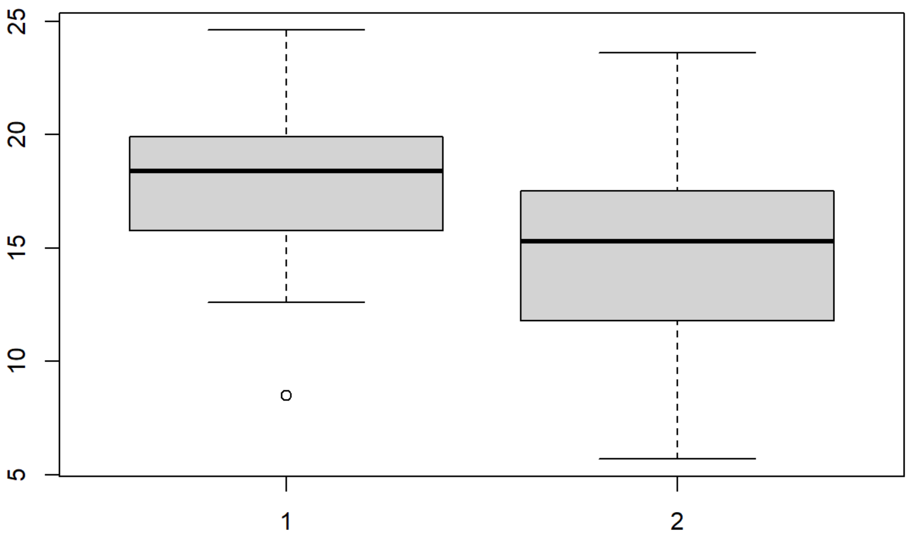
Verzani Problem Set

Next are considered the problems from Verzani's book on page 72.

Problem 11.1

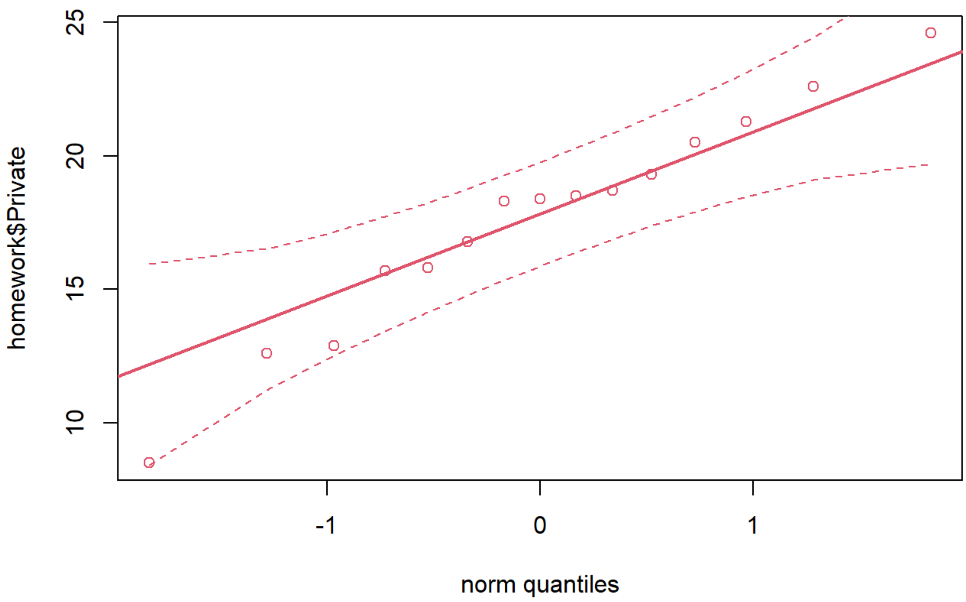
Consider the data set `homework`. This measures study habits of students from private and public high schools. Make a side-by-side boxplot. Use the appropriate test to test for equality of centers.

```
> library(UsingR)
Warning: package 'UsingR' was built under R version 4.0.3
Loading required package: MASS
Loading required package: HistData
Loading required package: Hmisc
Loading required package: lattice
Loading required package: survival
Loading required package: Formula
Loading required package: ggplot2

Attaching package: 'Hmisc'
The following objects are masked from 'package:base':

    format.pval, units

Attaching package: 'UsingR'
The following object is masked from 'package:survival':

    cancer
> head(homework)
  Private Public
1    21.3   15.3
2    16.8   17.4
3     8.5   12.3
4    12.6   10.7
5    15.8   16.4
6    19.3   11.3
> summary(homework)
    Private          Public
 Min.   : 8.50   Min.   : 5.70
 1st Qu.:15.75   1st Qu.:11.80
 Median :18.40   Median :15.30
 Mean   :17.63   Mean   :14.91
 3rd Qu.:19.90   3rd Qu.:17.50
 Max.   :24.60   Max.   :23.60
> boxplot(homework$Private, homework$Public)
```

