

Practice: FA

Jay Wei

2023-10-17

Contents

1	Exercise 1	2
1.1	Question a	2
1.2	Question b	5
1.3	Question c	11
1.4	Question d	14

1 Exercise 1

This problem comes from Chapter 6 of Johnson's textbook. Below is the description from the book:

The data consist of anthropometric and physical fitness measurements that were taken on 50 applicants to the police department of a major metropolitan city. The variables include:

1. Reaction time in seconds to a visual stimulus (REACT)
2. The applicant's height in centimeters (HEIGHT)
3. The applicant's weight in kilograms (WEIGHT)
4. The applicant's shoulder width in centimeters (SHLDR)
5. The applicant's pelvic width in centimeters (PELVIC)
6. The applicant's minimum chest circumference in centimeters (CHEST)
7. The applicant's thigh skinfold thickness in millimeters (THIGH)
8. The applicant's resting pulse rate (PULSE)
9. The applicant's diastolic blood pressure (DIAST)
10. The number of chin-ups the applicant was able to complete (CHNUP)
11. The applicant's maximum breathing capacity in liters (BREATH)
12. The applicant's pulse rate after 5 minutes of recovery from treadmill running (RECVR)
13. The applicants maximum treadmill speed (SPEED)
14. The applicant's treadmill endurance time in minutes (ENDUR)
15. The applicant's total body fat measurement (FAT)

The data is in the file `PoliceApplicant.csv`. Using this data, complete the following.

1.1 Question a

Use PCA with the correlation matrix to help choose an initial number of common factors.

```
police <- read.csv("PoliceApplicant.csv")
head(police)
```

```
##   ID REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH PULSE DIAST CHNUP BREATH
## 1  1 0.310  179.6  74.20  41.7   27.3  82.4  19.0   64    64     2    158
## 2  2 0.345  175.6  62.04  37.5   29.1  84.1   5.5   88    78    20    166
## 3  3 0.293  166.2  72.96  39.4   26.8  88.1  22.0  100    88     7    167
## 4  4 0.254  173.8  85.92  41.2   27.6  97.6  19.5   64    62     4    220
## 5  5 0.384  184.8  65.88  39.8   26.1  88.2  14.5   80    68     9    210
## 6  6 0.406  189.1 102.26  43.3   30.1 101.2  22.0   60    68     4    188
##   RECVR SPEED ENDUR  FAT
## 1   108   5.5     4 11.91
## 2   108   5.5     4  3.13
## 3   116   5.5     4 16.89
## 4   120   5.5     4 19.59
## 5   120   5.5     5  7.74
## 6    91   6.0     4 30.42
```

