Math 3070 § 1.	**Simulating the Sampling**	Name: Example
Treibergs	**Distribution of the t-Statistic**	May 20, 2011

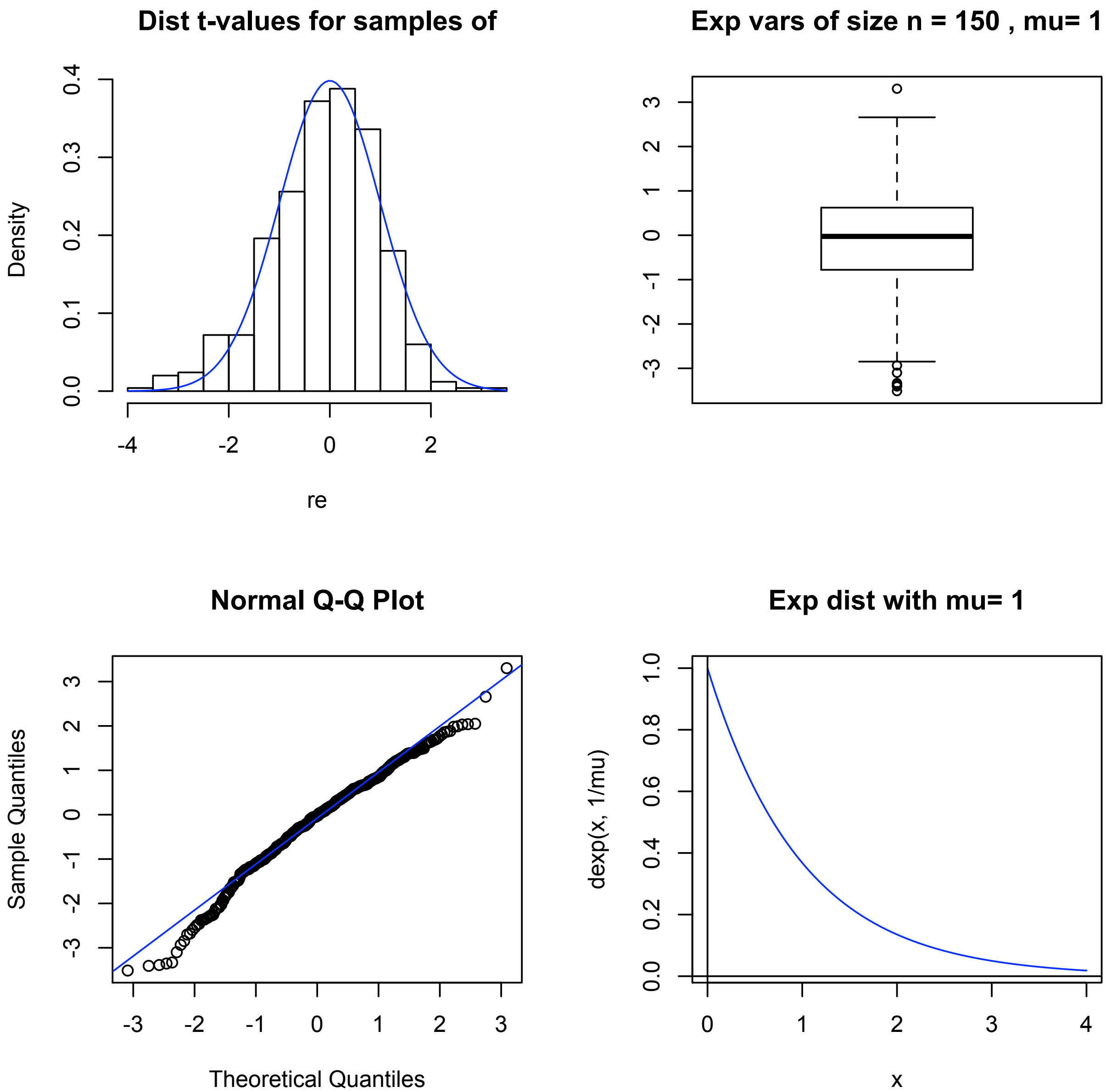
How well does a statistic behave when the assumptions about the underlying population fail to hold? We shall explore the t-statistic, for small and large samples. For small samples, the population is assumed to be approximately normal, so that the statistic itself follows a t-distribution and that hypothesis tests can me made using critical t-scores. We simulate random samples taken from exponential and Cauchy distributions and look at the empirical distributions of the t-statistic.

