**Two-sample Hypothesis Testing**

**Задача 1**

Проведено е допитване с въпрос: приемате ли увеличаване на цената на цигарите с $25\,\%$, като начин за намаляване на тютюнопушенето. $351$ от $605$ непушачи са отговорили с да, докато при пушачите $71$ от $195$ отговарят с да. Може ли да се приеме, че мнението на пушачите и непушачите съвпада.

**Решение:**

| | Non-Smokers | Smokers |
|---|---|---|
| Yes: | 351 | 71 |
| No: | 254 | 124 |
| | 605 | 195 |

$p_1$ - the proportion of the non-smokers who reply "yes"
$p_2$ - the proportion of the smokers who reply "yes"

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

One way to solve the task is with the **rejection region approach**

```
> n1 <- 605
> p1_hat <- 351 / n1
> n2 <- 195
> p2_hat <- 71 / n2
> p_hat <- (71 + 351) / (n1 + n2)
> z <-  ((p1_hat - p2_hat) - 0) / sqrt(p_hat * (1 - p_hat) * (1/n1 + 1/n2))
> z
[1] 5.255544
```

Let's compute the **critical area**

```
> alpha <- 0.05
> qnorm(alpha / 2, 0, 1)
[1] -1.959964
> qnorm(1 - alpha/2, 0, 1)
[1] 1.959964
```

So the critical area is $W_\alpha = \{Z \leq -1.96 \cup Z \geq 1.96\}$. In our case $z = 5.26$ and is much higher than $1.96$, so it is in the critical area for $H_0$ and we reject $H_0$.

Another way to solve the task is with the **p-value approach**

```
> 2 * pnorm(z, 0, 1, lower.tail = FALSE)
```

```
[1] 1.475872e-07
```

The $p-value = 0.000000148 < 0.05 = \alpha$, so we reject $H_0$.
Or we can use the prop.test function

```
> prop.test(c(351, 71), c(605, 195))

	2-sample test for equality of proportions with continuity correction

data:  c(351, 71) out of c(605, 195)
X-squared = 26.761, df = 1, p-value = 2.303e-07
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1345204 0.2976051
sample estimates:
   prop 1    prop 2 
0.5801653 0.3641026 
```

$$\text{The } p-value = 0.0000002303 < 0.05 = \alpha, \text{ so we reject } H_0.$$

**Задача 2:**
Измерено е времето (в дни) за излекуване от дадена болест след прием на ново лекарство 15 10 13 7 9 8 21 9 14 8. Извършено е измерване и на контролна група приемаща пласебо: 15 14 12 8 14 10 7 16 10 15 12. Може ли да се приеме, че новото лекарство подобрява състоянието на пациентитите?