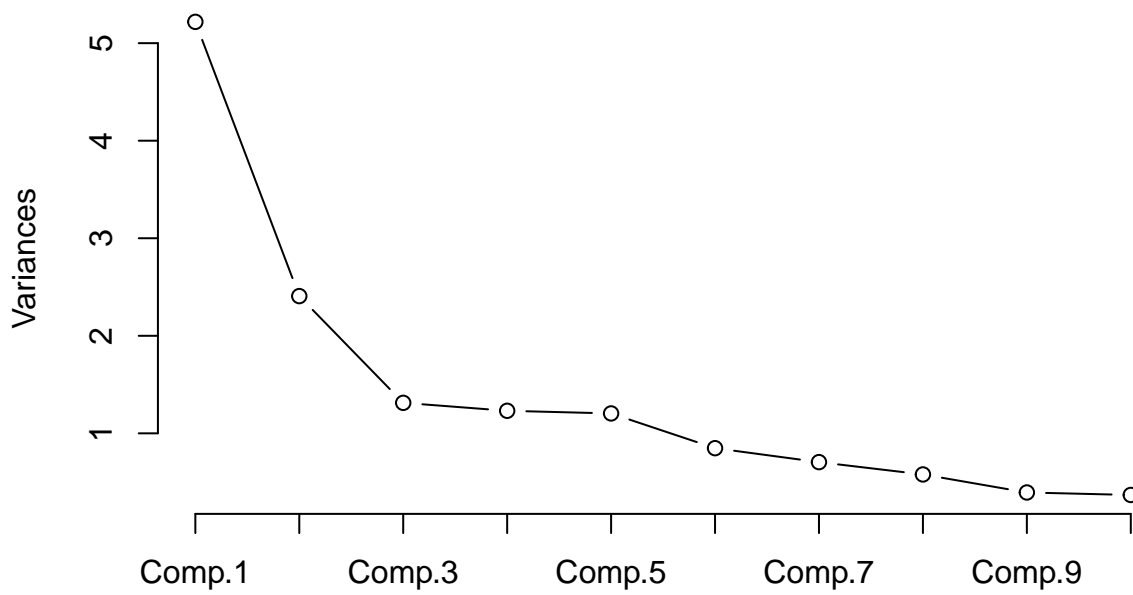
```
pca.cor <- princomp(formula = ~.-ID,
                     data = police, cor = TRUE, scores = TRUE)

summary(pca.cor, loadings = TRUE, cutoff = 0.0)
```

```
## Importance of components:
##               Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## Standard deviation 2.2844092 1.5513858 1.14572172 1.10953830 1.09719900
## Proportion of Variance 0.3479017 0.1604532 0.08751188 0.08207168 0.08025638
## Cumulative Proportion 0.3479017 0.5083549 0.59586678 0.67793847 0.75819484
##               Comp.6   Comp.7   Comp.8   Comp.9   Comp.10
## Standard deviation 0.92081715 0.8394927 0.76053179 0.62733307 0.60678801
## Proportion of Variance 0.05652695 0.0469832 0.03856057 0.02623645 0.02454611
## Cumulative Proportion 0.81472179 0.8617050 0.90026557 0.92650202 0.95104813
##               Comp.11  Comp.12  Comp.13  Comp.14  Comp.15
## Standard deviation 0.57147648 0.4322846 0.37256730 0.209478794 0.195281644
## Proportion of Variance 0.02177236 0.0124580 0.00925376 0.002925424 0.002542328
## Cumulative Proportion 0.97282049 0.9852785 0.99453225 0.997457672 1.000000000
##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
## REACT  0.051  0.152  0.111  0.056  0.821  0.112  0.040  0.196  0.253  0.044
## HEIGHT 0.305 -0.217  0.366  0.054  0.208 -0.154  0.192 -0.088  0.098  0.205
## WEIGHT 0.417 -0.029 -0.067  0.091 -0.056 -0.004  0.050  0.217 -0.146  0.018
## SHLDR  0.300 -0.211  0.275  0.106 -0.211 -0.061  0.298  0.178  0.398 -0.442
## PELVIC 0.294 -0.193  0.077  0.439 -0.099  0.106 -0.126 -0.226  0.198  0.403
## CHEST  0.361  0.004 -0.128  0.143 -0.172 -0.005 -0.139  0.489 -0.303 -0.035
## THIGH  0.284  0.307 -0.230 -0.205  0.012  0.106  0.348 -0.347  0.010  0.213
## PULSE -0.119  0.381  0.463  0.018 -0.003  0.186  0.329  0.330 -0.332 -0.021
## DIAST -0.036  0.277 -0.043  0.698  0.048  0.240  0.012 -0.354 -0.147 -0.401
## CHNUP -0.292 -0.236  0.240  0.216 -0.115  0.341 -0.012  0.142 -0.143  0.500
## BREATH 0.252 -0.032  0.439 -0.155  0.162 -0.239 -0.428 -0.327 -0.473 -0.113
## RECVR -0.025  0.424  0.400 -0.104 -0.380 -0.128  0.092 -0.184  0.219  0.142
## SPEED -0.029 -0.491  0.027 -0.198  0.041  0.314  0.469 -0.262 -0.314 -0.181
## ENDUR -0.204 -0.054 -0.131  0.326  0.084 -0.735  0.408  0.007 -0.233  0.160
## FAT    0.368  0.224 -0.233 -0.065  0.014  0.130  0.165  0.003 -0.194  0.224
##      Comp.11 Comp.12 Comp.13 Comp.14 Comp.15
## REACT  0.275  0.239  0.159  0.007  0.090
## HEIGHT 0.011 -0.692 -0.236 -0.155 -0.030
## WEIGHT 0.067 -0.083  0.031  0.837  0.158
## SHLDR -0.011  0.387 -0.227 -0.055 -0.226
## PELVIC -0.364  0.222  0.412 -0.103  0.162
## CHEST  0.328 -0.013  0.025 -0.479  0.337
## THIGH -0.067  0.240 -0.474 -0.083  0.371
## PULSE -0.497 -0.009  0.099 -0.034  0.076
## DIAST  0.156 -0.158 -0.119  0.010 -0.039
## CHNUP  0.295  0.216 -0.415  0.095 -0.141
## BREATH 0.009  0.303 -0.108  0.012 -0.071
## RECVR  0.514 -0.021  0.320  0.051  0.088
## SPEED  0.215  0.032  0.360 -0.039  0.140
## ENDUR  0.049  0.185  0.041  0.002 -0.013
## FAT    0.069  0.020  0.175 -0.102 -0.760
```

```
plot(pca.cor, type = "lines", main = "Scree plot for
      police applicants data")
```

