Math 3070 § 1. **Naphthalene Data: Two-Sample t-Test.** Name: Example
Treibergs December 9, 2010

Data File Used in this Analysis:

```
# M 3070 - 1           Naphthalene Data       Nov. 27, 2010
# Treibergs
#
# From Navidi, Statistics for Engineers and Scientists, 2nd ed
# McGraw Hill, 2008
#
# Taken from T. Ebihara & P. Bishop, "Influence of supplemental acetate on
# bioremediation of dissolved polycyclic aromatic hydrocarbons," J. Environ.
# Engineering, 2002.
#
# Experiment measured naphthalene concentrations in mg/L (PAH) in water measured
# in a sand column at various depths and after as number of days
#
# Assume all measurements are independent.
#
# We test whether there is strong evidence that PAH has decreased more than
# 1.2 mg/L from 45 days to 90 days.
#
Depth Days PAH
0 25 11.15
0 25 11.39
0 25 11.36
0 45 9.28
0 45 8.15
0 45 8.59
0 90 7.68
0 90 7.59
0 90 7.41
5 25 14.4
5 25 11.78
5 25 11.92
5 45 9.44
5 45 9.34
5 45 9.33
5 90 7.53
5 90 7.92
5 90 7.12
15 25 11.51
15 25 11.01
15 25 11.09
15 45 9.34
15 45 9.11
15 45 8.94
15 90 7.43
15 90 7.47
15 90 7.53
30 25 12.77
```

```
30 25 12.18
30 25 11.65
30 45 9.37
30 45 9.27
30 45 9.05
30 90 7.6
30 90 7.48
30 90 7.84
50 25 11.71
50 25 11.29
50 25 11.2
50 45 9.25
50 45 8.97
50 45 9.29
50 90 7.76
50 90 7.84
50 90 7.68
75 25 11.18
75 25 11.45
75 25 11.27
75 45 9.09
75 45 8.86
75 45 8.78
75 90 7.72
75 90 7.61
75 90 7.74
```

R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]

