Principal Component Analysis (PCA) dan Factor Analysis (FA) pada Data Penjualan Mobil

Muhammad Radifan Asyauri & Muchammad Sandy Kurniawan

2025-03-26

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
df <- read.csv("Clean_Car_sales.csv")
head(df)
```

```
##   Sales_in_thousands X__year_resale_value Price_in_thousands Engine_size
## 1             16.919               16.360           21.50000         1.8
## 2             39.384               19.875           28.40000         3.2
## 3             14.114               18.225           27.39075         3.2
## 4              8.588               29.725           42.00000         3.5
## 5             20.397               22.255           23.99000         1.8
## 6             18.780               23.555           33.95000         2.8
##   Horsepower Wheelbase Width Length Curb_weight Fuel_capacity Fuel_efficiency
## 1        140     101.2  67.3  172.4       2.639          13.2              28
## 2        225     108.1  70.3  192.9       3.517          17.2              25
## 3        225     106.9  70.6  192.0       3.470          17.2              26
## 4        210     114.6  71.4  196.6       3.850          18.0              22
## 5        150     102.6  68.2  178.0       2.998          16.4              27
## 6        200     108.7  76.1  192.0       3.561          18.5              22
##   Power_perf_factor
## 1          58.28015
## 2          91.37078
## 3          77.04359
## 4          91.38978
## 5          62.77764
## 6          84.56511
```

```
str(df)
```

```
## 'data.frame':    157 obs. of  12 variables:
##  $ Sales_in_thousands  : num  16.92 39.38 14.11 8.59 20.4 ...
##  $ X__year_resale_value: num  16.4 19.9 18.2 29.7 22.3 ...
##  $ Price_in_thousands  : num  21.5 28.4 27.4 42 24 ...
##  $ Engine_size         : num  1.8 3.2 3.2 3.5 1.8 2.8 4.2 2.5 2.8 2.8 ...
##  $ Horsepower          : num  140 225 225 210 150 200 310 170 193 193 ...
##  $ Wheelbase           : num  101 108 107 115 103 ...
##  $ Width               : num  67.3 70.3 70.6 71.4 68.2 76.1 74 68.4 68.5 70.9 ...
```

```
##  $ Length              : num  172 193 192 197 178 ...
##  $ Curb_weight         : num  2.64 3.52 3.47 3.85 3 ...
##  $ Fuel_capacity       : num  13.2 17.2 17.2 18 16.4 18.5 23.7 16.6 16.6 18.5 ...
##  $ Fuel_efficiency     : num  28 25 26 22 27 22 21 26 24 25 ...
##  $ Power_perf_factor   : num  58.3 91.4 77 91.4 62.8 ...
```

```r
summary(df)
```

```
##  Sales_in_thousands X__year_resale_value Price_in_thousands  Engine_size
##  Min.   :  0.11     Min.   : 5.16        Min.   : 9.235     Min.   :1.000
##  1st Qu.: 14.11     1st Qu.:12.54        1st Qu.:18.145     1st Qu.:2.300
##  Median : 29.45     Median :17.71        Median :23.400     Median :3.000
##  Mean   : 53.00     Mean   :18.07        Mean   :27.391     Mean   :3.061
##  3rd Qu.: 67.96     3rd Qu.:18.14        3rd Qu.:31.930     3rd Qu.:3.500
##  Max.   :540.56     Max.   :67.55        Max.   :85.500     Max.   :8.000
##    Horsepower       Wheelbase          Width           Length
##  Min.   : 55.0   Min.   : 92.6   Min.   :62.60   Min.   :149.4
##  1st Qu.:150.0   1st Qu.:103.0   1st Qu.:68.40   1st Qu.:177.6
##  Median :180.0   Median :107.0   Median :70.60   Median :187.8
##  Mean   :185.9   Mean   :107.5   Mean   :71.15   Mean   :187.3
##  3rd Qu.:215.0   3rd Qu.:112.2   3rd Qu.:73.40   3rd Qu.:196.1
##  Max.   :450.0   Max.   :138.7   Max.   :79.90   Max.   :224.5
##   Curb_weight    Fuel_capacity   Fuel_efficiency Power_perf_factor
##  Min.   :1.895   Min.   :10.30   Min.   :15.00   Min.   : 23.28
##  1st Qu.:2.975   1st Qu.:15.80   1st Qu.:21.00   1st Qu.: 60.73
##  Median :3.368   Median :17.20   Median :24.00   Median : 72.29
##  Mean   :3.378   Mean   :17.95   Mean   :23.84   Mean   : 77.04
##  3rd Qu.:3.778   3rd Qu.:19.50   3rd Qu.:26.00   3rd Qu.: 89.40
##  Max.   :5.572   Max.   :32.00   Max.   :45.00   Max.   :188.14
```

