DSA 8070 R Session 9: Factor Analysis

Whitney Huang, Clemson University

Contents

Intelligence Tests Example

Load the data summary

```
data(ability.cov)
(cor <- cov2cor(ability.cov$cov))
```

```
##           general   picture    blocks      maze   reading     vocab
## general 1.0000000 0.4662649 0.5516632 0.3403250 0.5764799 0.5144058
## picture 0.4662649 1.0000000 0.5724364 0.1930992 0.2629229 0.2392766
## blocks  0.5516632 0.5724364 1.0000000 0.4450901 0.3540252 0.3564715
## maze    0.3403250 0.1930992 0.4450901 1.0000000 0.1839645 0.2188370
## reading 0.5764799 0.2629229 0.3540252 0.1839645 1.0000000 0.7913779
## vocab   0.5144058 0.2392766 0.3564715 0.2188370 0.7913779 1.0000000
```

Factor analysis

We will use the *factanal* command to perform factor analysis. The usage can be found below:

```
(ability.FA <- factanal(factors = 1, covmat = ability.cov))
```

```
factanal(x, factors, data = NULL, covmat = NULL, n.obs = NA,
         subset, na.action, start = NULL,
         scores = c("none", "regression", "Bartlett"),
         rotation = "varimax", control = NULL, ...)
```

Figure 1: factanal usage

```
##
## Call:
## factanal(factors = 1, covmat = ability.cov)
##
## Uniquenesses:
## general picture  blocks    maze reading   vocab
##   0.535   0.853   0.748   0.910   0.232   0.280
##
## Loadings:
##         Factor1
## general 0.682
## picture 0.384
## blocks  0.502
## maze    0.300
## reading 0.877
## vocab   0.849
##
##                Factor1
## SS loadings      2.443
## Proportion Var   0.407
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 75.18 on 9 degrees of freedom.
## The p-value is 1.46e-12
```

The output suggests that the test of the hypothesis that one factor is sufficient has resulted in a p-value of 1.46e-12, which indicates strong evidence against the null hypothesis, suggesting that one factor is not sufficient to explain the data.

```
update(ability.FA, factors = 2)
```

```
##
## Call:
## factanal(factors = 2, covmat = ability.cov)
##
## Uniquenesses:
## general picture  blocks    maze reading   vocab
##   0.455   0.589   0.218   0.769   0.052   0.334
##
## Loadings:
##         Factor1 Factor2
## general 0.499   0.543
```

```
## picture 0.156   0.622
## blocks  0.206   0.860
## maze    0.109   0.468
## reading 0.956   0.182
## vocab   0.785   0.225
##
##                Factor1 Factor2
## SS loadings      1.858   1.724
## Proportion Var   0.310   0.287
## Cumulative Var   0.310   0.597
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 6.11 on 4 degrees of freedom.
## The p-value is 0.191
```

Stock Price Example

Load the stock price data

```
url <- "http://users.stat.umn.edu/~sandy/courses/8053/Data/Wichern_data/T8-4.DAT"
stock <- read.table(url, sep = "\t", header = F)
colnames(stock) <- c("JP Morgan", "City bank", "Wells Fargo", "Royal Dutch", "Exxon")
head(stock)
```

```
##    JP Morgan  City bank Wells Fargo Royal Dutch      Exxon
## 1  0.0130338 -0.0078431  -0.0031889  -0.0447693  0.0052151
## 2  0.0084862  0.0166886  -0.0062100   0.0119560  0.0134890
## 3 -0.0179153 -0.0086393   0.0100360   0.0000000 -0.0061428
## 4  0.0215589 -0.0034858   0.0174353  -0.0285917 -0.0069534
## 5  0.0108225  0.0037167  -0.0101345   0.0291900  0.0409751
## 6  0.0101713 -0.0121978  -0.0083768   0.0137083  0.0029895
```

