Stat 577 Homework 1

Name: [redacted] Total points: 60

Due Saturday, Sept 25 in Canvas using 'SubmitHW1' in Homework/Assignment module

Note: R code should be provided as an Appendix at the very end of your completed homework.

In this homework, you will be using Baumann data from carData package. First install the package and load it using

```
install.packages("carData")
library(carData)
```

Type Baumann in the Help window and read detail information about the data. To see all the variables in the data and to attach Baumann data, type

```
names(Baumann)
attach(Baumann)
```

Now create a new data considering all rows and columns 2 through 6 from Baumann data and let's call this new data newBaumann.

Question 1

For newBaumann data,

a. [5 points] Find the mean of each variable.

The mean of each variable is as follows:

pretest.1	pretest.2	post.test.1	post.test.2	post.test.3
9.787879	5.106061	8.075758	6.712121	44.015152

b. [5 points] Find the covarinace matrix. Round your answers to 3 decimal places. What is the dimension of this matrix?

The covariance matrix is as follows; the dimension of the covariance matrix is 5 x 5.

	pretest.1	pretest.2	post.test.1	post.test.2	post.test.3
pretest.1	9.124	2.238	5.801	0.707	-0.751
pretest.2	2.238	4.896	2.592	1.200	2.660
post.test.1	5.801	2.592	11.517	0.576	10.599
post.test.2	0.707	1.200	0.576	6.947	-0.734
post.test.3	-0.751	2.660	10.599	-0.734	44.138

c. [2 points] Report the covariance value between pretest.1 and pretest.2 variables.

The covariance value between pretest.1 and pretest.2 variables is 2.238.

d. [5 points] Find the correlation matrix. Round your answers to 3 decimal places. What is the dimension of this matrix?

The correlation matrix is as follows; the dimension of the correlation matrix is 5 x 5.

	pretest.1	pretest.2	post.test.1	post.test.2	post.test.3
pretest.1	1.000	0.335	0.566	0.089	-0.037
pretest.2	0.335	1.000	0.345	0.206	0.181
post.test.1	0.566	0.345	1.000	0.064	0.470
post.test.2	0.089	0.206	0.064	1.000	-0.042
post.test.3	-0.037	0.181	0.470	-0.042	1.000

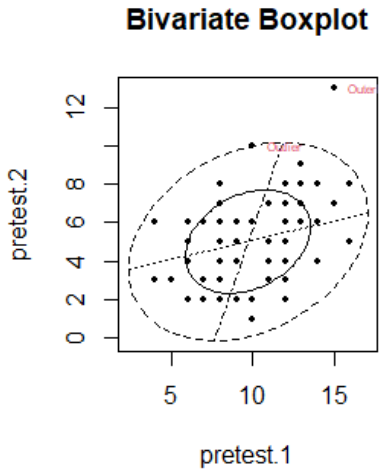
e. [2 points] Report the correlation value between post.test.1 and post.test.2 variables.

The correlation value between post.test.1 and post.test.2 variables is 0.064.

Question 2

[10 points] Draw a bivariate boxplot for pretest.1 and pretest.2 variables. In your graph, pretest.1 and pretest.2 variables will go along x and y axis, respectively. Find the index of any outlier(s) that is outside the fence (the outer ellipse). Write down the text "Outlier" to the right of the outlier point. Make sure to show x and y axes titles as well as the graph title.

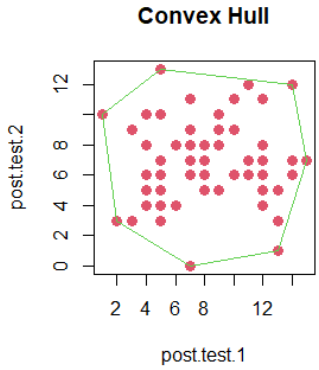
There are two outliers in the bivariate boxplot, and the index of them are 16 & 6.



Question 3

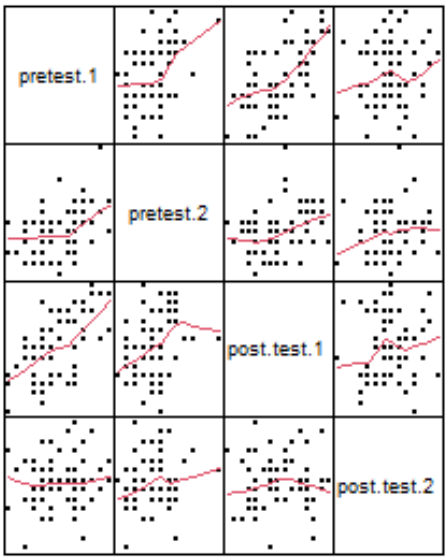
[10 points] Draw a convex hull for post.test.1 and post.test.2 variables. In your graph, post.test.1 and post.test.2 variables will go along x and y axis, respectively. Make sure to show x and y axes titles as well as the graph title. Report the index of extreme points in each direction.

