Siwei Zhang(Roger)
20335254
sep 9

HW2

1. I understand the program that simulates it by generating the sampling distribution of the sample means.

2. The probability is 0.1827883 .

2a. We can not use this information to compute the theoretical probability of getting a sample mean smaller than 7. Because the sample size 9 isn't large enough, so the results based on that will lead to imprecise.

3. The probability is 0.00844774 .

3a. We can use this information to compute the theoretical probability of getting a sample mean smaller than 7 and the theoretical probability is 0.008197536 . Because the sample size 64 is large enough to calculate the theoretical results.

4b. 95%

5b.

df	P(t < -1)	P(Z < -1)	-t*	t*
5	0.1816	0.1587	-2.57	+2.57
10	0.1704	0.1587	-2.23	+2.23
100	0.1599	0.1587	-1.98	+1.98

5c. 39.49-55.45

5d. If know that sd=9, then the interval will be 41.89-50.29. Result shows that interval with sd=9 is smaller than using the t distribution. Because using t distribution, we need to estimate, in other word, approximate calculate the confidence interval, but we can use sd to calculate the theoretical confidence interval.

If the sample size become 100. The confidence intervals will be 47.73-51.00 and 47.606-51.134. They are really close.

If the sample size are small(i.e. 1,2,3,4), the t interval will larger than the z interval. With the sample size become larger and larger, the t and z intervals will become very close to each other.

5e. The confidence interval become larger.

6. I run this with sample size equals 10 and 100.

sample size	interval-t.test	interval-sec
10	40.79-52.18	41.55-51.42
100	46.30-49.84	46.32-49.82

The results show that there is a slightly different between two confidence intervals when size=10 which isn't large enough, and with the size goes larger, for example size = 100, the two confidence intervals are almost same.

7.

```
#lets just use the normal distribution in the last question
#N(50,9)  mean=50, sd=9, var=81
#we will try to get sample variance in different sample size 10,100,1000,10000
for(n in c(10,100,1000,10000)){
 samp<-rnorm(n,50,9)
 s_samp<-sqrt(sum((samp-mean(samp))^2)/(n-1))#sqrt((Xi-x)^2/n-1)
 print(s_samp^2)
}
#the results of variances are
#10,    100,    1000,     10000
#100.12 83.87   82.22     80.78

#then we try it with exponent distribution
for(n in c(10,100,1000,10000)){
  samp<-rexp(n,rate=.1)
  s_samp<-sqrt(sum((samp-mean(samp))^2)/(n-1))#sqrt((Xi-x)^2/n-1)
  print(s_samp^2)
}
#the variance should be around 100
#the results are
#10,    100,    1000,     10000
#95.99  90.05   112.40    101.87

#the results show that in both shapes of distributions.
#with the size become larger,
#the variance of sample is more close to the variance of population
```