Summary statistics

```
(Xbar <- colMeans(stock))
```

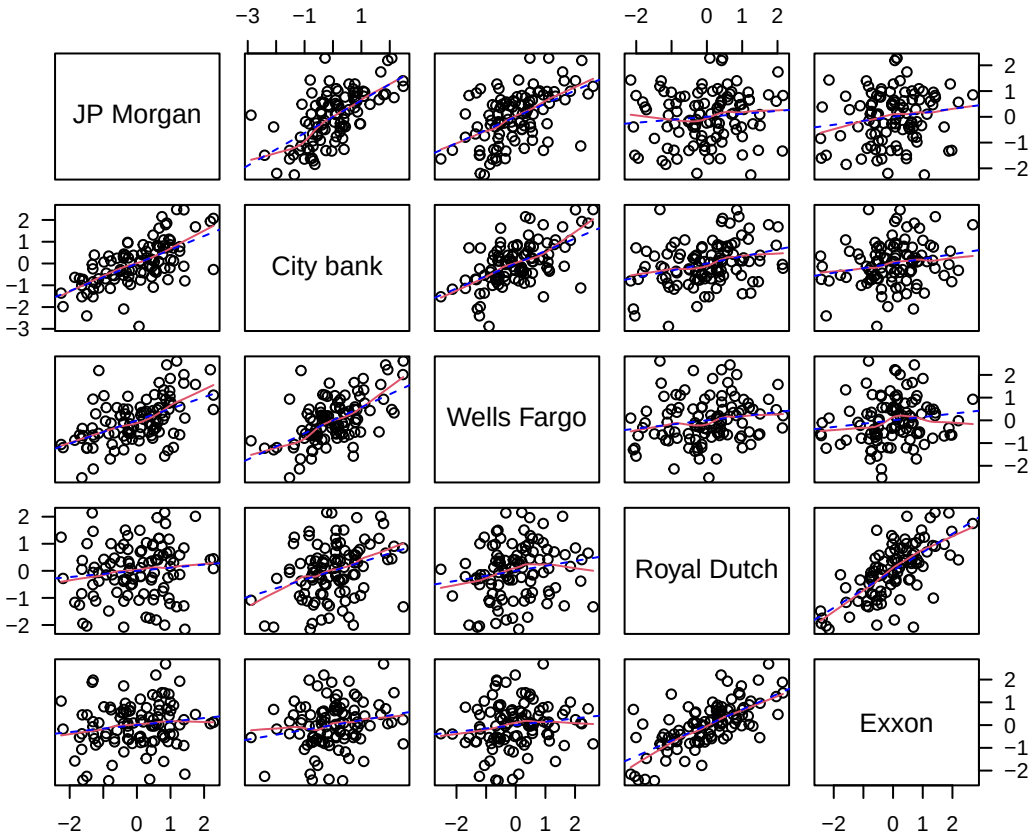
```
##    JP Morgan    City bank  Wells Fargo  Royal Dutch        Exxon
## 0.0010627806 0.0006554204 0.0016260816 0.0040491252 0.0040386417
```

```
(S <- cov(stock))
```

```
##                JP Morgan    City bank  Wells Fargo  Royal Dutch        Exxon
## JP Morgan   4.332695e-04 0.0002756679 1.590265e-04 6.411929e-05 8.896616e-05
## City bank   2.756679e-04 0.0004387172 1.799737e-04 1.814512e-04 1.232623e-04
## Wells Fargo 1.590265e-04 0.0001799737 2.239722e-04 7.341348e-05 6.054612e-05
## Royal Dutch 6.411929e-05 0.0001814512 7.341348e-05 7.224964e-04 5.082772e-04
## Exxon       8.896616e-05 0.0001232623 6.054612e-05 5.082772e-04 7.656742e-04
```

Scatter plot matrix

```
stock_std <- scale(stock, center = T, scale = T)
stock_std <- as.data.frame(stock_std)
par(las = 1, mgp = c(2.4, 1, 0), mar = c(3.5, 3.5, 1, 0.6))
pairs(stock_std, labels = names(stock),
      panel = function(x, y){panel.smooth(x, y)
abline(lsfit(x, y), lty = 2, col = "blue")})
```



