STAT 488: Multivariate Statistical Analysis — Homework 5

Charles Hwang

4/22/2022

Problem 9.1

```
rm(list=ls())
matrix(c(0.9,0.7,0.5))%*%t(matrix(c(0.9,0.7,0.5)))+matrix(c(0.19,0,0,0,0.51,0,0,0,0.75),nrow=3)
```

```
##      [,1] [,2] [,3]
## [1,] 1.00 0.63 0.45
## [2,] 0.63 1.00 0.35
## [3,] 0.45 0.35 1.00
```

We can see that $\rho = \mathbf{LL}' + \Psi = \begin{bmatrix} .9 & .7 & .5 \end{bmatrix} \begin{bmatrix} .9 \\ .7 \\ .5 \end{bmatrix} + \begin{bmatrix} .19 & 0 & 0 \\ 0 & .51 & 0 \\ 0 & 0 & .75 \end{bmatrix} = \begin{bmatrix} 1 & .63 & .45 \\ .63 & 1 & .35 \\ .45 & .35 & 1 \end{bmatrix}$, as intended by (9-5) on page 484 of the textbook.

Problem 9.2

```
(matrix(c(0.9,0.7,0.5))[1,1])^2 # (9-6), page 484  # Problem 9.2(a)
```

```
## [1] 0.81
```

```
(matrix(c(0.9,0.7,0.5))[2,1])^2
```

```
## [1] 0.49
```

```
(matrix(c(0.9,0.7,0.5))[3,1])^2
```

```
## [1] 0.25
```

```
# We can see that approximately 0.81 of the variance is explained by the first common
# factor, approximately 0.49 of the variance is explained by the second common factor,
# and approximately 0.25 of the variance is explained by the third common factor.
matrix(c(0.9,0.7,0.5))[1,1]     # (9-5), page 484  # Problem 9.2(b)
```

```
## [1] 0.9
```

```
matrix(c(0.9,0.7,0.5))[2,1]
```

```
## [1] 0.7
```

```
matrix(c(0.9,0.7,0.5))[3,1]
```

```
## [1] 0.5
```

We can see that Z_1 would likely carry the greatest weight because it has the strongest correlation (0.9).

