Math 3070 § 1. **Benzene Example:** Name: Example
Treibergs **Bootstrap Power Calculation** July 23, 2011

Today's example was motivated from Example 6.41 of Navidi, *Statistics for Engineers and Scientists, 2nd ed.*, McGraw Hill 2008, that discusses how to find power using bootstrapping. The t-statistic

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \tag{1}$$

where the sample is $X_1, X_2, \ldots, X_n \sim \mathrm{N}(\mu, \sigma)$ taken from a normal population, $\bar{X}$ is the sample mean and S is the sample standard deviation, satisfies the t distribution with $n-1$ degrees of freedom. This fact is used to construct the t-test of hypotheses for the mean. The case that the background distribution is normal but satisfies the alternative hypothesis $H_a : \mu = \mu_1$ where $\mu_1 \neq \mu_0$ is dealt with by looking at beta curves in the Devore text, but is not covered theoretically. It turns out to be that T satisfies the t-distribution with nonzero centrality parameter.

But, bootstrap simulation works here as it does for most statistical analyses. It is conceptually easy: use the data to determine the distribution, and then sample from the distribution to see the sampling distribution of the statistic. Thus, we wish to estimate the probability of type-II error $\beta(\mu_1)$, that is of accepting the null hypothesis given that $\mu = \mu_1$, *i.e.,* that the null hypothesis fails.

The data we use Navidi's problem 6.9.3 is from an article by Navidi, Mukherjee *et al.,* "Reaction Modeling" *Industrial and Engineering Chemistry Research,* 2002. They list $n = 24$ benzene conversions (in mole percent) for different benzenehydrozylation reactions. Although intended for the Wilcoxson Signed-Rank test, the data is plausibly normal (at 5%), and we handle it using parametric bootstrapping. The QQ-plot shows the data is somewhat light-tailed although lines up OK with normal, and the Shapiro-Wilk test does not reject normality.

We compute $\bar{x}$ and s for the data. We test whether the mean conversion is less than 45. The t-test yields a p-value of 0.383 so we are unable to reject the null hypothesis $H_0 : \mu = \mu_0 = 45$. The corresponding 95% one-sided CI is $(-\infty, 51.03756)$. Since 45 is in this interval we fail to reject H_0.

We test the hypothesis using non-parametric bootstrapping. We repeat $B = 100,000$ times taking a random sample of size n from the observed data with replacement. We plot the histogram of the bootstrapped means. The B observed means are sorted and the 5% quantile yields the bootsrapped lower-CI $(-\infty, 50.67937)$. Since 45 is in this interval, we fail to reject the null hypothesis $H_0 : \mu \geq 45$ at 5%.

In order to find the power of the t-test we do parametric bootstrapping as follows. Let $B = 10,000$. For one hundred numbers μ_1 between 45 and 30, we compute the t-statistic (1) for B random samples from $\mathrm{N}(\mu_1, s)$, and then work out the proportion that exceed the critical value $-t_{.05}$ (*i.e.,* the probability that the test accepts H_0 even though $\mu_1 < \mu_0$), where $\mathrm{P}(T < -t_{.05}) = .05$ for a random variable T satisfying the t-distribution with $n-1$ degrees of freedom. The plot of these proportions is the operating characteristic curve showing that the probability of a type II error, β, decreases as μ_1 decreases from 45 to 30.

Data Set Used in this Analysis :

```
# Math 3070 - 1     Benzene Data      July 23, 2011
# Treibergs
#
# From Navidi, "Statistics for Engineers and Scientists, 2nd ed.,"  McGraw
# Hill 2008. From an article by Navidi, Mukherjee et al, "Reaction Modeling..."
```

