21 parts $\approx$ 15 pts each

Math 583 Final, Fall 2012 Name___________

```
y<-USJudgeRatings[,c(9,10,12)]
x<-USJudgeRatings[,-c(9,10,12)]
mltreg(x,y,indices=c(2,5,6,7,8))
$partial
     partialF      Pval
[1,] 1.649415 0.1855314

$MANOVA
      MANOVAF         pval
[1,] 340.1018 1.121325e-14
```

1) The above output is for the R judge ratings data set consisting of lawyer ratings for $n = 43$ judges. $Y_1 = oral$ = sound oral rulings, $Y_2 = writ$ = sound written rulings, and $Y_3 = rten$ = worthy of retention. The predictors were $x_2 = cont$ = number of contacts of lawyer with judge, $x_3 = intg$ = judicial integrity, $x_4 = dmnr$ = demeanor, $x_5 = dilg$ = diligence, $x_6 = cfmg$ = case flow managing, $x_7 = deci$ = prompt decisions, $x_8 = prep$ = preparation for trial, $x_9 = fami$ = familiarity with law, and $x_{10} = phys$ = physical ability.

a) Do the MANOVA F test.

Ho the nontrivial predictors are not needed in the mreg model
H_1 at least one nontrivial predictor is needed

$F_0 = 340.1018$
pval = 0
reject Ho, at least one nontrivial predictor is needed
(or there is an mreg relationship between Y_1, Y_2, Y_3 and the predictors $x_2 - x_{10}$)

b) Do the MANOVA partial F test for the reduced model that deletes x_2, x_5, x_6, x_7 and x_8.

Ho the reduced model is good H_1 use the full model

$F_R = 1.649$

pval = .1855

fail to reject Ho, the reduced model is good

-30

2) The Johnson and Wichern (1988, p. 262) turtle data has $X_1 = length$, $X_2 = width$ and $X_3 = height$ for painted turtle shells. There is a sample of $n_1 = 24$ female and a sample of $n_2 = 24$ male turtles. Suppose $t_0 = 23.078$ and pvalue = 0.000. Do a 4 step one way MANOVA test.

Ho $\mu_1 = \mu_2$ H_1 not Ho ($\mu_1 \neq \mu_2$ since there are 2 groups)

$t_0 = 23.078$

pval = 0

reject Ho, the mean shell measurements are not the same (so depend on gender)

15

```
Rotation:  PC1         PC2         PC3
length 0.5771831 -0.5884323 -0.5662218
width  0.5811769 -0.1910978  0.7910215
height 0.5736663  0.7856393 -0.2316848

> summary(out$out)
Importance of components:PC1     PC2     PC3
Standard deviation      1.7065 0.25601 0.14961
Proportion of Variance 0.9707 0.02185 0.00746
Cumulative Proportion  0.9707 0.99254 1.00000
```

3) Now consider the turtle data of problem 2 with 48 cases (so gender is ignored). Principal component analysis output is shown above based on the (robust) correlation matrix.

a) How many principal components are needed?

1

b) Interpret the first principal component.

an average of length, width and height

30

4) Suppose that

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim N_4 \left(\begin{pmatrix} 3 \\ 4 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 & 2 & 1 & 1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 0 & 0 & 3 \end{pmatrix} \right).$$

a) Find the distribution of $(X_1, X_3)^T$.

$$N_2\left(\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \right)$$

b) Which pairs of random variables X_i and X_j are independent?

$X_2 \perp\!\!\!\perp X_4$, $X_3 \perp\!\!\!\perp X_4$

c) Find the correlation $\rho(X_1, X_3)$. $\dfrac{\sigma_{13}}{\sqrt{\sigma_{11}\sigma_{33}}} = \dfrac{1}{\sqrt{2}\sqrt{3}} = \dfrac{1}{\sqrt{6}} = \boxed{0.4082}$

$\overline{4}5$

5) Suppose $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid $p \times 1$ random vectors where $E(\boldsymbol{x}_i) = e^{0.5}\mathbf{1}$ and $\text{Cov}(\boldsymbol{x}_i) = (e^2 - e)\boldsymbol{I}_p$. Find the limiting distribution of $\sqrt{n}(\overline{\boldsymbol{x}} - \mathbf{c})$ for appropriate vector $\mathbf{c}$.

$$\sqrt{n}\,(\underline{\bar{X}} - \underbrace{e^{0.5}\,\underset{\sim}{1}}_{\underset{\sim}{c}}) \xrightarrow{D} N_p(\underset{\sim}{0}, (e^2 - e)\,I_p)$$

$\overline{1}5$

6) Let $\boldsymbol{\beta}_i$ be $p \times 1$ and suppose

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \sigma_{11}(\boldsymbol{X}^T\boldsymbol{X})^{-1} & \sigma_{12}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\ \sigma_{21}(\boldsymbol{X}^T\boldsymbol{X})^{-1} & \sigma_{22}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \end{bmatrix} \right).$$

