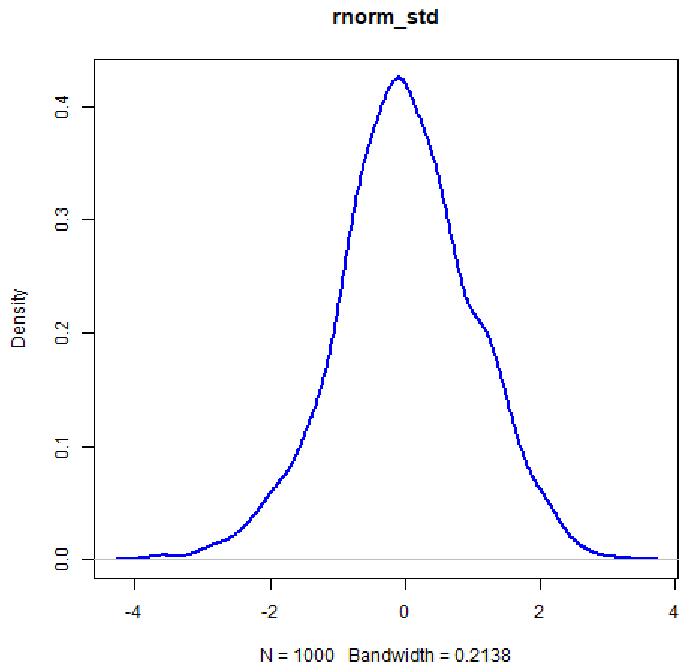
BACS HW3
Student ID: 110065508

Due: 2022/03/06

**Question 1.** (a) According to basic statistic theory, `rnorm_std` should have zero mean and standard deviation 1. The distribution should be bell-shaped. We called this **Z-distribution**. This can be verified by the following code

```
# Question 1 (a)
rnormal <- rnorm(n=1000, mean=940, sd=190)
rnorm_std <- (rnormal - mean(rnormal))/sd(rnormal)
cat("mean of rnorm_std=", mean(rnorm_std), " sd=", sd(rnorm_std), "\n")

png(filename = "1a.png")
plot(density(rnorm_std), col="blue", lwd=2, main = "rnorm_std") # plot pdf
dev.off()
```

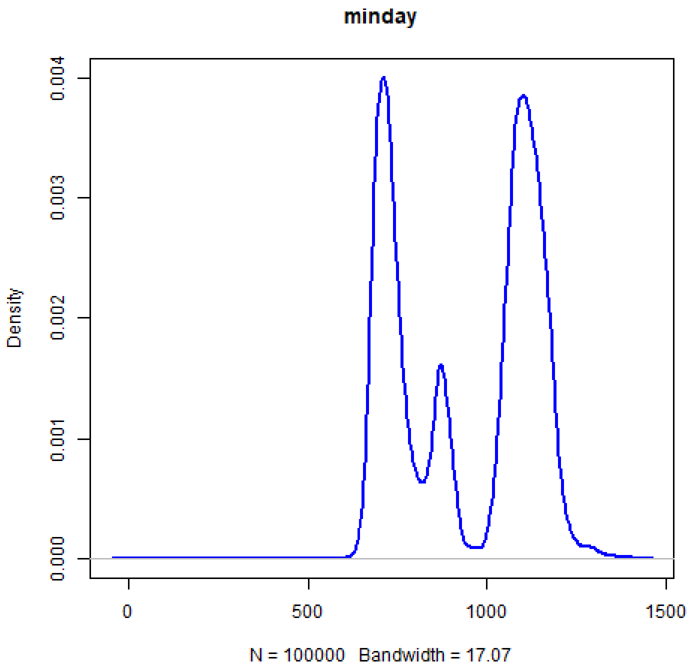
The code returns mean of `rnorm_std=-1.684449e-17, sd=1`, which meets our expectation.