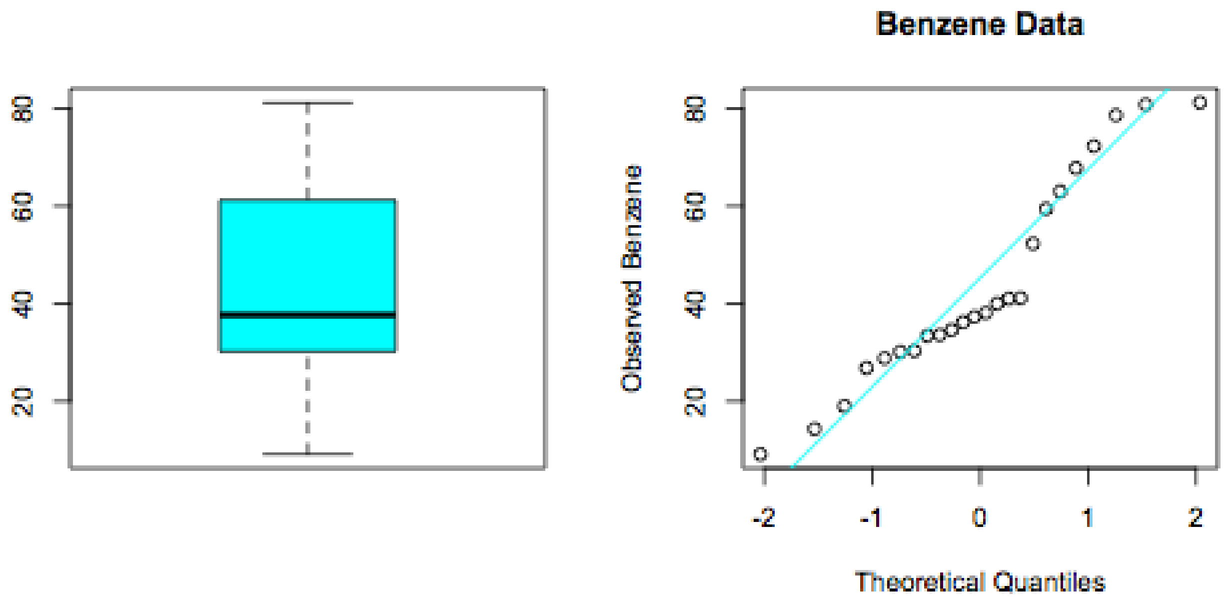
```
# Industrial and Engineering Chemistry Research, 2002.
#
# Benzene conversions (in mole percent) for different benzenehydrozylation
# reactions
#
Conversion
52.3
41.1
28.8
67.8
78.6
72.3
9.1
19
30.3
41
63
80.8
26.8
37.3
38.1
33.6
14.3
30.1
33.4
36.2
34.6
40
81.2
59.4
```

R Session:

```
R version 2.13.1 (2011-07-08)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

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 1.41 (5874) i386-apple-darwin9.8.0]
[History restored from /Users/andrejstreibergs/.Rapp.history]


> tt <- read.table("M3074BenzeneData.txt",header=T)
> tt
   Conversion
1        52.3
2        41.1
3        28.8
4        67.8
5        78.6
6        72.3
7         9.1
8        19.0
9        30.3
10       41.0
11       63.0
12       80.8
13       26.8
14       37.3
15       38.1
16       33.6
17       14.3
18       30.1
19       33.4
20       36.2
21       34.6
22       40.0
23       81.2
24       59.4
> attach (tt)
```

