# Question 2

**data** census;

input population professional over16 gov medhomevalue;

datalines;

2.67 5.71 69.02 30.3 1.48

2.25 4.37 72.98 43.3 1.44

3.12 10.27 64.94 32.0 2.11

5.14 7.44 71.29 24.5 1.85

5.54 9.25 74.94 31.0 2.23

5.04 4.84 53.61 48.2 1.60

3.14 4.82 67.00 37.6 1.52

2.43 2.40 67.20 36.8 1.40

5.38 4.30 83.03 19.7 2.07

7.34 2.73 72.60 24.5 1.42

4.94 4.66 64.32 27.7 1.42

4.82 4.26 82.64 20.3 1.46

5.02 4.17 84.25 20.6 1.42

3.37 1.00 69.93 16.4 1.17

3.63 6.40 70.31 29.0 2.00

7.43 6.00 70.53 37.7 1.44

2.20 10.59 69.85 41.7 2.01

7.16 4.71 79.44 33.0 1.55

6.33 2.88 66.24 38.1 1.73

2.57 1.85 67.25 33.4 1.18

6.38 1.56 63.00 18.2 0.93

5.34 3.41 72.57 20.1 1.66

4.87 5.20 75.13 16.5 3.64

2.04 4.83 67.78 17.4 1.49

5.48 1.34 77.43 21.6 1.32

7.77 5.32 58.57 31.2 3.21

6.29 2.60 64.32 27.4 1.78

6.38 3.71 78.61 34.1 1.30

5.76 4.06 83.77 31.4 1.52

6.03 3.10 76.04 25.0 1.08

5.09 1.85 74.65 24.1 0.97

4.36 1.67 65.43 23.7 1.07

3.07 2.00 68.03 26.2 1.19

1.82 1.13 49.50 21.9 1.62

3.31 0.94 74.75 26.5 1.12

3.45 0.72 65.99 22.0 1.20

1.74 0.97 60.24 22.0 1.17

1.81 1.54 70.05 24.4 1.00

5.59 1.66 77.96 17.1 1.30

3.72 1.69 82.40 16.3 1.52

3.39 1.24 67.17 27.7 1.03

2.25 2.80 70.81 23.4 1.14

3.31 1.30 71.30 19.2 1.21

5.27 1.20 73.08 30.3 1.35

3.26 1.02 74.36 16.5 1.23

6.76 1.53 78.37 22.6 1.33

2.92 4.42 58.50 68.5 2.25

1.64 16.70 64.61 49.4 3.13

1.36 14.26 66.42 22.5 2.80

3.58 3.38 65.57 26.1 1.31

3.38 2.17 66.10 22.6 1.44

7.25 1.16 78.52 23.6 1.50

5.44 2.93 73.59 22.3 1.65

5.83 4.47 77.33 26.2 2.16

3.74 2.26 79.70 20.2 1.58

9.21 2.36 74.58 21.8 1.72

2.14 6.30 86.54 17.4 2.80

6.62 4.79 78.84 20.0 2.33

4.24 5.82 71.39 27.1 1.69

4.72 4.71 78.01 20.6 1.55

6.48 4.93 74.23 20.9 1.98

;

**run**;

**proc** **princomp** cov data=census out = result;

var population professional over16 gov medhomevalue;

**run**;

|  |
| --- |
| The SAS System |

The PRINCOMP Procedure

|  |  |
| --- | --- |
| **Observations** | 10 |
| **Variables** | 5 |

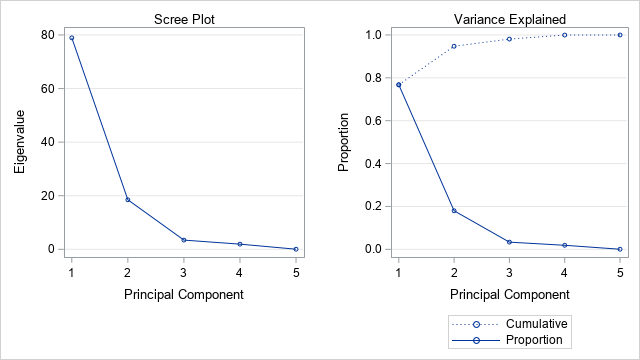
| **Simple Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **population** | **professional** | **over16** | **gov** | **medhomevalue** |
| **Mean** | 4.292000000 | 6.359000000 | 74.21800000 | 26.71000000 | 1.946000000 |
| **StD** | 1.682417573 | 2.019809947 | 5.95631467 | 7.76265275 | 0.433364101 |

| **Covariance Matrix** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **population** | **professional** | **over16** | **gov** | **medhomevalue** |
| **population** | 2.83052889 | -0.25276444 | 0.69893778 | -5.92813333 | 0.09380889 |
| **professional** | -0.25276444 | 4.07963222 | -4.91700222 | 2.73556667 | 0.30634000 |
| **over16** | 0.69893778 | -4.91700222 | 35.47768444 | -27.64275556 | 1.50375778 |
| **gov** | -5.92813333 | 2.73556667 | -27.64275556 | 60.25877778 | -1.79717778 |
| **medhomevalue** | 0.09380889 | 0.30634000 | 1.50375778 | -1.79717778 | 0.18780444 |

|  |  |
| --- | --- |
| **Total Variance** | 102.83442778 |

| **Eigenvalues of the Covariance Matrix** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 78.9378961 | 60.4478629 | 0.7676 | 0.7676 |
| **2** | 18.4900332 | 15.0509295 | 0.1798 | 0.9474 |
| **3** | 3.4391038 | 1.5110357 | 0.0334 | 0.9809 |
| **4** | 1.9280681 | 1.8887415 | 0.0187 | 0.9996 |
| **5** | 0.0393266 |  | 0.0004 | 1.0000 |

| **Eigenvectors** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **Prin1** | **Prin2** | **Prin3** | **Prin4** | **Prin5** |
| **population** | -.070269 | -.160797 | -.249730 | 0.952212 | -.011558 |
| **professional** | 0.066156 | -.174702 | 0.945479 | 0.221540 | -.148632 |
| **over16** | -.540785 | 0.814992 | 0.148038 | 0.135882 | -.054406 |
| **gov** | 0.835095 | 0.528473 | 0.005112 | 0.152331 | 0.010104 |
| **medhomevalue** | -.029211 | 0.011320 | 0.147522 | 0.050429 | 0.987275 |