```r
sum(is.na(df))
```

```
## [1] 0
```

```r
p <- ncol(df)
print(p)
```

```
## [1] 12
```

```r
library(psych)
r <- cor(df)
KMO(r)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = r)
## Overall MSA =  0.83
## MSA for each item =
##   Sales_in_thousands X__year_resale_value   Price_in_thousands
##                 0.72                 0.93                 0.67
##          Engine_size           Horsepower            Wheelbase
##                 0.95                 0.72                 0.82
##                Width               Length          Curb_weight
##                 0.97                 0.84                 0.88
##        Fuel_capacity      Fuel_efficiency    Power_perf_factor
##                 0.90                 0.92                 0.71
```

```r
bartlett.test(df)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df
## Bartlett's K-squared = 4692, df = 11, p-value < 2.2e-16
```

```r
df<- scale(df)
```

```r
pca_result <- prcomp(df, scale. = TRUE)


summary(pca_result)
```

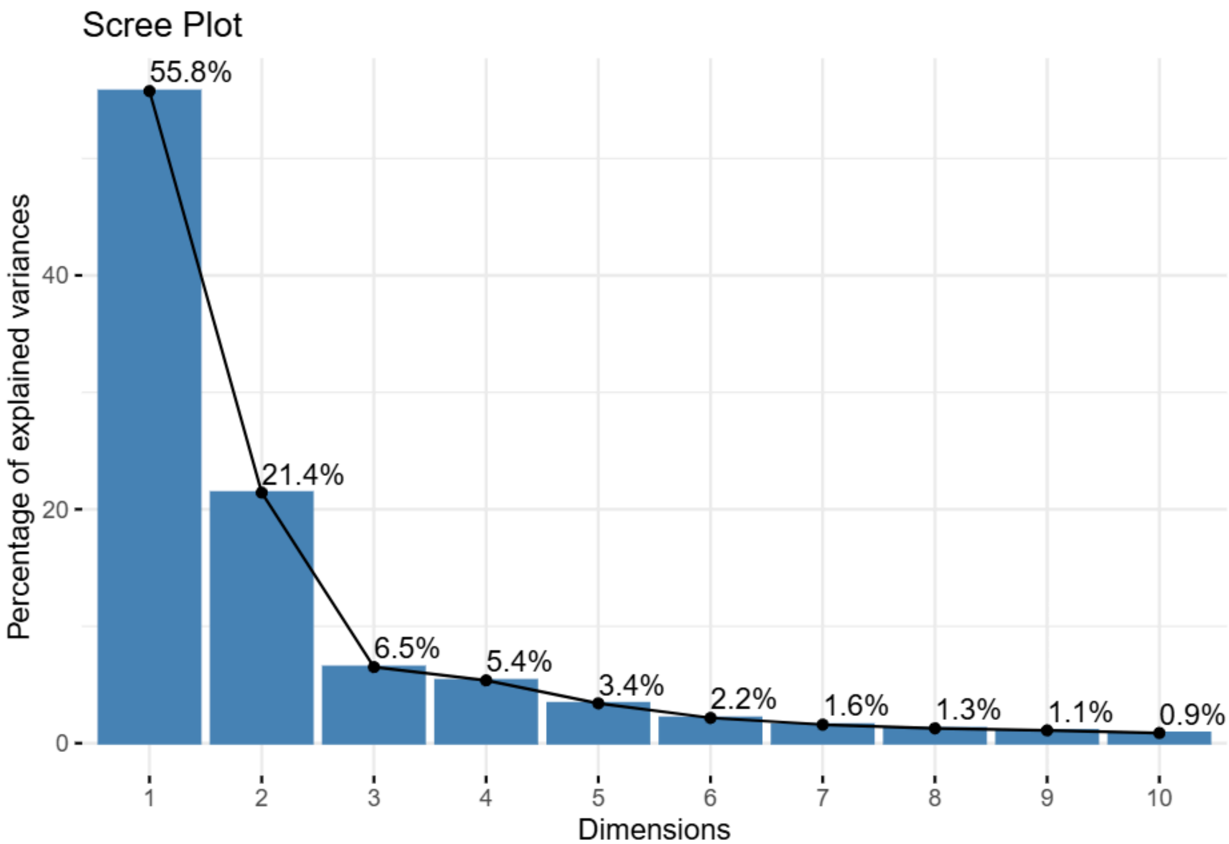
```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5    PC6     PC7
## Standard deviation     2.5867 1.6035 0.88398 0.80266 0.63921 0.5091 0.43652
## Proportion of Variance 0.5576 0.2143 0.06512 0.05369 0.03405 0.0216 0.01588
## Cumulative Proportion  0.5576 0.7719 0.83699 0.89067 0.92472 0.9463 0.96220
##                            PC8     PC9    PC10    PC11    PC12
## Standard deviation     0.39003 0.36294 0.32362 0.25277 0.03383
## Proportion of Variance 0.01268 0.01098 0.00873 0.00532 0.00010
## Cumulative Proportion  0.97488 0.98585 0.99458 0.99990 1.00000
```