Compute PCs from the correlation matrix

```
(s.cor <- var(stock_std))
```

```
##             JP Morgan City bank Wells Fargo Royal Dutch     Exxon
## JP Morgan   1.0000000 0.6322878   0.5104973   0.1146019 0.1544628
## City bank   0.6322878 1.0000000   0.5741424   0.3222921 0.2126747
## Wells Fargo 0.5104973 0.5741424   1.0000000   0.1824992 0.1462067
## Royal Dutch 0.1146019 0.3222921   0.1824992   1.0000000 0.6833777
## Exxon       0.1544628 0.2126747   0.1462067   0.6833777 1.0000000
```

```
s.pca <- prcomp(stock, scale = T, center = T)

s.pca$rotation
```

```
##                    PC1        PC2         PC3        PC4         PC5
## JP Morgan   -0.4690832 -0.3680070 -0.60431522 -0.3630228  0.38412160
## City bank   -0.5324055 -0.2364624 -0.13610618  0.6292079 -0.49618794
## Wells Fargo -0.4651633 -0.3151795  0.77182810 -0.2889658  0.07116948
## Royal Dutch -0.3873459  0.5850373  0.09336192  0.3812515  0.59466408
## Exxon       -0.3606821  0.6058463 -0.10882629 -0.4934145 -0.49755167
```

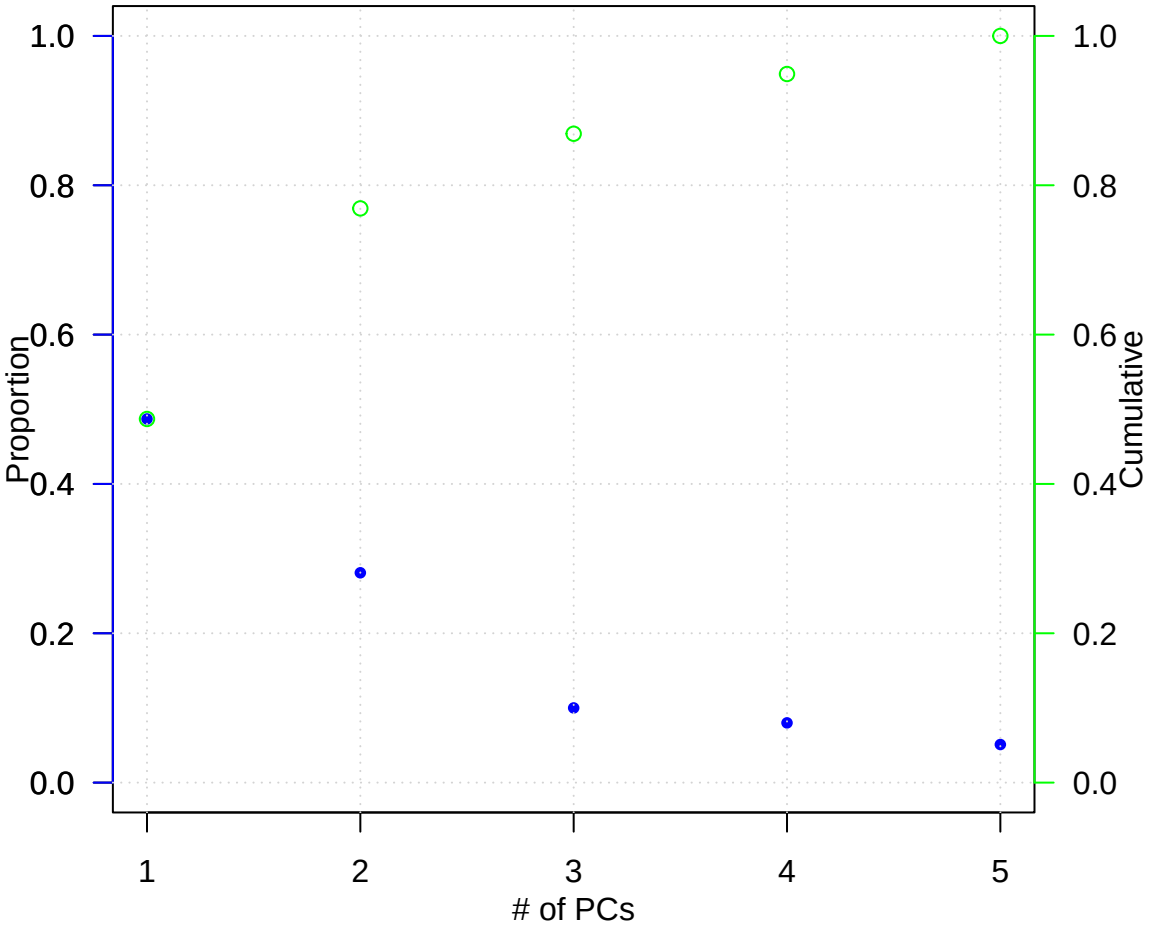
```
s <- var(s.pca$x)
(Proportion.std <- round(diag(s) / sum(diag(s)), 3))
```

```
##   PC1   PC2   PC3   PC4   PC5
## 0.487 0.281 0.100 0.080 0.051
```

```
(Cumulative.std <- round(cumsum(diag(s)) / sum(diag(s)), 3))
```

```
##   PC1   PC2   PC3   PC4   PC5
## 0.487 0.769 0.869 0.949 1.000
```

```
p <- 5
par(las = 1, mgp = c(2, 1, 0), mar = c(3, 3, 1, 3))
plot(1:p, Proportion.std, xlab = "# of PCs", ylim = c(0, 1),
ylab = "Proportion", pch = 16, cex = 0.8, xaxt = "n", col = "blue")
axis(1, at = 1:p)
mtext("Cumulative", 4, las = 0, line = 2)
axis(4, col = "green"); axis(2, col = "blue")
grid()
points(1:p, Cumulative.std, cex = 1, col = "green")
```