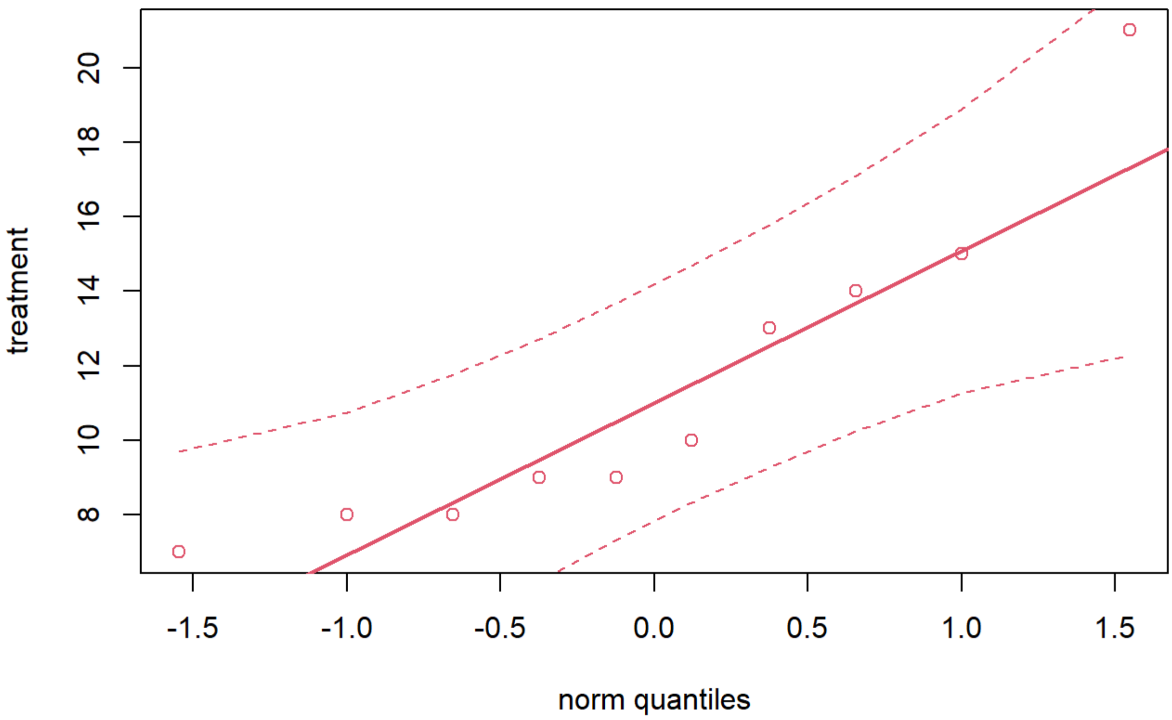
**Решение:**

```
> treatment = c(15, 10, 13, 7, 9, 8, 21, 9, 14, 8)
> placebo = c(15, 14, 12, 8, 14, 10, 7, 16, 10, 15, 12)
```

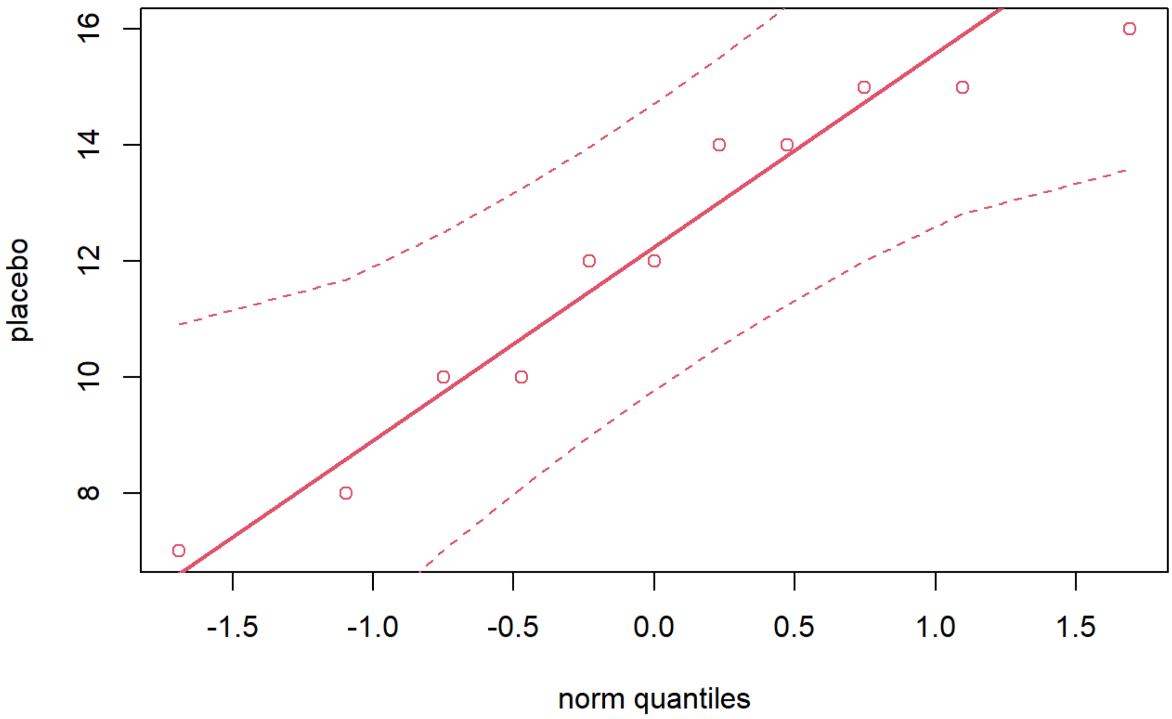
First we check if the data is normally distributed.

```
> 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(treatment)
```




```
> qqplot.das(placebo)
```



```
> shapiro.test(treatment)

	Shapiro-Wilk normality test

data:  treatment
W = 0.86663, p-value = 0.09131
```

