# Mini Project #3

## Yunchao Liu

#### Session1:

Exercise1:

1. use the QQplot and pearson test. There is no specific step because those two functions are in R, just call them.
2. use QQplot, histogram and chi-square test. The first two are built in functions in R so I just called them. For chi-square test, I set 10 bins. And set test statistics as sum((O-E)^2/E), where O stands for Observed value and E stands for Expected value. Because the chi-square is one sided in this case, so we get p-value by calculating 1-pchisq( test statistics )

Exercise2:

1. In one experiment, create data from standard normal distribution
2. get test statistics from (xbar – x0) / sqrt( v/ n ), where xbar is the mean of data, x0 is the mean of population, which is 0, variance of data, n is the size of data.
3. get the p-value by 2\*(1-pt( abs( tobs ), n-1 ), where tobs is the observed value of test statistics and n-1 is the degree of freedom.
4. Do the experiment 1000 times to get 1000 p-values
5. Do another chi-square test to check whether it is from Uniform(0,1), the steps of doing so are the same as in Exercise 1
6. We can also use histogram or QQ plot to check whether it is from Uniform(0,1 )
7. Change n value and conduct the above experiments to see whether n can affect the result.

#### Session2:

Exercise1:

1. Below is the result of Pearson test:

Pearson chi-square normality test

data: data

P = 8.94, p-value = 0.5378

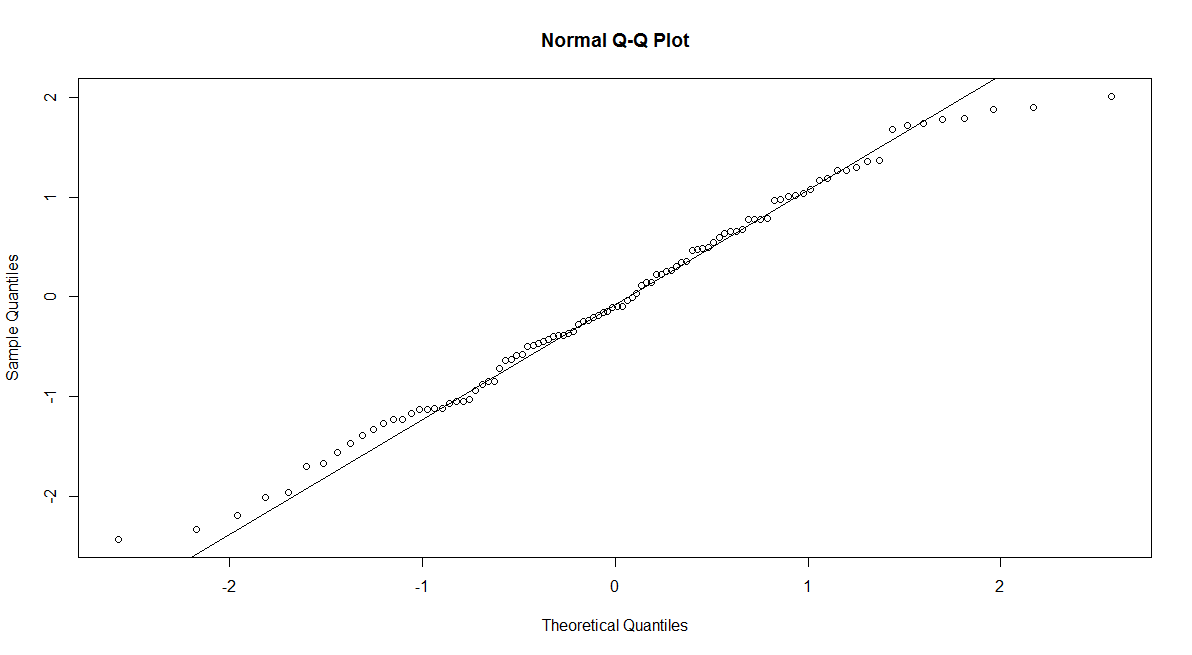
Pearson chi-square normality test

data: data

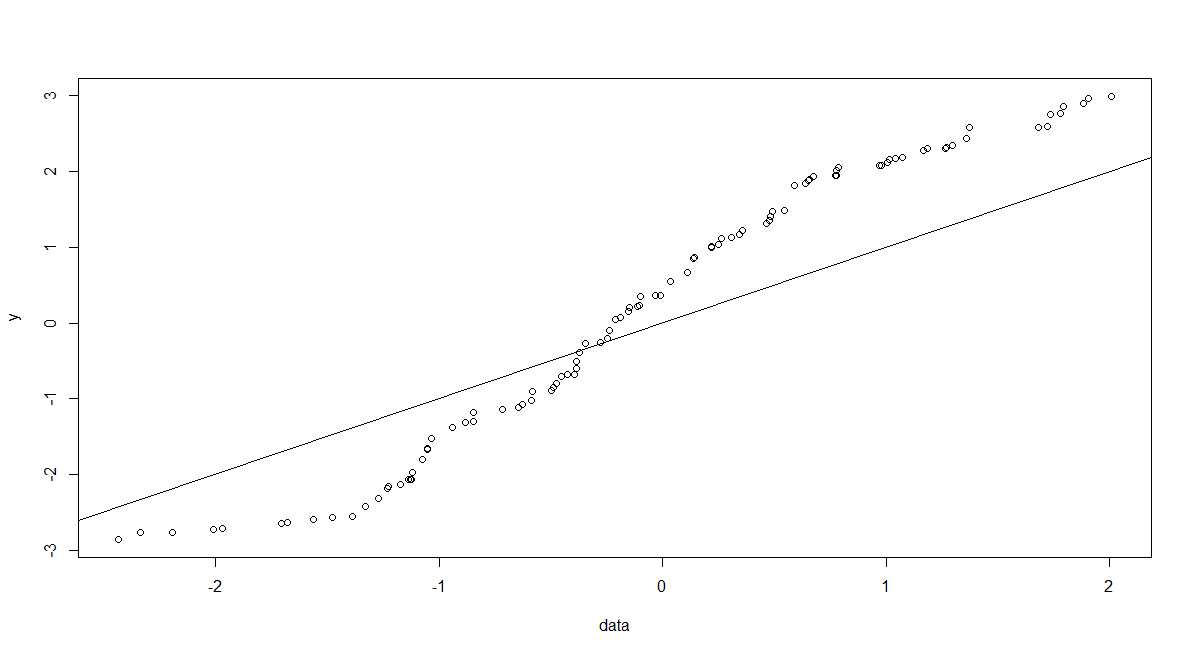
P = 8.94, p-value = 0.5378

So from the above result we know that there is not enough evidence to reject null hypothesis. So we conclude that it is from Standard Normal Distribution.

