Homework 5 Answer Key

1. (10 pts) Ex. 5.1 Show how the result $\sum = \Lambda\Lambda^T + \psi$ arises from the assumptions of uncorrelated factors, independence of the specific variates, and independence of common factors and specific variances. What form does Σ take if the factors are allowed to be correlated?

   a. uncorrelated factors, independence of the specific variates, and independence of common factors and specific variances.
   $\text{cov}(x) = \text{cov}(\Lambda f + u) = \Lambda\, cov\,(f)\Lambda^T + \text{cov}(u) = \Lambda\Lambda^T + \Psi$, $\Sigma = \Lambda\Lambda^T + \Psi$

   b. factors are allowed to be correlated

   $\text{cov}(x) = \text{cov}(\Lambda f + u) = \Lambda\, cov\,(f)\Lambda^T + \text{cov}(u) = \Lambda cov\,(f)\Lambda^T + \Psi$,
   $\Sigma = \Lambda cov\,(f)\Lambda^T + \Psi$

2. (10 pts) Ex. 5.2 Show that the communalities in a factor analysis model are unaffected by the transformation $\Lambda^* = \Lambda M$, where M is orthonormal matrix.

- Suppose that the $\boldsymbol{k}$-factor model holds for $\mathbf{x} = \Lambda\boldsymbol{f} + \boldsymbol{u}$, then the covariance matrix of X: $\Sigma = \Lambda\Lambda^T + \Psi$.

- Let M be a $k \times k$ orthonormal matrix.
  $x = \Lambda M M^T f + u \Rightarrow x = \Lambda^* f^* + u,$ where $\Lambda^* = \Lambda M$ and $f^* = M^T f$.
  Compute the covariance of x

$$\Sigma = \Lambda^*\Lambda^{*\mathrm{T}} + \Psi$$
$$\Sigma = \Lambda M(\Lambda M)^T + \Psi$$
$$\Sigma = \Lambda M M^T\Lambda^{\mathrm{T}} + \Psi$$
$$\Sigma = \Lambda\Lambda^T + \Psi$$

  Since the covariance matrix of X are the same before and after orthonormal transformation, therefore, the communalities are the same before and after the transformation. That is, the communalities in a factor analysis model are unaffected by the transformation.

3. (10 pts) Ex. 5.3 Give a formula for the proportion of variance explained by the jth factor estimated by the principal factor approach. (**This question Page 159 is not very clear**)

The variance of the ith variable σ_i^2, $\sigma_i^2 = var(x_i) = \sum_{j=1}^{k}\lambda_{ij}^2 + \psi_i$.
$h_i^2 = \sum_{j=1}^{k}\lambda_{ij}^2$ contributed by all k common factors is called the ith communality.

The portion of the variance of the ith variable σ_i^2, contributed by the common factors $\boldsymbol{f}_j$ is

$$cov(x_i, f_j) = \lambda_{ij}$$

The proportion of variance σ_i^2 explained by the jth factor $\boldsymbol{f}_j$ estimated by the principal factor approach

$$P_i = \frac{\lambda_{ij}}{\sum_{j=1}^{k}\lambda_{ij}^2 + \psi_i}$$

4. (20 pts) Ex. 5.4 Apply the factor analysis model separately to the life expectancies (life.csv on canvas) of men and women and compare the results.

   Method 1 Principal Factor Analysis

```
For men
> cor<-cor(lifem)
> eig<-eigen(cor)
> vectors<-as.matrix(eig$vectors)
> lam<-eig$values
> a<-diag(sqrt(lam))
> lamda<-vectors%*%a
> sum<-cumsum(lam)
> p<-sum/4
> values<-as.matrix(lam)
> f1<-sqrt(values[1,])*(eig$vectors[,1])
> f2<-sqrt(values[2,])*(eig$vectors[,2])
> phi
[1] 0.11048298 0.14691000 0.09568262 0.04207694
> p
[1] 0.6905263 0.9012119 0.9635866 1.0000000
```