Scree plot for police applicants data



There are 5 PCs with eigenvalues greater than 1, and they account for 76% of the total variance of 15 variables. For an initial choice, this appears to be a good place to start.

1.2 Question b

Using the initial number of common factors from Question a, examine the appropriate measures to judge their adequacy.

1.2.1 LRT involving 5 common factors

```
mod.fit5 <- factanal(x = ~ .-ID, factors = 5, data = police, rotation = "none")
print(x = mod.fit5, cutoff = 0.0)
```

```
##
## Call:
## factanal(x = ~. - ID, factors = 5, data = police, rotation = "none")
##
## Uniquenesses:
## REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH PULSE DIAST CHNUP BREATH
## 0.370 0.109 0.028 0.313 0.485 0.081 0.055 0.621 0.870 0.465 0.587
## RECVR SPEED ENDUR FAT
## 0.005 0.522 0.826 0.058
##
## Loadings:
## Factor1 Factor2 Factor3 Factor4 Factor5
## REACT 0.096 -0.127 -0.172 0.679 0.336
## HEIGHT 0.584 -0.142 0.451 0.050 0.569
## WEIGHT 0.956 -0.143 0.187 -0.010 -0.045
## SHLDR 0.591 -0.046 0.491 -0.249 0.183
## PELVIC 0.577 -0.225 0.333 -0.098 0.107
## CHEST 0.852 -0.104 0.240 0.051 -0.350
## THIGH 0.730 0.205 -0.578 -0.096 0.163
## PULSE -0.186 0.511 -0.075 0.276 0.048
## DIAST -0.018 0.158 -0.118 0.252 -0.165
## CHNUP -0.662 -0.056 0.290 -0.071 -0.060
## BREATH 0.444 0.077 0.297 0.130 0.323
## RECVR 0.017 0.997 0.024 -0.001 -0.001
## SPEED -0.137 -0.439 0.160 -0.389 0.299
## ENDUR -0.408 -0.067 0.041 0.037 -0.017
## FAT 0.915 0.058 -0.309 0.066 -0.049
##
##
## Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings 4.897 1.652 1.327 0.866 0.862
## Proportion Var 0.326 0.110 0.088 0.058 0.057
## Cumulative Var 0.326 0.437 0.525 0.583 0.640
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 53.8 on 40 degrees of freedom.
## The p-value is 0.0712
```

The LRT information is given at the bottom of the `print(x = mod.fit5, cutoff = 0.0)` output. Since $0.05 < 0.071 < 0.1$, there is marginal evidence that more factors are needed.