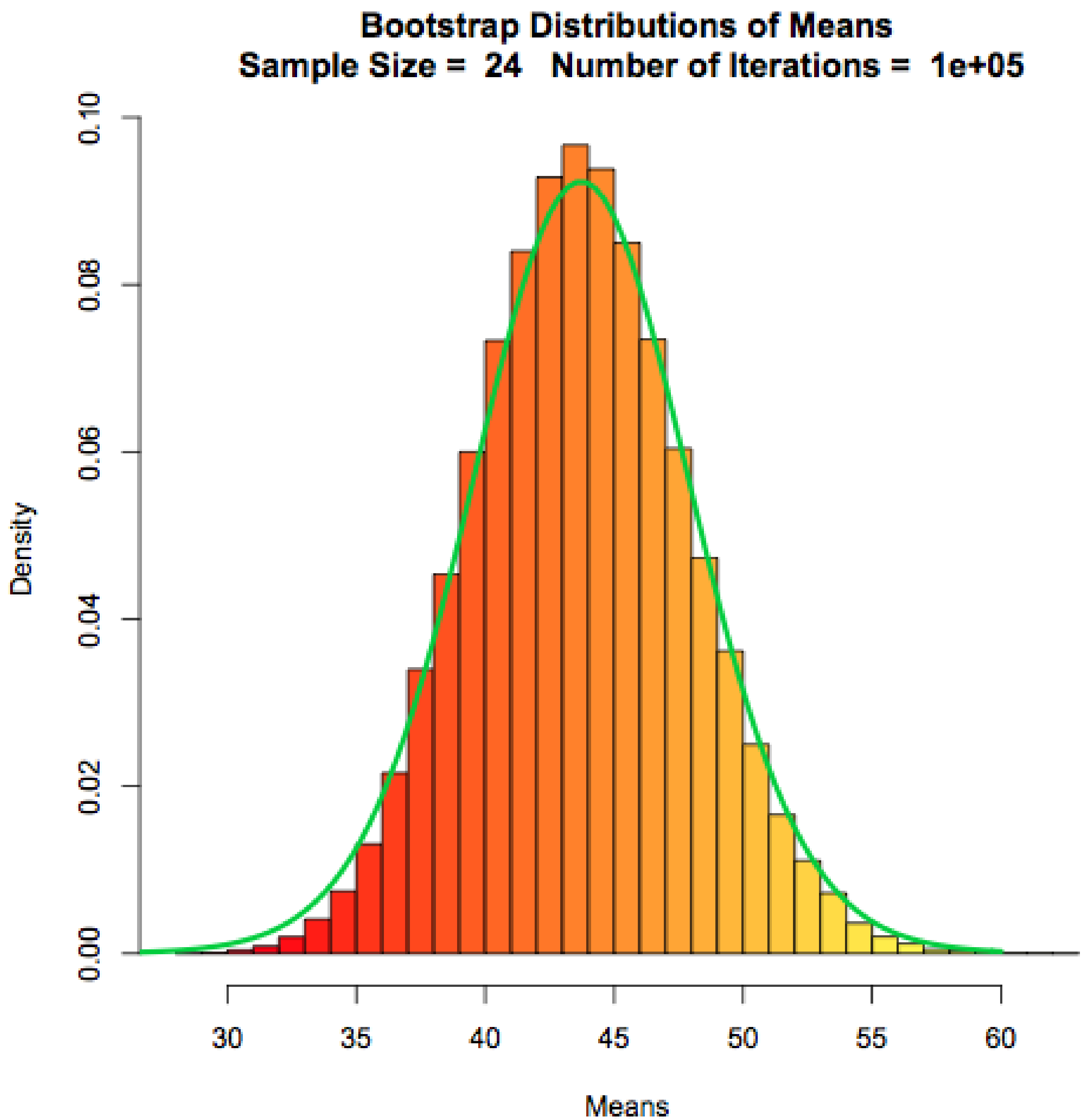
```
> ############### PLOT BOXPLOT & QQ-PLOT. NORMALITY PLAUSIBLE. ################
> layout(matrix(c(1,3,2,4),ncol=2))
> boxplot(Conversion,col=5)
> qqnorm(Conversion,main="Benzene Data",ylab="Observed Benzene")
> qqline(Conversion,col=5)
> # M3074Benzene3.pdf
```




```
> # Hmmm. A bit light tailed, perhaps, but normality not eggregiously violated.
> # The Shapiro-Wilk test does not reject normality at 5%
> shapiro.test(Conversion)

Shapiro-Wilk normality test

data:  Conversion
W = 0.9255, p-value = 0.0774


> ################ T-TEST FOR mu < 45 #########################################
> t.test(Conversion,alternative="less",mu=45)

One Sample t-test

data:  Conversion
t = -0.3012, df = 23, p-value = 0.383
alternative hypothesis: true mean is less than 45
95 percent confidence interval:
     -Inf 51.03756
sample estimates:
mean of x
  43.7125

> # So there in not sufficiently strong evidence to suggest mu < 45.
```