For men, based off of the above p, we can see that if $k = 2$ that will account for 90.1% of the total variation. However, two factors are too many if there are only four variables so we will use k=1.

```
> f1
[1] -0.8142875 -0.8547770 -0.9236369 -0.7178386
> f2
[1]  0.4758707  0.3499233 -0.2263014 -0.6653050
```

Running the same code for the women:

```
[1] 0.8085233 0.9777846 0.9970334 1.0000000
```

Based off these values, we could choose k=1.

```
> f1
[1] -0.8295070 -0.9826253 -0.9723135 -0.7969098
> f2
[1]  0.5407566  0.1540058 -0.1377002 -0.5847633
> phi
[1] 0.01950038 0.01072965 0.03564501 0.02298666
```

- Compare factors

Age (in years)	Men: F1	Women: F1
0	0.814	0.83
25	.855	0.982
50	.924	.972
75	0.718	0.797

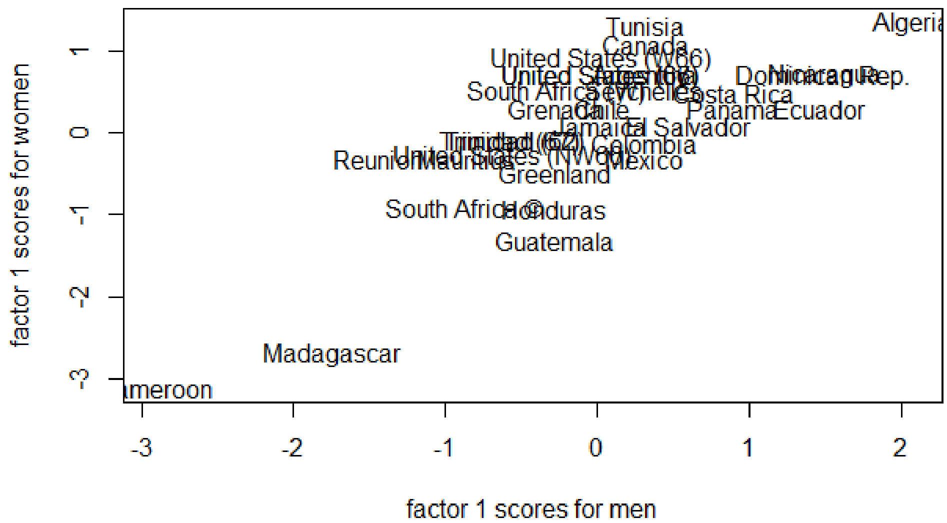
Comparing Factor1: Factor 1 is similar for both men and women with high factor loadings on all variables.

Age (in years)	Men: F1	Women: F1
0	-.476	.54
25	-.35	.154
50	.226	-.138
75	.665	-.585

Comparing Factor2: Factor 2 has highest factor loadings for 0 and 75 for both men and women, but the signs are flipped.

Since the factor 1 for both men and women covers the most of variations (70% for men, 81% for women) of the data, so the structures of the life expectancies for both men and women are similar overall.

- Compare life expectancies of countries by using factor 1 for both men and women



a. The countries are located along the diagonal, which means that the life expectancies for both men and women in the same country have the similar structure.

