**Timon Otter S2683301 & Ka Ho So S1994441 - Group 54**

**Assignment 5 Statistical Data Analysis**

**Exercise 5.1**

As hinted in the Assignment Description, the sign test will be used for the tests in this exercise, by using the binomial test in R. Denote for each observation in the grades sample. Note that the sample size for the given sample.

a)

The sign test assumes that the underlying distribution has a unique median, however, since , we delete from the sample, and apply the sign test to the rest of the data.

The **test statistic** is

The distribution of the test statistic under is

The test score for the sample is 16. Using the corresponding binomial test (see R code), this gives an **p-value** of 0.941.

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore, we conclude that we do not have evidence that shows that the median from the underlying distribution is (with significant probability) greater than 6.

b)

The sign test assumes that the underlying distribution has a unique median. Since none of the sample data is equal to 6.5, we do not have to adjust the data and sign test.

Testing this is the same as testing both one-sided variants for . is now rejected, if either of the one-sided tests rejects.

The **test statistic** is

The distribution of the test statistic under is

*Right-tailed part*

Note: for the sake of brevity, we will not list the test statistic and distribution again.

The test score for the sample is 8. Using the corresponding binomial test, this gives an p-value of 1.000.

Since this p-value is greater than the significance level, we do not reject the right-tailed part of the (original) null hypothesis.

*Left-tailed part*

The test score for the sample is 8. Using the corresponding binomial test, this gives an p-value of 0.000.

Since this p-value is lower than the significance level, we do reject the left-tailed part of the (original) null hypothesis.

**Overall conclusion:** Since one of the one-tailed versions (with ) of the two-tailed test rejects the null hypothesis, we reject the original null hypothesis ). Therefore, we conclude that (with significant probability) the median of the underlying distribution is not equal to 6.5.

c)

We will use the one-sided binomial test for this:

Since we have a total of (=42) observations, the distribution of the binomial test statistic under is

We have a total of observed successes (amount of grades larger or equal than 7).

Using the binomial test in R, the corresponding one-sided **p-value** is 0.000 (rounded).

**Conclusion:** Since this p-value is lower than the significance level, we reject the null hypothesis. Therefore, we conclude that the probability to get a grade of at least 7 is (probably) lower than 35%.

**Exercise 5.2**

a)

To explore the distribution of the cloud seeded data graphically we used some different methods. First we looked at the symmetry and plotted a histogram and a box plot of the data, as you can see in the box plot, it’s not the same on each side of the median so there could be no symmetry. In the symplot, the data is roughly on the median but with larger values the points diverge from this line. Thus, we can’t know for certain that the data is symmetrical. Also, since the data is small there’s no certainty to assume whether the data is symmetric or not. After looking at a few QQ-plots, it’s still not clear what the distribution could be, this is also due the small data. Therefore, this data does not give a clear view of what the distribution could be.



The numerically data of Cloud seeded has a sample size of 26, a mean of 441.98, median of 221.60, standard deviation of 650.787 and a variance of 423523.9. The minimum value in this data is 4.1 and the maximum value is 2745.6. The IQR = 307.9.

b)

Accuracy is the standard deviation of the data, so accuracy is 650.787.

c)

The bootstrap method we used is the empirical bootstrap, when you are not certain of the distribution it is safer to use the empirical bootstrap method and the empirical bootstrap method makes no assumptions about how your observations are distributed.

Pn= (X1, . . . ,Xn)~clouds.txt$seeded of a unknown distribution and the Tn = accuracy(x) as the statistics. The standard deviation we got from this sample is 177.93.



d) Accuracy is the MAD of the data, so accuracy is 229.9513.

Pn= (X1, . . . ,Xn)~clouds.txt$seeded of a unknown distribution and the Tn = accuracy(x) as the statistics. The standard deviation we got from this sample is 71.371.



e)

We prefer the MAD for the accuracy because the MAD is a more robust estimator for the spread.

f)

Since the data is small and looking at the graphical data we don’t assume that there is normality, so we don’t want to use t-test. We want to use a non parametric test, because these test make no assumption about the underlying distribution. The signed rank test assumes that the data is continues and symmetric around the median. In exercise a, the data doesn’t look symmetric so we won’t use the signed rank test, however the data is small so it’s sceptical that the data is not symmetric. For the signed test, it only makes the assumption that the median is unique. For small data is better to trust the p-value of the signed test.

g)

first we look at the data if there is a value that is equal to 119.0, if this is true we remove it and do a conditionally test on the data. Since there is a we use the following test.