```
## Loading required package: ggplot2
```

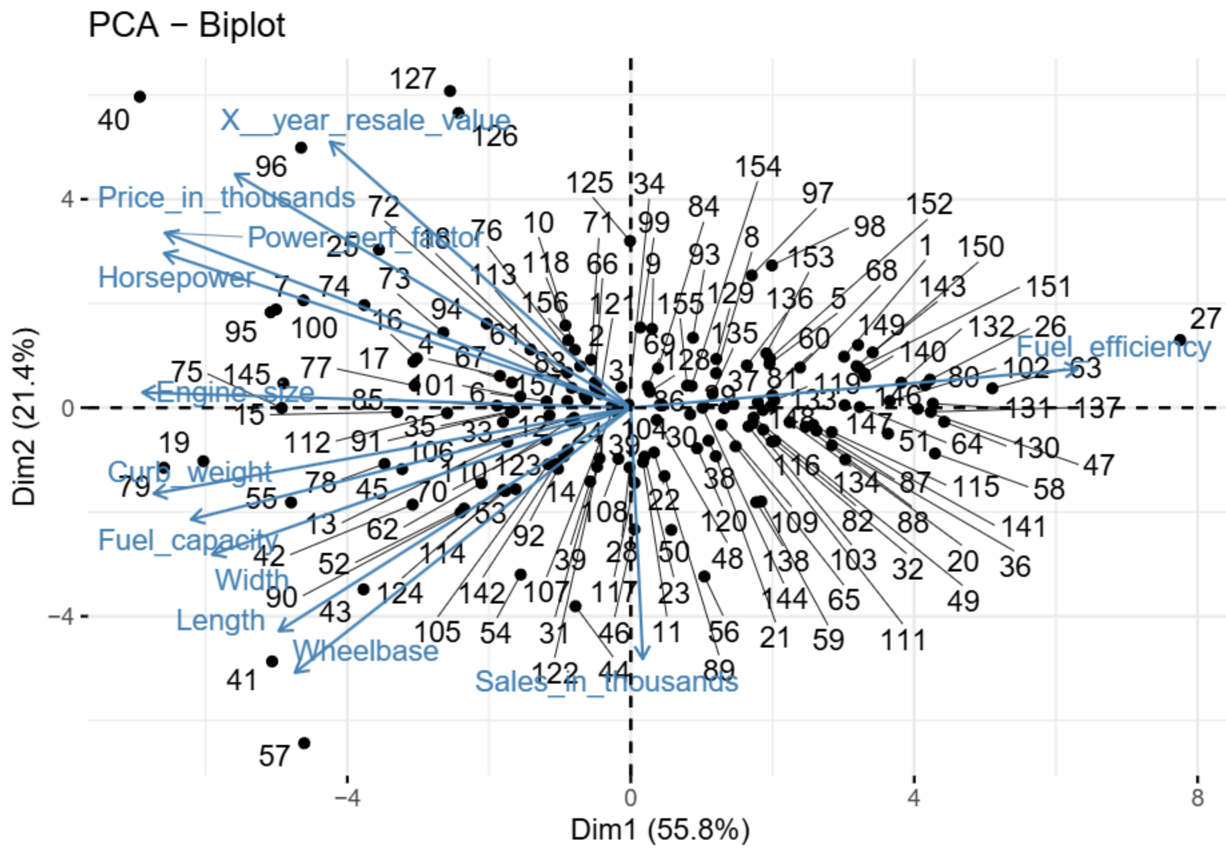
```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