The $p-value = 0.09131 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

```
> shapiro.test(placebo)

	Shapiro-Wilk normality test

data:  placebo
W = 0.93174, p-value = 0.4287
```

The $p-value = 0.4287 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

We can assume that both are normally distributed, so we can use t-test

$$H_0 : \mu_{treatment} = \mu_{placebo}$$
$$H_A : \mu_{treatment} \neq \mu_{placebo}$$

```
> t.test(treatment, placebo, alternative = "greater")

	Welch Two Sample t-test

data:  treatment and placebo
t = -0.41894, df = 15.853, p-value = 0.6596
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -3.571812       Inf
sample estimates:
mean of x mean of y 
 11.40000  12.09091 
```

The $p-value = 0.6596 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.
Or we can check if the variances are equal

$$H_0 : \sigma_{treatment} = \sigma_{placebo}$$
$$H_A : \sigma_{treatment} \neq \sigma_{placebo}$$

```
> var.test(treatment, placebo)

	F test to compare two variances

data:  treatment and placebo
F = 2.0827, num df = 9, denom df = 10, p-value = 0.2685
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
 0.5511213 8.2554098
sample estimates:
ratio of variances
         2.082667
```

The $p-value = 0.2685 > 0.05 = \alpha$, so we have no evidence to reject $H_0$. So we can assume that the variances are equal.

```
> t.test(treatment, placebo, alternative = "greater", var.equal = TRUE)

	Two Sample t-test

data:  treatment and placebo
t = -0.42639, df = 19, p-value = 0.6627
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -3.492741       Inf
sample estimates:
mean of x mean of y 
 11.40000  12.09091
```

The $p-value = 0.6627 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

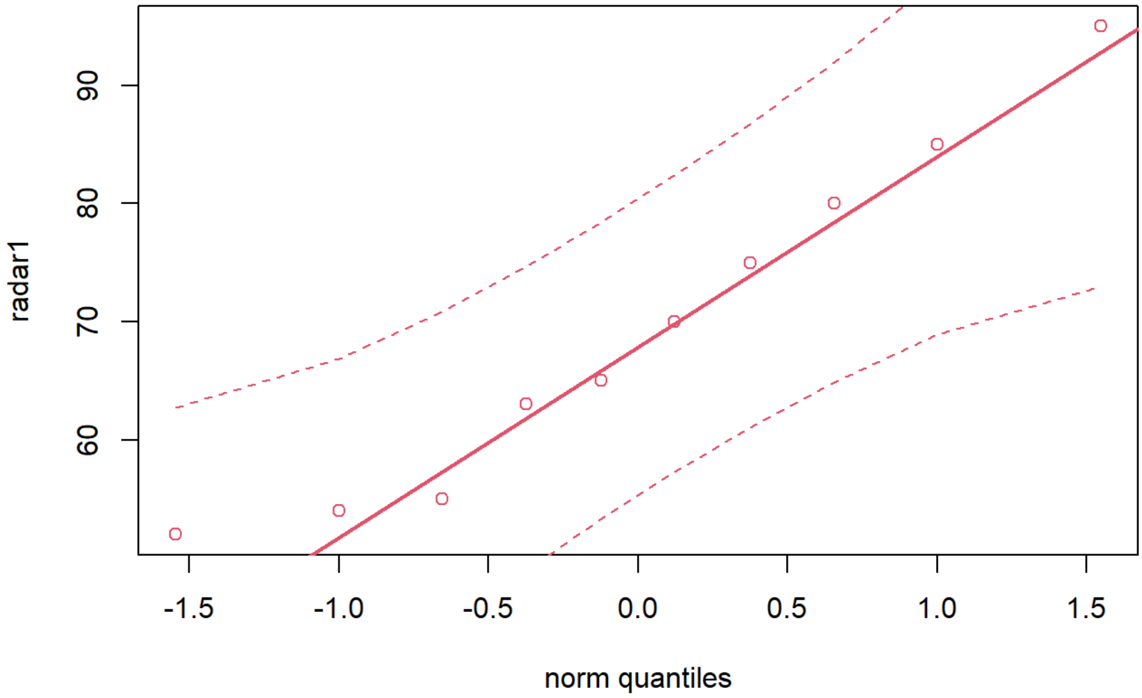
**Задача 3**
Сравняват се два радара за определяне скоростта на автомобил. Направени са десет наблюдения, измерванията на първия са: 70 85 63 54 65 80 75 95 52 55, а на втория: 72 86 62 55 63 80 78 90 53 57. Да се провери дали двата радара са еднакви.

