THE UNIVERSITY OF CHICAGO
Graduate School of Business
Business 41912-01, Spring Quarter 2008, Mr. Ruey S. Tsay

Solutions to Homework Assignment #5

1. Problem 9.20 of the textbook, p. 535

   Answer: For the PCA method, the results are given below:

```
> x=read.table("T1-5.DAT")
> S=cov(x)

> m1=princomp(x,cor=F)
> ei=m1$sdev
> vt=m1$loadings
> L=cbind(ei[1]*vt[,1])
> Er=S-as.matrix(L)%*%t(as.matrix(L))
> Psi=diag(Er)
> tmp=cbind(L,Psi,diag(S))
> print(tmp,digits=3)
      L       Psi   Var(x_i)
V1   0.1730  2.470   2.500
V2 -17.1169  7.528 300.516
V3  -0.2424  1.463   1.522
V4   0.0812  1.176   1.182
V5  -0.4180 11.189  11.364
V6  -1.9376 27.224  30.979
V7  -0.0403  0.477   0.479
*** Normalized loadings
> print(L,digits=3)
       [,1]
V1  0.01004
V2 -0.99320
V3 -0.01406
V4  0.00471
V5 -0.02426
V6 -0.11243
V7 -0.00234

> L2=cbind(ei[1]*vt[,1],ei[2]*vt[,2])
> Er2=S-as.matrix(L2)%*%t(as.matrix(L2))
```

```
> Psi2=diag(Er2)
> tmp2=cbind(L2,Psi2,diag(S))
> print(tmp2,digits=3)
       L1    L2      Psi2   Var(x_i)
V1   0.1730  0.4005  2.310   2.500
V2 -17.1169  0.6103  7.156 300.516
V3  -0.2424 -0.5231  1.190   1.522
V4   0.0812  0.0694  1.171   1.182
V5  -0.4180 -0.7901 10.565  11.364
V6  -1.9376 -5.1139  1.072  30.979
V7  -0.0403 -0.1251  0.461   0.479
*** Normalized loadings
> print(L2,digits=3)
       [,1]    [,2]
V1  0.01004  0.0762
V2 -0.99320  0.1162
V3 -0.01406 -0.0996
V4  0.00471  0.0132
V5 -0.02426 -0.1504
V6 -0.11243 -0.9734
V7 -0.00234 -0.0238
```

If correlation matrix is used in the principal component analysis, then the result is as follows:

```
> m1=princomp(x,cor=T)
> m1
Call:
princomp(x = x, cor = T)

Standard deviations:
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6    Comp.7
1.5286539 1.1772853 1.0972994 0.8526937 0.8083790 0.7325905 0.3948404

 7  variables and  42 observations.
> summary(m1)
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
Standard deviation     1.5286539 1.1772853 1.0972994 0.8526937 0.8083790
Proportion of Variance 0.3338261 0.1980001 0.1720094 0.1038695 0.0933538
Cumulative Proportion  0.3338261 0.5318262 0.7038356 0.8077051 0.9010589
                           Comp.6     Comp.7
```

```
Standard deviation     0.73259047 0.39484041
Proportion of Variance 0.07666983 0.02227128
Cumulative Proportion  0.97772872 1.00000000
> ei=m1$sdev
> vt=m1$loadings
> L=ei[1]*vt[,1]
> S=cor(x)
> Er=S-as.matrix(L)%*%t(as.matrix(L))
> Psi=diag(Er)
> tmp=cbind(L,Psi)
> print(tmp,digits=3)
        L   Psi
V1  0.362 0.869
V2 -0.314 0.901
V3 -0.842 0.290
V4 -0.577 0.667
V5 -0.761 0.420
V6 -0.496 0.754
V7 -0.488 0.762
> L2=cbind(ei[1]*vt[,1],ei[2]*vt[,2])
> Er2=S-as.matrix(L2)%*%t(as.matrix(L2))
> Psi2=diag(Er2)
> tmp2=cbind(L2,Psi2)
> print(tmp2,digits=2)
                Psi2
V1  0.36  0.328 0.76
V2 -0.31 -0.620 0.52
V3 -0.84 -0.008 0.29
V4 -0.58  0.512 0.40
V5 -0.76  0.235 0.37
V6 -0.50 -0.667 0.31
V7 -0.49  0.362 0.63
>
```

