# Central Limit Theorem - Monte Carlo

### clear working memory

```
rm(list=ls())
```

### loading libraries

```
##
## Attaching package: 'rmutil'

## The following object is masked from 'package:stats':
##
##     nobs

## The following objects are masked from 'package:base':
##
##     as.data.frame, units

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x tidyr::nesting() masks rmutil::nesting()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#Suppose that x is drawn from the following "mixing distribution." Let y be a binary random variable wi

y <- ifelse(runif(100000) < 0.90, 1, 0)

x <- ifelse(y == 1, rnorm(100000), rnorm(100000, mean = 100, sd = 20))

print(paste("Mean of x: ", mean(x)))
```

```
## [1] "Mean of x:  9.94409115040783"
```

```r
#b) For this distribution, use 10,000 draws from each of the following sample sizes: n = 36, n = 64, n 

generate_simulated_means <- function(N){
   # Generate the mean and standard deviations of N observations from the specified distribution functi

   # y is a binary random variable with Pr(y = 1) = 0.9.
   # If y = 1, then x is drawn from a standard normal distribution.
   # If y = 0, then x is drawn from a normal distribution with mean = 100 and standard deviation  = 20.
   y <- ifelse(runif(N) < 0.90, 1, 0)
   x <- ifelse(y == 1, rnorm(N), rnorm(N, mean = 100, sd = 20))

   # put data into data_frame so it is easier to summarize
   data <- tibble(y, x)

   # get the means for each column
   means <- sapply(data, mean)

   # name the means appropriately
   names(means) <- c("muy", "mux")

   # get the sds for each column
   sds <- sapply(data, sd)

   # name the sds appropriately
   names(sds) <- c("sdy", "sdx")

   # return the means and standard deviation associated with sample x of size N.
   return(c(means, sds))
   }



get_zscores <-function(obs_mean, true_mean, obs_sd, N){
  zscores <- (obs_mean - true_mean) / (obs_sd / sqrt(N))
  return( zscores )
}

significance_test <- function(zscores, alpha){
  beyond_critical_point <- as.numeric( zscores > alpha | zscores < -alpha )
  percent_significantly_different <- mean( beyond_critical_point )
  return( percent_significantly_different )
}



monte_carlo <- function(N, reps = 10000){

  replicated_sims <- replicate(reps, generate_simulated_means(N))

  expected_mu_y <- 0.9
  expected_mu_x <- 10 # Derived from 0.9*0 + 0.1*100
```

```r
  z_y <- get_zscores(replicated_sims['muy', ], expected_mu_y, replicated_sims['sdy', ], N)


  sig1_y <- significance_test(z_y, 0.025)
  print(paste("Percentage of simulated means which were significantly different from"))
  print(paste("sampling distribution at critical point 0.025:", sig1_y))
  print(paste("                    "))

  sig2_y <- significance_test(z_y, 0.975)
  print(paste("Percentage of simulated means which were significantly different from"))
  print(paste("sampling distribution at critical point 0.975:", sig2_y))
  print(paste("                    "))



  z_x <- get_zscores(replicated_sims['mux', ], expected_mu_x, replicated_sims['sdx', ], N)


  sig1_x <- significance_test(z_x, 0.025)
  print(paste("Percentage of simulated means which were significantly different from"))
  print(paste("sampling distribution at critical point 0.025:", sig1_x))
  print(paste("                    "))

  sig2_x <- significance_test(z_x, 0.975)
  print(paste("Percentage of simulated means which were significantly different from"))
  print(paste("sampling distribution at critical point 0.975:", sig2_x))
  print(paste("                    "))

}

for (N in c(36, 64, 100, 225, 2500, 12100)){
  print(paste('Starting simulations with samples of size', N))
  monte_carlo(N, 10000)
  print('')
}
```

```
## [1] "Starting simulations with samples of size 36"
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 1"
## [1] "                    "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.435399999999999"
## [1] "                    "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9841"
## [1] "                    "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.359300000000001"
## [1] "                    "
## [1] ""
## [1] "Starting simulations with samples of size 64"
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 1"
```

```
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3264"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9785"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3441"
## [1] "                          "
## [1] ""
## [1] "Starting simulations with samples of size 100"
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.8745"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3374"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9799"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3419"
## [1] "                          "
## [1] ""
## [1] "Starting simulations with samples of size 225"
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 1"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3232"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9802"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.332800000000001"
## [1] "                          "
## [1] ""
## [1] "Starting simulations with samples of size 2500"
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9728"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3112"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9769"
## [1] "                          "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3207"
## [1] "                          "
## [1] ""
## [1] "Starting simulations with samples of size 12100"
```

```
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9884"
## [1] "                        "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.337500000000001"
## [1] "                        "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.025: 0.9829"
## [1] "                        "
## [1] "Percentage of simulated means which were significantly different from"
## [1] "sampling distribution at critical point 0.975: 0.3358"
## [1] "                        "
## [1] ""
```

**Observations**

###Central Limit theorem is interpreted here. #a)Mean of the random variable 'x' following the 'Mixed distribution' = 10 #b) #1. As we increase N, the percentage of sample means that have a z-score below -0.025 and above 0.025 is ~99%. #2. For critical point z = 0.975: As we increase N, the percentage of sample means that have a z-score below -0.975 and above 0.975 is ~34%, which means 66% of the sample means are between z score of 0.975. #These simulation results are in accordance with a typical normal distribution where almost 68% of sample means lie within a z-score of 1 and where many sample means fall outside the z-score of 0.025 as the interval defined by the same is very very small.