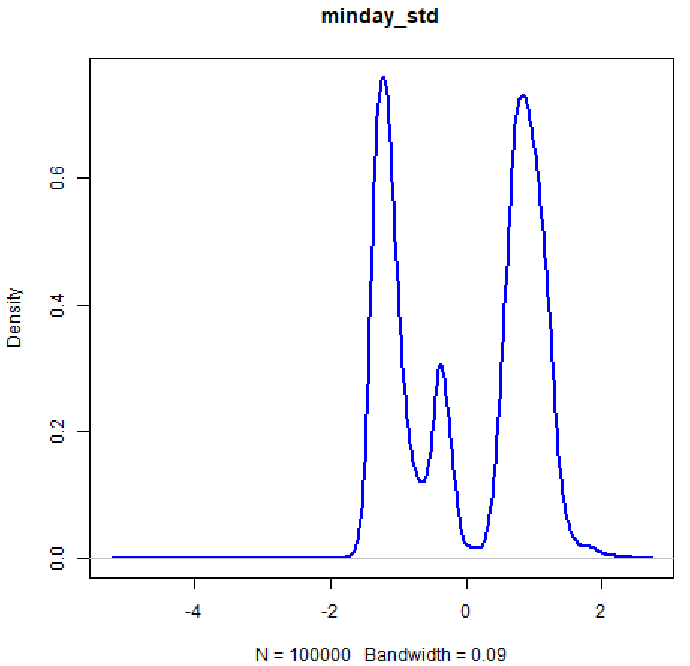



(b) The mean and standard deviation of `minday_std` should be 0 and 1. That's because standardization is a kind of "scaling" and "translation" which only depends on sample itself, not the kind of distribution it comes from. Hence, speaking of shapes, the distribution of `minday_std` should be same compared to `minday`. This can be verified by the following codes:

```
# Question 1 (b), Also partial code of Q3
bookings <- read.table("first_bookings_datetime_sample.txt", header=TRUE)
bookings$datetime[1:9]
hours <- as.POSIXlt(bookings$datetime, format="%m/%d/%Y %H:%M")$hour
mins <- as.POSIXlt(bookings$datetime, format="%m/%d/%Y %H:%M")$min
minday <- hours*60 + mins
plot(density(minday), main="Minute (of the day) of first ever booking",
     col="blue", lwd=2)

minday_std <- (minday - mean(minday))/sd(minday)
cat("mean of minday_std=", mean(minday_std), " sd=", sd(minday_std), "\n")

png(filename = "1b_1.png")
plot(density(minday), col="blue", lwd=2, main = "minday") # plot pdf
dev.off()
```

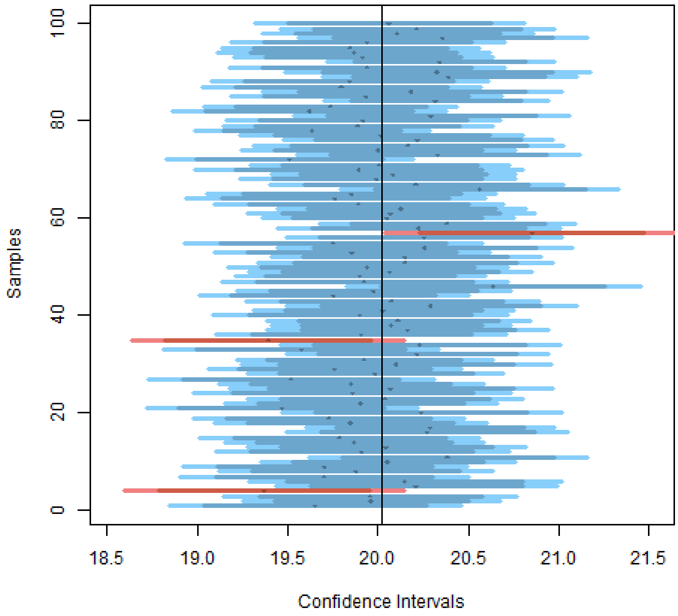
```r

png(filename = "1b_2.png")
plot(density(minday_std), col="blue", lwd=2, main = "minday_std") # plot pdf
dev.off()
```



