Spring

Math 585 HW 3 ~~Fall~~ 2024 Due Friday, ~~Sept. 14.~~ Feb. 9

MEET IN LAB, Neckers 258, Monday Feb. 5.

Q3 Wed. Feb. 7 MCLT, sample mean and coordinatewise median, DD Plot, 2 sheets of notes. Exam 1 Wed. Feb. 14 3 sheets of notes.

3.14

A) Suppose $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid $p \times 1$ random vectors from a multivariate t-distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with d degrees of freedom. Then $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{x}) = \frac{d}{d-2}\boldsymbol{\Sigma}$ for $d > 2$. Assuming $d > 2$, find the limiting distribution of $\sqrt{n}(\overline{\boldsymbol{x}} - \boldsymbol{c})$ for appropriate vector $\boldsymbol{c}$.

3.13

B) The table $\boldsymbol{W}$ shown below represents 4 measurements on 5 people.

age	breadth	cephalic	size
39.00	149.5	81.9	3738
35.00	152.5	75.9	4261
35.00	145.5	75.4	3777
19.00	146.0	78.1	3904
0.06	88.5	77.6	933

a) Find the sample mean $\overline{\boldsymbol{x}}$.

b) Find the coordinatewise median MED($\boldsymbol{W}$).

C) Shown below are 4 DD plots. Classify the data distribution as multivariate normal, elliptically contoured but not multivarite normal, not elliptically contoured or outliers are present. Explain your choices briefly.

a)

b)

c)

d)

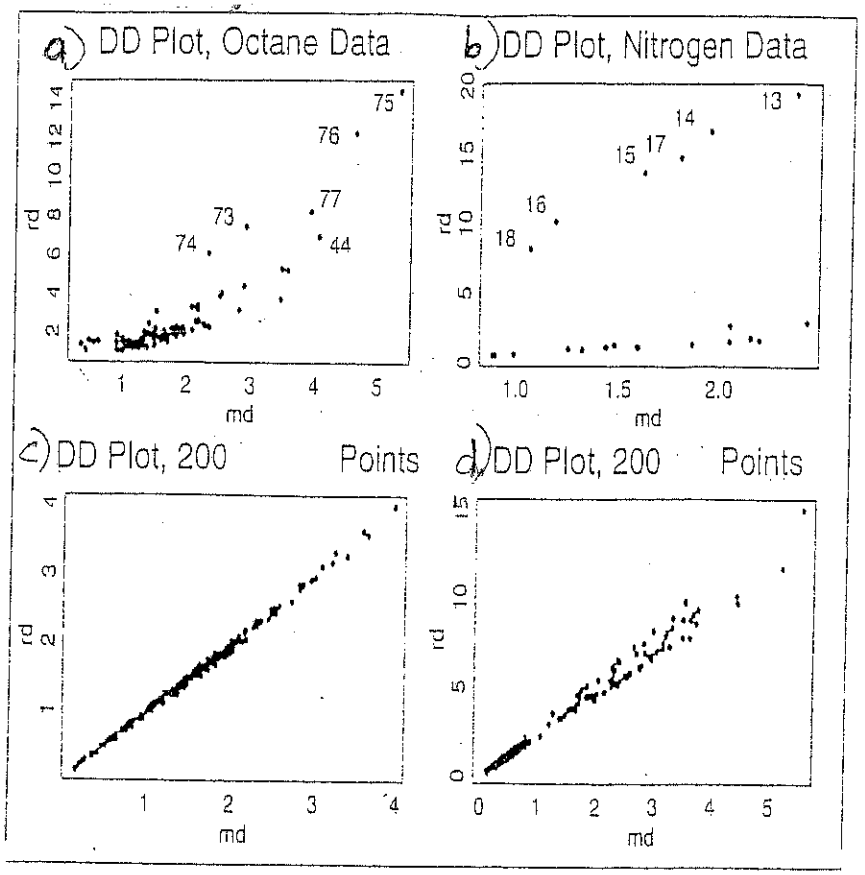

Figure 15.1: 4 DD Plots

The following 3 problems will be done in lab. Go to (http://parker.ad.siu.edu/Olive//multbk.htm). Click on *mpack.txt*. Then go to File>Save Page As and navigate to flash drive J. Similarly, save *mrobdata.txt* in flash drive J. Then enter R and wait for the cursor to appear. Then type the following commands (or copy and paste commands from www.math.siu.edu/olive/mrsashw.txt).

```
source("J:/mpack.txt")
source("J:/mrobdata.txt")
```

Alternatively, go to the *File* menu and drag down *Source R Code*. A window should appear. Navigate the *Look in* box until it says *Removable Disk (J:)*. In the *Files of type* box choose *All files(*.*)* and then select *mpack.txt*. Similarly, download *mrobdata.txt*. The above two lines should appear, following a prompt (>), should appear in the main R window.

You can also copy and paste the two files into R.

Also type the *library(MASS)* command.

D) 4.4c) Type the command *ddcomp(museum[,-1])*. This data is the Schaaffhausen (1878) skull measurements and cases 48–60 were apes while the first 47 cases were humans. This will make 4 DD plots based on the DGK, FCH, FMCD and median ball estimators. The DGK and median ball estimators are the two attractors used by the FCH estimator. With the leftmost mouse button, move the cursor to an outlier and click. After identifying at least 3 outliers in each plot, hold the rightmost mouse button down (and in R click on *Stop*) to advance to the next plot. When done, hold down the *Ctrl* and c keys to make a copy of the plot. Then paste the plot in *Word*.

E) 4.5 (Perform the *source("J:/mpack.txt")* command if you have not already done so.) The *concmv* function illustrates concentration with $p = 2$ and a scatterplot of X_1 versus X_2. The outliers are such that the MBA and FCH estimators can not always detect them. Type the command *concmv()*. Hold the rightmost mouse button down (and in R click on *Stop*) to see the DD plot after one concentration step. The start uses the coordinatewise median and $diag([MAD(X_i)]^2)$. Repeat 4 more times to see the DD plot based on the attractor. The outliers have large values of X_2 and the highlighted cases have the smallest distances. Repeat the command *concmv()* several times. Sometimes the start will contain outliers but the attractor will be clean (none of the highlighted cases will be outliers), but sometimes concentration causes more and more of the highlighted cases to be outliers, so that the attractor is worse than the start. Copy one of the DD plots where none of the outliers are highlighted into *Word*.

F) 4.6 (Perform the *source("J:/mpack.txt")* command if you have not already done so.) The *ddmv* function illustrates concentration with the DD plot. The outliers are highlighted. The first graph is the DD plot after one concentration step. Hold the rightmost mouse button down (and in R click on *Stop*) to see the DD plot after two concentration steps. Repeat 4 more times to see the DD plot based on the attractor. In this problem, try to determine the proportion of outliers *gam* that the DGK estimator can detect for $p = 2, 4, 10$ and 20. Make a table of p and *gam*. For example the command *ddmv(p=2,gam=.4)* suggests that the DGK estimator can tolerate nearly 40% outliers with $p = 2$, but the command *ddmv(p=4,gam=.4)* suggest that *gam* needs to be lowered (perhaps by 0.1 or 0.05). Try to make $0 < gam < 0.5$ as large as possible.