

**Stat 416 Homework 2**  
**Due Date:** 2:10 P.M., Tuesday, February 10

1. Try issuing each of the R commands below in the order that they appear. After doing so and studying the resulting output, explain in words what each command or set of commands does. In some cases you may wish to try some other examples to make sure you understand the command.

```
1:12
```

```
c(9, 5, 3)
```

```
x=c(1:6, 9, 5, 3)
```

```
x
```

```
y=2*x-3
```

```
y
```

```
y[2]
```

```
y[-2]
```

```
y[1:3]
```

```
y[-(1:3)]
```

```
sort(y)
```

```
y
```

```
m=matrix(y, nrow=3)
```

```
m
```

```
m[1, 3]
```

```
m[1, ]
```

```
m[, 1]
```

```
m[1:2, c(1, 3)]
```

```
m[, 1]>0
```

```
m[m[, 1]>0, ]
```

```
m=rbind(m, c(14, 15, 21))
```

```
m
```

```
apply(m, 1, sum)
```

```

apply(m, 2, mean)

plot(m[, 1], m[, 2])

plot(m[, 1], m[, 2], pch=16)

plot(m[, 1], m[, 2], pch=16, col=4)

lines(m[, 1], m[, 2], col=2)

```

2. The purpose of this problem is to provide you with the experience of examining images of cDNA microarray spots. The problem requires an Internet connection (preferably a fast one). Visit the Stanford Microarray Database at <http://genome-www5.stanford.edu/>. Along the left side of the web page under *Public Data*, click *Publications*. You should see a table with headings *Citation, Title, Taxonomy, . . . , Repository*. Under the *Data* heading, click the *SMD* icon for the first publication by Bhamre et al. Near the bottom of the next page under *Options*, click the *data* icon. You should see a table with headings *ExptID, Experiment, . . . , ExptDate*. For the first row of that table corresponding to ExptID 48926, click the icon that shows 9 circles in a 3-by-3 arrangement under the *Options* heading. You should see a slide image.
  - (a) Scroll down and to the right to get a view of the whole slide. Comment on the quality of this microarray slide based on the image.
  - (b) Based on the appearance of the array, how many pins were used to print this slide?
  - (c) Click on the spot in the top row and third column. (Be precise with your positioning of the cursor or the site won't be able to retrieve the data.) You should see a close up view of the spot in which the individual pixels are clearly visible. Tables give information about the spot and the pixel intensities. Use this information to answer the following questions.
    - i. This spot is a probe for what gene? Give the gene name.
    - ii. How many pixels are determined to be in this spot according to the image analysis software? (Don't try to count. Just look for the relevant table entry.)
    - iii. How many pixels were considered background for this spot?
    - iv. The table refers to Channel 1 and Channel 2 as well as R (for red) and G (for green). Can you figure out how to match the channel numbers with red and green based on the numbers in the table?
    - v. Does this spot have a higher red signal intensity or green signal intensity?
    - vi. Give the mean signal intensity for Channel 1.
    - vii. Give the median background intensity for Channel 1.
  - (d) Return to the image of the entire array and click on the fourth spot in the top row. Give the Channel 2 signal mean and background median for this spot.
  - (e) Return to the list of publications. Find a *Genome Biology* paper by Popper et al. (7th from the bottom). Click the *SMD* icon and then the *data* icon for *Kawasaki disease Cohort2 Acute*. Click the icon with nine circles to see the slide image corresponding to the first experiment in the list (ExptID 53533). Comment on the slide quality. Do the same for ExptID 53981.