Sign test

The **test statistic** is

The distribution of the test statistic under is

The test score for the sample is 16 and number of trials is 25. Using the corresponding binomial test (see R code), this gives an **p-value** of 0.9461. For the confidence interval of the sign test we have to make the interval manually, because R does not provide us the correct interval.

is rejected if



T = 9, then 302.8, 334.1, … exceed m\_0, so m\_0 < 302.8. T = 17, then 118.3, 129.6,… exceed m\_0, but 115.3 does not, so 115.3 < m\_0. Now we have to look if 115.3 and 302.8 are in the confidence interval and this has to be checked separately, in a conditional test with n = 24.

For H\_0 : m = 92.4 and T=18, check P(T>=17) = 1- P(T<=16):

1-pbinom(17-1,24,0.5)

[1] 0.03195733

For H\_0 : m = 302.8 and T=9, check P(T>=9) = 1- P(T<=8):

> 1-pbinom(9-1,24,0.5)

[1] 0.9242052

The 95-percent confidence interval of this sign test is [115.3, 302.8].

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore, we conclude that the median from the underlying distribution is (with significant probability) less then 119.

h)

**Sign test**

The **test statistic** is

The distribution of the test statistic under is .

The test score for the sample is 16 and number of trials is 25. Using the corresponding binomial test (see R code), this gives an **p-value** of 0.9461. For the confidence interval of the sign test we have to make the interval manually, because R does not provide us the correct interval.

is rejected if



T = 7, then 430.0, 489.1, … exceed m\_0, so m\_0 < 430.0. T = 20, then 40.6, 92.4, … exceed m\_0, but 32.7 does not, so 32.7 < m\_0. Now we have to look if 32.7 and 430.0 are in the confidence interval and this has to be checked separately, in a conditional test with n = 24.

For H\_0 : m = 31.4 and T=20, check P(T>=20) = 1- P(T<=19):

1-pbinom(20-1,24,0.5)

[1] 0.0007719398

For H\_0 : m = 302.8 and T=7, check P(T>=7) = 1- P(T<=6):

1-pbinom(7-1,24,0.5)

[1] 0.9886721

The 99-percent confidence interval of this sign test is (32.7, 430.0].

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore, we conclude that the median from the underlying distribution is (with significant probability) less then 119.

**Signed rank test**

The **test statistic** is ,

The distribution of the test statistic under is ,

The signed rank test gives V = 263 for the sample. Using the corresponding Wilcoxon signed rank test (see R code), this gives an **p-value** of 0.9967 and the 99-percent confidence interval (-, 704).

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore, we conclude that the median from the underlying distribution is (with significant probability) less then 119.

**t-test**

The **test statistic** is ,

The distribution of the test statistic under is

The t-test for the sample gives t = 2.5306 and df = 25. Using the corresponding t-test (see R code), this gives an **p-value** of 0.991 and the 99-percent confidence interval (-, 784.265).

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore, we conclude that the median from the underlying distribution is (with significant probability) less then 119.

i)