The values obtained in SAS are the same as in the example

# Question 6

8.11.a)

**data** census;

input x1 x2 x3 x4 x5;

x5 = x5\***10**;

datalines;

2.67 5.71 69.02 30.3 1.48

2.25 4.37 72.98 43.3 1.44

3.12 10.27 64.94 32.0 2.11

5.14 7.44 71.29 24.5 1.85

5.54 9.25 74.94 31.0 2.23

5.04 4.84 53.61 48.2 1.60

3.14 4.82 67.00 37.6 1.52

2.43 2.40 67.20 36.8 1.40

5.38 4.30 83.03 19.7 2.07

7.34 2.73 72.60 24.5 1.42

4.94 4.66 64.32 27.7 1.42

4.82 4.26 82.64 20.3 1.46

5.02 4.17 84.25 20.6 1.42

3.37 1.00 69.93 16.4 1.17

3.63 6.40 70.31 29.0 2.00

7.43 6.00 70.53 37.7 1.44

2.20 10.59 69.85 41.7 2.01

7.16 4.71 79.44 33.0 1.55

6.33 2.88 66.24 38.1 1.73

2.57 1.85 67.25 33.4 1.18

6.38 1.56 63.00 18.2 0.93

5.34 3.41 72.57 20.1 1.66

4.87 5.20 75.13 16.5 3.64

2.04 4.83 67.78 17.4 1.49

5.48 1.34 77.43 21.6 1.32

7.77 5.32 58.57 31.2 3.21

6.29 2.60 64.32 27.4 1.78

6.38 3.71 78.61 34.1 1.30

5.76 4.06 83.77 31.4 1.52

6.03 3.10 76.04 25.0 1.08

5.09 1.85 74.65 24.1 0.97

4.36 1.67 65.43 23.7 1.07

3.07 2.00 68.03 26.2 1.19

1.82 1.13 49.50 21.9 1.62

3.31 0.94 74.75 26.5 1.12

3.45 0.72 65.99 22.0 1.20

1.74 0.97 60.24 22.0 1.17

1.81 1.54 70.05 24.4 1.00

5.59 1.66 77.96 17.1 1.30

3.72 1.69 82.40 16.3 1.52

3.39 1.24 67.17 27.7 1.03

2.25 2.80 70.81 23.4 1.14

3.31 1.30 71.30 19.2 1.21

5.27 1.20 73.08 30.3 1.35

3.26 1.02 74.36 16.5 1.23

6.76 1.53 78.37 22.6 1.33

2.92 4.42 58.50 68.5 2.25

1.64 16.70 64.61 49.4 3.13

1.36 14.26 66.42 22.5 2.80

3.58 3.38 65.57 26.1 1.31

3.38 2.17 66.10 22.6 1.44

7.25 1.16 78.52 23.6 1.50

5.44 2.93 73.59 22.3 1.65

5.83 4.47 77.33 26.2 2.16

3.74 2.26 79.70 20.2 1.58

9.21 2.36 74.58 21.8 1.72

2.14 6.30 86.54 17.4 2.80

6.62 4.79 78.84 20.0 2.33

4.24 5.82 71.39 27.1 1.69

4.72 4.71 78.01 20.6 1.55

6.48 4.93 74.23 20.9 1.98

;

**run**;

**proc** **princomp** cov data=census out=census\_cor;

var x1 x2 x3 x4 x5;

**run**;

| **Covariance Matrix** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** | **x4** | **x5** |
| **x1** | 3.39689902 | -1.10213937 | 4.30555484 | -2.07828525 | 0.27203907 |
| **x2** | -1.10213937 | 9.67277464 | -1.51323628 | 10.95323224 | 12.03063661 |
| **x3** | 4.30555484 | -1.51323628 | 55.62591164 | -28.93746421 | -0.43559071 |
| **x4** | -2.07828525 | 10.95323224 | -28.93746421 | 89.06661202 | 9.57299727 |
| **x5** | 0.27203907 | 12.03063661 | -0.43559071 | 9.57299727 | 31.86250820 |

b)

|  |  |
| --- | --- |
| **Total Variance** | 189.62470552 |

| **Eigenvalues of the Covariance Matrix** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 108.271939 | 65.132265 | 0.5710 | 0.5710 |
| **2** | 43.139674 | 11.872547 | 0.2275 | 0.7985 |
| **3** | 31.267127 | 26.669029 | 0.1649 | 0.9634 |
| **4** | 4.598098 | 2.250230 | 0.0242 | 0.9876 |
| **5** | 2.347868 |  | 0.0124 | 1.0000 |

| **Eigenvectors** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **Prin1** | **Prin2** | **Prin3** | **Prin4** | **Prin5** |
| **x1** | -.037629 | 0.062309 | -.039979 | 0.555532 | 0.827338 |
| **x2** | 0.118930 | 0.249301 | 0.260525 | -.768392 | 0.515175 |
| **x3** | -.479673 | 0.759677 | -.430649 | -.028079 | -.080986 |
| **x4** | 0.858912 | 0.316400 | -.393644 | 0.068674 | -.049898 |
| **x5** | 0.128935 | 0.506704 | 0.768189 | 0.308955 | -.202630 |

= -0.0376+ 0.11893-0.4797+0.8591+0.1289

= 0.0623+ 0.2493+0.7597+0.3164+0.5067

c)

**proc** **corr** data=census\_cor pearson;

**run**;

79.85% of variance is explained by the first two principal components.