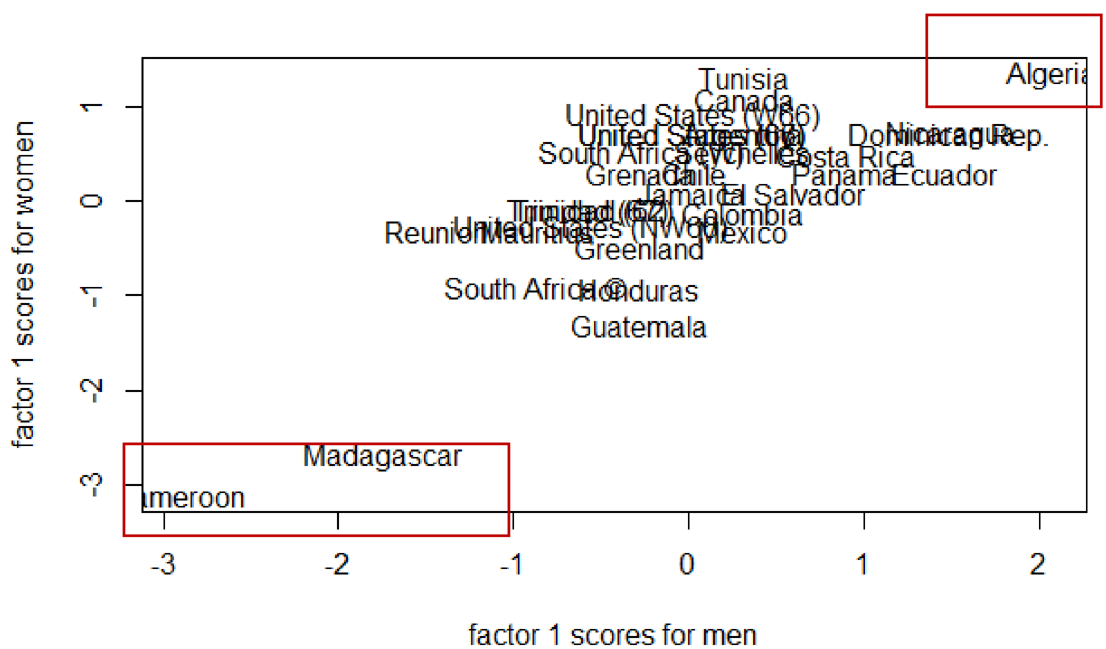
b. We see that the single factor for women measures life expectancy of younger and middle-aged women. From the plot of factor scores below, it appears Algeria does well for both life expectancy for middle-aged and elderly men and life expectancy of younger and middle-aged women. Cameroon and Madagascar does poorly for both. No country does really poorly for one factor and well for the other. Tunisia does very well for life expectancy of younger and middle-aged women and just medium for life expectancy for middle-aged and elderly men.

Method 2: Maximum likelihood estimate factor analysis

```
> factanal(life_men, factors = 1, method ="mle", rotation="none")

Call:
factanal(x = life_men, factors = 1, rotation = "none", method = "mle")

Uniquenesses:
   m0   m25   m50   m75 
0.594 0.552 0.005 0.434 

Loadings:
    Factor1
m0  0.638  
m25 0.669  
m50 0.998  
m75 0.752  

               Factor1
SS loadings      2.415
Proportion Var   0.604

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 14.45 on 2 degrees of freedom.
The p-value is 0.000728 
> factanal(life_women, factors = 1, method ="mle", rotation="none")
```

The ‘factanal’ function will only allow the 1-factor solution for this 4-variable data set. We see the single factor mainly measures life expectancy for middle-aged and elderly men.

```
> factanal(life_women, factors = 1, method ="mle", rotation="none")

Call:
factanal(x = life_women, factors = 1, rotation = "none", method = "mle")

Uniquenesses:
   w0   w25   w50   w75 
0.220 0.005 0.115 0.526 

Loadings:
    Factor1
w0  0.883  
w25 0.998  
w50 0.941  
w75 0.689  

               Factor1
SS loadings      3.134
Proportion Var   0.784

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 52.15 on 2 degrees of freedom.
The p-value is 4.74e-12 
> 
```



c. **The countries are located along the diagonal, which means that the life expectancies for both men and women in the same country have the similar structure.**
d. **We see that the single factor for women measures life expectancy of younger and middle-aged women. From the plot of factor scores below, it appears Algeria does well for both life expectancy for middle-aged and elderly men and life expectancy of younger and middle-aged women. Cameroon and**