1.2.2 $R - (\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})$

```
resid5 <- mod.fit5$correlation - (mod.fit5$loadings[,]%*%t(mod.fit5$loadings[,]) +
                                diag(mod.fit5$uniquenesses))
resid5
```

##		REACT	HEIGHT	WEIGHT	SHLDR	PELVIC
##	REACT	8.539986e-06	2.714330e-04	5.118995e-04	3.546090e-02	-5.199363e-02
##	HEIGHT	2.714330e-04	-5.500328e-07	-1.396064e-03	-1.005446e-02	1.124051e-02
##	WEIGHT	5.118995e-04	-1.396064e-03	-5.256590e-07	8.051305e-03	4.629001e-03
##	SHLDR	3.546090e-02	-1.005446e-02	8.051305e-03	-3.934016e-06	2.413874e-02
##	PELVIC	-5.199363e-02	1.124051e-02	4.629001e-03	2.413874e-02	1.182686e-06
##	CHEST	-2.558584e-03	2.027441e-03	-1.096650e-03	5.423900e-03	-3.040201e-02
##	THIGH	-8.243637e-04	-9.586955e-04	-3.794022e-05	1.273631e-02	-1.824721e-03
##	PULSE	2.937015e-02	-8.341640e-03	5.802675e-03	6.116761e-02	-5.444920e-02
##	DIAST	3.326710e-02	-1.923744e-02	5.283770e-03	2.365846e-02	2.750147e-01
##	CHNUP	1.611057e-02	9.063911e-03	-8.187994e-03	-3.443334e-02	1.139800e-01
##	BREATH	-1.914137e-02	1.508887e-02	-3.208967e-03	-6.382092e-02	-6.051125e-03
##	RECVR	-2.392037e-04	9.864313e-05	-4.431670e-05	8.723471e-05	-4.938439e-04
##	SPEED	-3.837498e-04	1.069986e-02	-4.513798e-03	3.271263e-02	-1.014055e-01
##	ENDUR	-3.416962e-02	2.840699e-02	4.436254e-03	-1.495370e-02	-1.743654e-02
##	FAT	2.788000e-03	3.812331e-03	-4.782110e-04	-2.962523e-02	1.344981e-02
##		CHEST	THIGH	PULSE	DIAST	CHNUP
##	REACT	-2.558584e-03	-8.243637e-04	2.937015e-02	3.326710e-02	1.611057e-02
##	HEIGHT	2.027441e-03	-9.586955e-04	-8.341640e-03	-1.923744e-02	9.063911e-03
##	WEIGHT	-1.096650e-03	-3.794022e-05	5.802675e-03	5.283770e-03	-8.187994e-03
##	SHLDR	5.423900e-03	1.273631e-02	6.116761e-02	2.365846e-02	-3.443334e-02
##	PELVIC	-3.040201e-02	-1.824721e-03	-5.444920e-02	2.750147e-01	1.139800e-01
##	CHEST	-7.209827e-07	-1.950650e-03	-1.417948e-02	-1.879484e-02	1.766974e-02
##	THIGH	-1.950650e-03	-1.903187e-07	9.797642e-05	1.244222e-02	-3.710228e-03
##	PULSE	-1.417948e-02	9.797642e-05	-3.433513e-06	8.018811e-02	1.046268e-01
##	DIAST	-1.879484e-02	1.244222e-02	8.018811e-02	-9.478682e-06	9.287287e-02
##	CHNUP	1.766974e-02	-3.710228e-03	1.046268e-01	9.287287e-02	-5.072080e-06
##	BREATH	1.194158e-02	-1.798349e-03	-5.043100e-02	-1.133017e-01	-1.170180e-01
##	RECVR	-2.294222e-05	-7.648859e-05	-3.663588e-05	-9.957368e-05	4.232549e-04
##	SPEED	-5.109718e-03	-1.171691e-02	3.825663e-03	-8.811350e-02	1.521814e-01
##	ENDUR	-4.630895e-03	5.629066e-03	-4.059049e-02	1.296886e-01	-7.139768e-02
##	FAT	5.520577e-03	9.913805e-05	6.466862e-03	-9.153696e-03	9.411676e-03
##		BREATH	RECVR	SPEED	ENDUR	FAT
##	REACT	-1.914137e-02	-2.392037e-04	-3.837498e-04	-3.416962e-02	2.788000e-03
##	HEIGHT	1.508887e-02	9.864313e-05	1.069986e-02	2.840699e-02	3.812331e-03
##	WEIGHT	-3.208967e-03	-4.431670e-05	-4.513798e-03	4.436254e-03	-4.782110e-04
##	SHLDR	-6.382092e-02	8.723471e-05	3.271263e-02	-1.495370e-02	-2.962523e-02
##	PELVIC	-6.051125e-03	-4.938439e-04	-1.014055e-01	-1.743654e-02	1.344981e-02
##	CHEST	1.194158e-02	-2.294222e-05	-5.109718e-03	-4.630895e-03	5.520577e-03
##	THIGH	-1.798349e-03	-7.648859e-05	-1.171691e-02	5.629066e-03	9.913805e-05
##	PULSE	-5.043100e-02	-3.663588e-05	3.825663e-03	-4.059049e-02	6.466862e-03
##	DIAST	-1.133017e-01	-9.957368e-05	-8.811350e-02	1.296886e-01	-9.153696e-03
##	CHNUP	-1.170180e-01	4.232549e-04	1.521814e-01	-7.139768e-02	9.411676e-03
##	BREATH	5.297805e-06	1.514684e-04	-2.994609e-02	-1.399702e-01	-1.260707e-02
##	RECVR	1.514684e-04	-1.865728e-06	-1.146346e-04	-1.504836e-04	1.969954e-04

```
## SPEED -2.994609e-02 -1.146346e-04 -9.919192e-07 -9.509287e-02 3.525481e-02
## ENDUR -1.399702e-01 -1.504836e-04 -9.509287e-02 1.273403e-05 -1.913862e-02
## FAT -1.260707e-02 1.969954e-04 3.525481e-02 -1.913862e-02 -8.722174e-07
```