| **Pearson Correlation Coefficients, N = 61  Prob > |r| under H0: Rho=0** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** | **x4** | **x5** | **Prin1** | **Prin2** | **Prin3** | **Prin4** | **Prin5** |
| **x1** | |  | | --- | | 1.00000 | |  | | |  | | --- | | -0.19227 | | 0.1377 | | |  | | --- | | 0.31322 | | 0.0140 | | |  | | --- | | -0.11948 | | 0.3590 | | |  | | --- | | 0.02615 | | 0.8415 | | |  | | --- | | -0.21244 | | 0.1002 | | |  | | --- | | 0.22205 | | 0.0854 | | |  | | --- | | -0.12129 | | 0.3518 | | |  | | --- | | 0.64633 | | <.0001 | | |  | | --- | | 0.68783 | | <.0001 | |
| **x2** | |  | | --- | | -0.19227 | | 0.1377 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | -0.06524 | | 0.6174 | | |  | | --- | | 0.37317 | | 0.0031 | | |  | | --- | | 0.68529 | | <.0001 | | |  | | --- | | 0.39790 | | 0.0015 | | |  | | --- | | 0.52649 | | <.0001 | | |  | | --- | | 0.46840 | | 0.0001 | | |  | | --- | | -0.52978 | | <.0001 | | |  | | --- | | 0.25381 | | 0.0484 | |
| **x3** | |  | | --- | | 0.31322 | | 0.0140 | | |  | | --- | | -0.06524 | | 0.6174 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | -0.41112 | | 0.0010 | | |  | | --- | | -0.01035 | | 0.9369 | | |  | | --- | | -0.66921 | | <.0001 | | |  | | --- | | 0.66900 | | <.0001 | | |  | | --- | | -0.32287 | | 0.0112 | | |  | | --- | | -0.00807 | | 0.9508 | | |  | | --- | | -0.01664 | | 0.8987 | |
| **x4** | |  | | --- | | -0.11948 | | 0.3590 | | |  | | --- | | 0.37317 | | 0.0031 | | |  | | --- | | -0.41112 | | 0.0010 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.17970 | | 0.1658 | | |  | | --- | | 0.94700 | | <.0001 | | |  | | --- | | 0.22020 | | 0.0882 | | |  | | --- | | -0.23323 | | 0.0705 | | |  | | --- | | 0.01560 | | 0.9050 | | |  | | --- | | -0.00810 | | 0.9506 | |
| **x5** | |  | | --- | | 0.02615 | | 0.8415 | | |  | | --- | | 0.68529 | | <.0001 | | |  | | --- | | -0.01035 | | 0.9369 | | |  | | --- | | 0.17970 | | 0.1658 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.23768 | | 0.0651 | | |  | | --- | | 0.58959 | | <.0001 | | |  | | --- | | 0.76098 | | <.0001 | | |  | | --- | | 0.11737 | | 0.3677 | | |  | | --- | | -0.05500 | | 0.6737 | |
| **Prin1** | |  | | --- | | -0.21244 | | 0.1002 | | |  | | --- | | 0.39790 | | 0.0015 | | |  | | --- | | -0.66921 | | <.0001 | | |  | | --- | | 0.94700 | | <.0001 | | |  | | --- | | 0.23768 | | 0.0651 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | |
| **Prin2** | |  | | --- | | 0.22205 | | 0.0854 | | |  | | --- | | 0.52649 | | <.0001 | | |  | | --- | | 0.66900 | | <.0001 | | |  | | --- | | 0.22020 | | 0.0882 | | |  | | --- | | 0.58959 | | <.0001 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | |
| **Prin3** | |  | | --- | | -0.12129 | | 0.3518 | | |  | | --- | | 0.46840 | | 0.0001 | | |  | | --- | | -0.32287 | | 0.0112 | | |  | | --- | | -0.23323 | | 0.0705 | | |  | | --- | | 0.76098 | | <.0001 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | |
| **Prin4** | |  | | --- | | 0.64633 | | <.0001 | | |  | | --- | | -0.52978 | | <.0001 | | |  | | --- | | -0.00807 | | 0.9508 | | |  | | --- | | 0.01560 | | 0.9050 | | |  | | --- | | 0.11737 | | 0.3677 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.00000 | | 1.0000 | |
| **Prin5** | |  | | --- | | 0.68783 | | <.0001 | | |  | | --- | | 0.25381 | | 0.0484 | | |  | | --- | | -0.01664 | | 0.8987 | | |  | | --- | | -0.00810 | | 0.9506 | | |  | | --- | | -0.05500 | | 0.6737 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 1.00000 | |  | |

The first principal component is strongly correlated with x3 and x4

There also exists a strong correlation between the second component and x3.

The third component has a strong correlation with x5.

The fourth and fifth principal components are highly correlated with x1.

In Example 8.3 the first two principal components explain 92.8% of the total variation, whereas here they only explain 79.85% of the total variation.

# Exercise 8.20

a)

**data** track;

input country $15. x1 x2 x3 x4 x5 x6 x7 x8;

datalines;