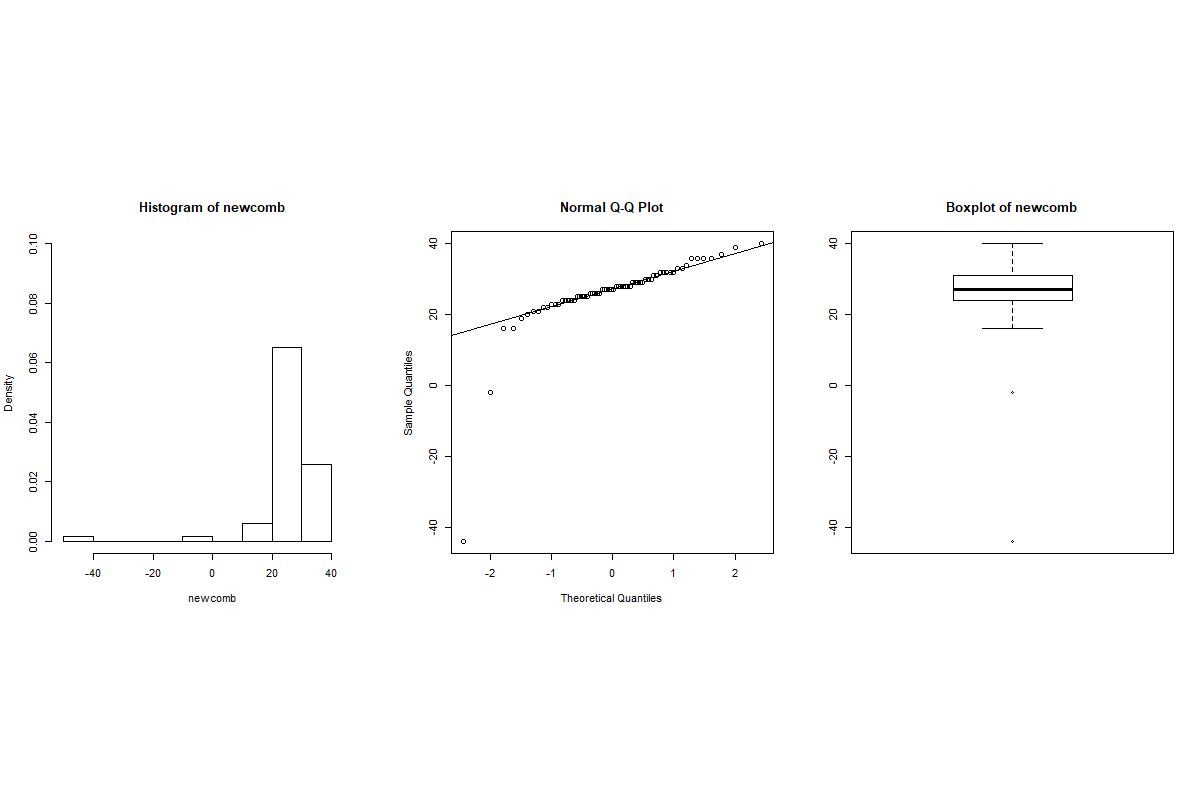
As in exercise f, we prefer the signed test. The data is small and there is no certainty for the distribution. Also the interval for the signed test is narrower so there is more precision of the true mean of the data.

**Exercise 5.3**

a)

