

# BACS-hw03-107070004

## Question 2)

Copy and run the code we used in class to create simulations of confidence intervals. Run `visualize_sample_ci()`, which simulates samples drawn randomly from a population. Each sample is a horizontal line with a dark band for its 95% CI, and a lighter band for its 99% CI, and a dot for its mean. The population mean is a vertical black line. Samples whose 95% CI includes the population mean are blue, and others are red.

```
# Visualize the confidence intervals of samples drawn from a population
visualize_sample_ci <- function(num_samples = 100, sample_size = 100,
                                pop_size=10000, distr_func=rnorm, ...) {
  # Simulate a large population
  population_data <- distr_func(pop_size, ...)
  pop_mean <- mean(population_data)
  pop_sd <- sd(population_data)

  # Simulate sample
  samples <- replicate(num_samples,
                      sample(population_data, sample_size, replace=FALSE))

  # Calculate descriptives of samples
  sample_means = apply(samples, 2, FUN=mean)
  sample_stdevs = apply(samples, 2, FUN=sd)
  sample_stderrs <- sample_stdevs/sqrt(sample_size)
  ci95_low <- sample_means - sample_stderrs*1.96
  ci95_high <- sample_means + sample_stderrs*1.96
  ci99_low <- sample_means - sample_stderrs*2.58
  ci99_high <- sample_means + sample_stderrs*2.58

  # Visualize confidence intervals of all samples
  plot(NULL, xlim=c(pop_mean-(pop_sd/2), pop_mean+(pop_sd/2)),
       ylim=c(1,num_samples), ylab="Samples", xlab="Confidence Intervals")
  add_ci_segment(ci95_low, ci95_high, ci99_low, ci99_high,
                sample_means, 1:num_samples, good=TRUE)

  # Visualize samples with CIs that don't include population mean
  bad = which(((ci95_low > pop_mean) | (ci95_high < pop_mean)) |
             ((ci99_low > pop_mean) | (ci99_high < pop_mean)))
  add_ci_segment(ci95_low[bad], ci95_high[bad], ci99_low[bad], ci99_high[bad],
                sample_means[bad], bad, good=FALSE)

  exclude_ci95<-which((ci95_low > pop_mean) | (ci95_high < pop_mean))
  exclude_ci99<-which((ci99_low > pop_mean) | (ci99_high < pop_mean))
  return(list(ci95_not = length(exclude_ci95), ci99_not = length(exclude_ci99), ci95_width=mean(abs(ci95
```

```

# Draw true population mean
abline(v=mean(population_data))
}

add_ci_segment <- function(ci95_low, ci95_high, ci99_low, ci99_high,
                           sample_means, indices, good=TRUE) {
  segment_colors <- list(c("lightcoral", "coral3", "coral4"),
                        c("lightskyblue", "skyblue3", "skyblue4"))
  color <- segment_colors[[as.integer(good)+1]]

  segments(ci99_low, indices, ci99_high, indices, lwd=3, col=color[1])
  segments(ci95_low, indices, ci95_high, indices, lwd=3, col=color[2])
  points(sample_means, indices, pch=18, cex=0.6, col=color[3])
}

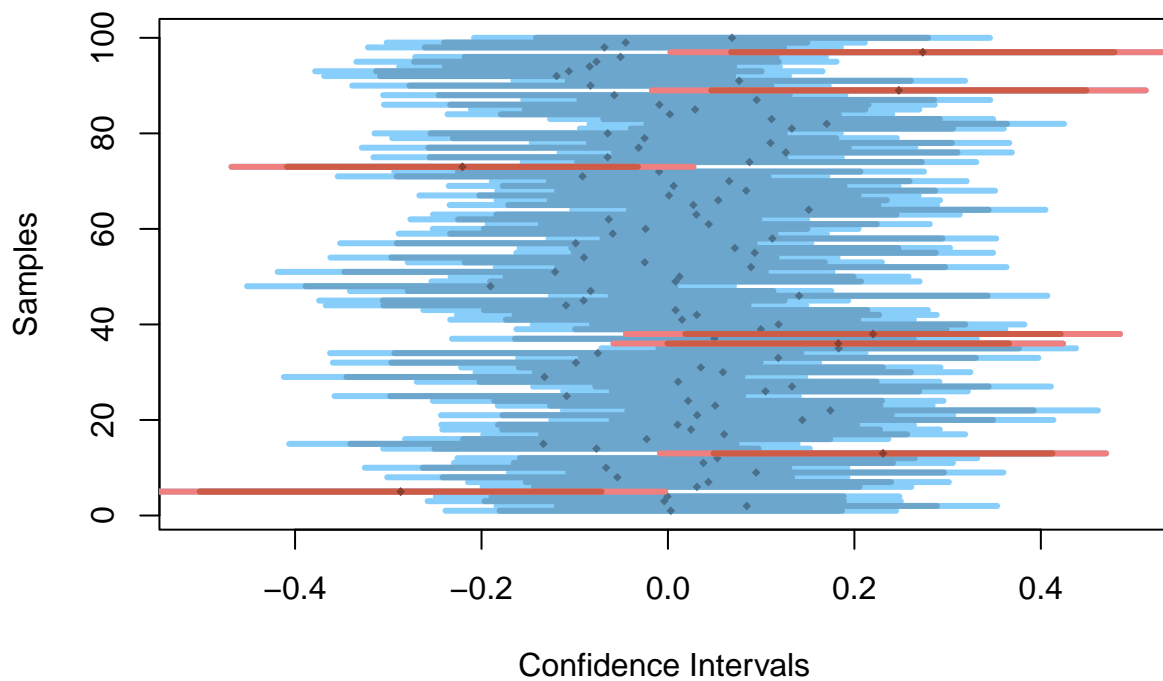
```