This study follows the description in Verzani's "SimpleR: Using **R** for Introductory Statistics," from the Appendix "A sample **R** session: a simulation example."

R Session:

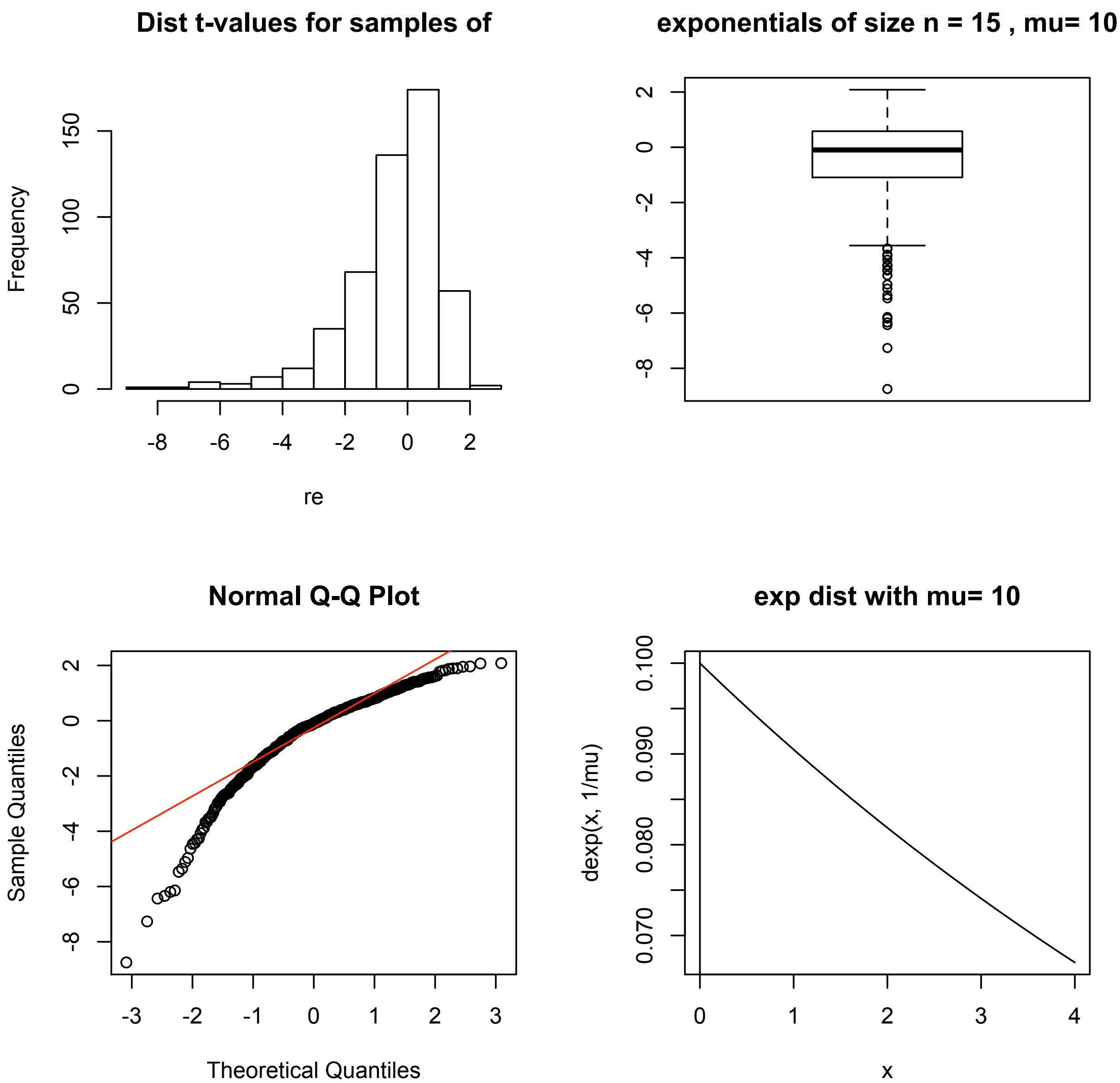
```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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]

> ####################################################################
> ### SIMULATION TO SEE HOW ROBUST THE t-STATISTIC IS TO CHANGES OF ###
> ### DISTRIBUTION                                                  ###
> ####################################################################
>
> # This study follows the description in Verzani's "SimpleR: Using R for
> # Introductory Statistics", from the Appendix "A sample R session: a
> # simulation example."