**Решение:**

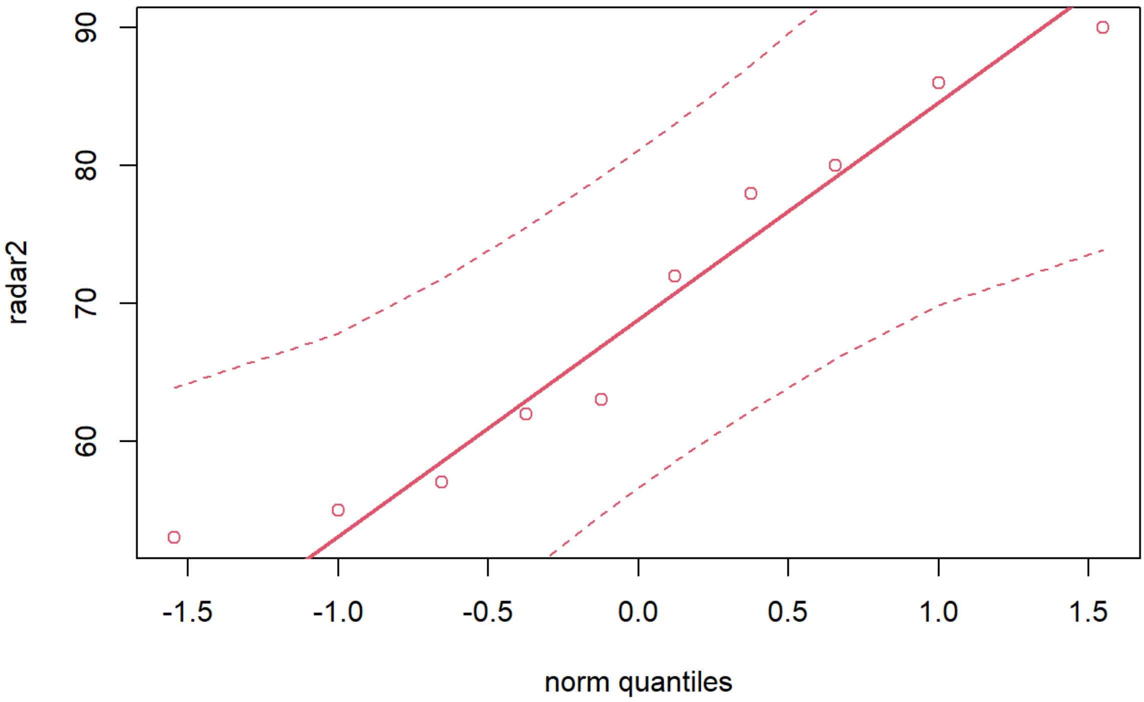
```
> radar1 <- c(70, 85, 63, 54, 65, 80, 75, 95, 52, 55)
> radar2 <- c(72, 86, 62, 55, 63, 80, 78, 90, 53, 57)
```

First we check if the data is normally distributed.