Find the distribution of

$$[\boldsymbol{L} \;\; \mathbf{0}] \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{L}\hat{\boldsymbol{\beta}}_1$$

where $\boldsymbol{L}\boldsymbol{\beta}_1 = \mathbf{0}$ and $\boldsymbol{L}$ is $r \times p$ with $r \leq p$. Simplify.

$$\hat{\beta}_1 - \beta_1 \sim N_p[\underset{\sim}{0}, \sigma_{11} (X^TX)^{-1}] \quad \text{so}$$

$$L\hat{\beta}_1 \underset{\substack{\uparrow \\ L\beta_1 = \underset{\sim}{0}}}{=} L(\hat{\beta}_1 - \beta_1) \sim N_r[\underset{\sim}{0}, \sigma_{11} \, L (X^TX)^{-1} L^T]$$

$\overline{15}$

7) The table $\boldsymbol{W}$ shown below represents 4 measurements on 5 trees. Find a) the sample mean $\overline{\boldsymbol{x}}$ and b) the coordinatewise median MED$(\boldsymbol{W})$.

north	east	south	west
72	66	76	77
60	53	66	63
56	57	64	58
41	29	36	38
32	32	35	36
261	237	277	272

sum

ordered data

32	41	56	60	72
29	32	53	57	66
35	36	64	66	76
36	38	58	63	77

$\underbrace{}_{\text{MED}(W)}$

$$\frac{1}{5}\begin{pmatrix} 261 \\ 237 \\ 277 \\ 272 \end{pmatrix} = \boxed{\begin{pmatrix} 52.5 \\ 47.4 \\ 55.4 \\ 54.4 \end{pmatrix} = \overline{X}}$$

$\overline{30}$

8) The output below is for a canonical correlations anlysis on the R Seatbelts data set where $y_1 = drivers$ = number of drivers killed or seriously injured, $y_2 = front$ = number of front seat passengers killed or seriously injured, and $y_3 = \cancel{back}$ rear = number of back seat passengers killed or seriously injured, $x_1 = kms$ = distance driven, $x_2 = \cancel{Price}$ Petrol = petrol price and $x_3 = Van$ killed = number of van drivers killed. The data consists of 192 monthly totals in Great Britain from January 1969 to December 1984.

a) What is the first canonical correlation $\hat{\rho}_1$?

0.8117

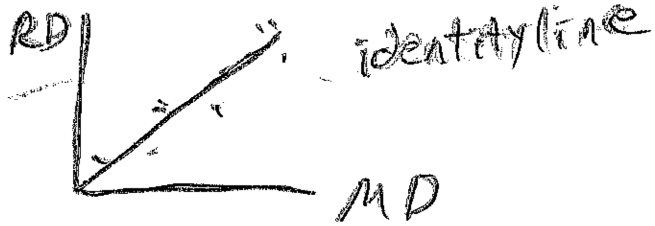
b) What is $\hat{\boldsymbol{a}}_1$?

$$\begin{pmatrix} -2.0802\ (10^{-5}) \\ -1.8480 \\ 1.5976\ (10^{-3}) \end{pmatrix}$$

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

$$\begin{pmatrix} 1.6788\ (10^{-6}) \\ 5.5947\ (10^{-4}) \\ -9.9650\ (10^{-4}) \end{pmatrix}$$

d) Let $\boldsymbol{z} = (\boldsymbol{x}^T, \boldsymbol{y}^T)^T$. The from the DD plot, the $\boldsymbol{z}_i$ appeared to follow a multivariate normal distribution. Sketch the DD plot.




```
> rcancor(x,y)
$out
$out$cor
[1] 0.8116953 0.5064619 0.1376399

$out$xcoef
                       [,1]          [,2]          [,3]
x.kms         -2.080206e-05 -0.0000233873 -2.259723e-06
x.PetrolPrice -1.847967e+00  3.7173715818  5.292041e+00
x.VanKilled    1.597620e-03 -0.0168450843  1.673662e-02

$out$ycoef
                   [,1]          [,2]          [,3]
y.drivers  1.678751e-06 -2.487259e-05  0.0004717902
y.front    5.594715e-04 -7.797027e-05 -0.0008157585
y.rear    -9.964980e-04 -7.521578e-04  0.0005045756
```

45

```
factanal(marry,factors=2,rotation="promax")

Uniquenesses:  pop    mmen    mwmn    mmilmen  milwmn
              0.010   0.005   0.005   0.005    0.005

Loadings:Factor1 Factor2
pop      0.986
mmen     1.003
mwmn     1.003
mmilmen          0.965
milwmn           0.958

               Factor1 Factor2
SS loadings      2.995   1.850
Proportion Var   0.599   0.370
Cumulative Var   0.599   0.969
```

9) The above output is for a factor analysis of the Hebbler (1847) data from the the 1843 Prussia census. Sometimes if the wife or husband was not at the household, then s/he would not be counted. $X_1 = pop$ = population of the district in 1843, $X_2 = mmen$ = number of married civilian men in the district, $X_3 = mwmn$ = number of women married to civilians in the district, $X_4 = mmilmen$ = number of married military men in the district, and $x_5 = milwmn$ = number of women married to military men in the district.

a) What is the cumulative variance explained by the 2 factors?