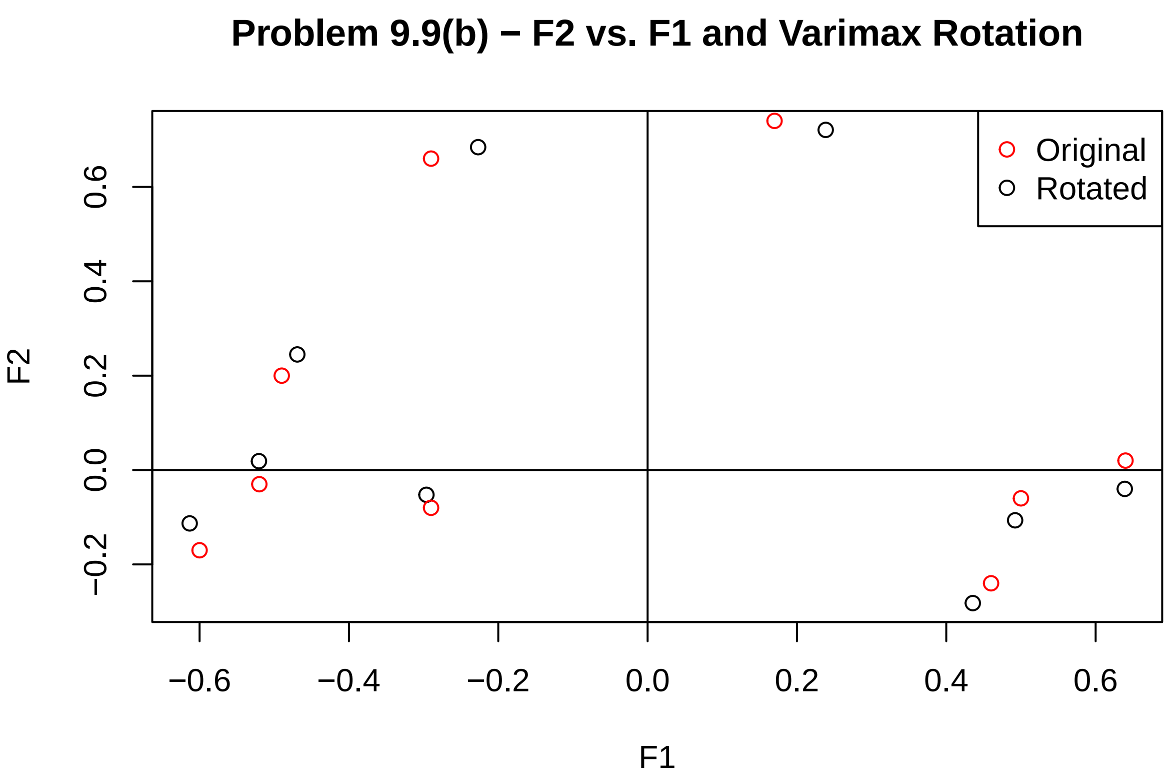
Problem 9.9

Problem 9.9(a)

I personally know nothing about liquor or alcohol in general, but these three factors seem to be reasonable interpretations of the common motivations and incentives for purchasing liquor.

Problem 9.9(b)

```
F1<-c(0.64,0.5,0.46,0.17,-0.29,-0.29,-0.49,-0.52,-0.6)
F2<-c(0.02,-0.06,-0.24,0.74,0.66,-0.08,0.2,-0.03,-0.17)
plot(varimax(matrix(c(F1,F2),ncol=2))$loadings,xlab="F1",ylab="F2",main="Problem 9.9(b) - F2 vs. F1 and
points(F1,F2,col="red")
abline(h=0,v=0)
legend("topright",c("Original","Rotated"),pch=1,col=c("red","black"))
```




```
varimax(matrix(c(F1,F2),ncol=2))$loadings
```

```
##
## Loadings:
##        [,1]   [,2]
##   [1,]  0.639
##   [2,]  0.492 -0.107
##   [3,]  0.436 -0.282
##   [4,]  0.239  0.721
##   [5,] -0.227  0.684
##   [6,] -0.296
##   [7,] -0.469  0.245
##   [8,] -0.521
##   [9,] -0.613 -0.113
##
```

```
##                  [,1]  [,2]
## SS loadings    1.904 1.156
## Proportion Var 0.212 0.128
## Cumulative Var 0.212 0.340
```

We can see that not much rotation of the factor axes was done, and so the interpretation of the rotated loadings for F_1 and F_2 is largely the same as Stoetzel's interpretation. Thus, the two interpretations agree.

Problem 9.20

```
ap<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T1-5
names(ap)<-c("Wind","Solar","NO_2","Ozone")
cov(ap)
```

```
##             Wind      Solar       NO_2     Ozone
## Wind   2.5000000  -2.780488 -0.5853659 -2.231707
## Solar -2.7804878 300.515679  6.7630662 30.790941
## NO_2  -0.5853659   6.763066 11.3635308  3.126597
## Ozone -2.2317073  30.790941  3.1265970 30.978513
```

```
L2<-t(prcomp(ap)$sdev[1:2]*t(prcomp(ap)$rotation[,c("PC1","PC2")])) # Problem 9.20(a)
L2
```

```
##               PC1        PC2
## Wind    0.1749782  0.4048141
## Solar -17.3246829  0.6085601
## NO_2   -0.4213923 -0.7421918
## Ozone  -1.9587473 -5.1867451
```

```
L<-factanal(factors=1,covmat=cov(ap))$loadings # Only doing m = 1   # Problem 9.20(b)
psi<-factanal(factors=1,covmat=cov(ap))$uniquenesses
L
```

```
##
## Loadings:
##       Factor1
## Wind  -0.324
## Solar  0.410
## NO_2   0.232
## Ozone  0.771
##
##                Factor1
## SS loadings      0.921
## Proportion Var   0.230
```

```
psi
```

```
##      Wind     Solar      NO_2     Ozone
## 0.8949140 0.8322114 0.9463160 0.4054956
```

```
L%*%t(L)+diag(psi) # There appears to be some rounding error, at least in the diagonal.
```

```
##              Wind       Solar        NO_2      Ozone
## Wind   1.00000075 -0.13278686 -0.07510881 -0.2499490
## Solar -0.13278686  1.00000003  0.09490699  0.3158339
## NO_2  -0.07510881  0.09490699  0.99999866  0.1786465
## Ozone -0.24994901  0.31583391  0.17864652  1.0000000
```

```r
data.frame(L2[,"PC1"])                                          # Problem 9.20(c)
```

```
##       L2....PC1..
## Wind    0.1749782
## Solar -17.3246829
## NO_2   -0.4213923
## Ozone  -1.9587473
```

```r
L
```

```
##
## Loadings:
##       Factor1
## Wind  -0.324
## Solar  0.410
## NO_2   0.232
## Ozone  0.771
##
##                Factor1
## SS loadings      0.921
## Proportion Var   0.230
```

We can see the factorization of the $m = 1$ model obtained by the principal component method is different than the factorization of the $m = 1$ model obtained by the maximum likelihood method. This illustrates how the two methods are different.