> qqplot.das(radar1)




> qqplot.das(radar2)



```
> shapiro.test(radar1)

	Shapiro-Wilk normality test

data:  radar1
W = 0.94879, p-value = 0.6543
```

The $p-value = 0.6543 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

```
> shapiro.test(radar2)

	Shapiro-Wilk normality test

data:  radar2
W = 0.92451, p-value = 0.3961
```

The $p-value = 0.3961 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

We can assume that both are normally distributed, the data is paired, so we can use t-test

$$H_0 : \mu_{radar_1} = \mu_{radar_2}$$
$$H_A : \mu_{radar_1} \neq \mu_{radar_2}$$

```
> t.test(radar1, radar2, paired = TRUE)

	Paired t-test

data:  radar1 and radar2
t = -0.26941, df = 9, p-value = 0.7937
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.879354  1.479354
sample estimates:
mean of the differences 
                   -0.2 
```

The $p-value = 0.7937 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

**Задача 4**

Разгледайте данните ewr от пакета UsingR. Сравнете времето за напускане на летището от такси обслужващо компаниите American airlines и Northwest airlines.

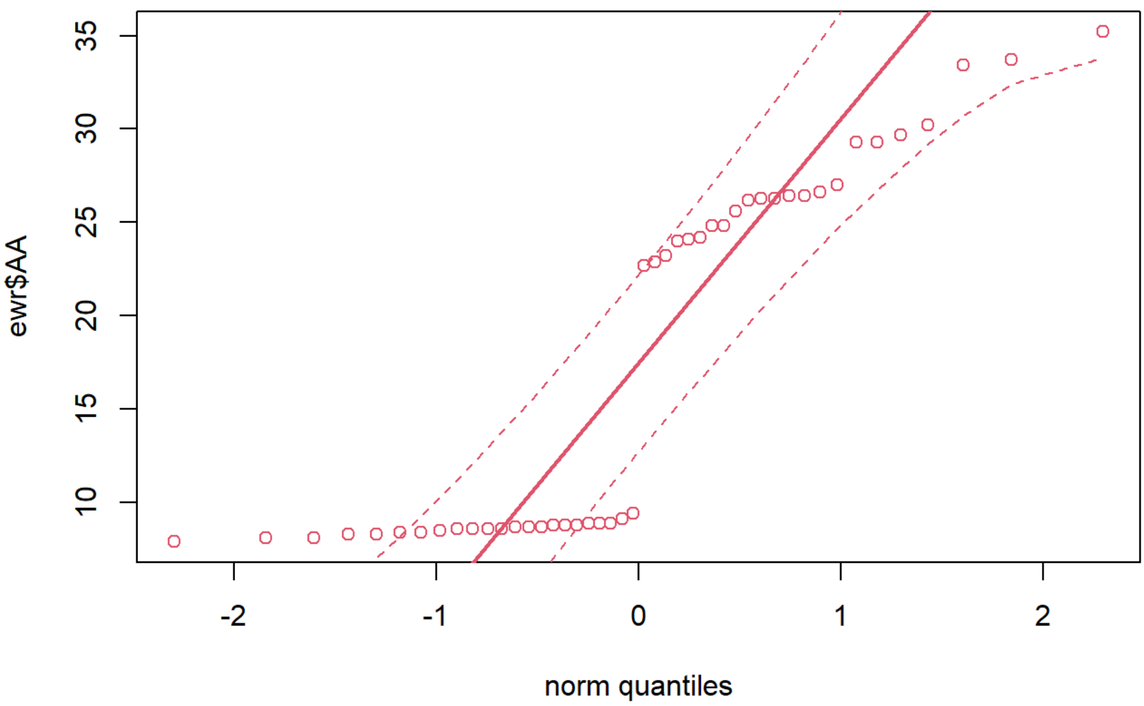
**Решение:**

```
> 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
```

