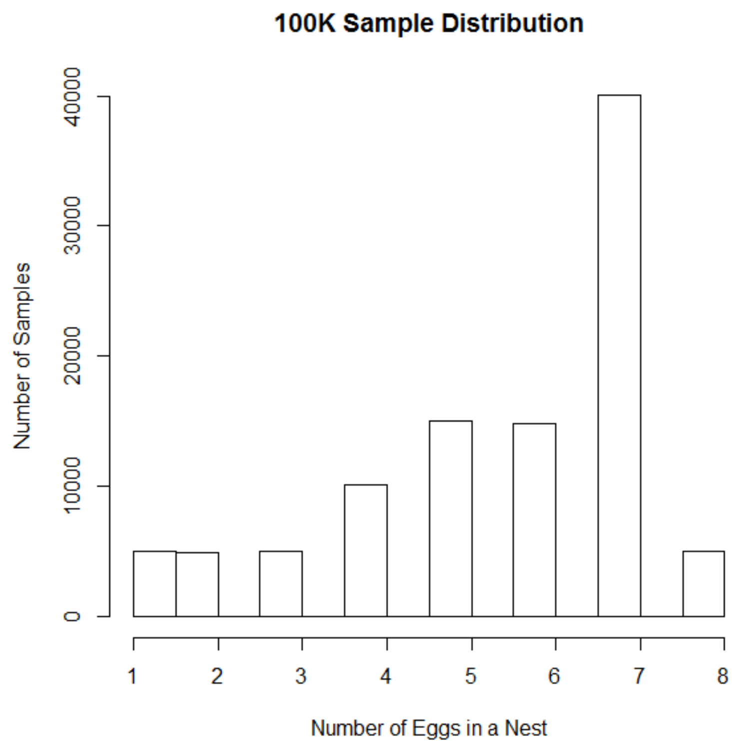
Derek Yang

Stats 67 Project 2

1. Mean of Eggs: 5.55

   Standard Deviation: 1.856744

2. Histogram of a sample size of 100,000 from same distribution



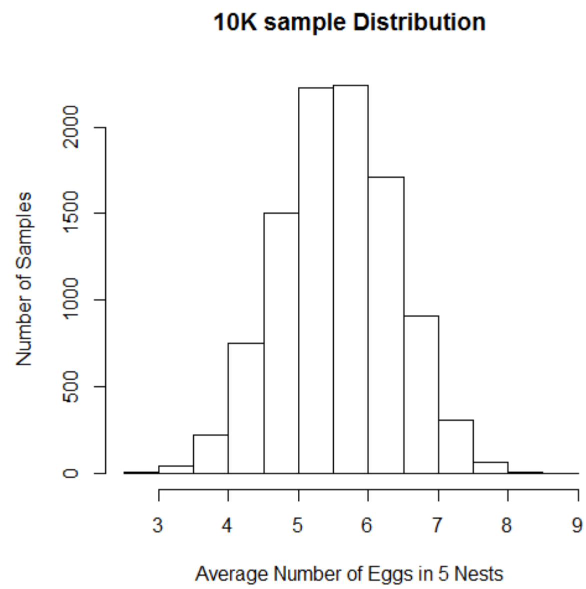
3.

   (a) A t-test is used for a normal distribution with a small number of samples, but we do not know if this distribution is normal or not.

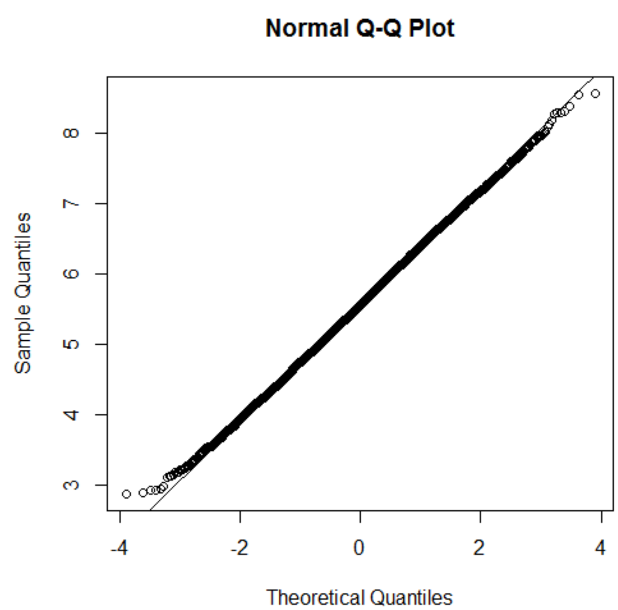
   (b) A z-test is used to approximate the distribution of sample means of a normal distribution, but we do not know that this distribution is normal and it is too small to assume normality.

ID: 63118832

(c) Histogram of a sample size of 10,000



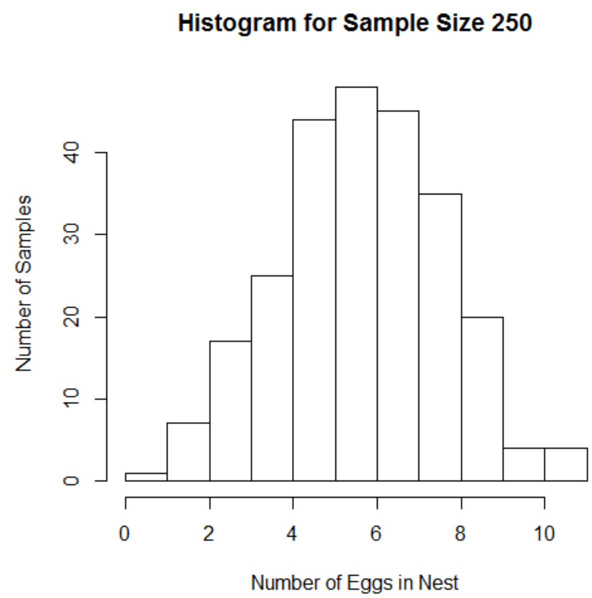
(d) A qqplot() of above distribution and it does look normal.