There are a lot of values to examine, so I used the following code to highlight those residuals that may be large in absolute value:

```
abs(resid5)>0.1
```

```
##      REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH PULSE DIAST CHNUP BREATH
## REACT FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## HEIGHT FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## WEIGHT FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## SHLDR FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## PELVIC FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE
## CHEST FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## THIGH FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## PULSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
## DIAST FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE FALSE TRUE
## CHNUP FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE TRUE
## BREATH FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE
## RECVR FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## SPEED FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE TRUE FALSE
## ENDUR FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE TRUE
## FAT FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##      RECVR SPEED ENDUR FAT
## REACT FALSE FALSE FALSE FALSE
## HEIGHT FALSE FALSE FALSE FALSE
## WEIGHT FALSE FALSE FALSE FALSE
## SHLDR FALSE FALSE FALSE FALSE
## PELVIC FALSE TRUE FALSE FALSE
## CHEST FALSE FALSE FALSE FALSE
## THIGH FALSE FALSE FALSE FALSE
## PULSE FALSE FALSE FALSE FALSE
## DIAST FALSE FALSE TRUE FALSE
## CHNUP FALSE TRUE FALSE FALSE
## BREATH FALSE FALSE TRUE FALSE
## RECVR FALSE FALSE FALSE FALSE
## SPEED FALSE FALSE FALSE FALSE
## ENDUR FALSE FALSE FALSE FALSE
## FAT FALSE FALSE FALSE FALSE
```

```
sum(abs(resid5)>0.1) # the number of true that comes up
```

```
## [1] 18
```

```
sum(abs(resid5)>0.2)
```

```
## [1] 2
```

```
max(abs(resid5))
```

```
## [1] 0.2750147
```

```
colMeans(abs(resid5)) # average residual for each column
```

```
##          REACT          HEIGHT          WEIGHT          SHLDR          PELVIC          CHEST
## 0.0151399408 0.0080465892 0.0031786248 0.0237552444 0.0471006952 0.0080886486
##          THIGH          PULSE          DIAST          CHNUP          BREATH          RECVR
## 0.0035935494 0.0306385132 0.0600750688 0.0500728309 0.0389654654 0.0001491723
##          SPEED          ENDUR          FAT
## 0.0380714711 0.0403803146 0.0098669264
```

