```
Math 585    HW 5 Spring 2024        Due Friday, March 1
Quiz 5 on Wed. Feb. 28, covers HW4 and 5, DD plots,
FCH, RFCH, RMVN, DGK, MB estimators, prediction regions, PCA, CCA.
2 pages, problems A)-F)
```

A), 6.3 Let $Y_j = \boldsymbol{e}_j^T \boldsymbol{x}$ be the first population principal component where $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$.

a) Using $\text{Cov}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{B}\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{B}^T$, show $\text{Cov}(\boldsymbol{x}, Y_j) = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{e}_j = \lambda_j \boldsymbol{e}_j$.

b) Now $V(Y_j) = \text{Cov}(\boldsymbol{e}_j^T \boldsymbol{x}, \boldsymbol{e}_j^T \boldsymbol{x})$. Show that $V(Y_j) = \lambda_j$.

c) Let $\boldsymbol{x} = (X_1, ..., X_p)^T$ where X_i is the ith random variable with $V(X_i) = \sigma_{ii}$ and by a) $\text{Cov}(X_i, Y_j) = \lambda_j e_{ij}$ where $\boldsymbol{e}_j = (e_{1j}, ..., e_{ij}, ..., e_{pj})^T$. Find $\text{corr}(X_i, Y_j)$.

For the following two problems perform the perform the *source("J:/mpack.txt")* command as described in homework 3.

B), 5.9 a) Type the *R* command `predsim()` and paste the output into *Word*.

This computes $\boldsymbol{x}_i \sim N_4(\boldsymbol{0}, diag(1, 2, 3, 4))$ for $i = 1, ..., 100$ and $\boldsymbol{x}_f = \boldsymbol{x}_{101}$. One hundred such data sets are made, and ncvr, scvr, mcvr counts the number of times $\boldsymbol{x}_f$ was in the nonparametric, semiparametric and parametric MVN 90% prediction regions. The volumes of the prediction regions are computed and voln, vols and volm are the average ratio of the volume of the ith prediction region over that of the semiparametric region. Hence vols is always equal to 1. For multivariate normal data, these ratios should converge to 1 as $n \to \infty$. Were the three coverages near 90%?

C), 6.9 a) Type the *R* command `pcasim()` and paste the output into *Word*.

This command computes the first 3 eigenvalues and eigenvectors for the classical and robust PCA using the $\boldsymbol{R}$ and $\boldsymbol{R}_U$. The multivariate normal data is such that the cases cluster tightly about the eigenvector $c(1, 1, ..., 1)^T$ corresponding to the largest eigenvalue. The term mncor gives the mean correlation between the classical and robust eigenvalues while the terms vexpl and rvexpl give the average variance explained by the largest 3 eigenvalues. The terms abscoreigvi give the absolute correlation between the i classical and robust eigenvector for $i = 1, ..., 3$ while the term abscorpc gives the absolute correlations of the first 3 principal components.

b) Are the robust and classical eigenvalues highly correlated? Is the absolute correlation for first classical principal component and the robust principal component high?

D), 6.4 The classical PCA output below is for the Buxton data described in homework 4 H) where 5 cases have massive outliers in the height and length variables. Interpret PC1 and PC2. (Use a simple linear combination such as in homework 4 Hb) or Exam 2 review 57).)

```
prcomp(z,scale=T)
[1] 1.431 1.074 0.964 0.926 0.106
        PC1    PC2    PC3    PC4    PC5
len   0.685  0.037  0.004 -0.189 -0.702
nas  -0.199  0.568  0.153 -0.783  0.047
big  -0.049 -0.569  0.783 -0.247 -0.007
ceph -0.100 -0.594 -0.603 -0.523  0.008
ht   -0.692 -0.000 -0.008  0.131 -0.710
```

E), 7.1 The *R* output below corresponds to Example 7.1. a) What is the first canonical correlation $\hat{\rho}_1$?

b) What is $\hat{\boldsymbol{a}}_1$?

c) What is $\hat{\boldsymbol{b}}_1$?

```
rcancor(x,y)
$out$cor
[1] 0.98596703 0.06797587

$out$xcoef
        [,1]       [,2]
S 0.14966183  0.6460117
M 0.03236328 -0.8543387

$out$ycoef
       [,1]       [,2]       [,3]
L 0.1625452  0.4237524 -2.8492678
W 0.2369692  1.5379681  0.9356495
H 0.2530324 -2.6806462  1.7785931
```

F) The SAS output below corresponds to the SAS handout that will be given in lab. Get the output using SAS if SAS works next week. What is the variance explained by the first principal component?

```
                  Eigenvalues of the Covariance Matrix
       Eigenvalue    Difference    Proportion    Cumulative
  1    154.310607    145.147647        0.9439        0.9439
  2      9.162960                      0.0561        1.0000
                       Eigenvectors
                             Prin1         Prin2
           July           0.343532      0.939141
           January        0.939141      -.343532
```