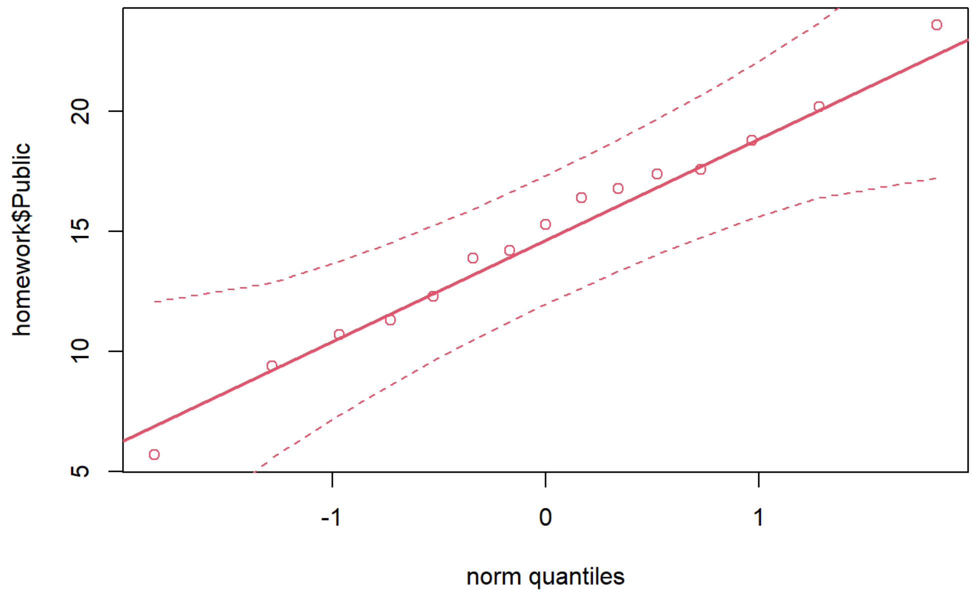


First let’s check for normality

```
> library(StatDA)
Warning: package 'StatDA' was built under R version 4.0.3
Loading required package: sgeostat
Warning: package 'sgeostat' was built under R version 4.0.3
Registered S3 method overwritten by 'geoR':
  method         from
  plot.variogram sgeostat
> qqplot.das(homework$Private)
```



> qqplot.das(homework$Public)



Shapiro test for normality

> shapiro.test(homework$Private)

Shapiro-Wilk normality test

data: homework$Private
W = 0.97017, p-value = 0.8606

The $p-value = 0.8606 > 0.05 = \alpha$, so we have no evidence to reject H_0.

> shapiro.test(homework$Public)

Shapiro-Wilk normality test

data: homework$Public
W = 0.99275, p-value = 0.9999

The $p-value = 0.9999 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Both have normal distributions, so we can use t-test

$H_0 : \mu_{public} = \mu_{private}$
$H_A : \mu_{public} \neq \mu_{private}$

They are independent and we don't know if the variances are equal or not. So we can assume that the variances are different.

```
> t.test(homework$Private, homework$Public)

	Welch Two Sample t-test

data:  homework$Private and homework$Public
t = 1.7134, df = 27.727, p-value = 0.09779
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5345123  5.9878456
sample estimates:
mean of x mean of y 
 17.63333  14.90667 
```

The $p-value = 0.09779 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Or we can check if the variances are equal or not

$H_0 : \sigma_{public} = \sigma_{private}$
$H_A : \sigma_{public} \neq \sigma_{private}$

```
> var.test(homework$Private, homework$Public)

	F test to compare two variances

data:  homework$Private and homework$Public
F = 0.81944, num df = 14, denom df = 14, p-value = 0.7146
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.275110 2.440771
sample estimates:
ratio of variances 
         0.8194392 
```