>
> # We are interested in how much the distribution of the t-statistic
> # is influenced by the shape of the population distribution.
> # For various background distributions of mean mu, we take 500
> # random samples of size  n. For each we compute the t-statistic and
> # then analyze the distribution by plotting its histogram, boxplot
> # QQ-normal plot.
```

```
> ######################## FUNCTION GIVES t-STATISTIC ################
> make.t <- function(x,mu) (mean(x)-mu)/(sqrt(var(x)/length(x)))
> mu <- 2; x <- rnorm(100,mu,1)
> make.t(x,mu)
[1] 1.365704
> mu <- 16; x <- rexp(100,1/mu);make.t(x,mu)
[1] 0.5150337
> mu <- 16; x <- rexp(100,1/mu);make.t(x,mu)
[1] -0.1985672
> mu <- 16; x <- rexp(100,1/mu);make.t(x,mu)
[1] 0.8135002
>
> ################# LOOP TO TAKE 500 SAMPLES OF SIZE n ################
>
> # Set up four plots per page. Save old graphics parameters. Restore at end.
> opar <- par(mfrow=c(2,2))
>
> n<- 150;mu<-1
> for(i in 1:500)re[i]<- make.t(rexp(n,1/mu),mu)
> hist(re,main="Dist t-values for samples of",freq=F)
> curve(dt(x,n-1),add=T,col=4)
> boxplot(re,main=paste("Exp vars of size n =",n,", mu=",mu))
> qqnorm(re);qqline(re,col=4)
> curve(dexp(x,1/mu),xlim=c(0,4),main=paste("Exp dist with mu=",mu),col=4)
> abline(h=0);abline(v=0)
```

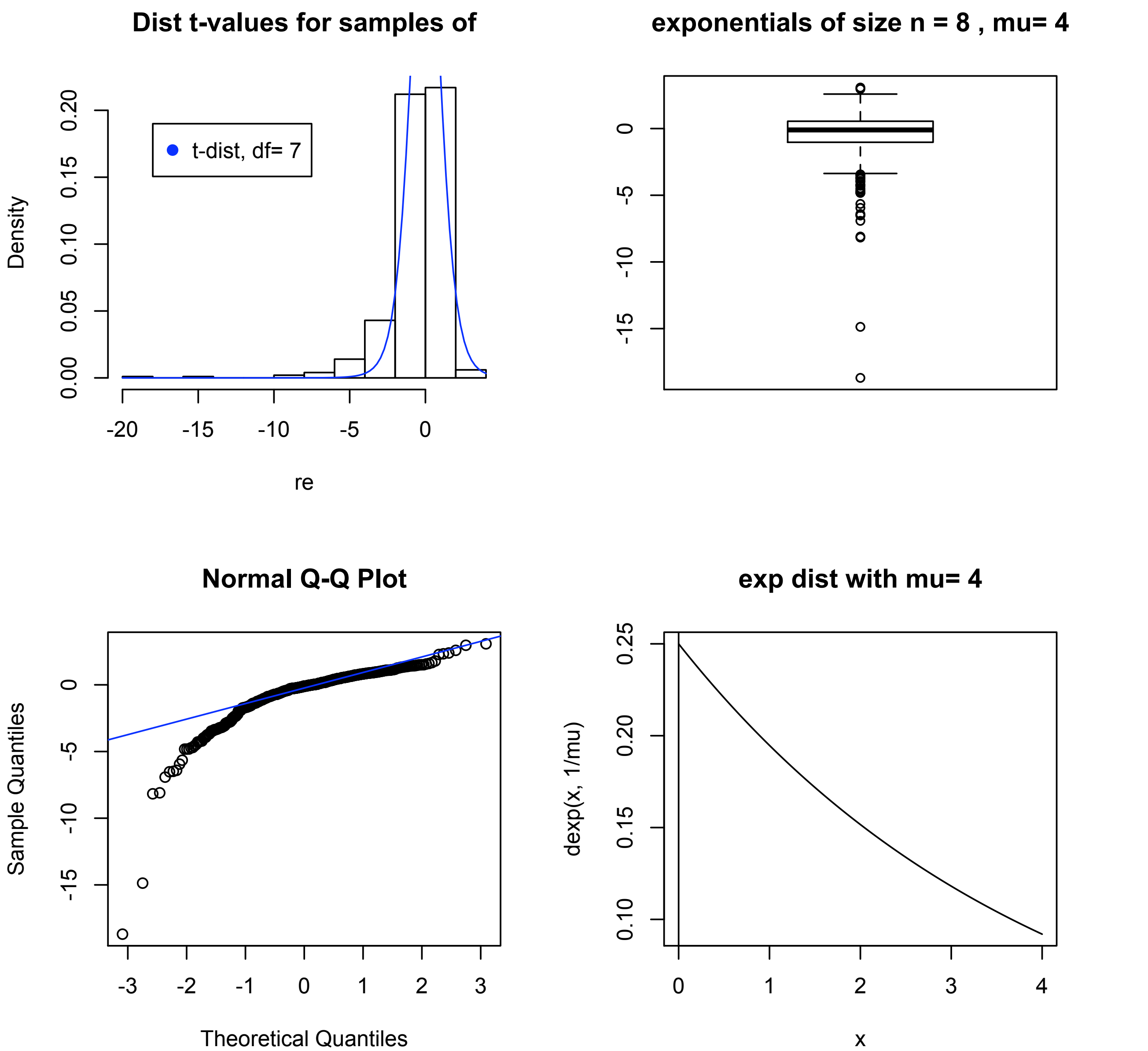



```
> # With a large sample selected from exponential distribution, the
> # t-statistic looke pretty nortmal: it is symmetric, and the QQ plot
> # is pretty much normal. We have added the standard t-dist with n-1 df
> # superimposed across the histogram.