Argentina 10.23 20.37 46.18 1.77 3.68 13.33 27.65 129.57

Australia 9.93 20.06 44.38 1.74 3.53 12.93 27.53 127.51

Austria 10.15 20.45 45.80 1.77 3.58 13.26 27.72 132.22

Belgium 10.14 20.19 45.02 1.73 3.57 12.83 26.87 127.20

Bermuda 10.27 20.30 45.26 1.79 3.70 14.64 30.49 146.37

Brazil 10.00 19.89 44.29 1.70 3.57 13.48 28.13 126.05

Canada 9.84 20.17 44.72 1.75 3.53 13.23 27.60 130.09

Chile 10.10 20.15 45.92 1.76 3.65 13.39 28.09 132.19

China 10.17 20.42 45.25 1.77 3.61 13.42 28.17 129.18

Columbia 10.29 20.85 45.84 1.80 3.72 13.49 27.88 131.17

CookIslands 10.97 22.46 51.40 1.94 4.24 16.70 35.38 171.26

CostaRica 10.32 20.96 46.42 1.87 3.84 13.75 28.81 133.23

CzechRepublic 10.24 20.61 45.77 1.75 3.58 13.42 27.80 131.57

Denmark 10.29 20.52 45.89 1.69 3.52 13.42 27.91 129.43

DominicanRepub 10.16 20.65 44.90 1.81 3.73 14.31 30.43 146.00

Finland 10.21 20.47 45.49 1.74 3.61 13.27 27.52 131.15

France 10.02 20.16 44.64 1.72 3.48 12.98 27.38 126.36

Germany 10.06 20.23 44.33 1.73 3.53 12.91 27.36 128.47

GreatBritain 9.87 19.94 44.36 1.70 3.49 13.01 27.30 127.13

Greece 10.11 19.85 45.57 1.75 3.61 13.48 28.12 132.04

Guatemala 10.32 21.09 48.44 1.82 3.74 13.98 29.34 132.53

Hungary 10.08 20.11 45.43 1.76 3.59 13.45 28.03 132.10

India 10.33 20.73 45.48 1.76 3.63 13.50 28.81 132.00

Indonesia 10.20 20.93 46.37 1.83 3.77 14.21 29.65 139.18

Ireland 10.35 20.54 45.58 1.75 3.56 13.07 27.78 129.15

Israel 10.20 20.89 46.59 1.80 3.70 13.66 28.72 134.21

Italy 10.01 19.72 45.26 1.73 3.55 13.09 27.28 127.29

Japan 10.00 20.03 44.78 1.77 3.62 13.22 27.58 126.16

Kenya 10.28 20.43 44.18 1.70 3.44 12.66 26.46 124.55

Korea,South 10.34 20.41 45.37 1.74 3.64 13.84 28.51 127.20

Korea,North 10.60 21.23 46.95 1.82 3.77 13.90 28.45 129.26

Luxembourg 10.41 20.77 47.90 1.76 3.67 13.64 28.77 134.03

Malaysia 10.30 20.92 46.41 1.79 3.76 14.11 29.50 149.27

Mauritius 10.13 20.06 44.69 1.80 3.83 14.15 29.84 143.07

Mexico 10.21 20.40 44.31 1.78 3.63 13.13 27.14 127.19

Myanmar(Burma) 10.64 21.52 48.63 1.80 3.80 14.19 29.62 139.57

Netherlands 10.19 20.19 45.68 1.73 3.55 13.22 27.44 128.31

NewZealand 10.11 20.42 46.09 1.74 3.54 13.21 27.70 128.59

Norway 10.08 20.17 46.11 1.71 3.62 13.11 27.54 130.17

PapuaNewGuinea 10.40 21.18 46.77 1.80 4.00 14.72 31.36 148.13

Philippines 10.57 21.43 45.57 1.80 3.82 13.97 29.04 138.44

Poland 10.00 19.98 44.62 1.72 3.59 13.29 27.89 129.23

Portugal 9.86 20.12 46.11 1.75 3.50 13.05 27.21 126.36

Romania 10.21 20.75 45.77 1.76 3.57 13.25 27.67 132.30

Russia 10.11 20.23 44.60 1.71 3.54 13.20 27.90 129.16

Samoa 10.78 21.86 49.98 1.94 4.01 16.28 34.71 161.50

Singapore 10.37 21.14 47.60 1.84 3.86 14.96 31.32 144.22

Spain 10.17 20.59 44.96 1.73 3.48 13.04 27.24 127.23

Sweden 10.18 20.43 45.54 1.76 3.61 13.29 27.93 130.38

Switzerland 10.16 20.41 44.99 1.71 3.53 13.13 27.90 129.56

Taiwan 10.36 20.81 46.72 1.79 3.77 13.91 29.20 134.35

Thailand 10.23 20.69 46.05 1.81 3.77 14.25 29.67 139.33

Turkey 10.38 21.04 46.63 1.78 3.59 13.45 28.33 130.25

U.S.A. 9.78 19.32 43.18 1.71 3.46 12.97 27.23 125.38

;

**run**;

**proc** **princomp** data=track out=result;;

title 'Principal component analysis (PCA) using the correlation matrix';

var x1 x2 x3 x4 x5 x6 x7 x8;

**run**;

| **Correlation Matrix** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** | **x4** | **x5** | **x6** | **x7** | **x8** |
| **x1** | 1.0000 | 0.9148 | 0.8041 | 0.7119 | 0.7658 | 0.7399 | 0.7148 | 0.6765 |
| **x2** | 0.9148 | 1.0000 | 0.8449 | 0.7969 | 0.7951 | 0.7613 | 0.7480 | 0.7211 |
| **x3** | 0.8041 | 0.8449 | 1.0000 | 0.7677 | 0.7716 | 0.7797 | 0.7657 | 0.7127 |
| **x4** | 0.7119 | 0.7969 | 0.7677 | 1.0000 | 0.8958 | 0.8607 | 0.8431 | 0.8070 |
| **x5** | 0.7658 | 0.7951 | 0.7716 | 0.8958 | 1.0000 | 0.9165 | 0.9013 | 0.8778 |
| **x6** | 0.7399 | 0.7613 | 0.7797 | 0.8607 | 0.9165 | 1.0000 | 0.9882 | 0.9441 |
| **x7** | 0.7148 | 0.7480 | 0.7657 | 0.8431 | 0.9013 | 0.9882 | 1.0000 | 0.9542 |
| **x8** | 0.6765 | 0.7211 | 0.7127 | 0.8070 | 0.8778 | 0.9441 | 0.9542 | 1.0000 |