a) Simulate 100 samples (each of size 100), from a normally distributed population of 10,000:

```

ans<-visualize_sample_ci(num_samples = 100, sample_size = 100, pop_size=10000,
distr_func=rnorm)

```



i) How many samples do we expect to NOT include the population mean in its 95% CI?

```
print(ans$ci95_not)
```

```
## [1] 7
```

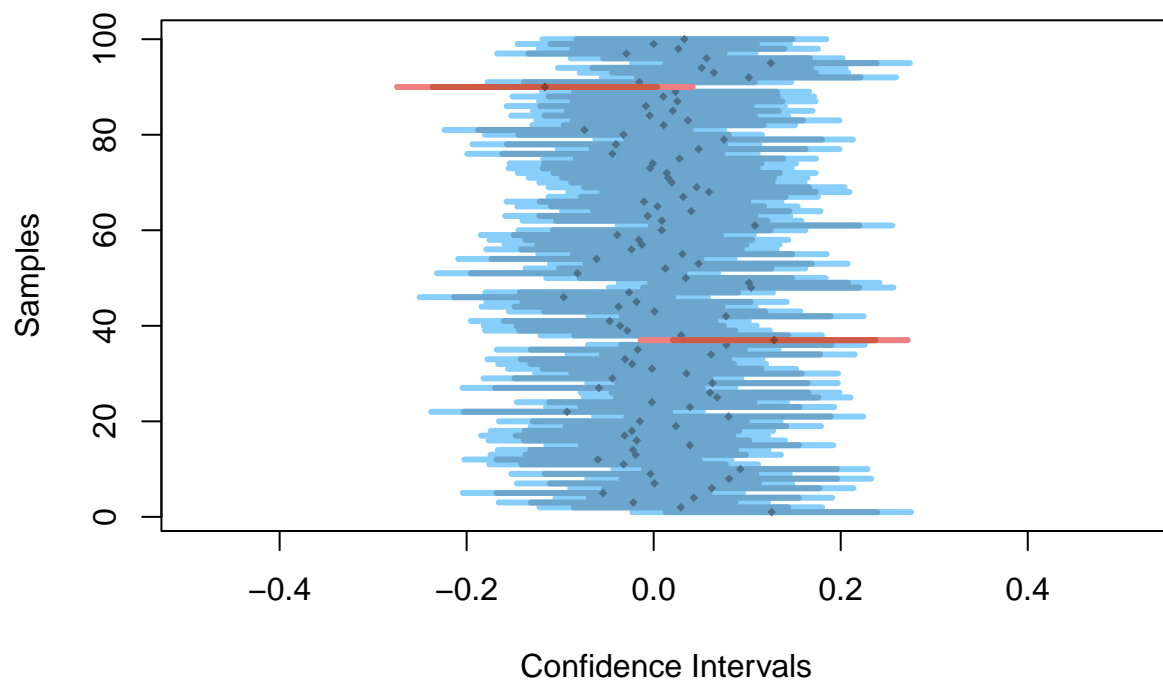
ii) How many samples do we expect to NOT include the population mean in their 99% CI?

```
print(ans$ci99_not)
```

```
## [1] 1
```