```

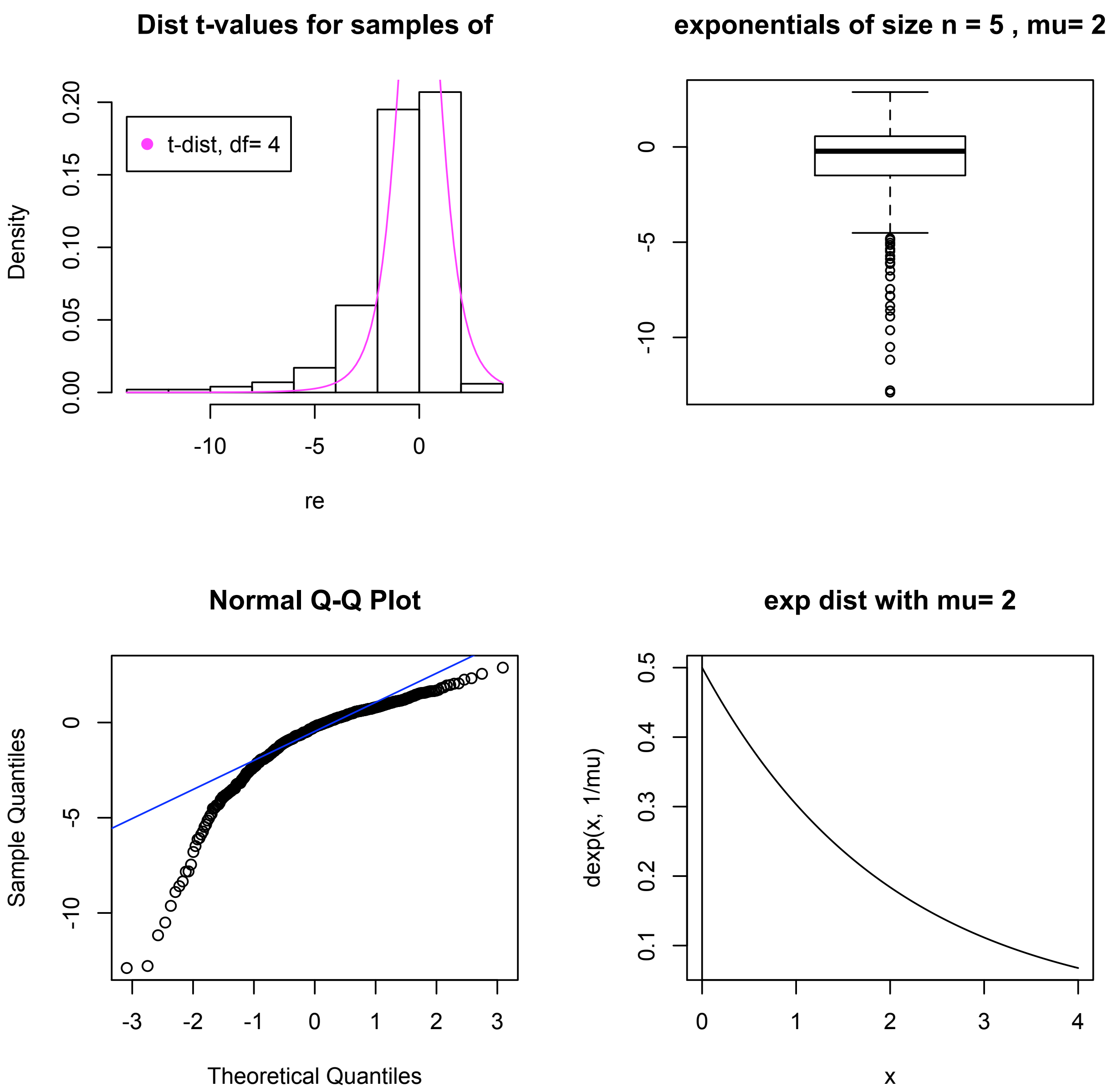
```
> # The exponential distribution is asymmetric and one tail is light,
> # the other s heavy. Do these properties show up in the simulation?
> # Let's try samples of size  15. The distribution of the statistic is
> # skewed and no longer normal.