| **Eigenvalues of the Correlation Matrix** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 6.70328995 | 6.06487984 | 0.8379 | 0.8379 |
| **2** | 0.63841011 | 0.41088562 | 0.0798 | 0.9177 |
| **3** | 0.22752449 | 0.02167531 | 0.0284 | 0.9462 |
| **4** | 0.20584918 | 0.10827174 | 0.0257 | 0.9719 |
| **5** | 0.09757744 | 0.02688953 | 0.0122 | 0.9841 |
| **6** | 0.07068791 | 0.02374586 | 0.0088 | 0.9929 |
| **7** | 0.04694205 | 0.03722319 | 0.0059 | 0.9988 |
| **8** | 0.00971886 |  | 0.0012 | 1.0000 |

| **Eigenvectors** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Prin1** | **Prin2** | **Prin3** | **Prin4** | **Prin5** | **Prin6** | **Prin7** | **Prin8** |
| **x1** | 0.332388 | 0.529399 | 0.343859 | -.380745 | -.299671 | 0.362037 | 0.347647 | 0.065701 |
| **x2** | 0.346051 | 0.470390 | -.003786 | -.217023 | 0.541434 | -.348592 | -.439897 | -.060755 |
| **x3** | 0.339124 | 0.345329 | -.067061 | 0.851300 | -.132986 | -.077084 | 0.113555 | 0.003470 |
| **x4** | 0.353013 | -.089455 | -.782711 | -.134279 | 0.227283 | 0.341308 | 0.258883 | 0.039274 |
| **x5** | 0.365985 | -.153652 | -.244270 | -.233020 | -.651624 | -.529780 | -.147036 | 0.039746 |
| **x6** | 0.369820 | -.294760 | 0.182863 | 0.054624 | -.071816 | 0.359144 | -.328320 | -.705685 |
| **x7** | 0.365949 | -.333606 | 0.243981 | 0.087069 | 0.061333 | 0.273086 | -.351113 | 0.697182 |
| **x8** | 0.354278 | -.386561 | 0.334633 | -.018121 | 0.337891 | -.375170 | 0.594157 | -.069317 |

b)

**proc** **print** data=result;

var prin1 prin2;

**run**;

| **Obs** | **Prin1** | **Prin2** |
| --- | --- | --- |
| **1** | -0.4163 | 0.39454 |
| **2** | -2.3525 | -0.55022 |
| **3** | -0.7306 | 0.18057 |
| **4** | -1.9798 | 0.37706 |
| **5** | 1.4861 | -1.64219 |

**proc** **corr** data=result;

var prin1 prin2 x1 x2 x3 x4 x5 x6 x7 x8;

**run**;

| **Pearson Correlation Coefficients, N = 54  Prob > |r| under H0: Rho=0** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Prin1** | **Prin2** | **x1** | **x2** | **x3** | **x4** | **x5** | **x6** | **x7** | **x8** |
| **Prin1** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 0.86058 | | <.0001 | | |  | | --- | | 0.89595 | | <.0001 | | |  | | --- | | 0.87802 | | <.0001 | | |  | | --- | | 0.91398 | | <.0001 | | |  | | --- | | 0.94756 | | <.0001 | | |  | | --- | | 0.95749 | | <.0001 | | |  | | --- | | 0.94747 | | <.0001 | | |  | | --- | | 0.91725 | | <.0001 | |
| **Prin2** | |  | | --- | | 0.00000 | | 1.0000 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.42299 | | 0.0014 | | |  | | --- | | 0.37584 | | 0.0051 | | |  | | --- | | 0.27592 | | 0.0434 | | |  | | --- | | -0.07148 | | 0.6075 | | |  | | --- | | -0.12277 | | 0.3765 | | |  | | --- | | -0.23551 | | 0.0865 | | |  | | --- | | -0.26655 | | 0.0514 | | |  | | --- | | -0.30886 | | 0.0231 | |

c)

The first component is strongly correlated with all variables (ρ > 0.8). This component relates to the athletic ability of the nations. One can see that all the variables have a similar size effect on this component.

The second principal component is not all that strongly correlated with the variables (|ρ| < 0.5). It relates to the relative strength of the nation at the different running distances. x1 and x2 has the biggest effect on this principal component.

d)

**data** rank;

set result;

keep country prin1;

**run**;

**proc** **sort** data=rank out=sorted;

by descending prin1 ;

**run**;

**proc** **print** data=sorted;

**run**;

| **Obs** | **country** | **Prin1** |
| --- | --- | --- |
| **1** | CookIslands | 10.7112 |
| **2** | Samoa | 8.4215 |
| **3** | Singapore | 3.6915 |
| **4** | PapuaNewGuinea | 3.6813 |
| **5** | Myanmar(Burma) | 3.2368 |
| **.**  **.**  **.** |  |  |
| **48** | Brazil | -2.2083 |
| **49** | Italy | -2.2439 |
| **50** | Australia | -2.3525 |
| **51** | France | -2.4020 |
| **52** | Kenya | -2.6083 |
| **53** | GreatBritain | -2.9150 |
| **54** | U.S.A. | -3.8284 |

The countries are not ordered as one would expect.

# Exercise 8.22

**data** exercise\_8;

input breed x1 x2 x3 x4 x5 x6 x7 x8;

datalines;