Problem 9.21

```r
L2
```

```
##               PC1        PC2
## Wind    0.1749782  0.4048141
## Solar -17.3246829  0.6085601
## NO_2   -0.4213923 -0.7421918
## Ozone  -1.9587473 -5.1867451
```

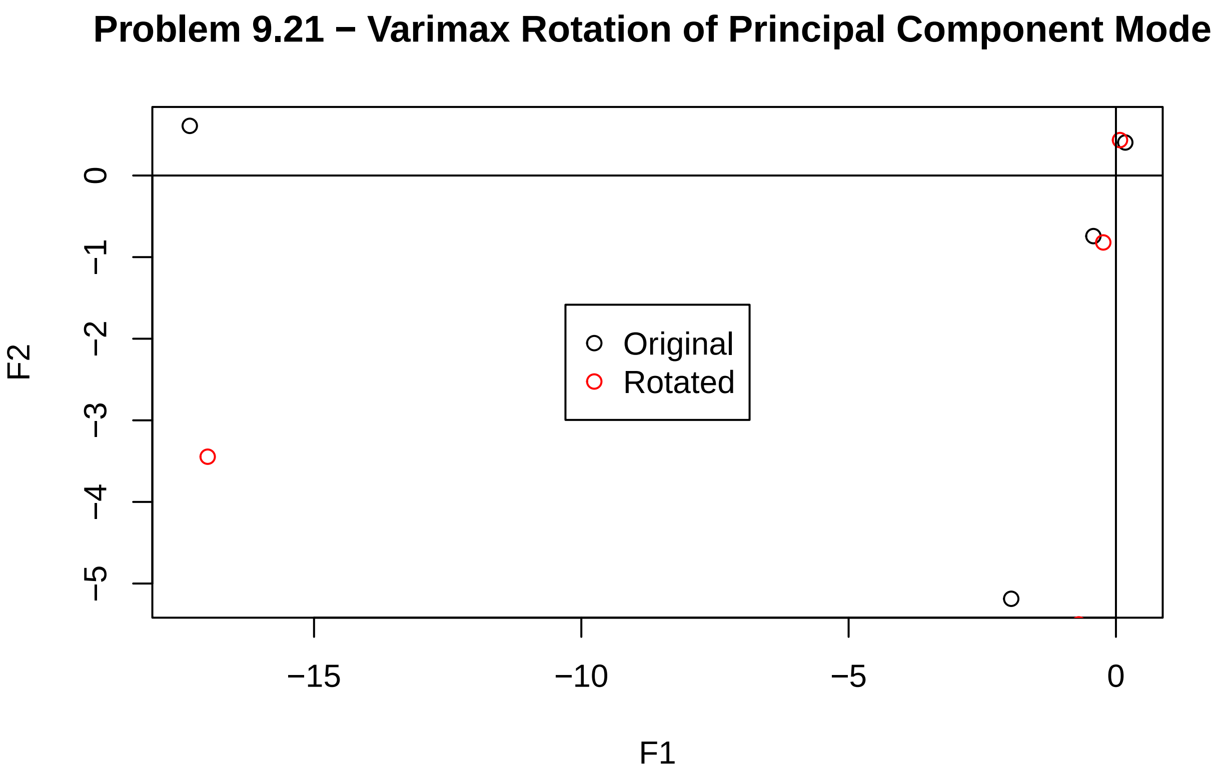
```r
varimax(L2)
```

```
## $loadings
##
## Loadings:
##       PC1     PC2
## Wind            0.434
## Solar -16.989  -3.446
## NO_2   -0.237  -0.820
## Ozone  -0.696  -5.500
##
##                    PC1    PC2
## SS loadings    289.188 42.988
## Proportion Var  72.297 10.747
## Cumulative Var  72.297 83.044
##
## $rotmat
##            [,1]      [,2]
## [1,]  0.9724650 0.2330489
## [2,] -0.2330489 0.9724650
```

```
varimax(L) # Since there is no m = 2 model with ML, a varimax rotation is not possible.
```

```
##
## Loadings:
##       Factor1
## Wind  -0.324
## Solar  0.410
## NO_2   0.232
## Ozone  0.771
##
##                Factor1
## SS loadings      0.921
## Proportion Var   0.230
```

```
plot(L2,xlab="F1",ylab="F2",main="Problem 9.21 - Varimax Rotation of Principal Component Model")
points(varimax(L2)$loadings,col="red")
abline(h=0,v=0)
legend("center",c("Original","Rotated"),pch=1,col=c("black","red"))
```



We can see the varimax rotation around the origin (0,0) when using the principal component method and the solar component of F_2 changed significantly. F_1 appears to be a solar factor and F_2 appears to be a ozone and solar factor.