b) Retrun the previous simulation with the same number of samples, but larger sample size (sample\_size=300):

```
ans <- visualize_sample_ci(num_samples = 100, sample_size = 300, pop_size=10000,  
distr_func=rnorm)
```



i) Now that the size of each sample has increased, do we expect their 95% and 99% CI to become wider or narrower than before?

narrower than before.

```
cat("size=100", c(ci95_width_old, ci99_width_old), "size=300", c(ans$ci95_width, ans$ci99_width))  
  
## size=100 0.3908968 0.5145479 size=300 0.2269754 0.2987737
```

ii) This time, how many samples (out of the 100) would we expect to NOT include the population mean in its 95% CI?

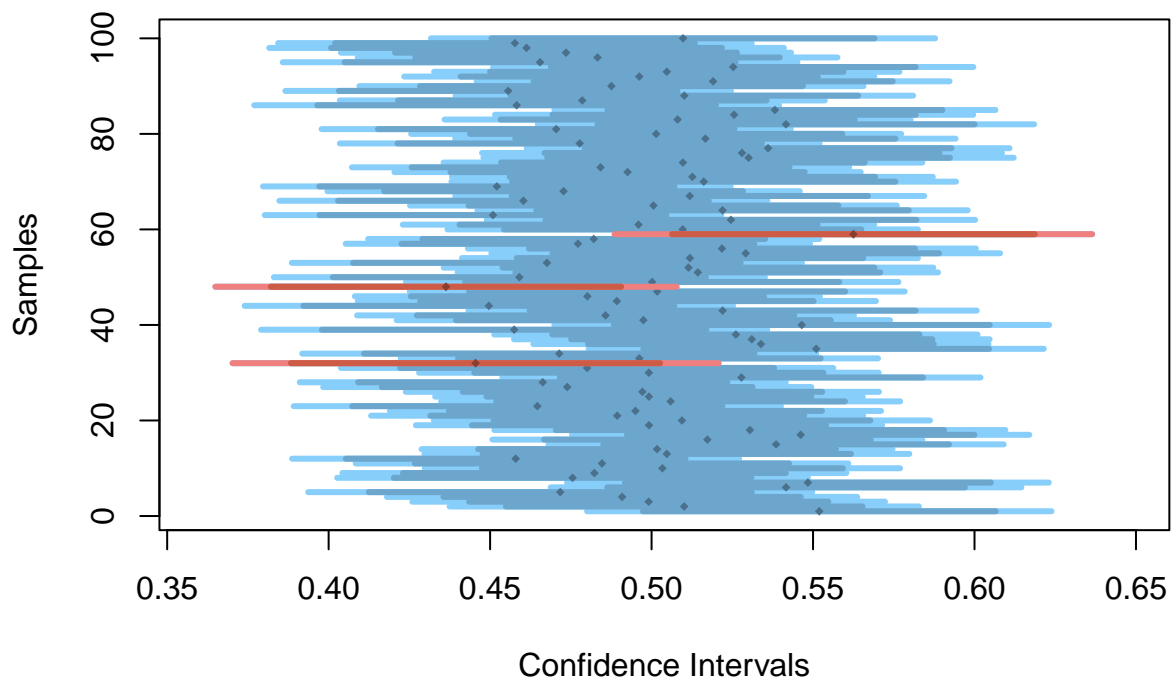
```
print(ans$ci95_not)
```

```
## [1] 2
```