Note that $R - (\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})$ will always be a symmetric matrix, so there are 105 (resid5 is a 15x15 data frame and thus $\frac{15*15-15}{2} = 105$) possible residuals that we need to investigate (diagonal elements of the matrix are always 0 due to the specific variances being added to the common factor part). The largest mean absolute deviation is for the DIAST variable with a value of 0.06. There are $18/2 = 9$ residuals that have an absolute value greater than 0.1, and $2/2 = 1$ residual that has an absolute value greater than 0.2. Overall, this is not too bad, but it does not hurt to investigate what would happen with 6 common factors.

1.2.3 LRT for a different number of common factors

```
mod.fit6 <- factanal(x = ~ .-ID, factors = 6, data = police, rotation = "none")
mod.fit6$STATISTIC # A
```

```
## objective
## 37.13529
```

```
mod.fit6$dof # deg of freedom
```

```
## [1] 30
```

```
mod.fit6$PVAL # p-value
```

```
## objective
## 0.1732341
```

```
resid6 <- mod.fit6$correlation - (mod.fit6$loadings[,]%*%t(mod.fit6$loadings[,]) +
                                diag(mod.fit6$uniquenesses))
sum(abs(resid6)>0.1)
```

```
## [1] 10
```



```
sum(abs(resid6)>0.2)
```

```
## [1] 2
```

```
max(abs(resid6))
```

```
## [1] 0.2092154
```

When using 6 common factors, the LRT gives a p-value of 0.17. There are $10/2=5$ residuals with an absolute value greater than 0.1, and $2/2=1$ residual with an absolute value greater than 0.2. The maximum possible residual in absolute value is 0.21. Overall, this is an improvement, but may be not enough to persuade me to use 6 common factors.

With respect to using a different number of common factors, below are the LRT results:

```
PVAL <- numeric(length = 7) # Pre-allocate PVAL vector
common.factor <- 1:7

mod.fit.list <- lapply(common.factor, function(i) {
  factanal(x = ~ . - ID, factors = i, data = police, rotation = "none")
})
PVAL <- sapply(mod.fit.list, function(mod) format(round(mod$PVAL, 6)))
result <- data.frame(common.factor, PVAL)
result
```

##	common.factor	PVAL
## 1	1	0
## 2	2	1e-05
## 3	3	0.000957
## 4	4	0.017685
## 5	5	0.071201
## 6	6	0.173234
## 7	7	0.615351

It appears that at least 4 are necessary. Note that with 4 common factors, there are $34/2 = 17$ residuals with an absolute value greater than 0.1, and $6/2 = 3$ residuals with an absolute value greater than 0.2. The maximum possible residual in absolute value is 0.26.

```
mod.fit4 <- factanal(x = ~ . - ID, factors = 4, data = police, rotation = "none")
mod.fit4$STATISTIC # A
```

```
## objective
## 74.46904
```

```
mod.fit4$dof # deg of freedom
```

```
## [1] 51
```

```
mod.fit4$PVAL # p-value
```

```
## objective  
## 0.01768521
```

```
resid4 <- mod.fit4$correlation - (mod.fit4$loadings[,]%*%t(mod.fit4$loadings[,]) +  
                                diag(mod.fit4$uniquenesses))  
sum(abs(resid4)>0.1)
```

```
## [1] 34
```

```
sum(abs(resid4)>0.2)
```

```
## [1] 6
```

```
max(abs(resid4))
```

```
## [1] 0.2613818
```

Overall, I will choose 5 common factors, but other choices are justifiable.