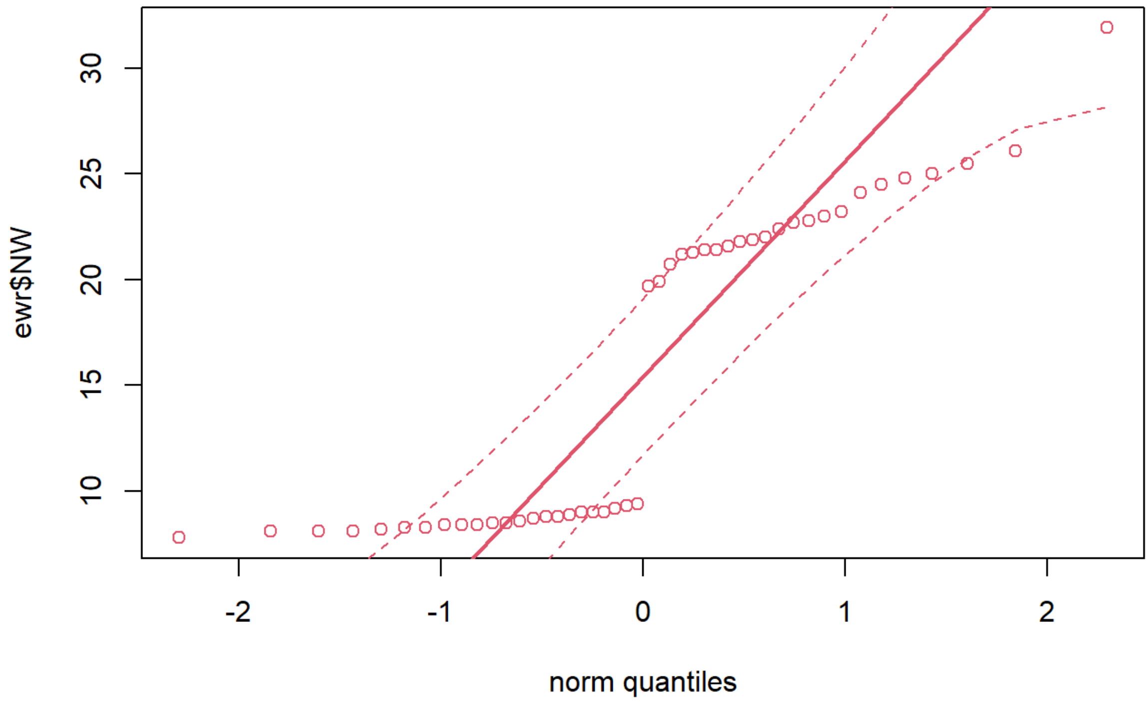
```
Loading required package: survival
Loading required package: Formula
Loading required package: ggplot2

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(ewr)
  Year Month  AA  CO  DL   HP  NW   TW  UA  US inorout
1 2000   Nov 8.6 8.3 8.6 10.4 8.1  9.1 8.4 7.6      in
2 2000   Oct 8.5 8.0 8.4 11.2 8.2  8.5 8.5 7.8      in
3 2000   Sep 8.1 8.5 8.4 10.2 8.3  8.6 8.2 7.6      in
4 2000   Aug 8.9 9.1 9.2 14.5 9.0 10.3 9.2 8.7      in
5 2000   Jul 8.3 8.9 8.2 11.5 8.8  9.1 9.2 8.2      in
6 2000   Jun 8.8 9.0 8.8 14.9 8.4 10.8 8.9 8.3      in
> ewr_out <- subset(ewr, inorout == "out", select = c("AA","NW"))
> ewr_AA <- ewr_out$AA
> ewr_NW <- ewr_out$NW
```

First we check if the data is normally distributed.

```
> qqplot.das(ewr$AA)
```



```
> qqplot.das(ewr$NW)
```




```
> shapiro.test(ewr$AA)

	Shapiro-Wilk normality test

data:  ewr$AA
W = 0.78856, p-value = 1.092e-06
```

The $p-value = 0.000001092 < 0.05 = \alpha$, so we reject $H_0$.

```
> shapiro.test(ewr$NW)

	Shapiro-Wilk normality test

data:  ewr$NW
W = 0.79173, p-value = 1.278e-06
```

The $p-value = 0.000001278 < 0.05 = \alpha$, so we reject $H_0$.