(b) The ML method

```
> k1=factanal(x,1,cor=F)
> k1
factanal(x = x, factors = 1, cor = F)

Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7
```

```
0.962 0.966 0.005 0.747 0.688 0.831 0.972

Loadings:
   Factor1
V1 -0.195
V2  0.183
V3  0.997
V4  0.503
V5  0.558
V6  0.411
V7  0.168

               Factor1
SS loadings      1.828
Proportion Var   0.261

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 34.11 on 14 degrees of freedom.
The p-value is 0.00199
> k2=factanal(x,2,cor=F)
> k2
factanal(x = x, factors = 2, cor = F)

Uniquenesses:
   V1    V2    V3    V4    V5    V6    V7
0.907 0.895 0.213 0.498 0.614 0.005 0.915

Loadings:
   Factor1 Factor2
V1 -0.176  -0.249
V2          0.319
V3  0.797   0.391
V4  0.692  -0.152
V5  0.602   0.152
V6          0.997
V7  0.251   0.147

               Factor1 Factor2
SS loadings      1.573   1.379
Proportion Var   0.225   0.197
Cumulative Var   0.225   0.422
```

```
Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 20.21 on 8 degrees of freedom.
The p-value is 0.00956
```

(c) The PCA approach based on the correlatin matrix gives results closer than that based on the covariance matrix to ML approach.

2. Problem 9.24 of the textbook, p. 535

Answer: For the PCA method, I chose $m = 3$, the results are given below:

```
> R=cor(x)
> m1=princomp(x,cor=T)
> summary(m1)
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4     Comp.5
Standard deviation     1.4113534 1.1694129 0.9296006 0.7314787 0.49126036
Proportion of Variance 0.3983837 0.2735053 0.1728315 0.1070122 0.04826735
Cumulative Proportion  0.3983837 0.6718890 0.8447204 0.9517327 1.00000000
> ei=m1$sdev
> vt=m1$loadings
> L=cbind(ei[1]*vt[,1],ei[2]*vt[,2],ei[3]*vt[,3])
> Er=R-as.matrix(L)%*%t(as.matrix(L))
> Psi=diag(Er)
> tmp=cbind(L,Psi)
> print(tmp,digits=3)
                            Psi
V1  0.371  0.541  0.729 0.0385
V2 -0.837  0.381 -0.153 0.1304
V3  0.460  0.708 -0.209 0.2442
V4 -0.676 -0.295  0.512 0.1933
V5 -0.696  0.584 -0.064 0.1699
```

On the other hand, the ML method only allows for $m = 2$, the results are given below:

```
> k1=factanal(x,2,scores=c("regression"))
> k1

Call:
factanal(x = x, factors = 2, scores = c("regression"))

Uniquenesses:
   V1    V2    V3    V4    V5
```

```
0.872 0.005 0.005 0.710 0.527

Loadings:
   Factor1 Factor2
V1 -0.135   0.331
V2  0.982  -0.176
V3  0.111   0.991
V4  0.299  -0.448
V5  0.682

                Factor1 Factor2
SS loadings       1.55   1.332
Proportion Var    0.31   0.266
Cumulative Var    0.31   0.576

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 4.08 on 1 degree of freedom.
The p-value is 0.0434
```

In the above, the factor scores are estimated by the regression method. Since the defacult of **R** program uses VARIMAX rotation. The results shown are rotated.