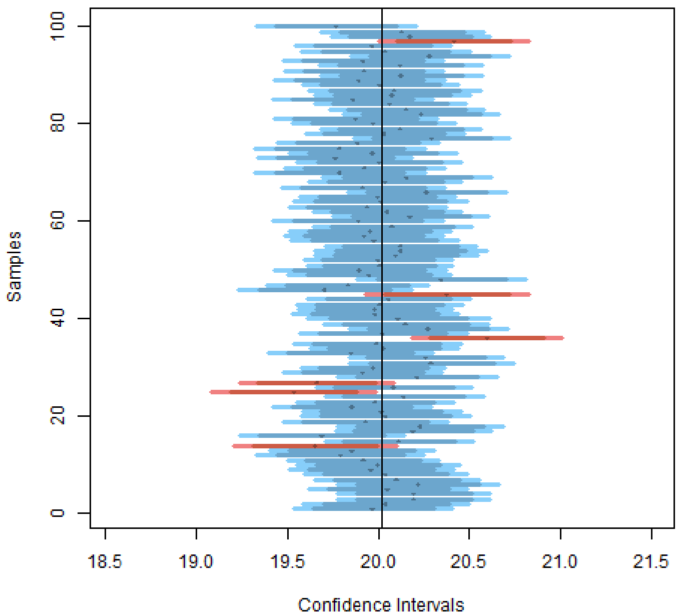

**Question 2.** (a) We expect 5 and 1 samples are not include in 95% and 99% CI, respectively. Of course, that might not be same as simulated. As the following code shows:

```r
# Question2 (a)
png(filename = "2a.png")
visualize_sample_ci(num_samples = 100, sample_size = 100, pop_size=10000,
                    distr_func=rnorm, mean=20, sd=3)
dev.off()
```




(b) Since 300 is 3 times of 100, we expect 15 and 3 samples are not include in 95% and 99% CI, respectively.

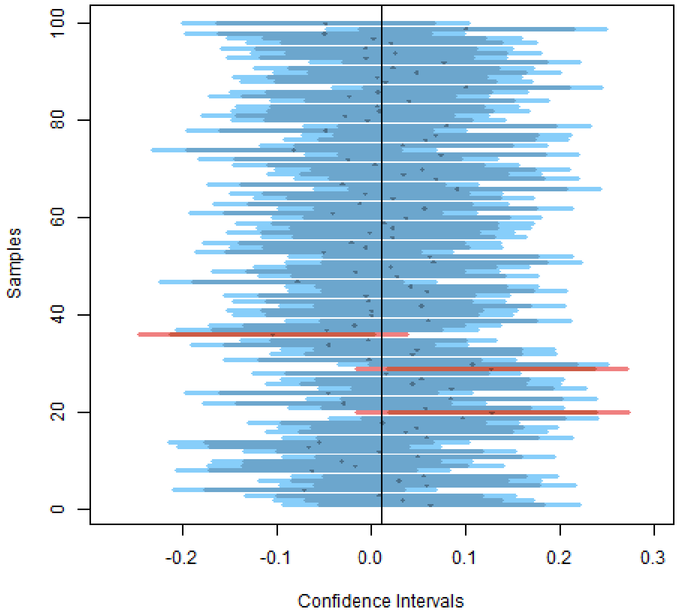
```r
# Question2 (b)
png(filename = "2b.png")
visualize_sample_ci(num_samples = 100, sample_size = 300, pop_size=10000,
                    distr_func=rnorm, mean=20, sd=3)
dev.off()
```