>
> n<- 15;mu<-10
> for(i in 1:500)re[i]<- make.t(rexp(n,1/mu),mu)
> for(i in 1:500)re[i]<- make.t(rexp(n,1/mu),mu)
> hist(re,main="Dist t-values for samples of")
> boxplot(re,main=paste("exponentials of size n =",n,", mu=",mu));qqnorm(re);qqline(re,col=2)
> curve(dexp(x,1/mu),xlim=c(0,4),main=paste("exp dist with mu=",mu));abline(h=0);abline(v=0)
```



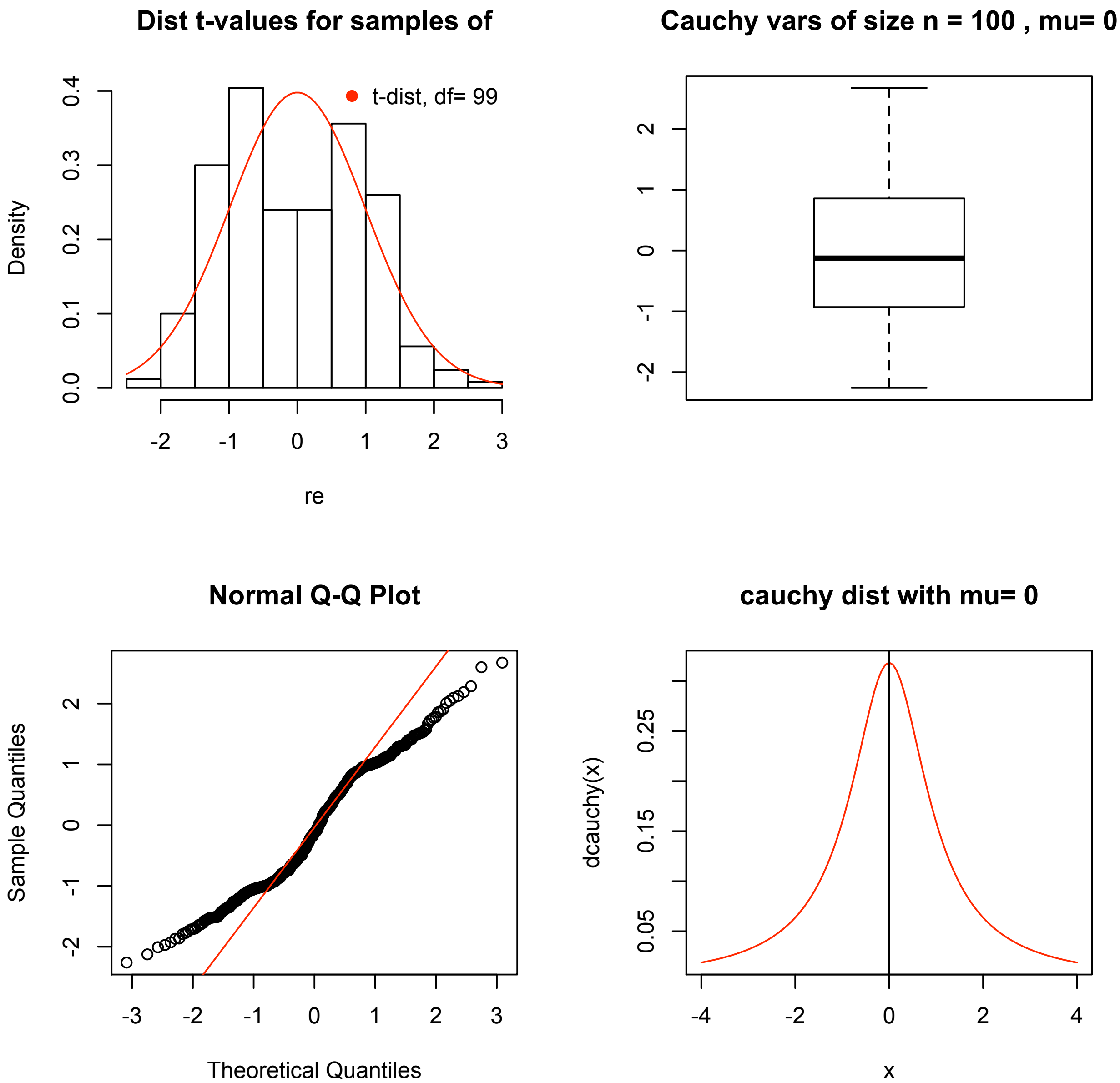
```
> # Let's try samples of size  8 and change the parameter. The
> # distribution of the statistic looks worse.