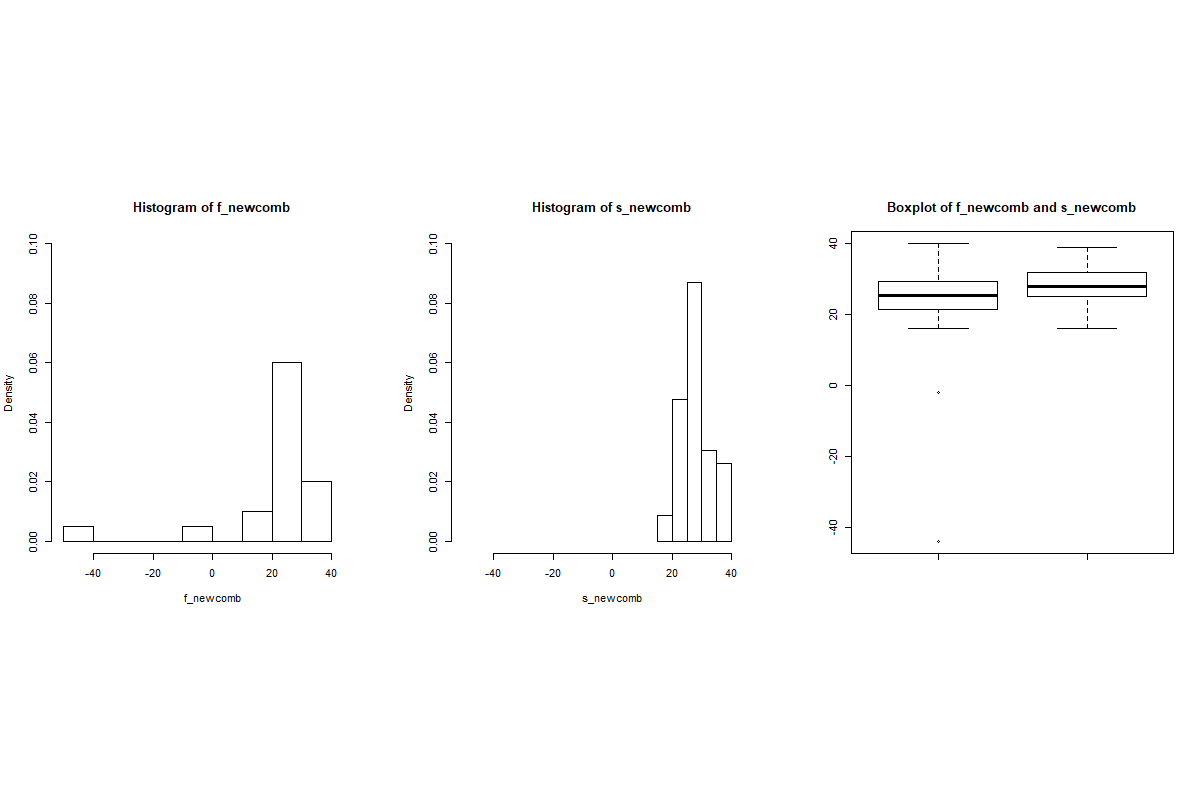
**Whole dataset (newcomb)**

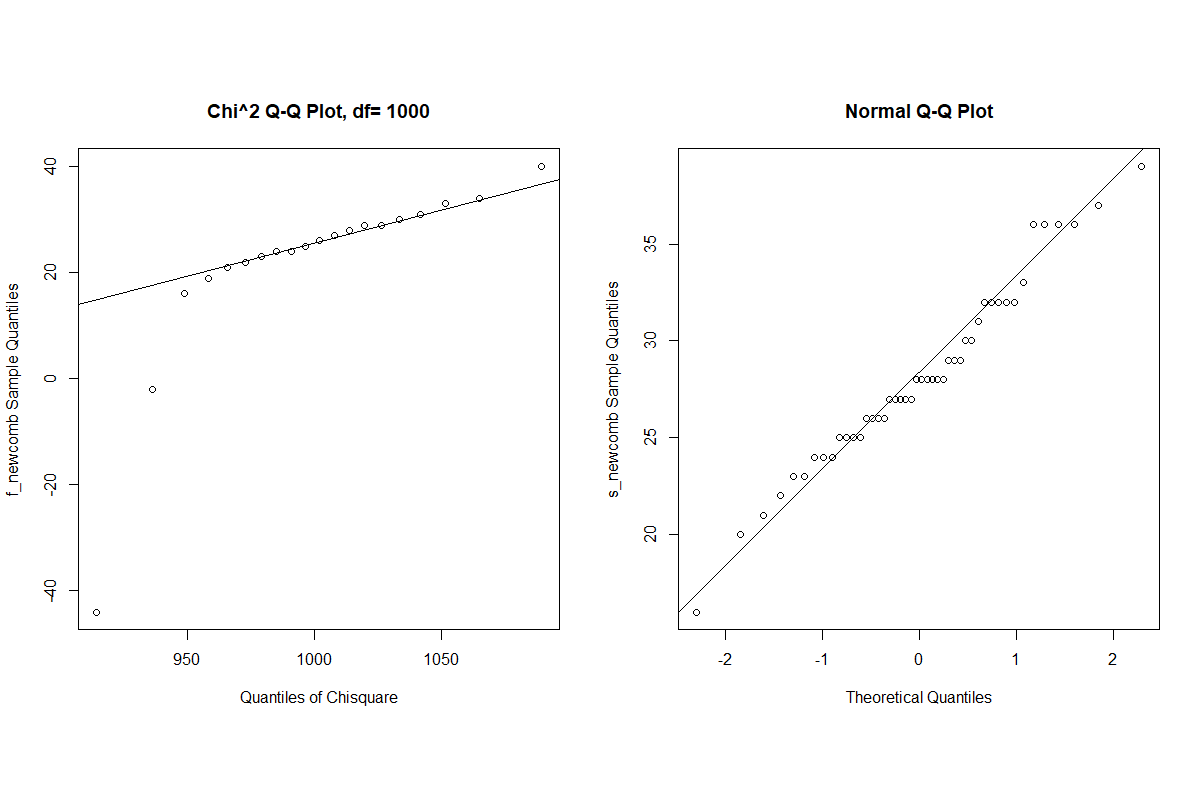
See below for a graphical summary of the data as a whole.



