Exercise 7.7.

We consider the data given in Exercise 7.1. where we exclude the variables 'Name' and 'points'. The data are available in the dataset heptathlon.csv.

1. Make a factor analysis with n=2 factors and rotate the factors by means of the Varimax criterion.
2. What characterizes an athlete with a high score of factor 1 and of factor 2? Are there big differences between the interpretation of the unrotated and the rotated factors?
3. What is the fraction of the total variation described by each of the 2 unrotated and the 2 rotated factors, and what is the fraction described by the 2 factors combined in both cases?
4. For the variable 'hurdles', the factor pattern values are $[-0.95273, 0.18394]$ in the unrotated and $[0.96905, -0.04985]$ in the rotated case. Determine the 2 sums of squares $(-0.95273)^2+0.18394^2$ and $0.96905^2+(-0.04985)^2$. How may those quantities be interpreted?

   Now perform a factor analysis with n=3 factors

5. The variance explained by rotated factor no 3 is 1.0559726. Which fraction is this of the total variance? How may this quantity be determined from the factor pattern values (factor loadings)?
6. Are there big differences between the interpretation of the unrotated and the rotated factors for n=3 factors?
7. Determine the (factor1, factor2) score plot, where you plot the scores against each other in the cases with 2 rotated and 3 rotated factors!

```
#Hint: use factor.scores() to get the rotated scores
```

8. Relate the performance of different athletes to their positions in the factor score plots.

Sample program for Exercise 7.7 illustrating the computation of factor scores:

```
Hep <-Heptathlon[,-c(1,2)]

rfa <-principal(cor(Hep),nfactors = 2,rotate = "varimax",scores = T)
rfal <- rfa$loadings[,1:2]
rfacom <- rfa$communality

print("Factor Loadings:")
rfal
print("Communality:")
rfacom
## [1] "Factor Loadings:"
##                   RC1         RC2
```

```
## Hurdles    0.96904538 -0.04986118
## High.Jump -0.83885384 -0.06041963
## Shot      -0.64656995  0.53116479
## Run200     0.76820184 -0.44140527
## Longjump  -0.94657391  0.18224443
## Javelin    0.07491091  0.92910851
## Run800     0.82419326  0.04040023
## [1] "Communality:"
##   Hurdles High.Jump      Shot    Run200  Longjump   Javelin    Run800
## 0.9415351 0.7073263 0.7001887 0.7849727 0.9292152 0.8688543 0.6809267
competitors = Heptathlon$Name

#Scores plot for 2 Factors
F2 <- matrix(c(rfal[,1],rfal[,2]),ncol=2)

#Thurstone: regression based weights
ScoringF2 <- factor.scores(Hep,F2,method = "Thurstone")
Scoring_PointsF2 <- ScoringF2$scores

scores = data.frame(Scoring_PointsF2)
ggplot(scores,aes(x = scores[,1],y = scores[,2]))+
  geom_point(col = 'red',size = 2)+
  geom_text_repel(aes(label = competitors),box.padding = 0.5,cex=3)+
  labs(x = 'Factor 1',y = 'Factor 2')+
  ggtitle('2 Rotated Factors')
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```