Factor Loadings and specific variances

```
# m = 2, factor Loadings
lambda <- s.pca$sdev^2
e <- s.pca$rotation
sqrt(lambda[1]) * e[, 1]
```

```
##   JP Morgan   City bank Wells Fargo Royal Dutch       Exxon
##  -0.7323218  -0.8311791  -0.7262022  -0.6047155  -0.5630885
```

```
sqrt(lambda[2]) * e[, 2]
```

```
##   JP Morgan   City bank Wells Fargo Royal Dutch       Exxon
##  -0.4365209  -0.2804859  -0.3738582   0.6939569   0.7186401
```

```
# specific variances
sVar <- diag(s.cor - (lambda[1] * e[, 1] %*% t(e[, 1]) + lambda[2] * e[, 2] %*% t(e[, 2])))
# residual matrix
round(s.cor - (lambda[1] * e[, 1] %*% t(e[, 1]) + lambda[2] * e[, 2] %*% t(e[, 2]) + diag(sVar)), 2)
```

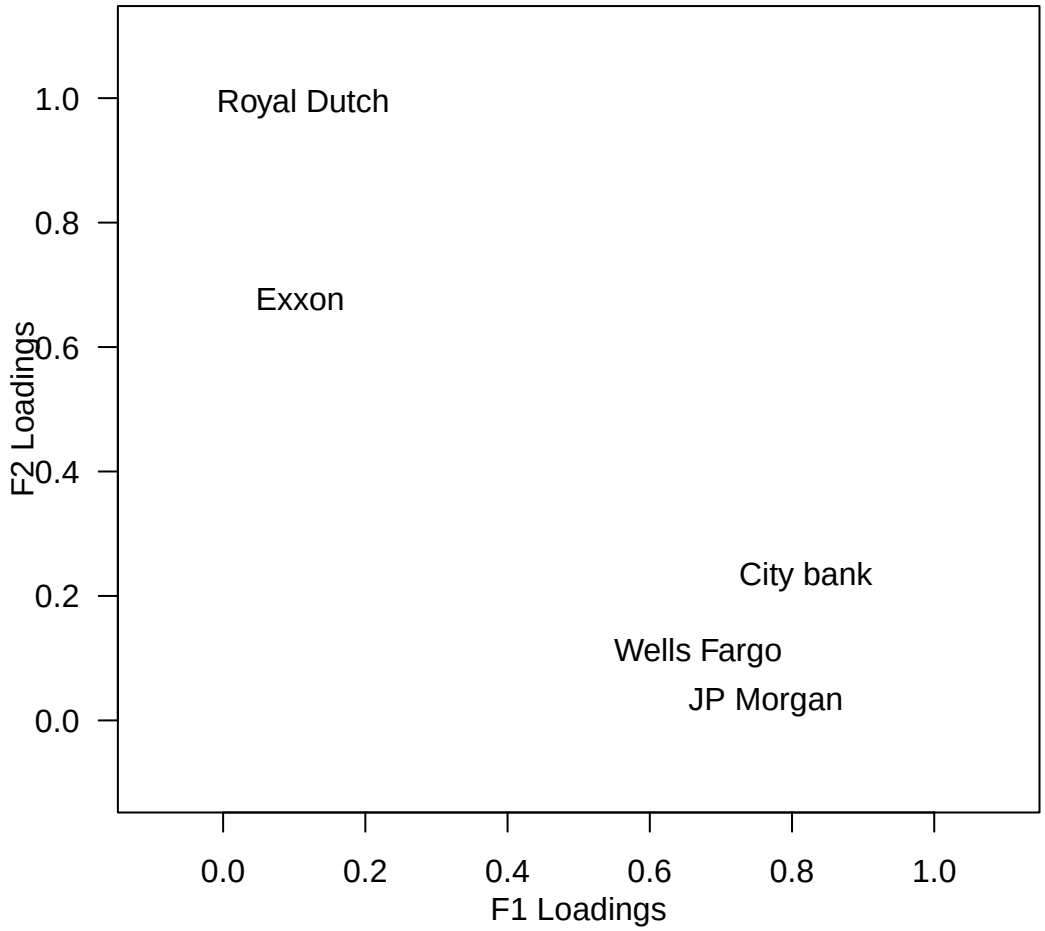
```
##             JP Morgan City bank Wells Fargo Royal Dutch Exxon
## JP Morgan        0.00     -0.10       -0.18       -0.03  0.06
## City bank       -0.10      0.00       -0.13        0.01 -0.05
## Wells Fargo     -0.18     -0.13        0.00        0.00  0.01
## Royal Dutch     -0.03      0.01        0.00        0.00 -0.16
## Exxon            0.06     -0.05        0.01       -0.16  0.00
```

MLE

```
(stock.fac <- factanal(stock, factors = 2, method = "mle", scale = T, center = T))
```

```
