Math 585 HW 10 Spring 2024 Due Friday, April 19
Quiz 10 on Wed. April 17, covers HW10. 2 pages, Problems A)-F)

For the following *R* problems perform the perform the *source("J:/mpack.txt")* and *source("J:/mrobdata.txt")* commands as described in homework 3. Also copy and paste commands from (http://parker.ad.siu.edu/Olive/mrsashw.txt) for the relevant problem into *R*.

A), 10.2 SAS Institute (1985, p. 498 - 501) describes a one way MANOVA model. There are two groups for gender: female and male. There were $p = 4$ (skull measurements) variables $X_1 = length$ $X_2 = basilar$, $X_3 = zygomat$ and $X_4 = postorb$. There were $n_1 = 18$ females and $n_2 = 22$ males measured. Suppose $t_0 = 0.9567$ and pvalue $= 0.6566$. Here t_o was Wilk's lambda, but the other three test statistics gave the same pvalue. Do a 4 step one way MANOVA test.

B) Suppose the 15 units are 1 Adatorwovor, 2 Adhikari, 3 Alanzi, 4 Alsibiani, 5 Al-Talib, 6 Fan, 7 Kuo, 8 Lamsal, 9 Liu, 10 Meyer, 11 Peiris, 12 Rathnayake, 13 Rupasinghe, 14 Schroeppel and 15 Watagoda. Use the following output to allocate the 15 units to three groups of 5. Show the three groups.

```
> sample(15)
 [1]  6  3  4  2  1 10  7  5 12 15 13  8 14 11  9
```

C), 11.4 The Buxton data has 5 massive outliers in variables len and buxy = height.

a) The *R* commands for this part do a factor analysis on the Buxton data using the sample covariance matrix. Copy and paste the output into *Word*.

i) Which variables have nonzero loadings for factor 1?

ii) Which variables have nonzero loadings for factor 2?

iii) What is the cumulative variance explained by the two factors?

b) The *R* commands for this part do a factor analysis on the Buxton data using the RMVN dispersion matrix. Copy and paste the output into *Word*.

i) Which variables have nonzero loadings for factor 1?

ii) Which variables have nonzero loadings for factor 2?

iii) What is the cumulative variance explained by the two factors?

D) Refer to the factor analysis handout where the factor analysis is applied using covariance matrix *ability.cov*.

a) Is one factor enough or are two factors needed?

b) For the analysis which used two factors, which factor corresponds to reading and vocabulary?

c) For the analysis which used two factors, which factor had the two smallest loadings on variables reading and vocabulary?

E), 12.1 Consider the Hotelling Lawley test statistic.
Let

$$T(\boldsymbol{W}) = n \ \ [vec(\boldsymbol{L}\hat{\boldsymbol{B}})]^T[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1} \otimes (\boldsymbol{L}\boldsymbol{W}\boldsymbol{L}^T)^{-1}][vec(\boldsymbol{L}\hat{\boldsymbol{B}})].$$

Let

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n} = \hat{\boldsymbol{W}}^{-1}.$$

Show $T(\hat{\boldsymbol{W}}) = [vec(\boldsymbol{L}\hat{\boldsymbol{B}})]^T[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1} \otimes (\boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T)^{-1}][vec(\boldsymbol{L}\hat{\boldsymbol{B}})]$.

F), 12.2 Consider the Hotelling Lawley test statistic. Let $T = [vec(\boldsymbol{L}\hat{\boldsymbol{B}})]^T[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1} \otimes (\boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T)^{-1}][vec(\boldsymbol{L}\hat{\boldsymbol{B}})]$. Let $\boldsymbol{L} = \boldsymbol{L}_j = [0, ..., 0, 1, 0, ..., 0]$ have a 1 in the jth position. Let $\hat{\boldsymbol{b}}_j^T = \boldsymbol{L}\hat{\boldsymbol{B}}$ be the jth row of $\hat{\boldsymbol{B}}$. Let $d_j = \boldsymbol{L}_j(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}_j^T = (\boldsymbol{X}^T\boldsymbol{X})^{-1}_{jj}$, the jth diagonal entry of $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. Then $T_j = \frac{1}{d_j}\hat{\boldsymbol{b}}_j^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1}\hat{\boldsymbol{b}}_j$. The Hotelling Lawley statistic $U = tr([(n-p)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}]^{-1}\hat{\boldsymbol{B}}^T\boldsymbol{L}^T[\boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T]^{-1}\boldsymbol{L}\hat{\boldsymbol{B}}])$. Hence if $\boldsymbol{L} = \boldsymbol{L}_j$, then $U_j = \frac{1}{d_j(n-p)} tr(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1}\hat{\boldsymbol{b}}_j\hat{\boldsymbol{b}}_j^T)$.

Using $tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB})$ and $tr(a) = a$ for scalar a, show the $(n-p)U_j = T_j$.

```
#Factor Analysis handout
> out1 <- factanal(x, factors = 3)
Uniquenesses:   v1    v2    v3    v4    v5    v6
               0.005 0.101 0.005 0.224 0.084 0.005
Loadings:
   Factor1 Factor2 Factor3
v1 0.944   0.182   0.267
v2 0.905   0.235   0.159
v3 0.236   0.210   0.946
v4 0.180   0.242   0.828
v5 0.242   0.881   0.286
v6 0.193   0.959   0.196
               Factor1 Factor2 Factor3
SS loadings      1.893   1.886   1.797
Proportion Var   0.316   0.314   0.300
Cumulative Var   0.316   0.630   0.929
#used varimax rotation
#factor 1 is almost an average of v1 and v2
#factor 2 is almost an average of v5 and v6
#factor 3 is almost an average of v3 and v4

> out2 <- factanal(x, factors = 3, rotation = "promax")
> out2
Uniquenesses:   v1    v2    v3    v4    v5    v6
               0.005 0.101 0.005 0.224 0.084 0.005
Loadings:   Factor1   Factor2   Factor3
            v1          0.985
            v2          0.951
            v3                   1.003
            v4                   0.867
            v5  0.910
            v6  1.033
               Factor1 Factor2 Factor3
SS loadings      1.903   1.876   1.772
Proportion Var   0.317   0.313   0.295
Cumulative Var   0.317   0.630   0.925
##promax rotation tries to give 0 loadings to lots of variables
##in the factor

##Factor analysis can also be performed by supplying
##a covariance matrix or a correlation matrix.
##As a diagnostic, supply the RMVN dispersion matrix or
##the RMVN generalized correlation matrix.

> out1 <- factanal(factors = 1, covmat=ability.cov)
```

```
> out1
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
>
> out2 <- factanal(factors = 2, covmat=ability.cov)
> out2

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
##Want pvalue > 0.05 to suggest that there are enough factors.
```