Madagascar does poorly for both. No country does really poorly for one factor and well for the other. Tunisia does very well for life expectancy of younger and middle-aged women and just medium for life expectancy for middle-aged and elderly men.

```
library("MVA")
life<-read.csv("D:/STAT 4400/Data/life.csv", row.names = 1, header=TRUE)
life<-as.matrix(life)
life_men<-life[,1:4]
life_women<-life[,5:8]
sapply(1, function(f) factanal(life_men, factors = f, method ="mle")$PVAL)
sapply(1, function(f) factanal(life_women, factors = f, method ="mle")$PVAL)
factanal(life_men, factors = 1, method ="mle", rotation="none")
factanal(life_women, factors = 1, method ="mle", rotation="none")
men.fa.1.scores <- factanal(life_men,factors=1, rotation="none",scores="regression")$scores
women.fa.1.scores <- factanal(life_women,factors=1, rotation="none",scores="regression")$scores
plot(men.fa.1.scores,women.fa.1.scores,type='n',xlab="factor 1 scores for men",ylab='factor 1 scores for women')
text(men.fa.1.scores,women.fa.1.scores,labels=rownames(life))
```

5. (15 pts) Ex. 5.5 The correlation matrix given below arises from the scores of 220 boys in six school subjects: (1) French, (2) English, (3) History, (4) Arithmetic, (5) Algebra, and (6) Geometry. Find the two-factor solution from a maximum likelihood factor analysis. By plotting the derived loadings, find an orthogonal rotation that allows easier interpretation of the results.

$$
\mathbf{R} = \begin{array}{l} \text{French} \\ \text{English} \\ \text{History} \\ \text{Arithmetic} \\ \text{Algebra} \\ \text{Geometry} \end{array}
\begin{pmatrix}
1.00 & & & & & \\
0.44 & 1.00 & & & & \\
0.41 & 0.35 & 1.00 & & & \\
0.29 & 0.35 & 0.16 & 1.00 & & \\
0.33 & 0.32 & 0.19 & 0.59 & 1.00 & \\
0.25 & 0.33 & 0.18 & 0.47 & 0.46 & 1.00
\end{pmatrix}.
$$

```
> corr <- matrix( c( 1,.44,.41,.29,.33,.25, .44,1,.35,.35,.32,.33, .41,.35,1,.16,.19,.18, .29,.35,
,.19,.59,1,.46, .25,.33,.18,.47,.46,1), nrow=6, ncol=6, byrow=T)
> factanal(covmat=corr, factors=2, rotation='varimax', n.obs=220)

Call:
factanal(factors = 2, covmat = corr, n.obs = 220, rotation = "varimax")

Uniquenesses:
[1] 0.508 0.595 0.644 0.377 0.440 0.628

Loadings:
     Factor1 Factor2
[1,] 0.233   0.661    French
[2,] 0.319   0.551    English
[3,]         0.591    History
[4,] 0.770   0.172    Arithmetic
[5,] 0.715   0.220    Algebra
[6,] 0.570   0.215    Geometry

               Factor1 Factor2
SS loadings      1.593   1.215
Proportion Var   0.265   0.202
Cumulative Var   0.265   0.468

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 2.18 on 4 degrees of freedom.
The p-value is 0.703
> 1-factanal(covmat=corr, factors=2, rotation='varimax', n.obs=220)$uniquenesses
[1] 0.4916376 0.4054227 0.3560394 0.6227685 0.5597374 0.3716486
```

- **We see that 2 factors are significant based on the chi-square test (p-value = 0.703).**
- **The first factor is mostly a "math ability" factor, with heavy loadings on arithmetic, algebra, and geometry.**
- **The second factor is a "humanities ability" factor, with higher loadings on French, English, and history. This is apparent to some degree in the book's unrotated solution, but it is much more apparent after we do a varimax rotation, as shown above.**

6. (15 pts) Ex. 5.6 The matrix below shows the correlations between ratings on nine statements about pain made by 123 people suffering from extreme pain. Each statement was scored on a scale from 1 to 6, ranging from agreement to disagreement. The nine pain statements were as follows:

    1) Whether or not I am in pain in the future depends on the skills of the doctors.
    2) Whenever I am in pain, it is usually because of something I have done or not done,
    3) Whether or not I am in pain depends on what the doctors do for me.
    4) I cannot get any help for my pain unless I go to seek medical advice.
    5) When I am in pain I know that it is because I have not been taking proper exercise or eating the right food.
    6) People's pain results from their own carelessness.
    7) I am directly responsible for my pain,
    8) Relief from pain is chief controlled by the doctors.
    9) People who are never 0 in pain are just plain lucky.