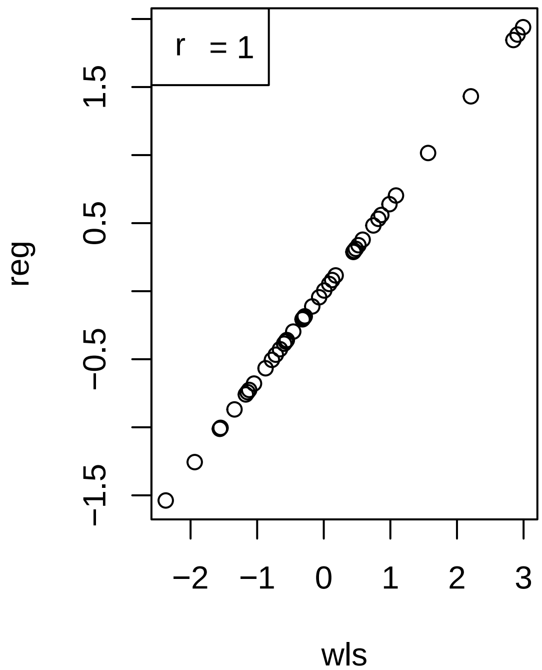
Problem 9.22

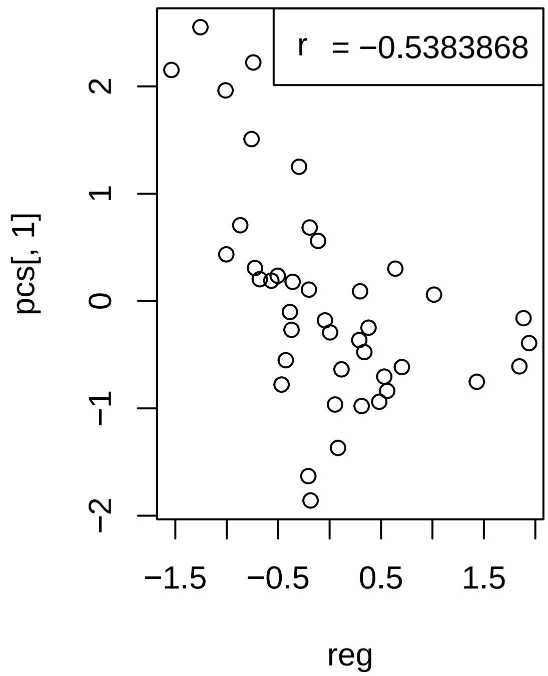
```
wls<-factanal(ap,1,scores="Bartlett")$scores                       # Problem 9.22(a)(i)
reg<-factanal(ap,1,scores="regression")$scores                     # Problem 9.22(a)(ii)
x<-matrix(rep(colMeans(ap),nrow(ap)),nrow=nrow(ap),byrow=TRUE) # Problem 9.22(b)
pcs<-t(t(L2)%*%t(ap-x))%*%diag(1/eigen(cov(ap))$values[1:2])
fs<-data.frame(wls,reg,pcs)
```

```r
names(fs)<-c("Weighted Least Squares","Regression","Princ. Comp. (F1)","Princ. Comp. (F2)")
fs
```

```
##    Weighted Least Squares   Regression Princ. Comp. (F1) Princ. Comp. (F2)
## 1             0.126409623  0.081928800       -1.36839363       0.740758048
## 2            -0.285210356 -0.184850977       -1.85807752       1.557519967
## 3            -0.319123412 -0.206830759       -1.63116124       1.391294470
## 4             0.864121039  0.560055464       -0.83724080      -0.638949618
## 5             0.480495665  0.311419593       -0.97820630       0.295047660
## 6             0.745008132  0.482855822       -0.93852190      -0.174491924
## 7             1.084574660  0.702936206       -0.61554288      -0.846465244
## 8             1.566538620  1.015307432        0.05957078      -1.218595579
## 9             0.521155447  0.337772074       -0.47564394      -0.151065586
## 10           -0.172795242 -0.111992319        0.56020365      -0.210383740
## 11           -1.047303636 -0.678780052        0.20316940       1.104398872
## 12            0.081787580  0.053008293       -0.96442978       0.744945517
## 13            0.457554361  0.296550840        0.09063405      -0.368813550
## 14           -0.778894836 -0.504818525        0.23528167       0.372695298
## 15           -0.309923227 -0.200867921        0.10621269      -0.060233190
## 16           -0.069031642 -0.044740895       -0.18051575       0.007504254
## 17           -0.595370927 -0.385872726       -0.10204910       0.580306601
## 18           -0.873989926 -0.566451702        0.18956885       0.788890126
## 19           -1.549292578 -1.004129901        0.43499276       1.166583009
## 20           -1.119754022 -0.725736708        0.30730908       0.760173309
## 21           -1.340357860 -0.868714808        0.70607214       0.675159316
## 22           -0.720250549 -0.466809900       -0.77847764       1.041024037
## 23            0.444808728  0.288290121       -0.36393214      -0.233127159
## 24           -1.937591313 -1.255794677        2.55105678       0.516345776
## 25            2.847281168  1.845384272       -0.60905826      -2.341158915
## 26           -0.657529139 -0.426158802       -0.55123853       0.939583564
## 27            0.583876857  0.378423171       -0.24840484      -0.252506059
## 28            0.007315077  0.004741059       -0.29222755       0.089565827
## 29           -0.298073317 -0.193187739        0.68494078       0.008599273
## 30           -2.372730322 -1.537817645        2.15248725       0.687837546