The $p-value = 0.7146 > 0.05 = \alpha$, so we have no evidence to reject H_0. So we can assume that the variances are equal.

```
> t.test(homework$Private, homework$Public, var.equal = TRUE)

	Two Sample t-test

data:  homework$Private and homework$Public
t = 1.7134, df = 28, p-value = 0.09769
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5330654  5.9863986
sample estimates:
mean of x mean of y 
 17.63333  14.90667 
```

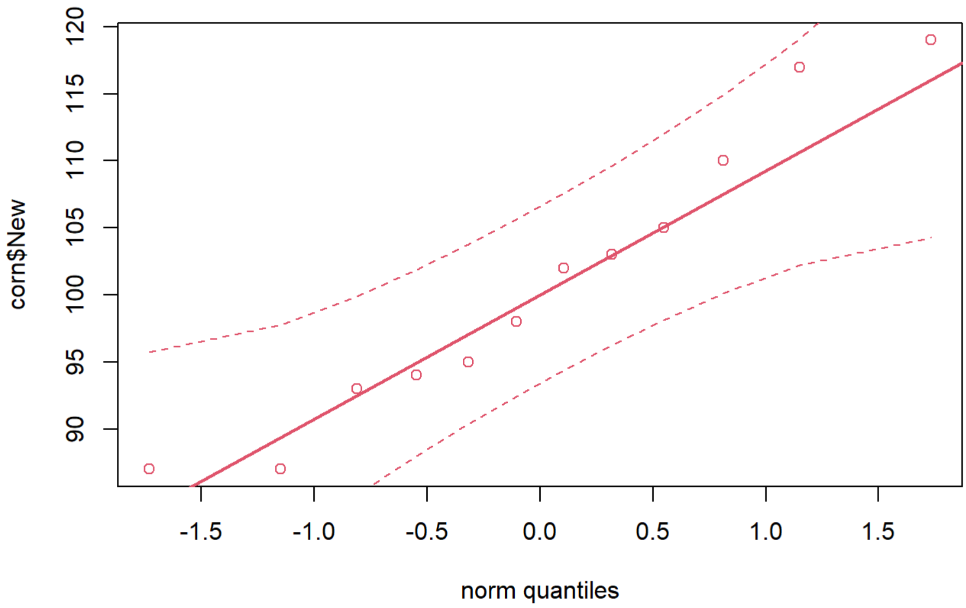
The $p-value = 0.09769 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Problem 11.2
Consider the data set corn. Twelve plots of land are divided into two and then one half of each is planted with a new corn seed, the other with the standard. Do a two-sample t-test on the data. Do the assumptions seems to be met. Comment why the matched sample test is more appropriate, and then perform the test. Did the two agree anyways?

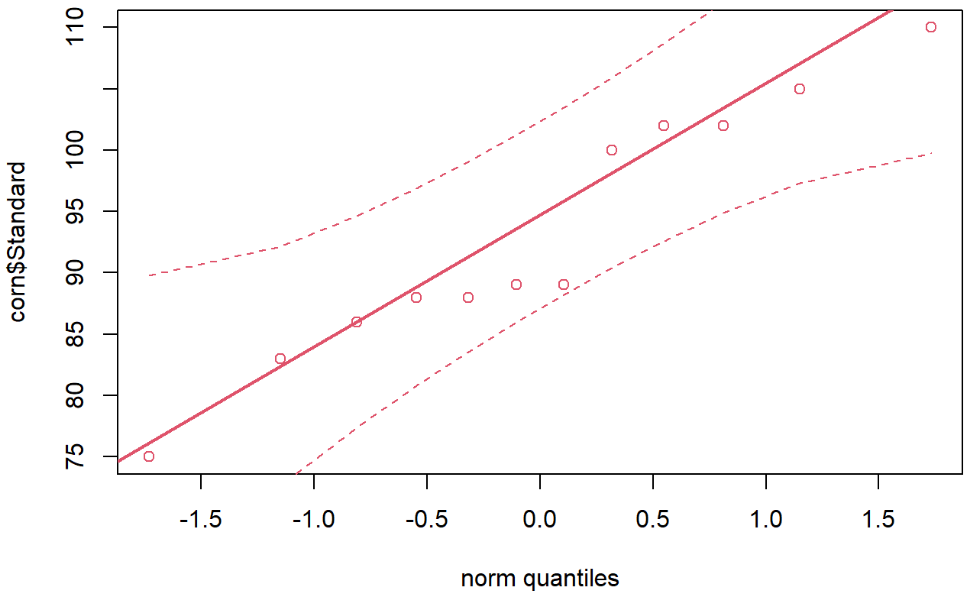
```
> head(corn)
  New Standard
1 110      102
2 103       86
3  95       88
4  94       75
5  87       89
6 119      102
```