$$\begin{pmatrix}
1.00 \\
-0.04 & 1.00 \\
0.61 & -0.07 & 1.00 \\
0.45 & -0.12 & 0.59 & 1.00 \\
0.03 & 0.49 & 0.03 & -0.08 & 1.00 \\
-0.29 & 0.43 & -0.13 & -0.21 & 0.47 & 1.00 \\
-0.30 & 0.30 & -0.24 & -0.19 & 0.41 & 0.63 & 1.00 \\
0.45 & -0.31 & 0.59 & 0.63 & -0.14 & -0.13 & -0.26 & 1.00 \\
0.30 & -0.17 & 0.32 & 0.37 & -0.24 & -0.15 & -0.29 & 0.40 & 1.00
\end{pmatrix}$$

(a) (5 pts) Apply maximum likelihood factor analysis, and use the test described in the chapter to select the necessary number of common factors.

```
> pain.corr <-
matrix( c(1,-.04,.61,.45,.03,-.29,-.3,.45,.3,-.04,1,-.07,-.12,.49,.43,.3,-.31,-.17,.61,-.07,1,.59,.03,-.13,-.24,
.59,.32,.45,-.12,.59,1,-.08,-.21,-.19,.63,.37,.03,.49,.03,-.08,1,.47,.41,-.14,-.24,-.29,.43,-.13,-.21,.47,1,.63,
-.13,-.15,-.3,.3,-.24,-.19,.41,.63,1,-.26,-.29,.45,-.31,.59,.63,-.14,-.13,-.26,1,.4,.3,-.17,.32,.37,-.24,-.15,-.
29,.4,1), nrow=9, ncol=9, byrow=T)

> factanal(covmat=pain.corr, factors=3, rotation='none', n.obs=123)

Call:

factanal(factors = 3, covmat = pain.corr, n.obs = 123, rotation = "none")

Uniquenesses:

[1] 0.404 0.518 0.336 0.455 0.499 0.171 0.496 0.239 0.754

Loadings:

      Factor1 Factor2 Factor3

 [1,]  0.607   0.297   0.374

 [2,] -0.458   0.288   0.435

 [3,]  0.610   0.506   0.189

 [4,]  0.621   0.400

 [5,] -0.408   0.441   0.375
```

```
 [6,] -0.677   0.591  -0.145
 [7,] -0.626   0.331
 [8,]  0.674   0.481  -0.275
 [9,]  0.446   0.169  -0.135
               Factor1 Factor2 Factor3
SS loadings      3.004   1.501   0.624
Proportion Var   0.334   0.167   0.069
Cumulative Var   0.334   0.501   0.570
Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 19.2 on 12 degrees of freedom. The p-value is 0.0838
```

$$K = 3$$

(b) (10 pts) Rotate the factor solution selected using an orthogonal procedure, and interpret the results.

```
Call:
factanal(factors = 3, covmat = pain.corr, n.obs = 123, rotation = "varimax")

Uniquenesses:
[1] 0.404 0.518 0.336 0.455 0.499 0.171 0.496 0.239 0.754

Loadings:
      Factor1 Factor2 Factor3
 [1,]  0.649  -0.372   0.190
 [2,] -0.126   0.194   0.655
 [3,]  0.794  -0.144   0.116
 [4,]  0.725  -0.106
 [5,]          0.292   0.645
 [6,]          0.825   0.377
 [7,] -0.225   0.590   0.325
 [8,]  0.815          -0.304
 [9,]  0.437          -0.221

               Factor1 Factor2 Factor3
SS loadings      2.507   1.331   1.291
Proportion Var   0.279   0.148   0.143
Cumulative Var   0.279   0.426   0.570

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 19.2 on 12 degrees of freedom.
The p-value is 0.0838
```

- **The first factor seems to measure how much of pain relief is controlled by doctors.**
- **The second factor is a kind of "personal blame for pain" factor.**
- **The third factor measures the belief that "pain results from some action of the person".**