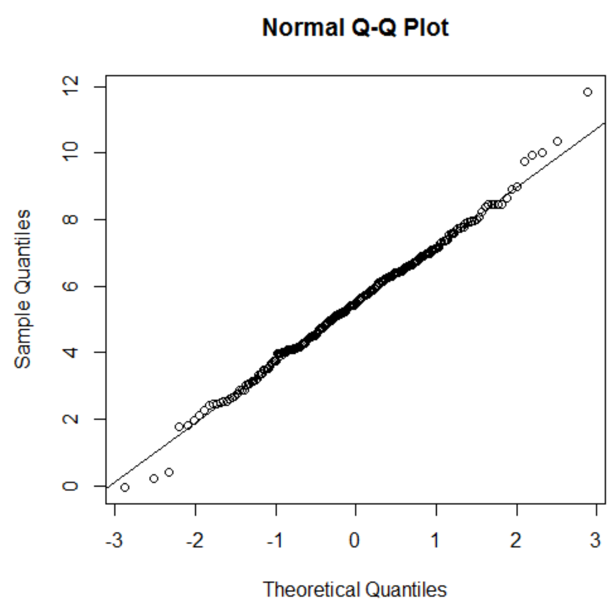


(e) It is unusual to have an average of 7.2 eggs per nest given 7.2 is not even on the confidence interval.

(f) 95% Confidence Interval: (3.929926,7.184942) .
This confidence interval is pretty wide, meaning that the sample size is small.

4. N=250





(a) I would use z-distribution because my standard deviation is not unknown and my sample size is greater than 30.

(b) Average Confidence Interval of 10,000: (4.984603, 6.115338)

i. 9,500 would contain the true mean.

ii. True Mean (1) is contained 9,454 times.

(c)

i. We expect to reject the null about 500 times.

ii. We rejected the null 491 times.

iii. I think that a Type I error has a small chance of occurring in our experiment data because it was computed by a computer. I have created conditions in my hypothesis tests to ensure that if the null hypothesis should be rejected, it is rejected, and if it should not be rejected, it will not be rejected. However if there were to be Type I errors, there would be 491 of them.

iv. I think that Type II error is highly unlikely because the mean of the 10,000 trials should have their means around the true mean. There should be no Type II errors.

(d) Central Limit Theorem

Appendix

```
#Derek Yang
#ID:63118832
#STATS 67
#Project 2

#1.
#Expected Value
e<-sum(c(1*.05,2*.05,3*.05,4*.1,5*.15,6*.15,7*.4,8*.05))
e
#Standard Deviation
sd<-sqrt(sum(((1-e)^2)*.05,((2-e)^2)*.05,((3-e)^2)*.05,((4-e)^2)*.1,((5-e)^2)*.15,((6-
e)^2)*.15,((7-e)^2)*.4,((8-e)^2)*.05))
sd

#2.
s100k<-sample(seq(1:8), size = 100000, replace = T,
prob=c(.05,.05,.05,.1,.15,.15,.4,.05))
hist(s100k,main="100K Sample Distribution",xlab="Number of Eggs in a
Nest",ylab="Number of Samples")

#3.
##c.
s10k<-rep(NA,10000)
for(i in 1:10000)
  s10k[i]<-mean(rnorm(5,5.55,1.856744))
hist(s10k,main="10K sample Distribution",xlab="Average Number of Eggs in 5
Nests",ylab="Number of Samples")

##d
qqnorm(s10k)
qqline(s10k)

##f
pci10k<-rep(NA,10000)
nci10k<-rep(NA,10000)
for(i in 1:10000){
  ci10k<-1.96*(1.856744/sqrt(5))
  pci[i]<-mean(s10k)+ci10k
  nci[i]<-mean(s10k)-ci10k
}
mean(pci)
mean(nci)

#4
s250<-rnorm(250,mean=5.55,sd=1.856744)
hist(s250,main="Histogram for Sample Size 250", xlab="Number of Eggs in Nest",
ylab="Number of Samples")
qqnorm(s250)
qqline(s250)

##b###ii
count<-0
nci<-rep(NA,10000)
pci<-rep(NA,10000)
for(i in 1:10000){
y<-rnorm(250,mean=5.55,sd=1.856744)
ci<-1.96*(1.856744/sqrt(250))
pci[i]<-mean(y)+ci
nci[i]<-mean(y)-ci
if(nci[i] <= 5.55 && pci[i] >=5.55)
  count<-count+1
}
mean(nci)
mean(pci)
count
```

```
##c
nreject<-0
reject<-0
for(i in 1:10000){
  hyptest<-mean(rnorm(250,mean=5.55,sd=1.856744))
  ztest<-(hyptest-5.55)/(1.856744/sqrt(250))
  if(ztest>1.96 | ztest< -1.96)
    {reject<-reject+1}
  else
    {nreject<-nreject+1}
}
reject
nreject
```