The data is not normally distributed, so we can use hypothesis for the median

$$H_0 : Me(AA) = Me(NW)$$
$$H_A : Me(AA) \neq Me(NW)$$

```
> wilcox.test(ewr$AA, ewr$NW)
Warning in wilcox.test.default(ewr$AA, ewr$NW): cannot compute exact p-value
```

with ties

Wilcoxon rank sum test with continuity correction

data:  ewr$AA and ewr$NW
W = 1263, p-value = 0.1101
alternative hypothesis: true location shift is not equal to 0

The $p-value = 0.1101 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

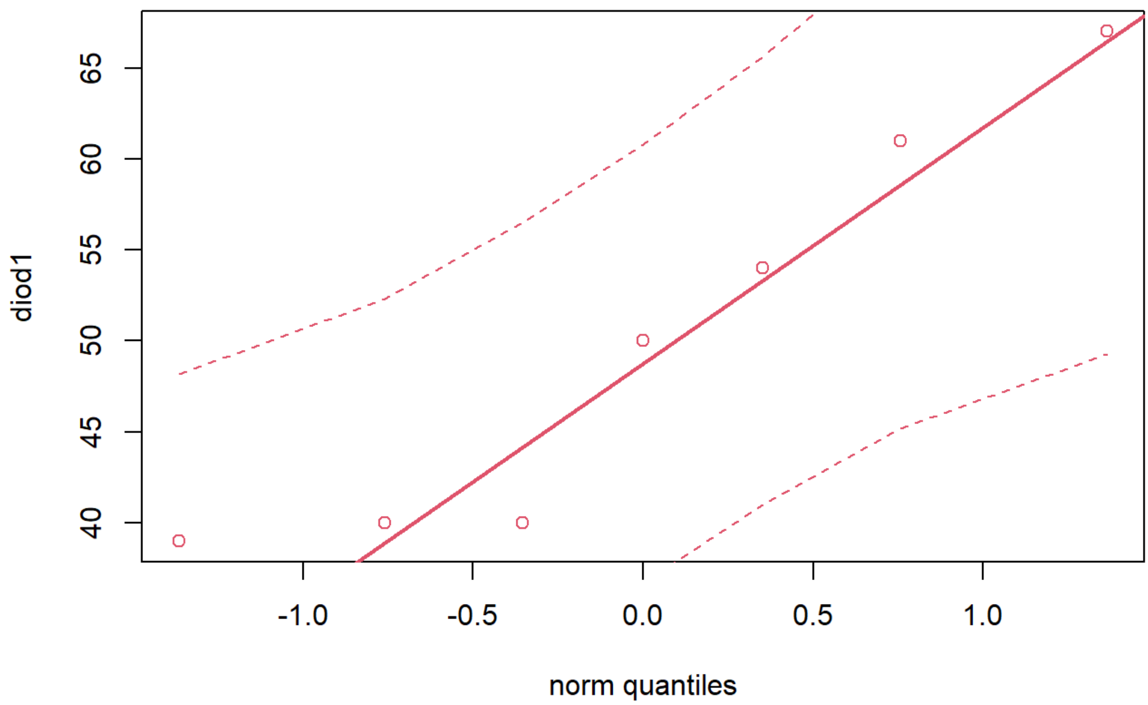
**Задача 6**
Измервани са напреженията на пробив на диоди от две партиди получени са следните наблюдения: 39, 50, 61, 67, 40, 40, 54 за първата партида и 60, 53, 42, 41, 40, 54, 63, 69 за втората. Може ли да се приеме че диодите имат еднакво напрежение на пробив.