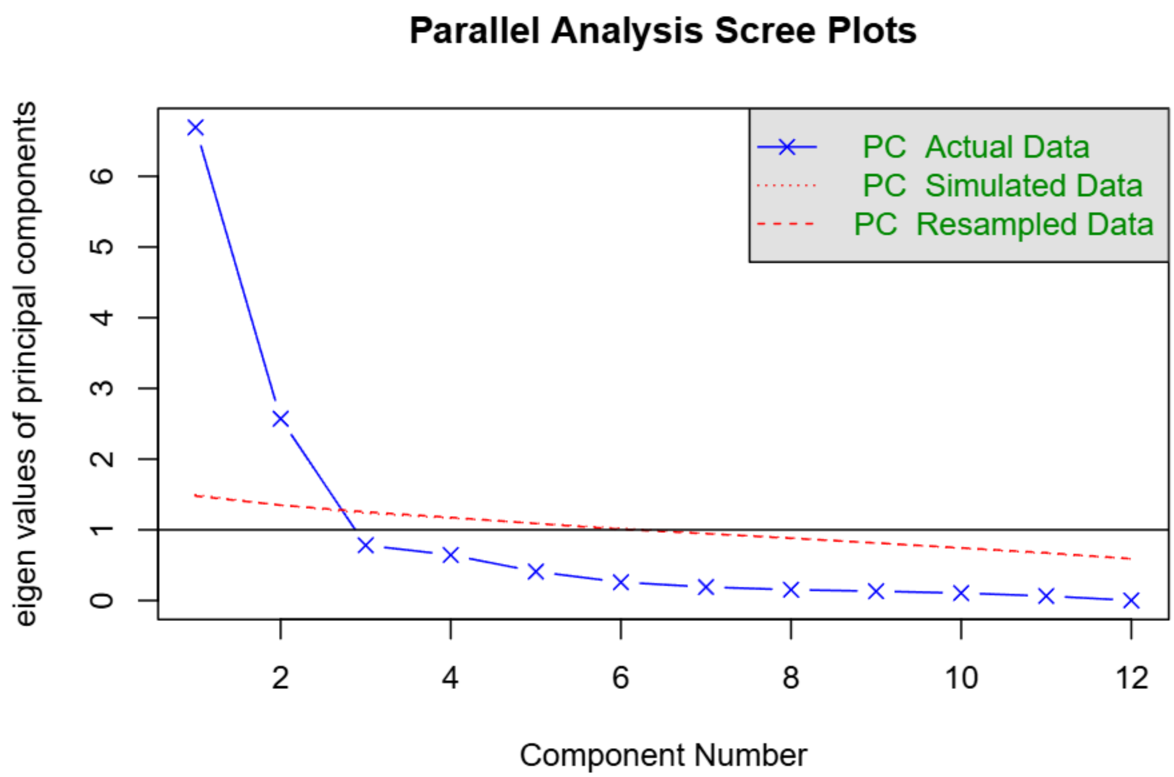
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```




```r
# Biplot PCA
fviz_pca_biplot(pca_result, repel = TRUE)
```




```r
# Parallel Analysis
#fa.parallel(df, fm = "ml", fa = "both")
fa.parallel(df, fa = "pc")
```




```
## Parallel analysis suggests that the number of factors =  NA  and the number of components =  2
```

```
#Factor Analysis
varcov <- cov(df)
pc <- eigen(varcov)
pc$values
```

```
##  [1] 6.691203830 2.571192763 0.781426944 0.644271014 0.408586245 0.259153970
##  [7] 0.190545442 0.152124933 0.131725757 0.104730296 0.063894102 0.001144704
```

```
pc$vectors
```

```
##               [,1]        [,2]         [,3]        [,4]        [,5]        [,6]
##  [1,]  0.008594942  0.39430485  0.744967314  0.49659316 -0.05791056 -0.03906725
##  [2,] -0.215153270 -0.41725016  0.256403822  0.12725271  0.48280190 -0.31930644
##  [3,] -0.283191520 -0.36679817  0.121181534  0.01806013  0.24902739 -0.05121043
##  [4,] -0.349360433 -0.02387282  0.104837881  0.01435761 -0.47660410  0.20660608
##  [5,] -0.333970559 -0.24240653  0.192115126 -0.15518349 -0.22100106  0.21228727
##  [6,] -0.240296794  0.41567976  0.059565578 -0.20189642  0.39896923  0.15971796
##  [7,] -0.299527557  0.23033279 -0.007239365 -0.27218856 -0.32314576 -0.80702331
##  [8,] -0.251818759  0.35034241  0.128237780 -0.48095448  0.23287613  0.25105300
##  [9,] -0.341327961  0.13486392 -0.292777115  0.15451925  0.06572817  0.02379266
## [10,] -0.314663595  0.17531663 -0.310570313  0.35037715  0.24910048 -0.11072841
## [11,]  