c) If we ran the above two examples (a and b) using a uniformly distributed population (specify `distr_func=runif` for `visualize_sample_ci`), how do you expect your answers to (a) and (b) to change, and why?

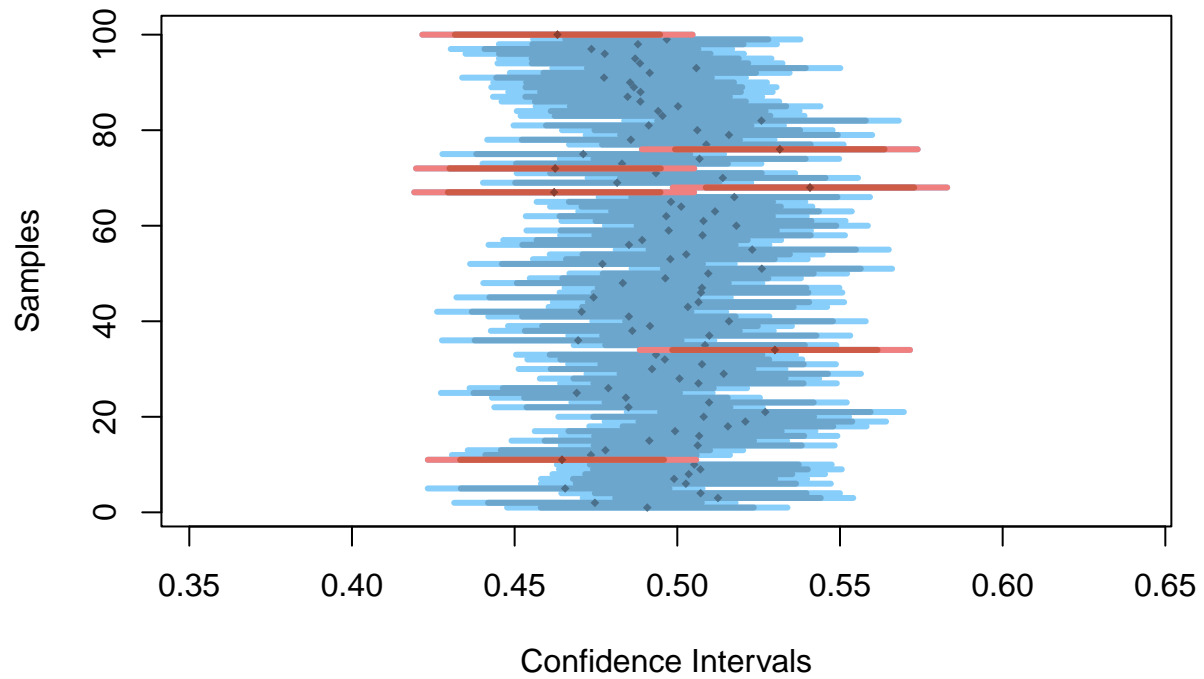
for (a) Become narrower.

```
ans_a <- visualize_sample_ci(num_samples = 100, sample_size = 100, pop_size=10000,  
distr_func=runif)
```



for (b)

```
ans_b <- visualize_sample_ci(num_samples = 100, sample_size = 300, pop_size=10000,  
distr_func=runif)
```



Compared to (a) and (b), they both have wider width when sample\_size is 100, and have narrower width when sample\_size is 300.