1 2200 51.0 1128 70.9 7 0.25 54.8 1720

1 2250 51.9 1108 72.1 7 0.25 55.3 1575

1 1625 49.9 1011 71.6 6 0.15 53.1 1410

1 4600 53.1 993 68.9 8 0.35 56.4 1595

1 2150 51.2 996 68.6 7 0.25 55.0 1488

1 1225 49.2 985 71.4 6 0.15 51.4 1500

1 2250 51.0 959 72.1 7 0.20 54.0 1522

1 4000 51.5 1060 69.3 7 0.30 55.6 1765

1 1600 50.1 979 71.2 6 0.25 51.5 1365

1 1525 49.6 1083 75.8 6 0.30 54.6 1640

1 1850 50.6 1036 69.2 6 0.15 54.8 1570

1 2850 51.1 870 70.9 7 0.15 52.9 1450

1 2650 51.1 998 65.5 7 0.40 54.6 1505

1 1550 50.2 973 69.5 6 0.40 53.0 1530

1 2000 49.0 893 73.9 6 0.20 51.9 1470

1 2300 49.6 975 68.2 6 0.50 52.9 1842

1 1900 49.1 997 67.9 6 0.30 54.0 1500

1 1400 48.4 946 68.6 5 0.20 51.2 1480

1 1650 50.9 928 67.2 6 0.25 54.1 1480

1 1500 49.5 963 69.4 6 0.35 53.1 1670

1 1375 49.2 911 67.4 6 0.20 53.4 1490

1 1500 48.1 1003 70.5 5 0.25 54.7 1748

1 2400 51.1 915 64.9 7 0.25 54.6 1725

1 1425 48.9 924 72.7 5 0.15 52.1 1374

1 1525 49.4 959 68.4 6 0.15 52.6 1565

1 1800 47.7 944 66.5 5 0.40 53.3 1556

1 2500 50.6 897 67.2 6 0.30 54.9 1688

1 1600 48.9 974 71.0 5 0.30 54.2 1722

1 1300 49.9 872 70.7 6 0.20 53.3 1325

1 1400 48.4 841 71.3 5 0.15 51.5 1365

1 1300 48.6 920 71.4 5 0.15 52.9 1450

1 1400 47.6 974 69.7 5 0.15 51.9 1570

5 2000 50.5 1002 68.8 6 0.20 54.4 1735

5 1300 50.2 998 68.7 6 0.15 52.9 1540

5 1300 49.0 1015 69.8 6 0.30 51.9 1550

5 1300 48.7 1056 72.9 5 0.15 52.6 1525

5 1500 49.6 984 71.4 6 0.15 53.4 1650

5 1225 48.9 934 66.0 5 0.20 52.1 1430

5 2750 49.7 929 66.9 6 0.25 53.3 1688

5 1500 49.9 919 67.1 6 0.20 54.3 1425

5 1325 47.8 931 67.1 5 0.25 51.5 1520

5 1800 49.6 952 69.4 6 0.25 52.3 1512

5 1375 51.0 1002 72.1 7 0.25 51.9 1410

5 975 48.6 936 65.3 5 0.35 51.4 1550

5 1325 48.3 870 65.6 5 0.30 52.5 1588

5 1850 50.1 853 67.9 6 0.15 52.9 1390

5 1025 48.8 843 67.3 5 0.20 50.4 1390

5 1000 47.7 913 68.2 5 0.15 49.4 1345

5 975 47.2 844 70.6 5 0.15 50.1 1285

8 1750 54.0 1252 76.5 8 0.15 56.9 1648

8 1450 53.3 1383 81.4 8 0.20 59.6 1904

8 1200 52.8 1076 74.0 7 0.15 55.5 1615

8 2000 53.5 1175 74.5 8 0.10 57.4 1686

8 1450 53.2 1027 71.2 8 0.10 56.9 1696

8 1800 52.3 1116 71.1 7 0.10 57.5 1620

8 1525 51.8 1095 71.1 7 0.15 54.6 1712

8 1925 52.7 1141 78.5 7 0.15 55.6 1572

8 3450 54.8 1039 70.6 8 0.10 58.7 1600

8 1650 52.8 981 74.1 7 0.10 56.9 1750

8 1900 52.4 933 71.5 7 0.10 56.2 1640

8 1850 51.2 1083 74.5 7 0.20 55.9 1752

8 1550 52.3 1143 77.7 7 0.10 56.1 1785

8 1825 53.0 1055 76.8 8 0.10 56.7 1526

8 1475 52.9 1037 75.0 7 0.10 55.5 1406

8 2200 51.8 1076 74.5 7 0.15 55.8 1475

8 1850 53.1 964 70.8 8 0.10 55.5 1535

8 1550 51.2 1057 74.8 7 0.10 55.5 1520

8 1250 50.8 1040 74.5 6 0.10 55.8 1516

8 1350 52.7 1079 75.5 7 0.15 56.1 1595

8 1725 51.4 1034 71.2 7 0.10 56.0 1655

8 1750 50.7 1012 71.6 6 0.10 54.3 1480

8 1450 51.4 997 73.4 7 0.10 55.2 1454

8 1200 49.8 991 70.8 6 0.15 54.6 1475

8 1425 50.0 928 70.8 6 0.10 53.9 1375

8 1250 50.1 990 71.0 6 0.10 54.9 1564

8 1500 51.7 992 70.6 7 0.15 55.1 1458

;

**run**;

a)

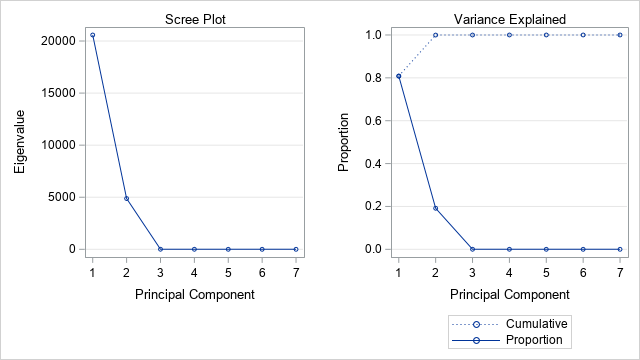
**proc** **princomp** cov data=exercise\_8 out = result;

var x2 x3 x4 x5 x6 x7 x8;

**run**;

|  |  |
| --- | --- |
| **Total Variance** | 25463.580231 |

| **Eigenvalues of the Covariance Matrix** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 20579.6126 | 15704.9378 | 0.8082 | 0.8082 |
| **2** | 4874.6748 | 4869.2456 | 0.1914 | 0.9996 |
| **3** | 5.4292 | 2.1129 | 0.0002 | 0.9998 |
| **4** | 3.3163 | 2.8475 | 0.0001 | 1.0000 |
| **5** | 0.4688 | 0.3948 | 0.0000 | 1.0000 |
| **6** | 0.0741 | 0.0695 | 0.0000 | 1.0000 |
| **7** | 0.0045 |  | 0.0000 | 1.0000 |