0.319921952 -0.06100794  0.285302120 -0.45429037  0.15884398 -0.17053283
## [12,] -0.333329732 -0.27445477  0.181883371 -0.11576563 -0.12691551  0.15623800
##              [,7]        [,8]         [,9]       [,10]       [,11]
##  [1,] -0.11639319  0.01099162 -0.114074540  0.09776134  0.04182092
##  [2,]  0.48716925 -0.20959311 -0.009735307 -0.22350018  0.17694475
##  [3,] -0.46868117  0.24514882 -0.228094637  0.01244330 -0.58471399
##  [4,]  0.07302561 -0.46665744  0.148174828 -0.48952182 -0.33398623
##  [5,] -0.03748809  0.10346328  0.197425917  0.27066942  0.42393388
```

```
##  [6,] -0.07910795  0.37993197  0.424633385 -0.45306158  0.06545991
##  [7,]  0.02573961  0.15644351 -0.006426979  0.03825343 -0.02699552
##  [8,]  0.30291337 -0.23724368 -0.374494003  0.34830488 -0.19804493
##  [9,] -0.34370379 -0.15814197 -0.561094016 -0.26680529  0.47075855
## [10,] -0.16988410 -0.39451345  0.453643716  0.42173305 -0.09256427
## [11,] -0.50996623 -0.49208122  0.143257130 -0.07304446  0.15403916
## [12,] -0.13031900  0.13956015  0.110408292  0.21217511  0.18878296
##               [,12]
##  [1,]  1.185220e-03
##  [2,] -1.107114e-03
##  [3,]  1.782721e-01
##  [4,]  1.032526e-02
##  [5,]  6.088813e-01
##  [6,]  1.272035e-03
##  [7,]  3.277369e-03
##  [8,] -1.244756e-03
##  [9,] -7.725002e-03
## [10,] -2.609167e-05
## [11,] -6.514083e-03
## [12,] -7.728243e-01
```

```r
sp = sum(pc$values[1:3])
```

```r
L1 = sqrt(pc$values[1])*pc$vectors[,1]
L2 = sqrt(pc$values[2])*pc$vectors[,2]
L3 = sqrt(pc$values[3])*pc$vectors[,3]

L = cbind(L1,L2,L3)
L
```