## 31           -0.554147947 -0.359155224        0.17856293       0.392029845
## 32           -0.458308870 -0.297039853        1.25089595      -0.273933716
## 33           -1.170857713 -0.758858111        1.50885794       0.457534150
## 34            2.909439380  1.885670349       -0.15986883      -2.700435879
## 35           -1.144633155 -0.741861410        2.22270508      -0.626402177
## 36            0.178672597  0.115801560       -0.63671964       0.166992721
## 37            0.819895931  0.531392218       -0.70412148      -0.145747987
## 38            2.208807893  1.431575987       -0.75245395      -1.443325522
## 39            2.992765016  1.939675489       -0.39216187      -2.757800954
## 40           -0.570789912 -0.369941241       -0.26843114       0.789880225
## 41            0.986164795  0.639154652        0.30181329      -1.025919852
## 42           -1.560712665 -1.011531506        1.96254364       0.194687241
```

```r
par(mfrow=c(1,2))                                              # Problem 9.22(c)
plot(wls,reg)
legend("topleft",legend="= 1",pch="r")
plot(reg,pcs[,1])
legend("topright",legend="= -0.5383868",pch="r")
```





We can see from the output that the factor scores from weighted least squares, regression, and the principal component method are different. The first plot shows a perfect positive linear relationship between the factor scores from weighted least squares and regression, negating the need to compare both to the principal component method. There is clearly a strong relationship between the two methods which makes sense because they originate from the same function. The second plot shows a moderate negative linear relationship between regression and the principal component method which indicates the principal component method is considerably different than the other two.

Problem 9.23

```
factanal(factors=1,covmat=cor(ap))$loadings # Using correlation matrix
```

```
##
## Loadings:
##       Factor1
## Wind  -0.324
## Solar  0.410
## NO_2   0.232
## Ozone  0.771
##
##                Factor1
## SS loadings      0.921
## Proportion Var   0.230
```

```
L
```

```
##
## Loadings:
##       Factor1
## Wind  -0.324
## Solar  0.410
## NO_2   0.232
## Ozone  0.771
```

```
##
##                Factor1
## SS loadings      0.921
## Proportion Var   0.230
```

```
factanal(factors=1,covmat=cor(ap))$uniquenesses
```

```
##      Wind     Solar      NO_2     Ozone
## 0.8949140 0.8322114 0.9463160 0.4054956
```

```
psi
```

```
##      Wind     Solar      NO_2     Ozone
## 0.8949140 0.8322114 0.9463160 0.4054956
```

We can see that it does not make a difference whether $\mathbf{R}$ or $\mathbf{S}$ is used for obtaining the maximum-likelihood estimates for $\mathbf{L}$ and Ψ. In the documentation for the `factanal()` function, the information for the `covmat` argument explains: "Of course, correlation matrices are covariance matrices." However, the loadings and subsequent interpretations may be different.