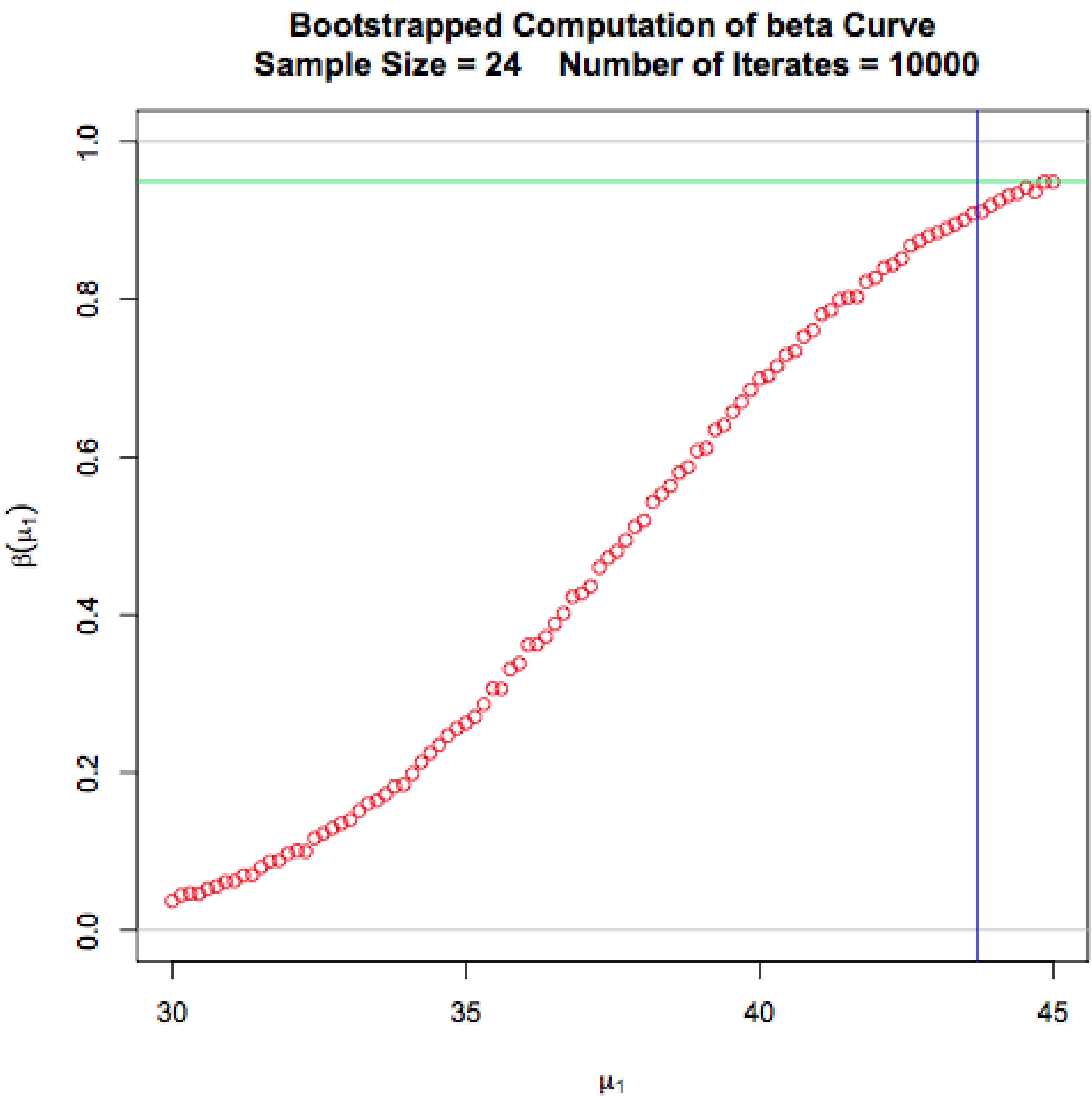
```
> # Navidi's exercise asks some other tests.

> t.test(Conversion,alternative="greater",mu=30)

One Sample t-test

data:  Conversion
t = 3.2084, df = 23, p-value = 0.00195
alternative hypothesis: true mean is greater than 30
95 percent confidence interval:
 36.38744      Inf
sample estimates:
mean of x
  43.7125

> t.test(Conversion,mu=55)

One Sample t-test

data:  Conversion
t = -2.641, df = 23, p-value = 0.0146
alternative hypothesis: true mean is not equal to 55
95 percent confidence interval:
 34.87109 52.55391
sample estimates:
mean of x
  43.7125

> ############# NON-PARAMETRIC BOOTSTRAP TO GET CI ON mu #####################
> B <- 100000
> n <- length(Conversion)
> # Sample from the dataset itself, as a proxy for sampling from the actual
> # distribution.
> y <- replicate(B,mean(sample(Conversion,n,replace=T)))
> # The quantiles give the CI bounds.
> quantile(y,c(.025,.05,.95,.975))
    2.5%       5%      95%    97.5%
35.80833 37.01250 50.67937 52.10417

> # Thus the bootstrapped 95% lower CI for mu is  (-Inf, 50.67937)
```

```
> ############ HISTOGRAM OF NON-PARAMETRIC BOOTSTRAPPED MEANS #################
> hist(y, breaks =4 0, col = heat.colors(40), freq = F, xlab = "Means",
+ main = paste("Bootstrap Distributions of Means\n Sample Size = ", n,
+ "  Number of Iterations = ", B))
> # Superimpose the theoretical distribution of means from normal samples.
> # Fits pretty good!
> curve(sqrt(n)*dt(sqrt(n)*(x-xbar)/s, n-1)/s, 25, 60, add = T, lwd=3, col=3)
> # M3074Benzene1.pdf
```



```
> ######## PARAMETRIC BOOTSTRAPPING TO FIND THE POWER CURVE #################
> # beta curve via simulation
> B <- 10000
> # This is done 100 times for the graph making 1,000,000 samples taken.
> xbar <- mean(Conversion)
> s <- sd(Conversion)
