Setu Namburu - Extra Credit Problem #1 (5 points)

This problem illustrates quartile calculations using random samples of different sizes from the standard normal distribution.

Use *set.seed(1237)* and *rnorm(n, mean = 0, sd = 1)* with *n* = 10, *n* = 30, *n* = 100 and *n* = 300 to draw four different random samples from the standard normal distribution. Reset *set.seed(1237)* prior to drawing each of the four samples.

For each sample, calculate the first, second and third quartile using *quantile()*. Use "type = 2" (method used in Business Statistics) and "type = 7" (R default) and generate quartiles for each.

Display the results. The quartiles for the standard normal distribution are -0.6745, 0.0 and +0.6745 as shown using *qnorm(c(0.25, 0.5, 0.75), mean = 0, sd = 1, lower.tail = TRUE)*. Note below.

```
qnorm(c(0.25, 0.5, 0.75), mean = 0, sd = 1, lower.tail = TRUE)
```

```
## [1] -0.6744898  0.0000000  0.6744898
```

Take note of the results for the first and third quartile. Compare the computed results between the two methods (type = 2 and type = 7). Comment on the rate of convergence for these estimates as the sample size is increased. What does this exercise indicate about describing a population distribution with samples?

```
# Add your set.seed(), rnorm() and quantile() code to this code 'chunk':
##Function for calculating quartiles using type =2 and type=7

x.quantile<-function(n){
  set.seed(1237)
  p <- c(0.25, 0.5, 0.75)
  x<-rnorm(n,mean=0,sd=1)
  cat("\n Quartiles for standard normal distribution using Type2 and n = ",n, ":",quantile(x,
probs = p, na.rm = FALSE, names = TRUE, type = 2))
  cat("\n Quartiles for standard normal distribution using Type7 and n = ",n, ":", quantile(x,
 probs = p, na.rm = FALSE, names = TRUE, type = 7))
}

###4 sample sizes
s<-c(10,30,100,300)

###Quartiles for 4 normal distribution samples
for (n in s){
x.quantile(n)
}
```

```
##
##  Quartiles for standard normal distribution using Type2 and n =  10 : -0.9256115 -0.376204
0.858124
##  Quartiles for standard normal distribution using Type7 and n =  10 : -0.9082042 -0.376204
0.6520158
##  Quartiles for standard normal distribution using Type2 and n =  30 : -0.8559826 0.03138278
```

```
 0.858124
##  Quartiles for standard normal distribution using Type7 and n =  30 : -0.7944214 0.03138278
 0.821137
##  Quartiles for standard normal distribution using Type2 and n =  100 : -0.7695685 0.1652762
 0.6962258
##  Quartiles for standard normal distribution using Type7 and n =  100 : -0.7477238 0.1652762
 0.6892506
##  Quartiles for standard normal distribution using Type2 and n =  300 : -0.6287819 0.0047916
36 0.6218793
##  Quartiles for standard normal distribution using Type7 and n =  300 : -0.6278615 0.0047916
36 0.617315
```

The first and third quartile estimates differ between the two methods (type 2 and type 7) at lower sample sizes. As the sample size increases, the estimates do not differ much between both the methods. So even though there are several methods to estimate the quartiles, as the sample size increases all of them tend to provide same results.

As the sample size increases, the 1st and 3rd quartile estimates get closer to the quartiles of standard normal distribution which is the population (they converge closer to the population parameters (1st and 3rd quartiles): -0.6745 and +0.6745). So the more samples we have the better chance we have at estimating population parameters (1st and 3rd quartiles in this case).

4 points Be sure to answer all the questions.

Regarding the second question, (What does this exercise indicate about describing a population distribution with samples?) It is important to make sure the sample size is large enough for your purposes to adequately represent the population. Sample size affects the precision with which parameters are estimated.