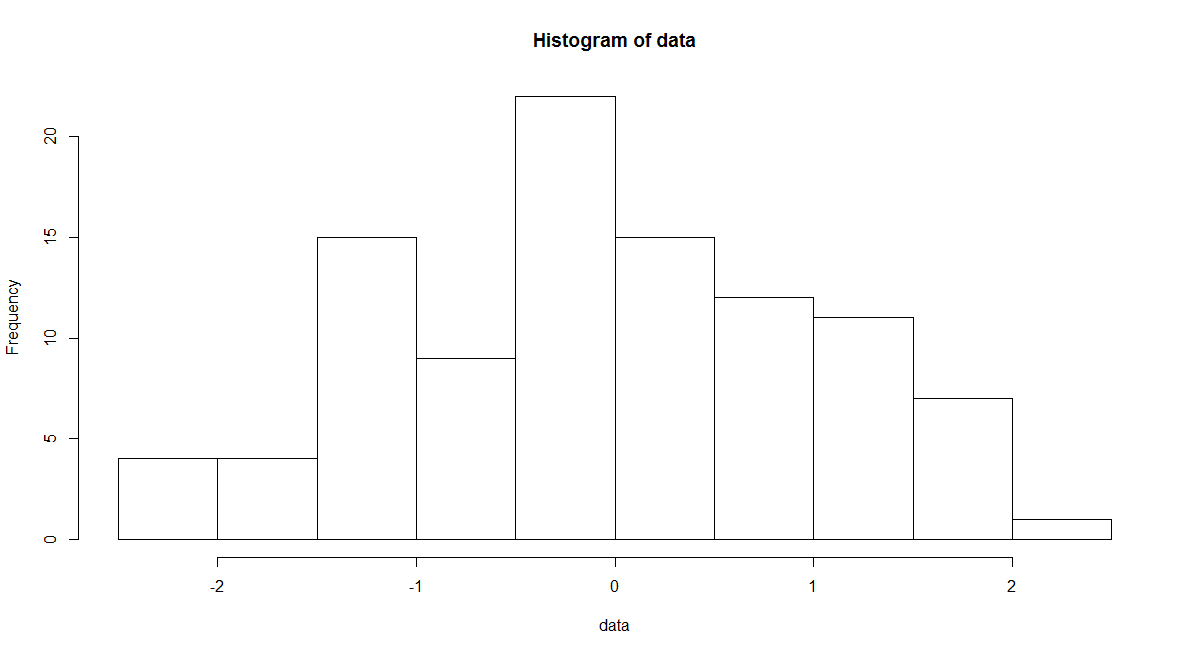
By examining QQ plot, we can also conclude that it is from Standard Normal Distribution



By making Uniform(-3,3), we can see that it is not from Uniform(-3, 3)



we can also see this fact from the histogram below



And by using chi-square test, the final result is 0, which supports our conclusion that it is not from Unifrom( -3, 3)

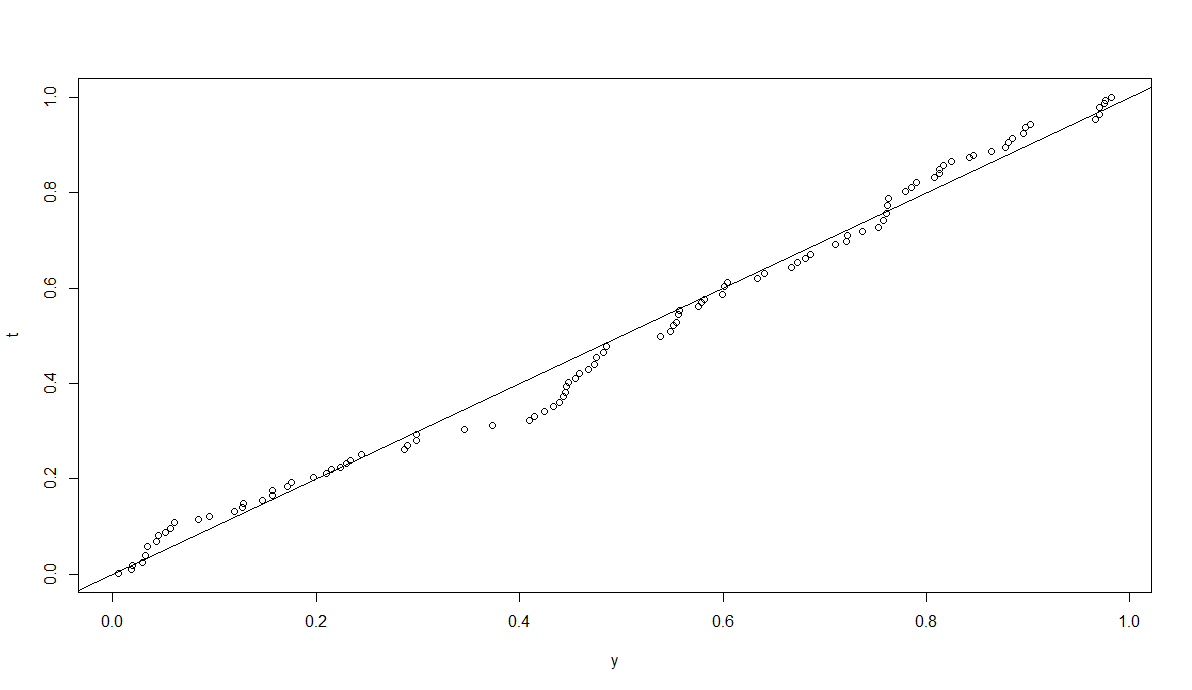
(C)