>
> n<- 8;mu<-4
> for(i in 1:500)re[i]<- make.t(rexp(n,1/mu),mu)
> hist(re,main="Dist t-values for samples of",freq=F);curve(dt(x,n-1),add=T,col=4)
> legend(-18, .19,legend=paste("t-dist, df=",n-1),col=4,pch=19)
> boxplot(re,main=paste("exponentials of size n =",n,", mu=",mu));qqnorm(re);qqline(re,col=4)
> curve(dexp(x,1/mu),xlim=c(0,4),main=paste("exp dist with mu=",mu));abline(h=0);abline(v=0)
```



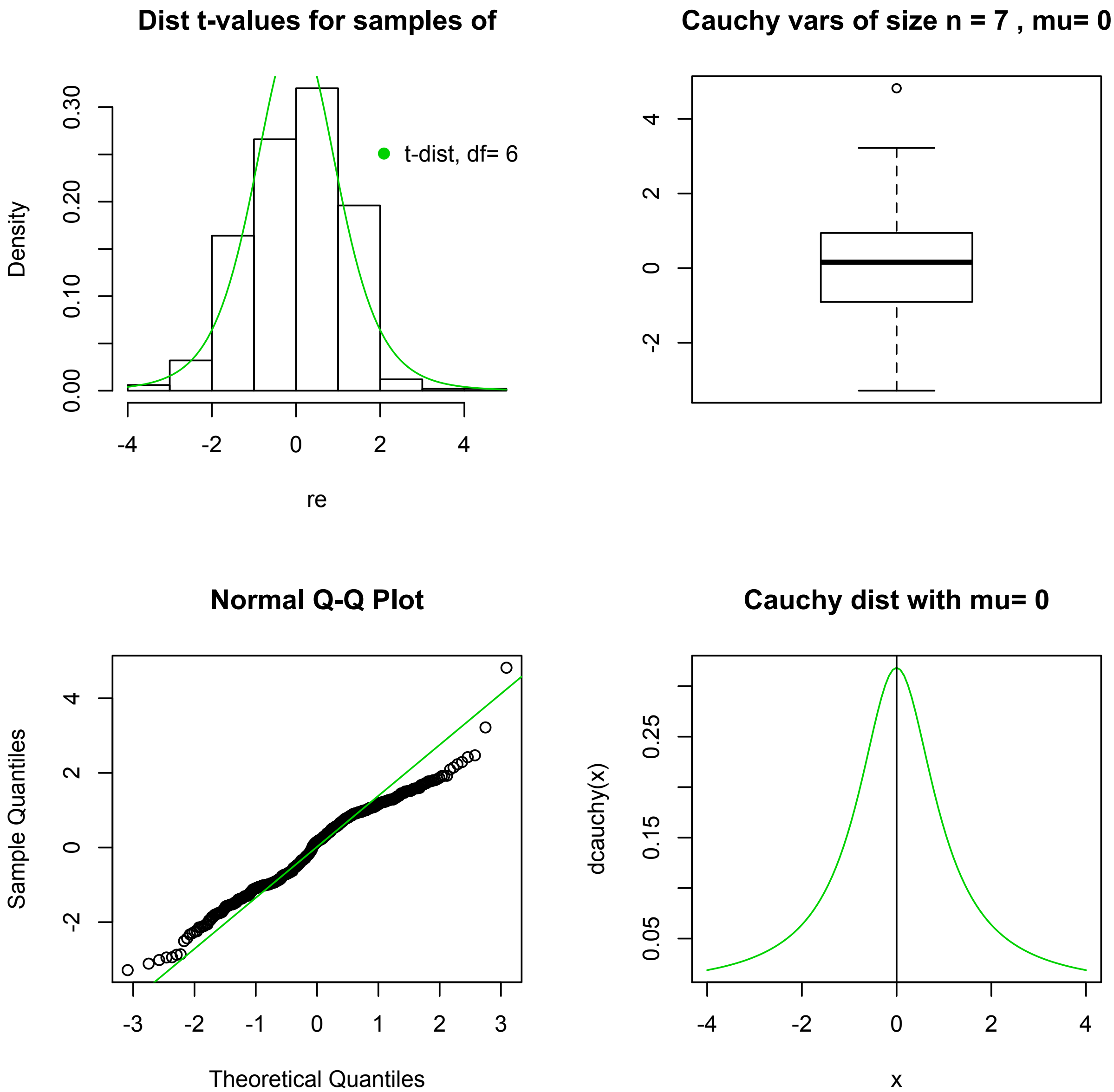
```
> # Let's try samples of size  5 and change the parameter. The
> # distribution of the statistic looks even worse.