##
## Call:
## factanal(x = stock, factors = 2, method = "mle", scale = T, center = T)
##
## Uniquenesses:
##   JP Morgan   City bank Wells Fargo Royal Dutch       Exxon
##       0.417       0.275       0.542       0.005       0.530
##
## Loadings:
##             Factor1 Factor2
## JP Morgan   0.763
## City bank   0.819   0.232
## Wells Fargo 0.668   0.108
## Royal Dutch 0.113   0.991
## Exxon       0.108   0.677
##
##                Factor1 Factor2
## SS loadings      1.725   1.507
## Proportion Var   0.345   0.301
## Cumulative Var   0.345   0.646
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 1.97 on 1 degree of freedom.
## The p-value is 0.16
```

```
par(las = 1, mgp = c(2, 1, 0), mar = c(3, 3, 1, 3))
plot(stock.fac$loadings, xlab = "F1 Loadings", ylab = "F2 Loadings",
     type = "n", xlim = c(-0.1, 1.1), ylim = c(-0.1, 1.1))
text(stock.fac$loadings, labels = colnames(stock))
```




```
# residual matrix
pred <- (stock.fac$loadings %*% t(stock.fac$loadings)) + diag(stock.fac$uniqueness)
(resid <- s.cor - pred)
```

```
##                JP Morgan     City bank   Wells Fargo   Royal Dutch
## JP Morgan    1.055860e-07  7.496780e-06 -2.564223e-03 -3.325561e-04
## City bank    7.496780e-06  3.255673e-08  1.608871e-03  2.116218e-04
## Wells Fargo -2.564223e-03  1.608871e-03  5.157373e-08 -9.518792e-06
## Royal Dutch -3.325561e-04  2.116218e-04 -9.518792e-06 -1.559500e-06
## Exxon        5.198222e-02 -3.307885e-02  5.547153e-04  1.218853e-04
##                      Exxon
## JP Morgan    5.198222e-02
## City bank   -3.307885e-02
## Wells Fargo  5.547153e-04
## Royal Dutch  1.218853e-04
## Exxon        2.670491e-07
```