It’s possible. Because from the above analysis methods, we can only say that we don’t have enough evidence to reject null hypothesis but there is still possibility that the null hypothesis is wrong. So in that case, both of the null hypothesis can be accepted but the fact is that one of them, or even both of them, is wrong.

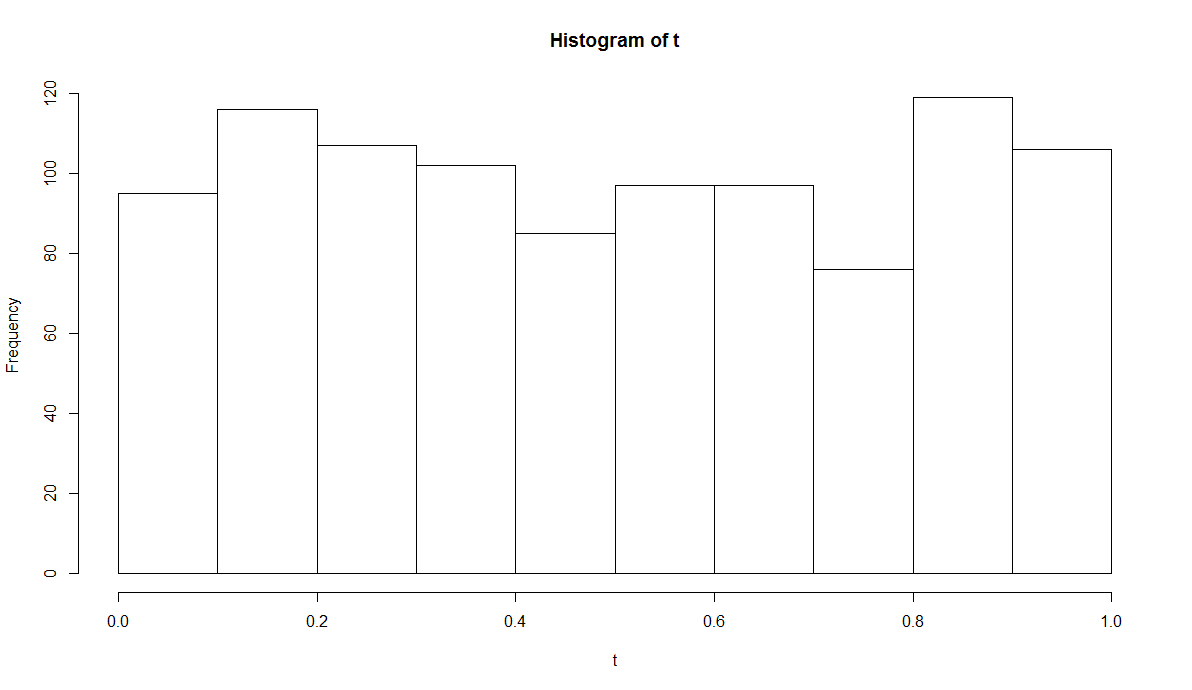
Exercise2:

By doing the chi-square test, the p-value we get is 0..6309(n=5), 0.6267(n=10), 0.6225(n=30), 0.7830(n=100). So we don’t have enough evidence to reject null hypothesis. So we conclude that the p-values follow Uniform(0,1)

Also by using QQ plot, we can see from below that it follow Uniform(0,1)



And from the histogram below, we can also conclude that it follows Uniform(0,1 )