1.3 Question c

Using the varimax method, state the FA model for the number of common factors chosen. Interpret the common factors.

Notice the alternative way to include the variables with the x argument.

```
mod.fit5v <- factanal(x = police[, -1], factors = 5, rotation = "varimax")
print(x = mod.fit5v, cutoff = 0.0)
```



```
##
## Call:
## factanal(x = police[, -1], factors = 5, rotation = "varimax")
##
## Uniquenesses:
## REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH PULSE DIAST CHNUP BREATH
## 0.370 0.109 0.028 0.313 0.485 0.081 0.055 0.621 0.870 0.465 0.587
## RECVR SPEED ENDUR FAT
## 0.005 0.522 0.826 0.058
##
## Loadings:
##      Factor1 Factor2 Factor3 Factor4 Factor5
## REACT  0.092  0.075  0.060 -0.011  0.782
## HEIGHT 0.176  0.888 -0.164 -0.096  0.189
## WEIGHT 0.614  0.615 -0.187  0.424 -0.040
## SHLDR  0.193  0.747 -0.146  0.100 -0.247
## PELVIC 0.238  0.585 -0.272  0.195 -0.066
## CHEST  0.488  0.458 -0.112  0.666 -0.121
## THIGH  0.957  0.060  0.104 -0.117  0.042
## PULSE -0.079 -0.114  0.575 -0.089  0.146
## DIAST  0.037 -0.166  0.230  0.166  0.142
## CHNUP -0.690 -0.175 -0.028 -0.109 -0.124
## BREATH 0.166  0.598  0.082  0.011  0.145
## RECVR  0.102  0.059  0.948 -0.127 -0.258
## SPEED -0.191  0.166 -0.534 -0.327 -0.147
## ENDUR -0.354 -0.198 -0.028 -0.088  0.038
## FAT    0.895  0.245  0.017  0.273  0.080
##
##
##      Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings  3.149  2.843  1.760  0.948  0.905
## Proportion Var 0.210  0.190  0.117  0.063  0.060
## Cumulative Var 0.210  0.399  0.517  0.580  0.640
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 53.8 on 40 degrees of freedom.
## The p-value is 0.0712
```

Part of the model:

$$\begin{aligned} z_1 &= 0.092f_1 + 0.075f_2 + 0.060f_3 - 0.011f_4 + 0.782f_5 + \eta_1 \\ &\vdots \\ z_{15} &= 0.895f_1 + 0.245f_2 + 0.017f_3 + 0.273f_4 + 0.080f_5 + \eta_5 \end{aligned}$$

where z_1 is the standardized REACT variable and z_{15} is the standardized FAT variable (This is automatically done by the `factanal` function).

Interpretation of this model is not necessarily easy! Below is what Johnson first says about interpreting the common factors:

Interpretation of the rotated factors requires researchers to possess knowledge, experience, discretion, and wisdom, while remaining objective, and while suppressing, as much as possible, their own biases and prejudices. Researchers must carefully consider the population being sampled when making interpretations and should always keep in mind that the underlying factors are measuring unique and independent characteristics of the population that was sampled.