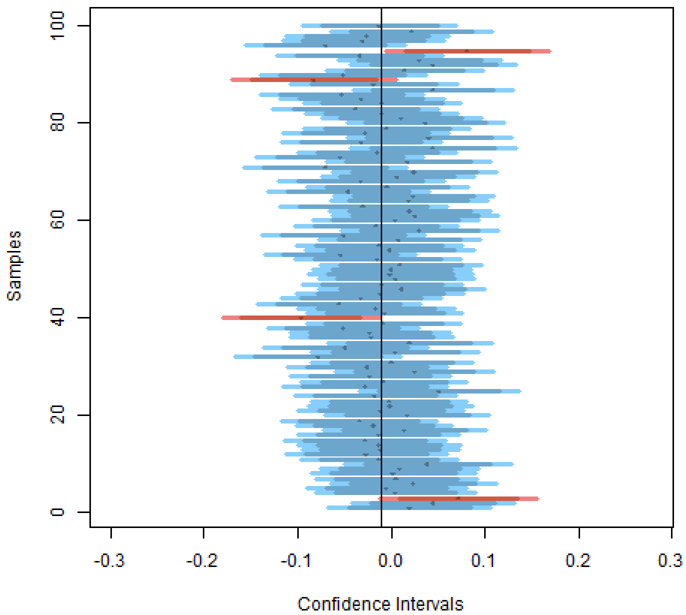



(c) Notice that by the CLT, the mean of any distribution converges to a normal random variable. So the answer is unlikely to change compared to (a) and (b). Notice that the sample size is 100 so we are confidant to say that.

```r
# Question 2 (c)
png(filename = "2c_1.png")
visualize_sample_ci(num_samples = 100, sample_size = 100, pop_size=10000,
                    distr_func=runif, min = -1, max = 1)
dev.off()
```

```r
png(filename = "2c_2.png")
visualize_sample_ci(num_samples = 100, sample_size = 300, pop_size=10000,
                    distr_func=runif, min = -1, max = 1)
dev.off()
```




**Question 3.** (a) This can be done by the following code:

```r
# Question 3

compute_sample_mean <- function(sample0) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  mean(resample)
}

plot_resample_density <- function(sample_i) {
  lines(density(sample_i), col=rgb(0.0, 0.4, 0.0, 0.01))
  return(mean(sample_i))
```

```
}

# Question 3 (a)
#(i)
ci95_trad <- mean(minday) + c(-1.96, 1.96)*sd(minday)
cat("mean of minday=", mean(minday), " sd of minday=", sd(minday),
    " ci95=", ci95_trad, "\n")

#(ii)
resamples <- replicate(2000, sample(minday, length(minday), replace=TRUE))
sample_means <- apply(resamples, 2, FUN=plot_resample_density)

#(iii)
png(filename = "3a.png")
plot(density(sample_means), lwd=0, main="bootstrapped samples")
dev.off()

#(iv)
ci95_boot <- mean(resamples) + c(-1.96, 1.96)*sd(resamples)
cat("mean of resamples=", mean(resamples), " sd of resamples=", sd(resamples),
    " ci95_boot=", ci95_boot, "\n")
```

We have:

(i) The population mean of `minday` is 942.4964, its standard error is 189.6631, and the 95% CI of the sampling means is [570.75671314.236].

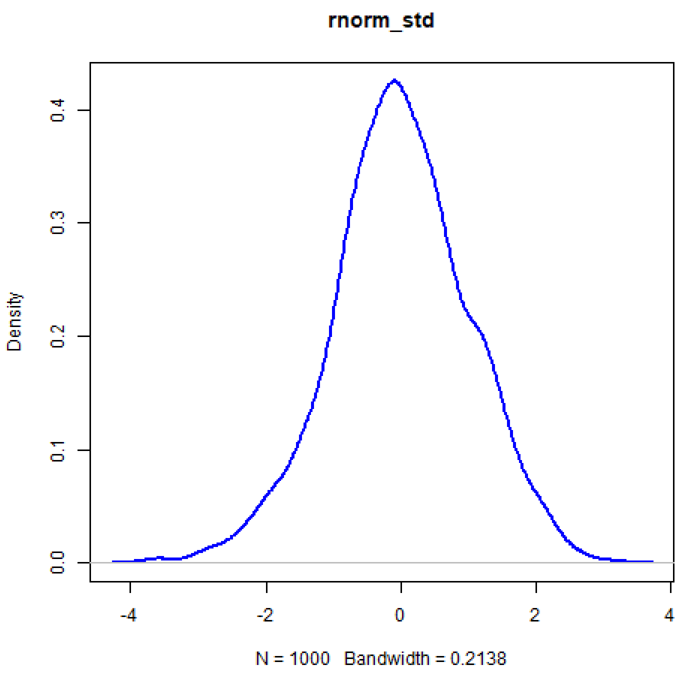
(iv) The 95% CI of the bootstrapped means is [570.79111314.232].

(b) This can be done by the following code:

```
# Question 3 (b)
# (i) (ii)
cat(" median of minday =", median(minday), "\n")
png(filename = "3b.png")
## Distribution of sampling
plot(density(resamples), lwd=2, xlim=c(0, 400))
## Confidence intervals of the sampling means
abline(v=median(resamples), lwd=2)
dev.off()

# (iii)
quantile(median(resamples), probs=c(0.025, 0.975))
```

(i) The median of `minday` is 1040

(iii) The 95% CI of the bootstrapped median is


rnorm_std
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
N = 1000   Bandwidth = 0.2138