First we need to check if the data is normally distributed.

```
> qqplot.das(corn$New)
```



```
> qqplot.das(corn$Standard)
```



Shapiro test for normality

```
> shapiro.test(corn$New)

	Shapiro-Wilk normality test

data:  corn$New
W = 0.94358, p-value = 0.5457
```

The $p-value = 0.5457 > 0.05 = \alpha$, so we have no evidence to reject H_0.

```
> shapiro.test(corn$Standard)

	Shapiro-Wilk normality test

data:  corn$Standard
W = 0.93955, p-value = 0.4923
```

The $p-value = 0.4923 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Both have normal distributions, so we can use t-test

$$H_0 : \mu_{new} = \mu_{standard}$$
$$H_A : \mu_{new} \neq \mu_{standard}$$

The data are paired as we divide the plots in two

```
> t.test(corn$New, corn$Standard, paired = TRUE)
```

```
	Paired t-test

data:  corn$New and corn$Standard
t = 3.8308, df = 11, p-value = 0.00279
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  3.297258 12.202742
sample estimates:
mean of the differences
                   7.75
```

The $p-value = 0.4923 > 0.05 = \alpha$, so we reject H_0. The new and the standard seed doesn't have the same mean.

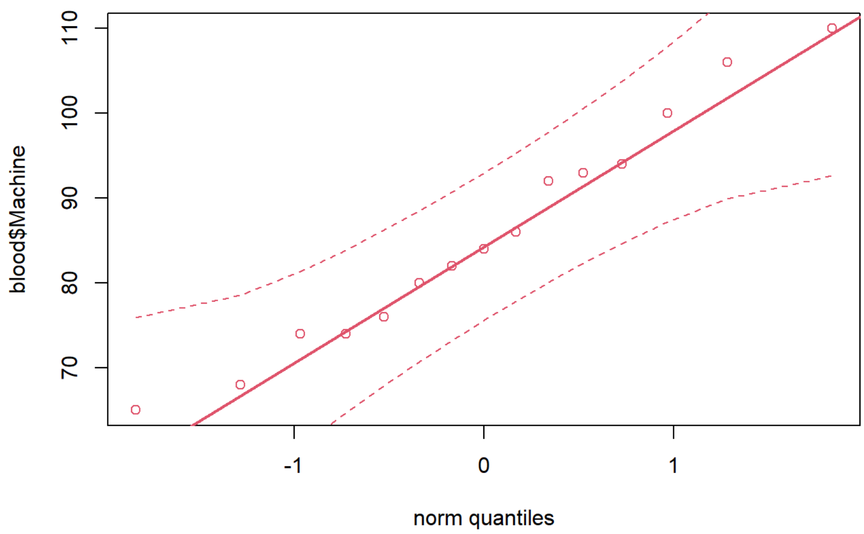
Problem 11.3
Consider the data set blood. Do a significance test for equivalent centers. Which one did you use and why? What was the p-value?

```
> head(blood)
  Machine Expert
1      68     72
2      82     84
3      94     89
4     106    100
5      92     97
6      80     88
```

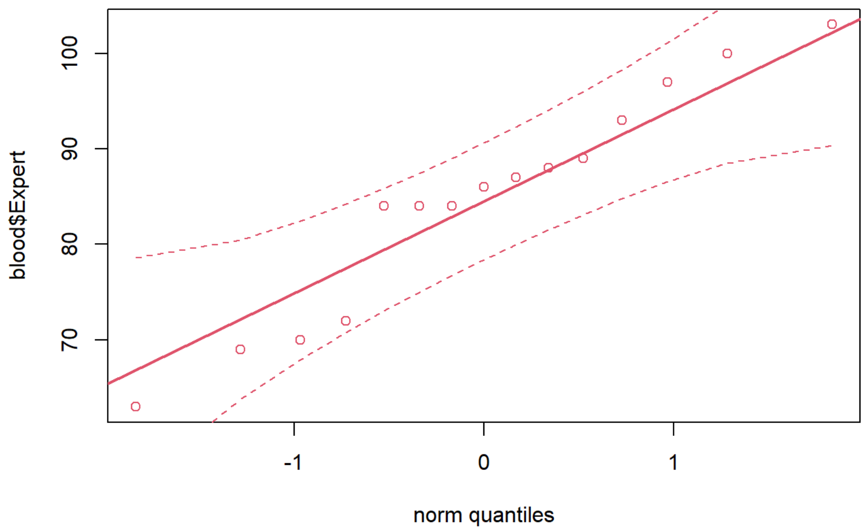
The data present the blood pressure of 15 males taken by machine and expert, so we have paired samples.