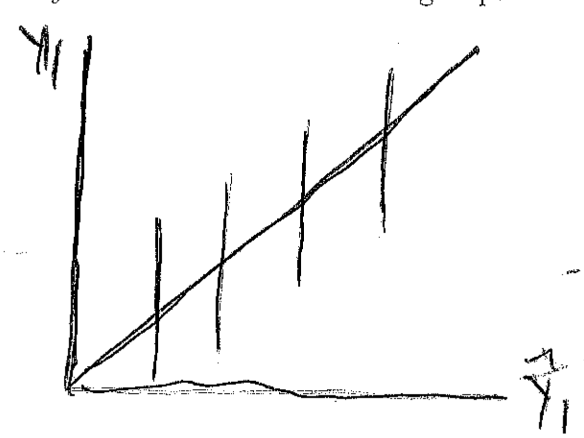
0.969

b) Explain Factor 1. ave of pop, mmen, mwmen

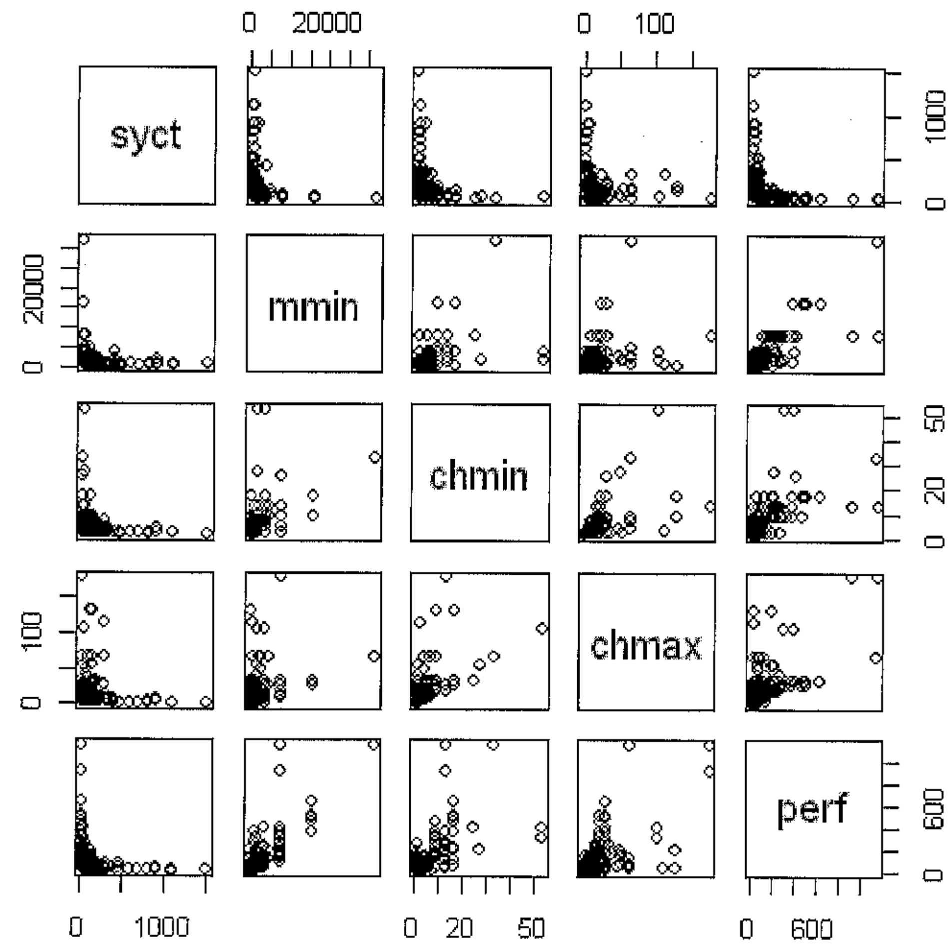
c) Explain Factor 2. ave of mmilmen, milwn

45

10) Sketch a response plot for Y_1 for the one way MANOVA model with 4 groups.



15



eldz4

The plot above is for the Venables and Ripley (2003) CPU data where syct = cycle time + 1,

mmin = minimum main memory + 1, chmin = minimum number of channels + 1,

chmax = maximum number of channels + 1, and perf = published performance + 1. Give transformations,

if any, that you would use to remove strong nonlinearities from the variables.

log (syct)
log (mmin)
log(chmin)
log(chmax)
log (perf)

$\frac{max}{min} > 10$

15

7