**Решение:**

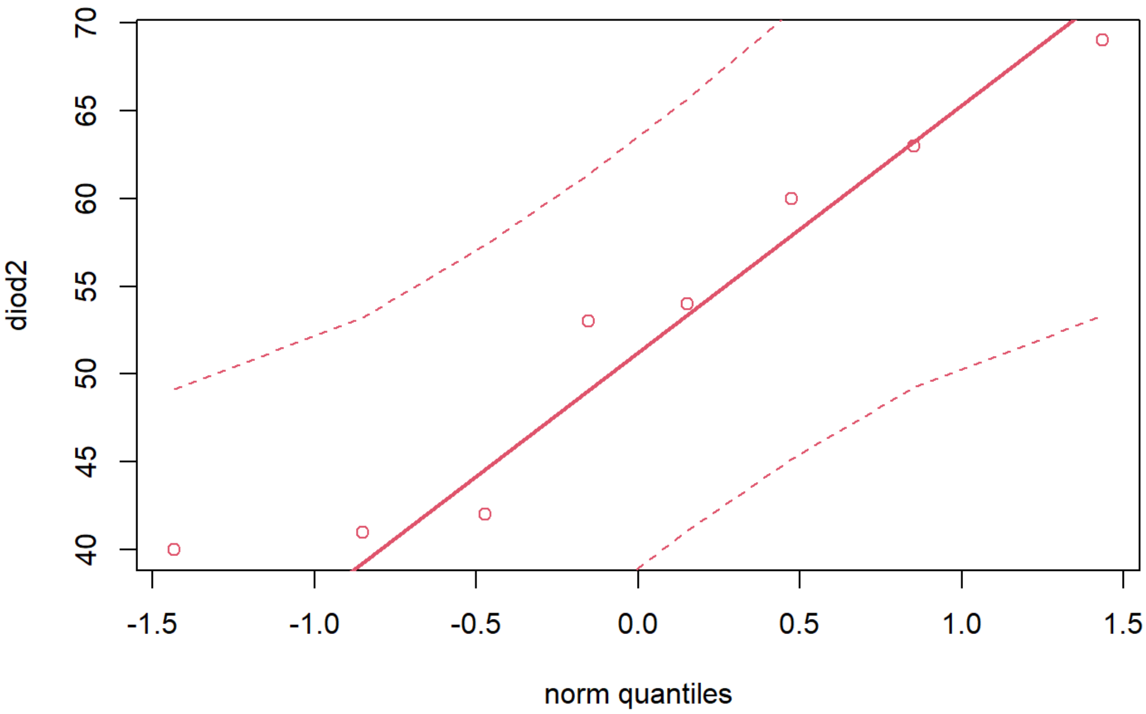
```
> diod1 <- c(39, 50, 61, 67, 40, 40, 54)
> diod2 <- c(60, 53, 42, 41, 40, 54, 63, 69)
```

First we check if the data is normally distributed.

```
> qqplot.das(diod1)
```



```
> qqplot.das(diod2)
```




```
> shapiro.test(diod1)

	Shapiro-Wilk normality test

data:  diod1
W = 0.8903, p-value = 0.2762
```

The $p-value = 0.2762 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

```
> shapiro.test(diod2)

	Shapiro-Wilk normality test

data:  diod2
W = 0.9149, p-value = 0.3899
```

The $p-value = 0.3899 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

We can assume that both are normally distributed, so we can use t-test

$$H_0 : \mu_{diod_1} = \mu_{diod_2}$$
$$H_A : \mu_{diod_1} \neq \mu_{diod_2}$$

```
> t.test(diod1, diod2)

	Welch Two Sample t-test

data:  diod1 and diod2
t = -0.45541, df = 12.665, p-value = 0.6565
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -15.008141   9.793855
sample estimates:
mean of x mean of y 
 50.14286  52.75000 
```

The $p-value = 0.6565 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.

Or we can check if the variances are equal

$$H_0 : \sigma_{diod_1} = \sigma_{diod_2}$$
$$H_A : \sigma_{diod_1} \neq \sigma_{diod_2}$$

```
> var.test(diod1, diod2)

	F test to compare two variances

data:  diod1 and diod2
F = 1.0379, num df = 6, denom df = 7, p-value = 0.9473
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2027742 5.9114393
sample estimates:
ratio of variances 
          1.037919 
```

The $p-value = 0.9473 > 0.05 = \alpha$, so we have no evidence to reject $H_0$. So we can assume that the variances are equal.

```
> t.test(diod1, diod2, var.equal = TRUE)

	Two Sample t-test

data:  diod1 and diod2
t = -0.45602, df = 13, p-value = 0.6559
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.958318   9.744033
sample estimates:
mean of x mean of y 
 50.14286  52.75000 
```

The $p-value = 0.6559 > 0.05 = \alpha$, so we have no evidence to reject $H_0$.