We need to check if the data is normally distributed.

```
> qqplot.das(blood$Machine)
```



```
> qqplot.das(blood$Expert)
```



Shapiro test for normality

```
> shapiro.test(blood$Machine)

	Shapiro-Wilk normality test

data:  blood$Machine
W = 0.96996, p-value = 0.8575
```

The $p-value = 0.8575 > 0.05 = \alpha$, so we have no evidence to reject H_0.

```
> shapiro.test(blood$Expert)

	Shapiro-Wilk normality test

data:  blood$Expert
W = 0.94816, p-value = 0.4959
```

The $p-value = 0.4959 = 0.05 = \alpha$, so we have no evidence to reject H_0.

Both have normal distributions, so we can use t-test

$$H_0 : \mu_{machine} = \mu_{expert}$$
$$H_A : \mu_{machine} \neq \mu_{expert}$$

```
> t.test(blood$Machine, blood$Expert, paired = TRUE)

	Paired t-test

data:  blood$Machine and blood$Expert
```

```
t = 0.68162, df = 14, p-value = 0.5066
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.146615  4.146615
sample estimates:
mean of the differences
                      1
```

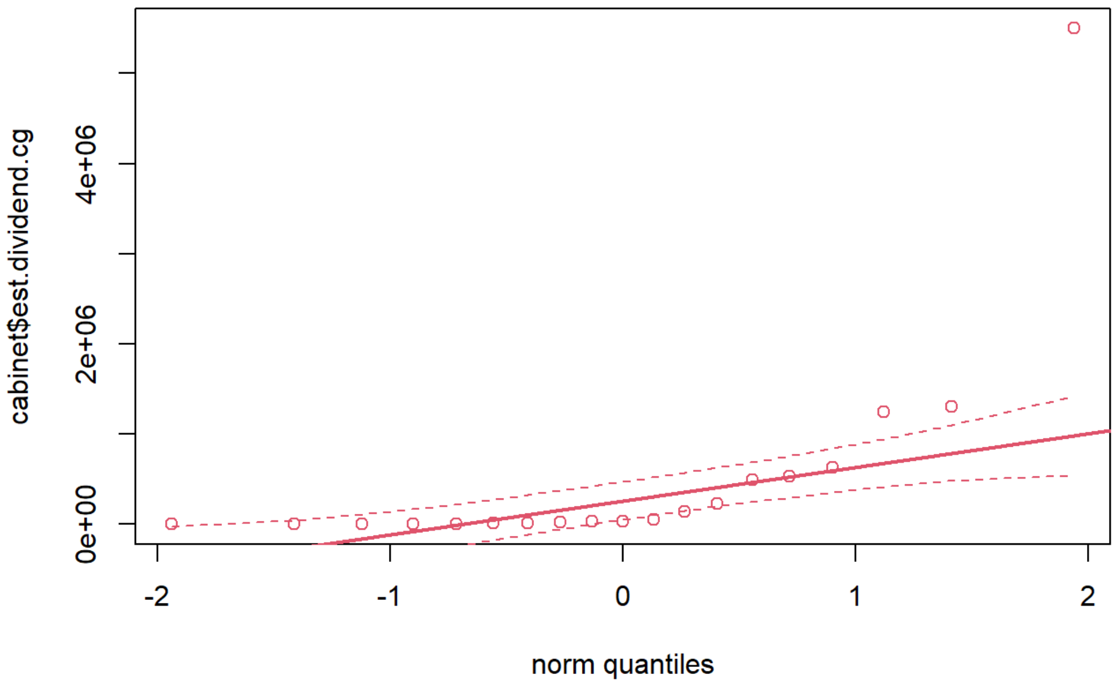
The $p-value = 0.5066 > 0.05 = \alpha$, so we have no evidence to reject H_0.

