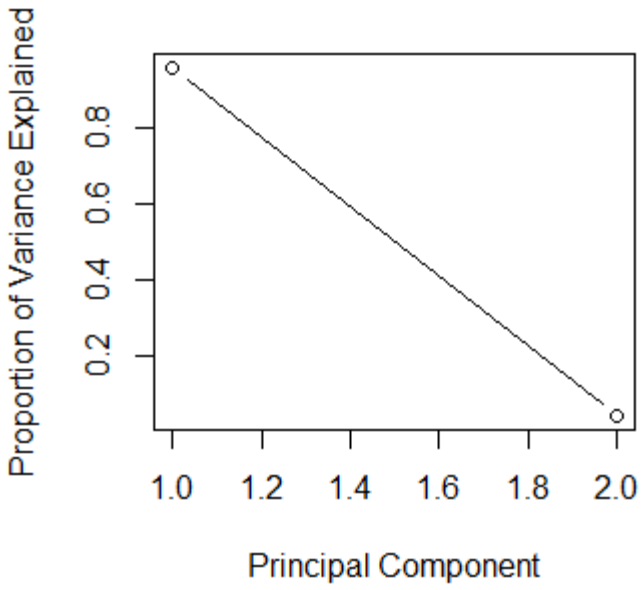
```
Importance of components:
                          PC1     PC2
Standard deviation     1.3843 0.28943
Proportion of Variance 0.9581 0.04189
Cumulative Proportion  0.9581 1.00000
```

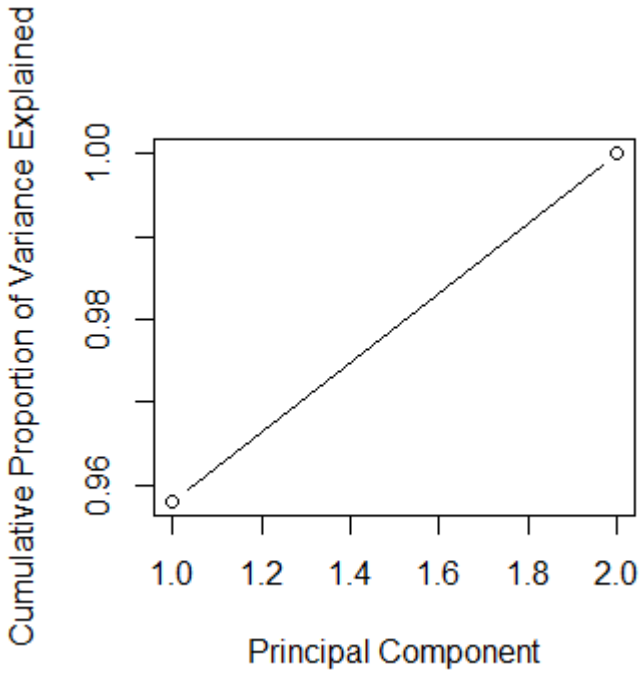
The proportion of variance after applying PCA is given below:

FFEV1, FVC

0.95811472, 0.04188528

This show that FFEV1 explains 95.81% of total variance.





Since FFEV1 explains 95% of variance, we can remove FVC without compromising on explained variance.

PCA on FEV1, FFVC and RATIO

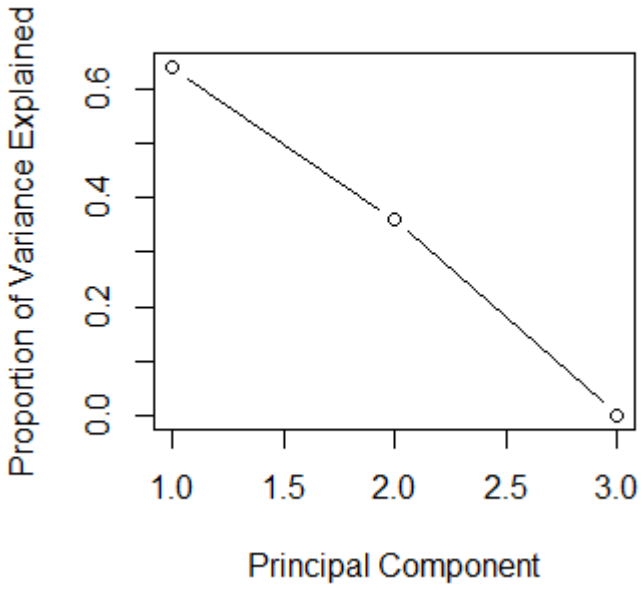
The proportion of variance after applying PCA is given below:

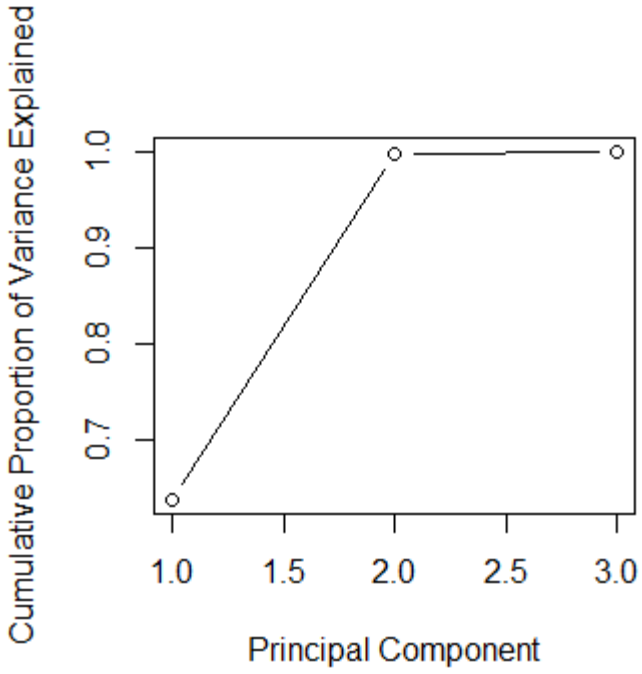
```
Importance of components:
                          PC1    PC2     PC3
Standard deviation     1.385 1.0395 0.05047
Proportion of Variance 0.639 0.3602 0.00085
Cumulative Proportion  0.639 0.9991 1.00000
```

'FFEV1', 'FFVC', 'RATIO'

0.6389900905, 0.3601606781, 0.0008492313

This shows that FFEV1 and FFVC explains 99.9% of total variance





The cumulative proportion graphs show the importance of FFEV1 and FVC which constitute 99.9% variance. Therefore, we can remove RATIO without compromising on explained variance.

QUESTION 3

Using the family lung function data, perform a principal components analysis on age, height, and weight for the oldest child.

Solution:

PCA on 'OCAGE', 'OCHEIGHT', 'OCWEIGHT'

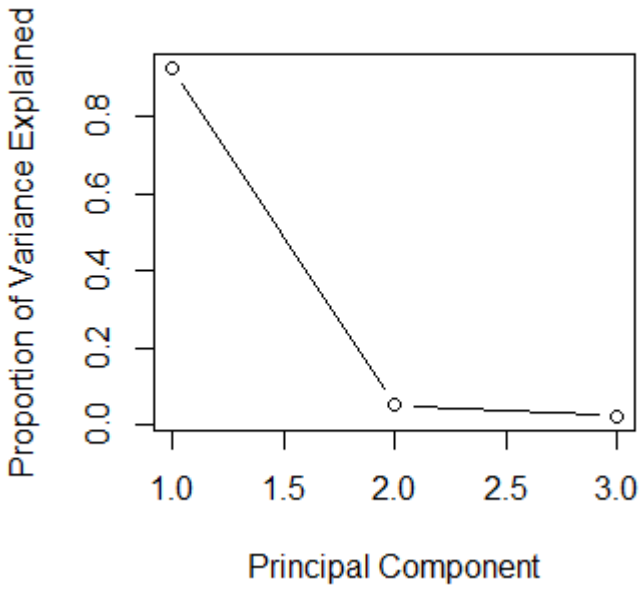
The proportion of variance after applying PCA is given below:

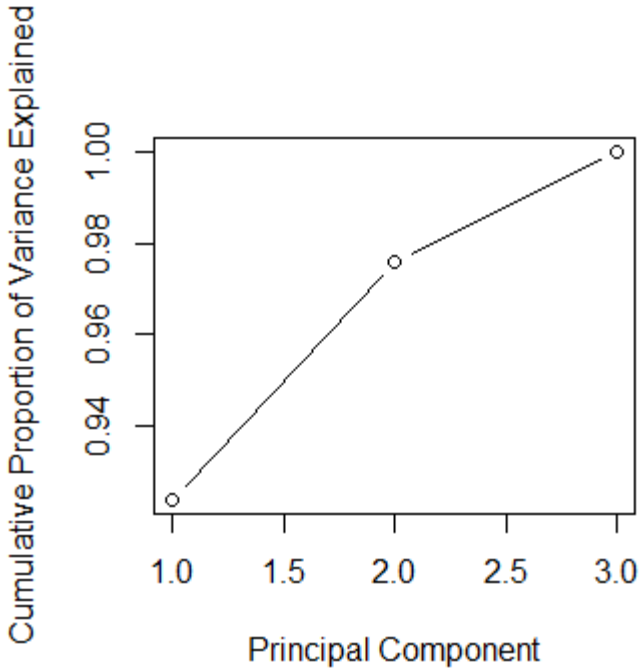
```
Importance of components:
                          PC1     PC2     PC3
Standard deviation     1.665 0.39538 0.26799
Proportion of Variance 0.924 0.05211 0.02394
```

```
Cumulative Proportion  0.924 0.97606 1.00000
```

‘OCAGE',	'OCHEIGHT',	'OCWEIGHT’
0.92395261	0.05210831	0.02393908

This shows that OCAGE explains 92.3% of total variance





The cumulative proportion of variance shows that 'OCAGE' explains 92% of total variance. Therefore, we can remove OCHEIGHT and OCWEIGHT without compromising on explained variance.