[Workspace restored from /Users/andrejstreibergs/.RData]
```

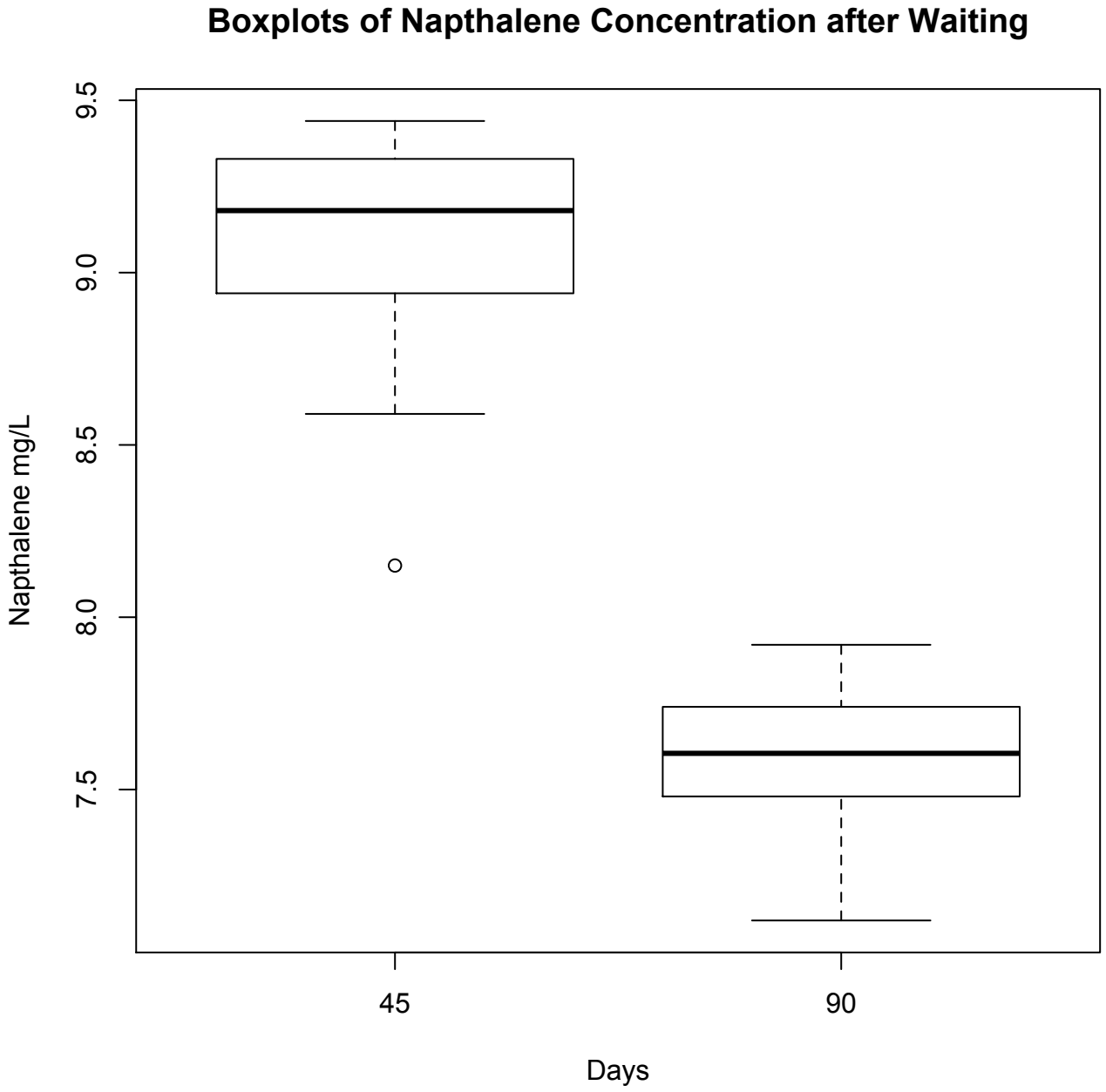
```
> tt <- read.table("M3073NaphthaleneData.txt",header=TRUE)
> tt
   Depth Days   PAH
1      0   25 11.15
2      0   25 11.39
3      0   25 11.36
4      0   45  9.28
5      0   45  8.15
6      0   45  8.59
7      0   90  7.68
8      0   90  7.59
9      0   90  7.41
10     5   25 14.40
11     5   25 11.78
12     5   25 11.92
13     5   45  9.44
14     5   45  9.34
15     5   45  9.33
16     5   90  7.53
17     5   90  7.92
18     5   90  7.12
19    15   25 11.51
20    15   25 11.01
21    15   25 11.09
22    15   45  9.34
23    15   45  9.11
24    15   45  8.94
25    15   90  7.43
26    15   90  7.47
27    15   90  7.53
28    30   25 12.77
29    30   25 12.18
30    30   25 11.65
31    30   45  9.37
32    30   45  9.27
33    30   45  9.05
34    30   90  7.60
35    30   90  7.48
36    30   90  7.84
37    50   25 11.71
38    50   25 11.29
39    50   25 11.20
40    50   45  9.25
41    50   45  8.97
42    50   45  9.29
43    50   90  7.76
44    50   90  7.84
45    50   90  7.68
46    75   25 11.18
47    75   25 11.45
48    75   25 11.27
49    75   45  9.09
```