> n <- 5; mu <- 2
> for(i in 1:500)re[i]<- make.t(rexp(n,1/mu),mu)
> hist(re,main="Dist t-values for samples of",freq=F);curve(dt(x,n-1),add=T,col=6)
> legend(-14, .19,legend=paste("t-dist, df=",n-1),col=6,pch=19)
> boxplot(re,main=paste("exponentials of size n =",n,", mu=",mu));qqnorm(re);qqline(re,col=4)
> curve(dexp(x,1/mu),xlim=c(0,4),main=paste("exp dist with mu=",mu));abline(h=0);abline(v=0)
```



```
> # Following Venables, we experiment with a symmetric, heavy-tailed
> # distribution, the Cauchy distribution. Take a look at large sample
> # first. Its t-statistic looks fairly normal, if not slightly short.
>
> n <- 100
> mu <- 0;
> for(i in 1:500)re[i]<- make.t(rcauchy(n),mu)
> hist(re,main="Dist t-values for samples of",freq=F)
> curve(dt(x,n-1),add=T,col=2)
> legend(.5, .43,legend=paste("t-dist, df=",n-1),col=2,pch=19,bty="n")
> boxplot(re,main=paste("Cauchy vars of size n =",n,", mu=",mu))
> qqnorm(re)
> qqline(re,col=2)
> curve(dcauchy(x),xlim=c(-4,4),main=paste("Cauchy dist with mu=",mu),col=2)
> abline(h=0)
> abline(v=0)
```


Dist t-values for samples of
Cauchy vars of size n = 100 , mu= 0
t-dist, df= 99
Density
0.0
0.1
0.2
0.3
0.4
-2
-1
0
1
2
3
re
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-3
cauchy dist with mu= 0
dcauchy(x)
0.05
0.15
0.25
-4
4
x

```
> # Take a look at a small sample (n=7). Its t-statistic still looks fairly
> # normal. Thus the t-statistic is fairly robust with respect to this
> # kind of population.
>
> n <- 7; mu <- 0
> for(i in 1:500)re[i]<- make.t(rcauchy(n),mu)
> hist(re,main="Dist t-values for samples of",freq=F)
> curve(dt(x,n-1),add=T,col=3)
> legend(1.6, .28,legend=paste("t-dist, df=",n-1),col=3,pch=19,bty="n")
> boxplot(re,main=paste("Cauchy vars of size n =",n,", mu=",mu))
> qqnorm(re)
> qqline(re,col=3)
> curve(dcauchy(x),xlim=c(-4,4),main=paste("Cauchy dist with mu=",mu),col=3)
> abline(h=0)
> abline(v=0)
>
> # Return to previous settings.
> par(opar)
```


Dist t-values for samples of
Cauchy vars of size n = 7 , mu= 0
Density
re
t-dist, df= 6
0.00
0.10
0.20
0.30
-4
-2
0
2
4
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-3
-2
-1
0
1
2
3
Cauchy dist with mu= 0
dcauchy(x)
x
0.05
0.15
0.25