Considering the scree plot above, visually it appears that there is an elbow at 2, so we will need 2 principal components.

b)

| **Eigenvectors** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Prin1** | **Prin2** | **Prin3** | **Prin4** | **Prin5** | **Prin6** | **Prin7** |
| **x2** | 0.005887 | 0.009680 | 0.286337 | 0.608787 | 0.535569 | -.509727 | 0.024592 |
| **x3** | 0.487047 | 0.872697 | -.034277 | -.003227 | 0.000444 | -.000457 | -.000253 |
| **x4** | 0.008526 | 0.029196 | 0.904389 | -.425175 | 0.008388 | 0.010389 | 0.014293 |
| **x5** | 0.003112 | 0.004886 | 0.133267 | 0.311194 | 0.390573 | 0.855204 | -.037984 |
| **x6** | 0.000069 | -.000493 | -.018864 | -.005278 | 0.011906 | 0.043786 | 0.998778 |
| **x7** | 0.009330 | 0.008577 | 0.284215 | 0.593037 | -.748598 | 0.082331 | 0.013820 |
| **x8** | 0.873259 | -.487193 | 0.004847 | -.005597 | 0.002665 | -.000341 | -.000256 |

**proc** **print** data=result;

var prin1 prin2;

**run**;

| **Obs** | **Prin1** | **Prin2** |
| --- | --- | --- |
| **1** | 208.162 | 35.010 |
| **2** | 71.819 | 88.247 |
| **3** | -119.552 | 83.925 |
| **4** | 33.267 | -21.924 |

c)

Both the first and the second principal component comprise mostly of FtFrBody (x3) and SaleWt (x8). It will not be possible to develop a body size or body configuration index using only the first 2 principal components.

d)

goptions reset = all;

symbol1 v=triangle color=black h=**1**;

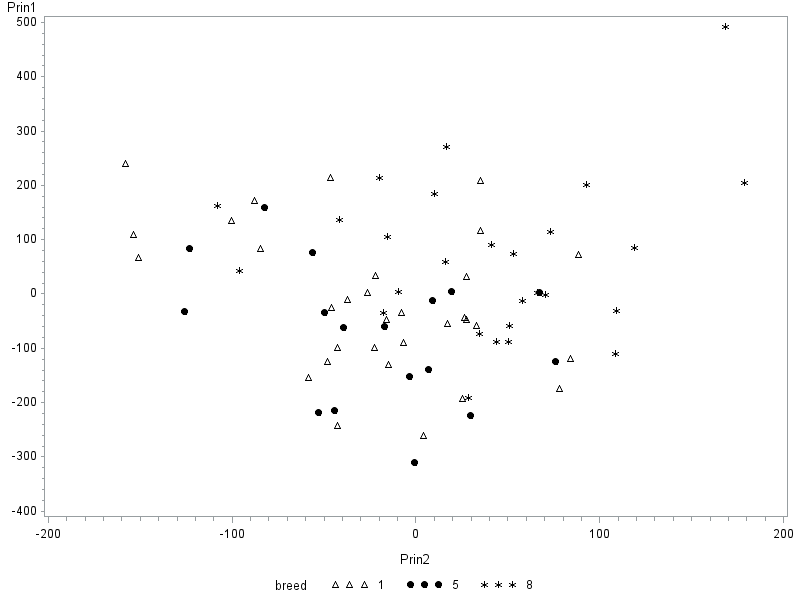
symbol2 v=dot color=black h=**1**;

symbol3 v=star color=black h=**1**;

**proc** **gplot** data = result;

plot prin1\*prin2=breed;

**run**;



There is no noticeable difference between the two groups. There does appear to be one extreme observation in terms of principle component 1 and 2.

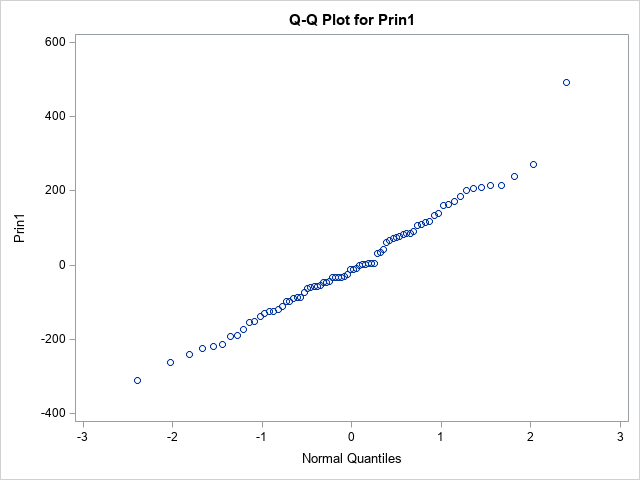
e)

**proc** **univariate** data=result;

var prin1;

qqplot ;

**run**;



The QQ plot follows an approximately straight line, which means that principle component 1 is normally distributed.