```
50    75   45  8.86
51    75   45  8.78
52    75   90  7.72
53    75   90  7.61
54    75   90  7.74
```

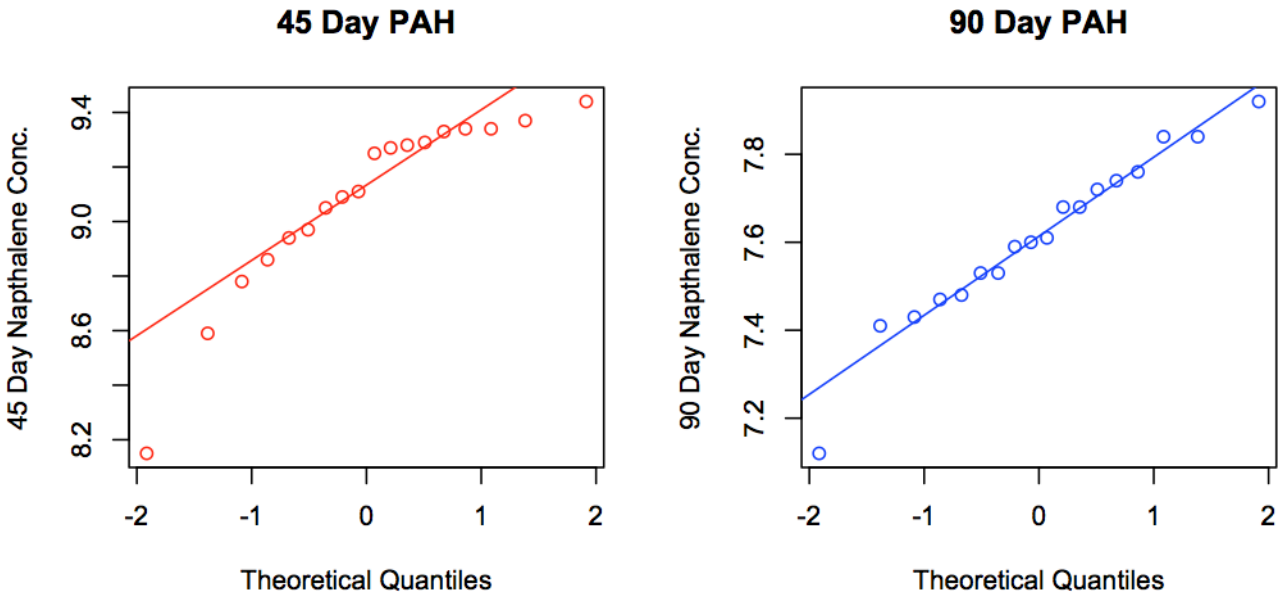
```
> # For this study, we exclude the data for 25 Days. Take  only tt with Days>25
```

```
> tt2 <- subset(tt, Days>25)
> attach(tt2)
> tt2
   Depth Days  PAH
4      0   45 9.28
5      0   45 8.15
6      0   45 8.59
7      0   90 7.68
8      0   90 7.59
9      0   90 7.41
13     5   45 9.44
14     5   45 9.34
15     5   45 9.33
16     5   90 7.53
17     5   90 7.92
18     5   90 7.12
22    15   45 9.34
23    15   45 9.11
24    15   45 8.94
25    15   90 7.43
26    15   90 7.47
27    15   90 7.53
31    30   45 9.37
32    30   45 9.27
33    30   45 9.05
34    30   90 7.60
35    30   90 7.48
36    30   90 7.84
40    50   45 9.25
41    50   45 8.97
42    50   45 9.29
43    50   90 7.76
44    50   90 7.84
45    50   90 7.68
49    75   45 9.09
50    75   45 8.86
51    75   45 8.78
52    75   90 7.72
53    75   90 7.61
54    75   90 7.74
```

```
> # LET  x  BE THE  45  DAY CONCENTRATIONS
> day <- factor(Days)
> x <- subset(PAH,day==45)

> # LET  y  BE THE  90  DAY CONCENTRATIONS
> y <- subset(PAH,day==90)

> # BOXPLOTS OF THE TWO DATA SETS.
> layout(1)
> boxplot(PAH~day,xlab="Days",ylab="Napthalene mg/L",
  main="Boxplots of Napthalene Concentration after Waiting")
```


Boxplots of Napthalene Concentration after Waiting
Napthalene mg/L
9.5
9.0
8.5
8.0
7.5
45
90
Days

