Non-parametric methods:
The Sign Test in R

In this document we introduce nonparametric tests. Especially we concentrate on the 'Sign test' that can be used as an alternative to matched pairs t-test.

Consider 'Schizophrenia' example where volume of left hippocampus region in brain was measured for 15 pairs of twins. One twin was affected by schizophrenia while other was not. Let's first construct two variables 'Unaffect' and 'Affected'. We have to do it manually since there is no data file with these observations:

```
Unaffect=c(1.94,1.44,1.56,1.58,2.06,1.66,1.75,1.77,1.78,
    1.92,1.25,1.93,2.04,1.62,2.08)
Affected=c(1.27,1.63,1.47,1.39,1.93,1.26,1.71,1.67,1.28,
    1.85,1.02,1.34,2.02,1.59,1.97)
```

Here 'c' function binds several numbers into one variable. We also find a variable of differences between 'Unaffect' and 'Affected':

```
diff=Unaffect-Affected
```

The goal is to test if the means of these two variables are the same or not (same as difference between two means is zero or not). Since this is a matched pair example, we can use 't.test' function with 'paired=TRUE' argument:

```
t.test(Unaffect,Affected,paired=TRUE)
```

```
        Paired t-test

data:  Unaffect and Affected
t = 3.2289, df = 14, p-value = 0.006062
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.0667041 0.3306292
sample estimates:
mean of the differences
              0.1986667
```

The p-value is small and therefore we reject null hypothesis and conclude that the means are not the same. However we should not use paired t-test without checking some conditions. In order to use t-test, the pairs have to be independent and distribution of differences approximately normal. We assume that the first condition is satisfied by design and check the second condition using boxplot and QQ-plot of 'diff' variable:

```
par(mfrow=c(1,2))
boxplot(diff,main='Boxplot of diff')
qqnorm(diff)
```
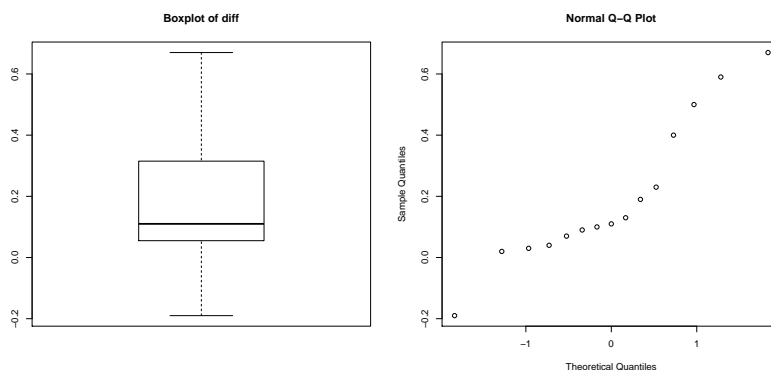
We observe that the boxplot is skewed and QQ-plot departs from linear relationship. Hence we cannot assume that the distribution of 'diff' is nearly normal and therefore t-test may not be appropriate.

We can however compare these two groups using the sign test rather than t-test. To use sign test we do not have to assume normality of differences. For the sign test we need to find number of observations with positive and negative difference, we can easily do it in R:

```
num.diff.pos=sum(diff>0)
num.diff.pos
```

```
[1] 14
```

```
num.diff.neg=sum(diff<0)
num.diff.neg
```

```
[1] 1
```

'diff>0' is just a variable of 1s and 0s such that it is one if 'diff' is positive and 0 otherwise. When we sum such a vector we find number of observations with positive difference. So we have 15 observations with 14 positive differences and 1 with negative difference. The null hypothesis here is that means of two groups are the same which is equivalent that number of positive (or negative) differences has Binomial(15,0.5) distribution. Now we have distribution under the null and the observed value is 14, we can easily get two sided p-value (as was discussed in previous documents 'dbinom(x,n,p)' calculates probability of getting x successes from n independent trials and with probability of success p) :

```
sign.p.value=dbinom(14,15,0.5)+dbinom(15,15,0.5)+
    dbinom(0,15,0.5)+dbinom(1,15,0.5)
sign.p.value
```

```
[1] 0.0009765625
```

The p-value is small and we reject null hypothesis and conclude that two groups have different means. Instead of finding the p-value manually as we did above, we could use 'binom.test' function with similar arguments:

```
binom.test(x=14,n=15,p=0.5,alternative="two.sided")
```

```
        Exact binomial test

data:  14 and 15
number of successes = 14, number of trials = 15, p-value = 0.0009766
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.6805154 0.9983136
sample estimates:
probability of success
             0.9333333
```

Note that the p-value is exactly the same as we got before.