> c(xbar, s)

[1] 43.71250 20.93816


> sn <- sqrt(n)
> t <- function(x,m){(mean(x)-m)*sn/sd(x)}
>
> t(Conversion,45)
[1] -0.3012412

> pt(t(Conversion,45),n-1)
[1] 0.3829686

> alpha <- .05
> ta <- qt(alpha, n-1, lower.tail = F); ta
[1] 1.713872

> xbar
[1] 43.7125
> s
[1] 20.93816
> # CI:
> c(-Inf, xbar+ta*s/sn, Inf)
[1]  -Inf   51.03756
> # So in the t-test, we have the correct  t-statistic and p-value and CI
```

```
> ########### FINDING THE FRACTION TYPE-II ERRORS ############################
> # If true, the value of  replicate(B,t(rnorm(n,xbar,s),45) > -ta  is  1
> # If false it is zero. Thus the sum of  B  such is the number true out of  B.
>
> sum(replicate(B,t(rnorm(n,xbar-.1,s),45)> -ta))/B
[1] 0.9078
> sum(replicate(B,t(rnorm(n,xbar-.2,s),45)> -ta))/B
[1] 0.9034
> sum(replicate(B,t(rnorm(n,xbar-.3,s),45)> -ta))/B
[1] 0.9016

> B
[1] 10000

> # 100 mu1 values
> mu1 <- seq(30,45,15/99)
> length(mu1)
[1] 100
> w <- numeric(100)
> # For each mu1, simulate B times and find the fraction the t-test accepts
> for(i in 1:100)
+                {
+                 w[i] <- sum(replicate(B, t(rnorm(n, mu1[i], s),45) > -ta))/B
+                }
> plot(xx, w, ylim = 0:1, col=2, main =
+ paste("Bootstrapped Computation of beta Curve\n Sample Size =", n,
+ "   Number of Iterates =", B), xlab = expression(mu[1]),
+ ylab = expression(beta(mu[1])))
> abline(h = c(0,1,.95), col = c(8,8,3))
> abline(v = xbar, col = 4)
> # The blue line is the observed value of $\bar X$.
> # The green line is 95%.
> # The the fraction of samples that satisfy  T < -ta
> # is 95% when mu1 = 45 (as expected.) This fraction falls off
> # as mu1 decreses.
```


Bootstrapped Computation of beta Curve
Sample Size = 24 Number of Iterates = 10000
1.0
0.8
0.6
0.4
0.2
0.0
β(μ1)
30
35
40
45
μ1