**Comparing First 20 observations (f\_newcomb) with last 46 observations (s\_newcomb)**

See below for a graphical summary (separated for improved readability)





**Estimate for the difference**

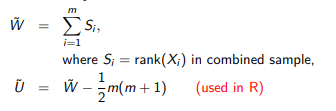
Our estimate of choice is: (f\_newcomb) (s\_newcomb). The computed value is -2.5. The reason for this is that the median is a better alternative than e.g. the mean for skewed distributions (f\_newcomb).

Test for (stochastic difference)

We have performed different tests for testing the difference between the observations (see the R code). However, we think a very suitable test is the Wilcoxon two-sample test. The reason for this being that other two-sample tests are less appropriate for various reasons (t-test: not appropriate because of non-normal data, median test: less powerful, KS-test: not appropriate because of large ties, permutation test: not appropriate because of dependent observations).

We choose a one-sided left-tailed test, inspired by the estimate above. Denote the underlying distribution for f\_newcomb and the underlying distribution for s\_newcomb.

The **test statistic** is 

The distribution of the test statistic under is 

The computed test score (called W in R) is 348. This corresponds to a (non-exact, because of ties) **p-value** of 0.059.

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore we conclude that we do not have evidence that shows that a significant (left-sided) difference, i.e. that the underlying distribution of f\_newcomb is stochastically smaller than s\_newcomb. However, since the p-value is rather close to the significance level, from our conclusion it does not follow that there is absolutely no “(stochastic) difference”.

b)

First we have to transform the newcomb.txt data into the original observation of time, by multiplying the data by and then adding 24.8.



After looking at the data, you can see symmetry on the boxplot and roughly every point is on the symplot line, however there’s no certainty that it is symmetric since the data is small. So to be safe we picked the signed test for the 95-percent confidence interval.

**Sign test**

The **test statistic** is

The distribution of the test statistic under is .

The test score for the sample is 58 and number of trials is 66. Using the corresponding binomial test (see R code), this gives an **p-value** of 0.9999. For the confidence interval of the sign test we have to make the interval manually, because R does not provide us the correct interval.

is rejected if

For the computed 95-percent confidence interval for the sign test is the following interval (24.826, 24.828) and as you can see, the true value of the traveling time 24.8332 is not in the interval.

**Appendix**

**Exercise 5.1**

> grades = scan("statgrades.txt")

> n = length(grades)

> grades==6

> adjgrades = grades[-26]

> t = sum(adjgrades>6)

> binom.test(t,n-1,alternative = "g")

> grades==6.5

> t = sum(grades>6.5)

> binom.test(t,n,alternative = "g")

> binom.test(t,n,alternative = "l")

> k = sum(grades>=7)

> binom.test(k,n,alternative = "l")

**Exercise 5.2**

> source("functions\_Ch3.txt")

> source("functions\_Ch5.txt")

> clouds = read.table("clouds.txt")

> seeded = clouds$seeded.clouds

> summary(seeded)

> par(mfrow =c(2,3),pty="s")

> hist(seeded,col="blue",main="Histogram of seeded",xlab="seeded")

> boxplot(seeded,col="blue",main="Boxplot seeded",ylab="Sorted Data",xlab="seeded")

> symplot(seeded,pch=20,col="blue",cex=1,main="Symplot of seeded")

> qqnorm(seeded,pch=20,col="blue",cex=1,ylab="Sorted Data",main="normal qqplot seeded")

> qqline(seeded)

> qqexp(seeded,pch=20,col="blue",cex=1,ylab="Sorted Data")

> qqcauchy(seeded,col="blue",cex=1,ylab="Sorted Data")

> sd\_seeded = sd(seeded)

> par(mfrow =c(2,2),pty="s")

> B=1000

> bo\_seed = bootstrap(seeded,sd,B=1000)

> hist(bo\_seed,col="green" ,main="Empirical bootstrap of seeded",xlab="accuracy(sd) seeded")

> sd(bo\_seed)

> summary(bo\_seed)

> mad\_s = mad(seeded)

> bo\_mad = bootstrap(seeded,mad,B=1000)

> sd(bo\_mad)

> hist(bo\_mad,col="blue" ,main="Empirical bootstrap of seeded",xlab="accuracy(MAD) seeded")

> summary(bo\_mad)

> l\_seed = length(seeded)

> s\_seed = sum(seeded>119)

> seed2 = seeded[seeded == 119]

> l\_seed2 = length(seed2)

> binom.test(s\_seed,l\_seed-l\_seed2,alt="l")