```
> # QQ - PLOTS OF THE TWO DATA SETS.  NORMALITY WORSE IN 45 DAY BUT NOT BAD.
> layout(matrix(c(1,3,2,4),ncol=2))

> qqnorm(x,col=2,main="45 Day PAH",ylab="45 Day Napthalene Conc.");qqline(x,col=2)
> qqnorm(y,col=4,main="90 Day PAH",ylab="90 Day Napthalene Conc.");qqline(y,col=4)
```




```
> #   TEST THE HYPOTHESES
> #      H0:    mu_x - mu_y = Delta_0 = 1.2     VS
> #      Ha:    mu_x - mu_y > Delta_0 = 1.2
> #
> ########## DO THE COMPUTATION "BY HAND" ##########################################
> #
> ###### SAMPLE MEAN, SD, N FOR THE TWO SAMPLES ####################################

> m <- length(x); mx <- mean(x); sx <- sd(x); c(m,mx,sx)
[1] 18.0000000  9.0805556  0.3287673

> n <- length(y); my <- mean(y); sy <- sd(y); c(n,my,sy)
[1] 18.0000000  7.6083333  0.1918103

> ########## STANDART ERROR OF XBAR - YBAR, T-STATISTIC ############################

> ss <- sqrt(sx^2/m + sy^2/n); ss
[1] 0.08971532

> t=(mx-my-1.2)/ss; t
[1] 3.034289
```

```
> # THIS IS THE COMPUTATION OF THE APPROXIMATE d.f. GIVEN IN THE TEXT WHEN THE
> # TWO VARIANCES CANNOT BE ASSUMED TO BE EQUAL.
> # IT IS KNOWN AS THE WELCH-SATTERTHWAITHE TEST.

> nu <- (sx^2/m+sy^2/n)^2/((sx^2/m)^2/(m-1)+(sy^2/m)^2/(n-1)); nu
[1] 27.37136

> ###### P-VALUE = AREA OF TAIL TO RIGHT OF  T IN T-DIST WITH nu DEGREES OF FREEDOM ##
>  pt(t,nu, lower.tail=F)
[1] 0.002617539
> #   P-VALUE IS SIGNIFICANT.  WE REJECT  H0  IN FAVOR OF  Ha.


> # ONE-SIDED  95 %  CONFIDENCE BOUND: mux - muy IS LESS THAN
> mx-my+qt(.05,nu,lower.tail=FALSE)*ss
[1] 1.624960
>

> #  RUN THE SAME TEST USING THE CANNED ROUTINE.

> t.test(x,y,alternative="greater",mu=1.2,paired=FALSE,var.equal=FALSE)

Welch Two Sample t-test

data:  x and y
t = 3.0343, df = 27.371, p-value = 0.002618
alternative hypothesis: true difference in means is greater than 1.2
95 percent confidence interval:
 1.319485      Inf
sample estimates:
mean of x mean of y
 9.080556  7.608333


> #  NOTE THAT  t.test  GIVES THE LOWER CONFIDENCE BOUND. HOWEVER, RECALL THAT "UPPER" AND
> #  "LOWER" SWAP WHEN PERFORMING A ONE-SIDED TEST OF HYPOTHESIS:
> #  WE REJECT  H0  AT THE  alpha  LEVEL IF
> #
> #                    mx - my - Delta_0 >= ss * tcrit(alpha,nu)
> #
> #  BUT THIS IS THE SAME AS THE TEST VALUE BEING OUTSIDE THE LOWER  1-alpha  CI FOR
> #  mu_x - mu_y:
> #
> #                    mx - my -  ss * tcrit(alpha,nu) >= Delta_0
> #
> mx-my-qt(.05,nu,lower.tail=FALSE)*ss
[1] 1.319485

> # THUS  Delta_0 = 1.2  IS OUTSIDE THE LOWER CI FOR  mu_x - mu_y  SO WE REJECT  H0.
```