- We discussed a measure of spot quality called “area to perimeter.” Find the area-to-perimeter value for a spot that forms a perfect half circle.
- Suppose the following are the pixel intensities associated with an Affymetrix probe cell. Determine the single value that would be assigned to this probe cell assuming the layout of the values matches the layout of the pixels in the probe cell. For your convenience, the numbers below have been saved in a text file on the course web page. To read it into R, type the commands:

```
d=read.table("http://www.public.iastate.edu/~dnett/microarray/pixelData.txt")
d=as.matrix(d)
```

(Note that if you copy and paste this code from this pdf file to the R prompt, you may need to retype the quotation marks and/or the tilde (~) because these do not always copy properly. If you get an error message when you run the command in R, that is probably the cause.)

```
30610 31343 26855 34881 26215 29016 35762 37741
31046 32474 27517 29427 24026 33187 30683 36603
31988 36411 29184 34416 32092 24087 32175 33624
29232 19270 33989 35218 30082 35236 28004 38053
23515 32037 17145 26992 29316 28634 25706 26930
26896 24833 29453 37174 28360 30951 31682 35808
25509 32228 26776 30999 27335 23802 26478 40689
29476 34116 29046 35431 31898 29921 28882 33119
```

- Evaluate `TrimMean(34, 25, 22, 31, 47, 98, 26, 18, 19, 24, 34, 25, 27, 33, 31; 0.2, 0.9)`. Make sure that your answer is consistent with the definition of trimmed mean and quantile given in class notes. (Note that the R function `quantile` uses a slightly different definition than the one given in the class notes.)
- Suppose the values below are PM and MM values that have been background adjusted using the MAS 5.0 background adjustment method.

	Probe Pair						
	1	2	3	4	5	6	7
PM	2234	3945	4232	2645	1787	1100	2345
MM	5895	2238	1887	300	1800	154	156

- Determine the ideal mismatch for each probe pair using the algorithm proposed by Affymetrix.
  - Determine the *probe value* for each probe pair using the algorithm proposed by Affymetrix.
- Suppose the matrix below contains background-adjusted (RMA method), base-2-logged, and quantile-normalized PM values from 10 GeneChips for an Affymetrix probe set with 11 probe pairs. There is one row for each GeneChip and one column for each probe.

```
7.58 8.23 8.87 9.80 10.85 10.44 9.23 10.52 9.06 9.92 10.67
9.67 11.52 9.85 11.64 12.46 12.24 11.39 12.20 11.33 12.39 12.70
8.67 10.63 9.20 10.67 11.80 11.37 10.28 11.37 10.12 11.35 11.52
8.55 10.61 9.16 10.80 11.68 11.28 10.43 11.48 10.37 11.64 11.85
8.59 10.72 9.19 11.11 11.93 11.39 10.75 11.73 10.49 11.76 12.05
7.76 9.74 8.64 9.69 10.72 10.17 9.18 10.33 8.90 10.20 10.48
```

```
8.25 10.37 8.94 10.60 11.44 11.12 10.27 11.25 9.92 11.37 11.72
8.02 10.17 8.87 10.06 11.33 10.82 9.67 10.93 9.40 10.70 11.04
8.79 10.98 9.12 10.76 11.88 11.37 10.39 11.30 10.20 11.40 11.60
9.26 11.09 9.52 11.26 11.86 11.59 10.97 11.70 10.68 11.92 12.07
```

For your convenience, the numbers above have been saved in a text file on the course web page. To read it into R, type the commands:

```
d=read.table("http://www.public.iastate.edu/~dnett/microarray/pmval.txt")
d=as.matrix(d)
```

(Note that if you copy and paste this code from this pdf file to the R prompt, you may need to retype the quotation marks and/or the tilde (~) because these do not always copy properly. If you get an error message when you run the command in R, that is probably the cause.)

- (a) Determine the GeneChip-specific RMA expression measures for this probe set.
- (b) Suppose the GeneChips corresponding to rows 1 through 5 correspond to five control experimental units while the GeneChips corresponding to rows 6 through 10 represent five treated experimental units. Conduct a  $t$ -test to determine if there is evidence for differential expression for this probe set. Provide a test statistic, its degrees of freedom, a  $p$ -value, and a brief conclusion.