3. Problem 10.18 of the textbook, p. 573

   Answer: The canonical correlations are 0.92, 0.82, 0.27, and 0.09, respectively. The canonical variates are given below:

```
$xcoef
          [,1]        [,2]       [,3]       [,4]
V1  0.06689193 -0.15532630 -0.2533441  0.2259421
V2  0.03787201 -0.27574614  0.6299487  0.1048487
V3 -0.17490257  0.09417236 -0.4125382 -0.5312127
V4 -0.12496918  0.69613608  1.3215807  0.1267527

$ycoef
           [,1]          [,2]         [,3]          [,4]
V5  0.081735822  0.3533117157 -0.263202034 -1.1970693046
V6 -0.005446751  0.0086370579  0.000664767  0.0187756778
V7 -0.002368852  0.0000365322 -0.012125613 -0.0001620718
V8 -3.550862729 -6.7807308300 -3.380184995 -0.3870054231
```

The first pair alone is not a good summary. The sequence tests show that the first two cnaonical correlations are significantly different from zero. The test statistics for

one, two or three zero canonical correlations are 0.48(0.49), 4.60(0.33), and 66.79(0), respectively, where the number in parentheses are p-values. For the pulp fiber characteristics, the first canonical variate is essential zero span tensile ($x_4^{(2)}$) whereas the second canonical variate is a weighted difference between $x_4^{(2)}$ and $x_1^{(2)}$.

4. Problem 11.3 of the textbook, p. 650

   Answer: Simply follows the hint.

5. Problem 11.23 of the textbook, p. 656

   Answer: I take the cube-root transformation. The **r-discri.txt** program is used to perform the analysis. From the estimate $\hat{a}$ vector, the x_3 has a small coefficient. The variance of x_3 is relative small too. Thus, x_3 might not be necessary. The error rate is 14.28%.

   The output is given below:

```
*** Cube-root transformation
> y2=y^(1/3)
> discri(y2,size)
[1] "mean vector of popu 1"
          [,1]
[1,] 3.2943470
[2,] 5.2781994
[3,] 0.9526323
[4,] 5.8018597
[5,] 0.9243345
[1] "mean vector of popu 2"
         [,1]
[1,] 3.450733
[2,] 5.612483
[3,] 1.818491
[4,] 6.174415
[5,] 1.795313
[1] "S1"
           V1          V2           V3           V4          V5
V1 0.22472017 0.030627234  0.065115219  0.027695106 0.036788961
V2 0.03062723 0.015437093  0.001287438  0.012206119 0.008861018
V3 0.06511522 0.001287438  0.292154088 -0.001314652 0.074448865
V4 0.02769511 0.012206119 -0.001314652  0.017583739 0.008071969
V5 0.03678896 0.008861018  0.074448865  0.008071969 0.343241076
[1] "S2"
             V1         V2           V3         V4          V5
V1  0.098957663 0.01634665 -0.009366823 0.01638227 -0.04443109
```

```
V2  0.016346647 0.09016279  0.177663464 0.08049326  0.02916889
V3 -0.009366823 0.17766346  1.177309569 0.17111707  0.93089564
V4  0.016382274 0.08049326  0.171117065 0.08552922  0.01113234
V5 -0.044431094 0.02916889  0.930895643 0.01113234  1.34960490
[1] "a-hat"
           V1         V2         V3        V4        V5
[1,] 0.5836422 -0.9348471 0.02724501 -9.276956 -1.241391
[1] " minus m-hat"
         [,1]
[1,] 60.32421
[1] "Mis-specifiction items in Popu 1:"
[1] 13.0000000 -0.3652136
[1] 42.000000 -1.001063
[1] 59.0000000 -0.8354214
[1] 68.0000000 -0.6195788
[1] "Mis-specification items in Popu 2:"
[1] 70.000000  1.281430
[1] 73.000000  1.371675
[1] 79.0000000  0.2369256
[1] 81.0000000  0.7032652
[1] 83.000000  2.803824
[1] 87.000000  3.551806
[1] 88.000000  1.849917
[1] 89.0000000  0.8065407
[1] 96.0000000  0.6088622
[1] 97.00000000  0.01951708
[1] "mis-specification rate of Popu 1"
[1]  4 69
[1] "mis-specification rate of Popu 2"
[1] 10 29
[1] "Error rate: "
[1] 0.1428571
[1] "mean and std error of transformed Popu 1"
[1] 2.367521 1.470271
[1] "mean and std error of transformed Popu 2"
[1] -2.367521  3.314302
```