Please remember that the loadings for the common factors represent correlations between the original variables and the common factors. The farther these loadings are away from zero, the more of an association exists between the common factors and the original variables. To help then see which of these loadings are “away from zero”, one could set the `cutoff` argument value to something other than 0.0. Johnson uses 0.40, but this value could easily be set to something lower. Below is the output from using 0.40.

```
print(x = mod.fit5v, cutoff = 0.4)
```

```
##
## Call:
## factanal(x = police[, -1], factors = 5, rotation = "varimax")
##
## Uniquenesses:
## REACT HEIGHT WEIGHT SHLDR PELVIC CHEST THIGH PULSE DIAST CHNUP BREATH
## 0.370 0.109 0.028 0.313 0.485 0.081 0.055 0.621 0.870 0.465 0.587
## RECVR SPEED ENDUR FAT
## 0.005 0.522 0.826 0.058
##
## Loadings:
## Factor1 Factor2 Factor3 Factor4 Factor5
## REACT 0.782
## HEIGHT 0.888
## WEIGHT 0.614 0.615 0.424
## SHLDR 0.747
## PELVIC 0.585
## CHEST 0.488 0.458 0.666
## THIGH 0.957
## PULSE 0.575
## DIAST
## CHNUP -0.690
## BREATH 0.598
## RECVR 0.948
## SPEED -0.534
## ENDUR
## FAT 0.895
##
## Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings 3.149 2.843 1.760 0.948 0.905
## Proportion Var 0.210 0.190 0.117 0.063 0.060
## Cumulative Var 0.210 0.399 0.517 0.580 0.640
```

##

Test of the hypothesis that 5 factors are sufficient.

The chi square statistic is 53.8 on 40 degrees of freedom.

The p-value is 0.0712

Please remember that the common factors are independent of each other. Thus, each common factor needs to have a distinct interpretation. Below are possible interpretations of the common factors where I draw heavily upon Johnson's thoughts about the data. A researcher could likely judge better why particular combinations of body size measurements make sense.

Factor 1: This may be a measurement of body size. In particular, this could be a measure of obesity level. The negative correlation with CHNUP would make sense because the larger one's obesity level, the smaller number of chin-ups that one could complete (generally speaking). Again, hopefully, the subject matter researcher could make more sense of this.

Factor 2: This again could be a measurement of body size. In this case, it may be more geared toward skeletal structure due to the inclusion of variables like height and pelvic.

Factor 3: Cardiovascular fitness level

Factor 4: This again could be a measurement of body size. Johnson acknowledged difficulty with interpreting it. He suggests In this case, it may be with respect to a "measure of upper body strength" and whether or not an individual lifts weights.

Factor 5: Reaction time

The fifth common factor represents only one original variable, REACT, and it does not appear in any of other common factors. It may make sense to consider this variable separately from the other variables.

Johnson also points out that DIAST and ENDUR do not appear as a substantial component of any common factor. Thus, he suggests that the data is truly 7-dimensional (5+2).

As a reminder, the judgments above are based on using a cut-off of 0.4 for the common factor loadings. Other cut-offs could lead to different interpretations. This is a problem with a factor analysis and other explanatory analysis methods. A key then is to state your assumptions and limitations for your analysis.

1.4 Question d

Examine and interpret the appropriate plots of the factor scores.

Examining the factor scores could be used in situations where one wants to identify applicants with desirable characteristics. For example, maybe large common factor #1 values would be desirable. In that case, applicant #15 may be the most desirable:

```
mod.fit5vreg <- factanal(x = ~.-ID, factors = 5, data = police,  
                        rotation = "varimax", scores = "regression")  
factor.scores <- mod.fit5vreg$scores  
factor.scores[factor.scores[,1] == max(factor.scores[,1]),]
```

```
##      Factor1      Factor2      Factor3      Factor4      Factor5  
## 2.48327059  0.23779632  1.90241144 -0.31495394  0.07053816
```

Obviously, simple scatter plots will not work well here due to the number of common factors. Instead, we could use plots that allow for a larger number of dimensions.

```
library(MASS)  
  
col.f1 <- ifelse(test = factor.scores[,1] == max(factor.scores[,1]),  
                yes = "red", no="skyblue")  
  
parcoord(x = factor.scores, col.f1, main = "Factor score plot (regression method)")
```

Factor score plot (regression method)