> t.test(seeded,mu=119,alt="l")

> wilcox.test(seeded,mu=119,alt="less",exact = TRUE)

> t.test(seeded,mu=119,alt="l",conf.level = 0.99)

> wilcox.test(seeded,mu=119,alt="l",conf.level = 0.99,exact = FALSE,conf.int = T)

> binom.test(s\_seed,l\_seed-l\_seed2,alt="l",conf.level = 0.99)

> rbind(0:25, round(pbinom(0:25,size = 25,p=0.5),3))

> rbind(0:25, round(1-pbinom((0:25)-1,size = 25,p=0.5),3))

> sort(seeded2)

> 1-pbinom(17-1,24,0.5)

> 1-pbinom(9-1,24,0.5)

> 1-pbinom(20-1,24,0.5)

> 1-pbinom(7-1,24,0.5)

**Exercise 5.3**

> source("functions\_Ch3.txt")

> source("functions\_Ch5.txt")

> newcomb = matrix(scan("newcomb.txt"))

> f\_newcomb = newcomb[1:20]

> s\_newcomb = newcomb[21:66]

> par(mfrow =c(1,4),pty="s")

> summary(newcomb)

> hist(newcomb, xlim=c(-50,50), ylim=c(0,0.1), prob=T)

> qqnorm(newcomb)

> qqline(newcomb)

> boxplot(newcomb, main="Boxplot of newcomb")

> par(mfrow =c(2,3),pty="s")

> hist(f\_newcomb, xlim=c(-50,50), ylim=c(0,0.1), prob=T)

> hist(s\_newcomb, xlim=c(-50,50), ylim=c(0,0.1), prob=T)

> boxplot(f\_newcomb,s\_newcomb, main="Boxplot of f\_newcomb and s\_newcomb")

> qqchisq(f\_newcomb, df = 1000, ylab="f\_newcomb Sample Quantiles")

> qqline(f\_newcomb, dist = function(p) qchisq(p, df = 1000))

> qqnorm(s\_newcomb, ylab="s\_newcomb Sample Quantiles")

> qqline(s\_newcomb)

> est = median(f\_newcomb) - median(s\_newcomb)

> data = c(f\_newcomb,s\_newcomb)

> nx = length(f\_newcomb)

> ny = length(s\_newcomb)

> myteststat = function(z,m,n) {median(z[1:m])-median(z[(m+1):(m+n)])}

> B=1000

> permutationtval = numeric(B)

> for (i in 1:B) permutationtval[i]=myteststat(sample(data),nx,ny)

> t=myteststat(data,nx,ny)

> sum(permutationtval<=t)/B

> ks.test(f\_newcomb,s\_newcomb,alternative = "greater",conf.int = "true")

> wilcox.test(f\_newcomb,s\_newcomb,alternative = "less",conf.int = "true")

> tt=sum(f\_newcomb<=median(data))

> 1-phyper(t-1,nx,ny,floor((nx+ny+1)/2))

> realnewcomb = 24.8 + (newcomb \* 10^-3)

> par(mfrow =c(2,2),pty="s")

> hist(realnewcomb,col="blue",main="Histogram of original newcomb",xlab="original newcomb")

> boxplot(realnewcomb,col="blue",main="Boxplot original newcomb",ylab="Sorted Data",xlab="original newcomb")

> symplot(realnewcomb,pch=20,col="blue",cex=1,main="Symplot of original newcomb")

> qqnorm(realnewcomb,pch=20,col="blue",cex=1,ylab="Sorted Data",main="normal qqplot original newcomb")

> qqline(realnewcomb)

> l\_new = length(realnewcomb)

> s\_new = sum(realnewcomb<24.8332)

> new2 = realnewcomb[realnewcomb == 24.8332]

> l\_new2 = length(new2)

> binom.test(s\_new,l\_new-l\_new2,alt="l")

> rbind(0:66, round(pbinom(0:66,size = 66,p=0.5),4))

> rbind(0:66, round(1-pbinom((0:66)-1,size = 66,p=0.5),4))

> sort(realnewcomb)

> 1-pbinom(26-1,65,0.5)

> 1-pbinom(42-1,65,0.5)

> t.test(realnewcomb,mu=24.8332,alt="l",conf.int = "true")

> wilcox.test(realnewcomb,mu=24.8332,alt="l",exact = TRUE,conf.int = "true")