There are 7 extreme points in the convex hull, and the index of them are 58, 34, 20, 57, 60, 52, 63.



Question 4

[6 points] Draw a scatterplot matrix of the following four variables: pretest.1, pretest.2, post.test.1, post.test.2 and provide your graph below.



Question 5

[15 points] Create a 6 by 5 matrix **z** whose elements are random numbers generated from a uniform $U(1,2)$ distribution. Use set.seed(1233):

```
set.seed(1233)
z=matrix(runif(30,1,2), nrow=6, ncol=5, byrow=F)
z
```

a. Now form **z'z**. Is **z'z** symmetric?

The **z'z** matrix is symmetric, as follows:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	13.51101	12.15237	13.34348	13.14311	13.80931
[2,]	12.15237	11.56014	12.37574	11.64739	12.58879
[3,]	13.34348	12.37574	14.39255	13.28758	14.12087
[4,]	13.14311	11.64739	13.28758	13.18173	13.49975
[5,]	13.80931	12.58879	14.12087	13.49975	14.59231

b. Find the eigenvalues of **z'z**.

The **z'z** matrix has 5 eigenvalues, which are 65.56187151, 0.71661607, 0.64177513, 0.26610296, and 0.05137859.

c. Based on the eigenvalues of **z'z**, is your **z'z** non-negative definite or positive definite?

Since all the eigenvalues of **z'z** are greater than 0, **z'z** is positive definite.

d. Find the determinant of **z'z**. Is the determinant of **z'z** equal to the product of the eigenvalues?

The determinant of **z'z** is 0.4122424.

Yes, it is equal to the product of the eigenvalues.

e. Find the trace of **z'z**. Is the trace of **z'z** equal to the sum of the eigenvalues?

The trace of **z'z** is 67.23774.

Yes, it is equal to the sum of the eigenvalues.

f. Find the generalized inverse matrix of **z'z**. You may use ginv() function from MASS package or write the code.

The inverse matrix of **z'z** is as follows:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	10.331510	-4.4898139	3.508333	-6.6113226	-3.1824412
[2,]	-4.489814	3.5069017	-1.983412	3.0188567	0.3499958
[3,]	3.508333	-1.9834117	2.852243	-2.7812946	-1.7960318
[4,]	-6.611323	3.0188567	-2.781295	5.8473834	0.9340608
[5,]	-3.182441	0.3499958	-1.796032	0.9340608	3.6521437

g. Find the square root matrix of **z'z**.

The square root matrix of **z'z** is as follows:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.991644	1.474879	1.385387	1.657999	1.643424
[2,]	1.474879	1.956797	1.435117	1.181682	1.449095
[3,]	1.385387	1.435117	2.298973	1.545300	1.655435
[4,]	1.657999	1.181682	1.545300	2.072399	1.534147
[5,]	1.643424	1.449095	1.655435	1.534147	2.167377

Appendix-- R Code

```
#install.packages("carData")
library(carData)
names(Baumann)
attach(Baumann)
newBaumann = data.frame(pretest.1, pretest.2, post.test.1, post.test.2,
post.test.3)

# Question 1
# a
colMeans(newBaumann)
# b
round(cov(newBaumann), digits = 3)
# d
round(cor(newBaumann), digits=3)

# Question 2
#a
#install.packages("MVA")
library(MVA)
n=nrow(newBaumann)
out=match(sort(pretest.2)[(n-1):n],pretest.2)
out
bvbox(cbind(pretest.1, pretest.2), xlab="pretest.1", ylab="pretest.2", main = "Bivariate Boxplot",
pch = 16, cex = 0.50, col = 1)
text(pretest.1[out], pretest.2[out], labels = c("Outlier", "Outerlier"), cex = 0.50, col=c(2,2), pos = c(4,4))

# Question 3
extr=chull(cbind(post.test.1, post.test.2))
extr
extr <- c(extr, extr[1])
plot(post.test.1, post.test.2, xlab="post.test.1", ylab="post.test.2", main ="Convex Hull",
pch = 16, col = 2, cex = 1.25)
lines(post.test.1[extr], post.test.2[extr], type = "l", col = 3, lwd = 1)

# Question 4
d = data.frame(pretest.1, pretest.2, post.test.1, post.test.2)
pairs(d, panel = panel.smooth, gap = 0, xaxt = "n", yaxt = "n", lwd = 1, pch = 20, cex = 0.5)

# Question 5
set.seed(1233)
z = matrix(runif(30,1,2), nrow=6, ncol=5, byrow=F)
z
# a
A = t(z) %*% z
A
# b
eigen(A)$values
# d
d1 = det(A)
d1
d2 = prod(eigen(A)$values)
```

```
d2
# e
sum(diag(A))
sum(eigen(A)$values)
# f
evl=eigen(A)$values
evc=eigen(A)$vectors
A.inv=evc %*% solve(diag(evl)) %*% t(evc)
# g
A.sq.root=evc%*%diag(sqrt(evl))%*%t(evc)
A.sq.root
```