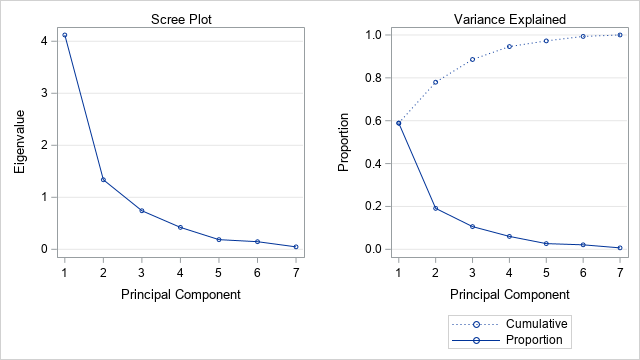
a)

**proc** **princomp** data=exercise\_8 out = result;

var x2 x3 x4 x5 x6 x7 x8;

**run**;

| **Eigenvalues of the Correlation Matrix** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 4.12069793 | 2.78356863 | 0.5887 | 0.5887 |
| **2** | 1.33712930 | 0.59574675 | 0.1910 | 0.7797 |
| **3** | 0.74138255 | 0.31995733 | 0.1059 | 0.8856 |
| **4** | 0.42142522 | 0.23561929 | 0.0602 | 0.9458 |
| **5** | 0.18580593 | 0.03930356 | 0.0265 | 0.9723 |
| **6** | 0.14650237 | 0.09944566 | 0.0209 | 0.9933 |
| **7** | 0.04705670 |  | 0.0067 | 1.0000 |



Considering the scree plot above, following a visual inspection it appears that there is an elbow at 3, so we will need 3 principal components.

b)

| **Eigenvectors** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Prin1** | **Prin2** | **Prin3** | **Prin4** | **Prin5** | **Prin6** | **Prin7** |
| **x2** | 0.449931 | -.042790 | -.415709 | 0.113356 | 0.065871 | -.072234 | 0.774926 |
| **x3** | 0.412326 | 0.129837 | 0.450292 | 0.247479 | -.719343 | -.177061 | 0.017768 |
| **x4** | 0.355562 | -.315508 | 0.568273 | 0.314787 | 0.579367 | 0.127800 | -.002397 |
| **x5** | 0.433957 | 0.007728 | -.452345 | 0.242818 | 0.142995 | -.434144 | -.582337 |
| **x6** | -.186705 | 0.714719 | -.038732 | 0.618117 | 0.160238 | 0.208017 | 0.042442 |
| **x7** | 0.452854 | 0.101315 | -.176650 | -.215769 | -.109535 | 0.799288 | -.236723 |
| **x8** | 0.269947 | 0.600515 | 0.253312 | -.582433 | 0.290547 | -.276561 | 0.047036 |

**proc** **print** data=result;

var prin1 prin2 prin3;

**run**;

| **Obs** | **Prin1** | **Prin2** | **Prin3** |
| --- | --- | --- | --- |
| **1** | 1.41746 | 1.39832 | 0.43499 |
| **2** | 1.50429 | 0.58676 | 0.00330 |
| **3** | -0.60115 | -1.13222 | 0.32865 |
| **4** | 1.50654 | 1.65924 | -1.98875 |
| **5** | 0.19496 | 0.36729 | -1.12426 |

The first principal component is determined almost equally by variables x2 to x7.

The second principal component is mostly determined by variables x5 and x7.

The third principal component is mostly determined by variables x1 to x4.

d)

goptions reset = all;

symbol1 v=triangle color=black h=**1**;

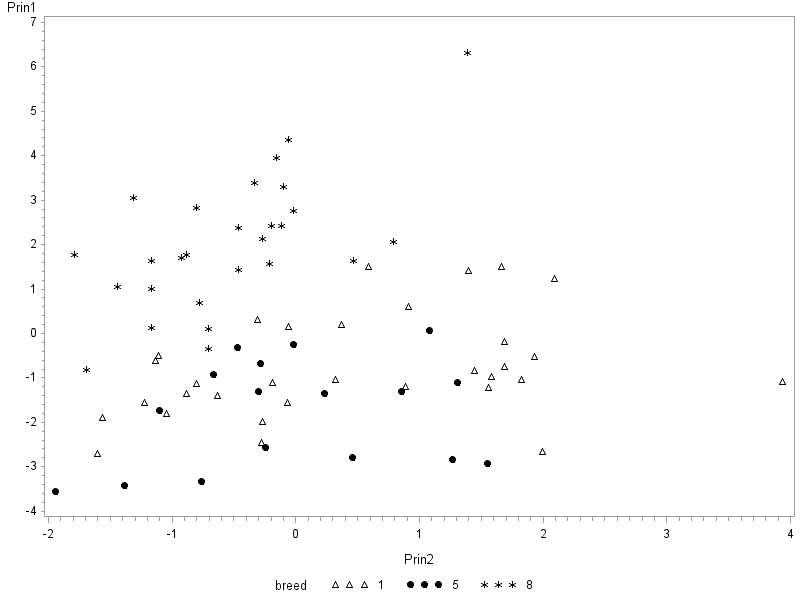
symbol2 v=dot color=black h=**1**;

symbol3 v=star color=black h=**1**;

**proc** **gplot** data = result;

plot prin1\*prin2=breed;

**run**;



The groups appear to be more clearly separated. There are two extreme observations, one in the principal component 1 space and 1 in the principal component 2 space

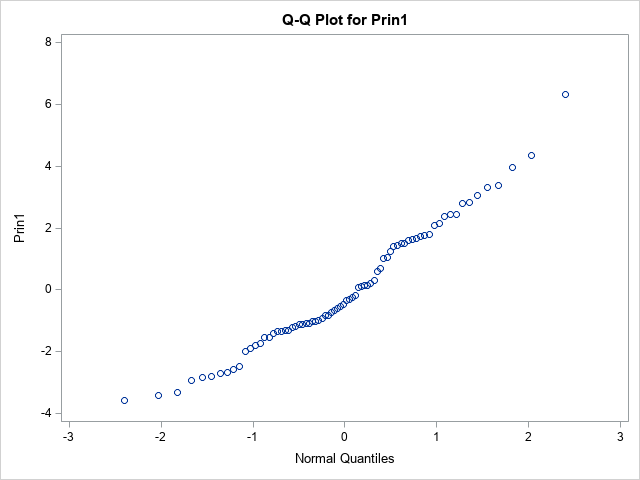
e)

**proc** **univariate** data=result;

var prin1;

qqplot ;

**run**;



The line appears to not be as linear as before and the assumption of normality does not hold all that well.