Problem 11.4

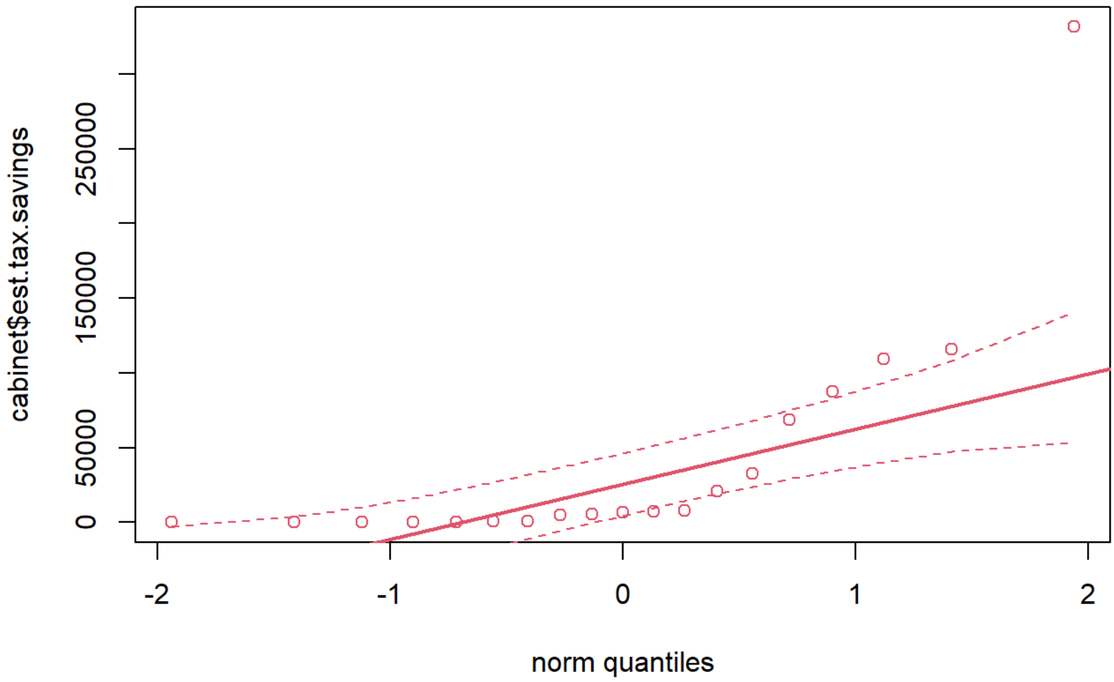
Do a test of equality of medians on the cabinets data set. Why might this be more appropriate than a test for equality of the mean or is it?

```
> head(cabinet)
             name         position est.dividend.cg est.tax.savings
1  George W. Bush        President           23947            5651
2     Dick Cheney   Vice President          493798          116002
3       John Snow Sec. of Treasury         5500000          331594
4    Colin Powell     Sec. of State        1250000          109506
5 Donald Rumsfeld  Sec. of Defense         1300000           87327
6    Donald Evans Sec. of Commerce          625448           68370
```

```
> qqplot.das(cabinet$est.dividend.cg)
```



> qqplot.das(cabinet$est.tax.savings)



Shapiro test for normality

> shapiro.test(cabinet$est.dividend.cg)

Shapiro-Wilk normality test

data: cabinet$est.dividend.cg
W = 0.46916, p-value = 2.706e-07

The $p-value = 0.0000002706 < 0.05 = \alpha$, so we reject H_0.

> shapiro.test(cabinet$est.tax.savings)

Shapiro-Wilk normality test

data: cabinet$est.tax.savings
W = 0.58517, p-value = 3.154e-06

The $p-value = 0.000003154 < 0.05 = \alpha$, so we reject H_0.

Both are not normally distributed, so we can make test for equality of medians

$$H_0 : Me_1 = Me_2$$
$$H_A : Me_1 \neq Me_2$$

```
> wilcox.test(cabinet$est.dividend.cg, cabinet$est.tax.savings)

	Wilcoxon rank sum exact test

data:  cabinet$est.dividend.cg and cabinet$est.tax.savings
W = 258, p-value = 0.02333
alternative hypothesis: true location shift is not equal to 0
```

The $p-value = 0.02333 < 0.05 = \alpha$, so we reject H_0.