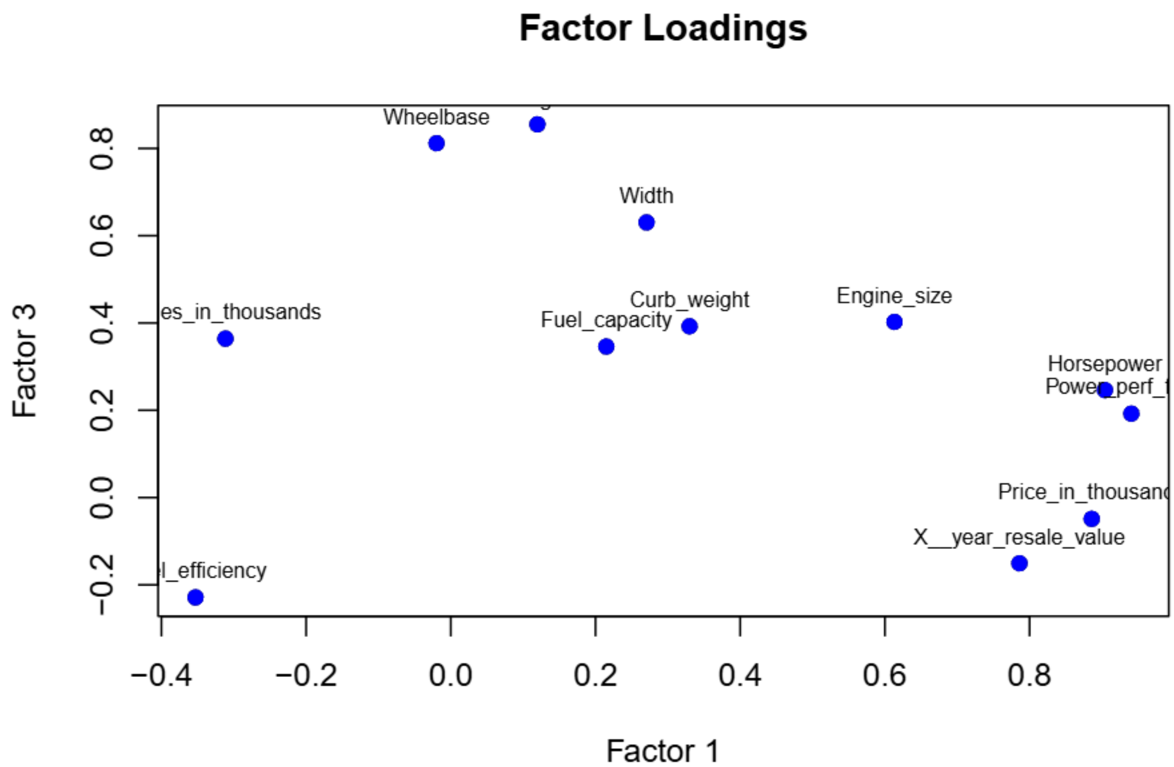
```
##                L1          L2          L3
##  [1,]  0.02223285  0.63226543  0.65853886
##  [2,] -0.55654474 -0.66905810  0.22665676
##  [3,] -0.73254174 -0.58815863  0.10712248
##  [4,] -0.90370326 -0.03827991  0.09267496
##  [5,] -0.86389371 -0.38869740  0.16982661
##  [6,] -0.62158440  0.66653998  0.05265499
##  [7,] -0.77479875  0.36933722 -0.00639948
##  [8,] -0.65138868  0.56177193  0.11336009
##  [9,] -0.88292537  0.21625347 -0.25881016
## [10,] -0.81395169  0.28111915 -0.27453905
## [11,]  0.82755367 -0.09782587  0.25220239
## [12,] -0.86223606 -0.44008656  0.16078191
```

```r
#Factor Analysis
library(psych)
fa <- fa(r = cor(df),
                nfactors = 3,
                rotate = "varimax")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect.  Try a
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected.  Examine the results carefully
```

```r
load <- fa$loadings
plot(load[,c(1,3)], type="p", xlab="Factor 1", ylab="Factor 3", main="Factor Loadings")
points(load[,c(1,3)], pch=19, col="blue")
text(load[,c(1,3)], labels=colnames(df), cex=.7, pos=3)
```




```r
fa.diagram(load)
```

Factor Analysis