#### Session3:

data = c(-2.434, -2.336, -2.192, -2.010, -1.967, -1.707, -1.678, -1.563, -1.476,-1.388, -1.331, -1.269, -1.229, -1.227, -1.174, -1.136, -1.127, -1.124, -1.120, -1.073, -1.052, -1.051, -1.032, -0.938, -0.884, -0.847, -0.846, -0.716, -0.644, -0.625, -0.588, -0.584, -0.496, -0.489, -0.473, -0.453, -0.427, -0.395, -0.386, -0.386, -0.373, -0.344, -0.280, -0.246, -0.239, -0.211, -0.188, -0.155, -0.149, -0.112, -0.103, -0.101, -0.033, -0.011, 0.033,0.110, 0.139, 0.143, 0.218, 0.218, 0.251, 0.261, 0.308, 0.343, 0.357, 0.463, 0.477, 0.48, 0.489, 0.545, 0.590, 0.638, 0.652, 0.656, 0.673, 0.772, 0.775, 0.776, 0.787, 0.969, 0.978, 1.005, 1.013, 1.039, 1.072, 1.168, 1.185, 1.263, 1.269, 1.297, 1.360, 1.370, 1.681, 1.721, 1.735, 1.779, 1.792, 1.881, 1.903, 2.009) **# data**

matrix( data, nrow = 10, byrow = TRUE ) **# format the data**

**#(A)using peason.test**

library(nortest)

pearson.test(data)

**#(A)using qqplot**

qqnorm( data )

qqline(data)

**#(B)using qqplot**

y = runif(100, -3,3)

qqplot(data, y)

abline(a=0,b=1)

**#(B)using chi-square test**

obs = table(cut( data,breaks=seq(from = min(data), to = max( data), length = 11),include.lowest=T,right=F))

exp = 100\* seq( 1/10, 10 )

pivet = sum(((obs-exp)^2)/exp)

p.value = 1- pchisq( pivet, df = 9)

**#(B)using histgram**

hist(data)

Exercise 2

t\_test = function( data, x0 ){ **# use two-sided t-test**

xbar = mean( data )

v = var( data )

n = length( data )

t = ( xbar - x0 )/ sqrt( v/ n )

return( t )

}

pvalue = function( tobs, n ){ **#get p-value**

result = 2\*( 1- pt( abs( tobs ), n-1 ) )

return( result )

}

exp = function( n, x0 ){ **#get a p- value from a singe experiment**

data = rnorm( n, 0, 1 )

t = t\_test( data, x0 )

result = pvalue( t, n )

return( result )

}

all\_exp = function( nsim, n ){ **#get p-value from each experiment**

result = replicate( nsim, exp( n, 0) )

return( result )

}

obs\_count = function( all\_exp ){**# get the observed count**

result = c( sum( all\_exp < 0.1), sum( all\_exp > 0.1 & all\_exp < 0.2), sum( all\_exp >0.2 & all\_exp < 0.3), sum( all\_exp> 0.3 & all\_exp< 0.4), sum(all\_exp > 0.4 & all\_exp< 0.5), sum(all\_exp>0.5 & all\_exp<0.6), sum( all\_exp > 0.6 & all\_exp < 0.7), sum( all\_exp >0.7 & all\_exp < 0.8), sum( all\_exp>0.8 & all\_exp < 0.9), sum( all\_exp > 0.9 & all\_exp < 1))

return( result )

}

exp\_count = function( nsim ){ **# get the expected count**

result = nsim\* rep( 1/10, 10)

return(result)

}

n = 100 **#This number should be change to see the effect of different n**

**##using p.value to see whether it is uniform(0,1)**

t = all\_exp( 1000, n)

obs = obs\_count(t)

exp = exp\_count( 1000)

pivet = sum( ( (obs-exp)^2 / exp ) )

p.value = 1- pchisq( pivet, df = 9)

**##using histgram to show whether it is uniform(0,1)**

hist( t )

**##using qqplot to show whether it is uniform**

y = runif( 100,0, 1)

qqplot( y, t )

abline(a=0, b= 1)