S520_033023_lab_CI_simul.R

adityamhaske

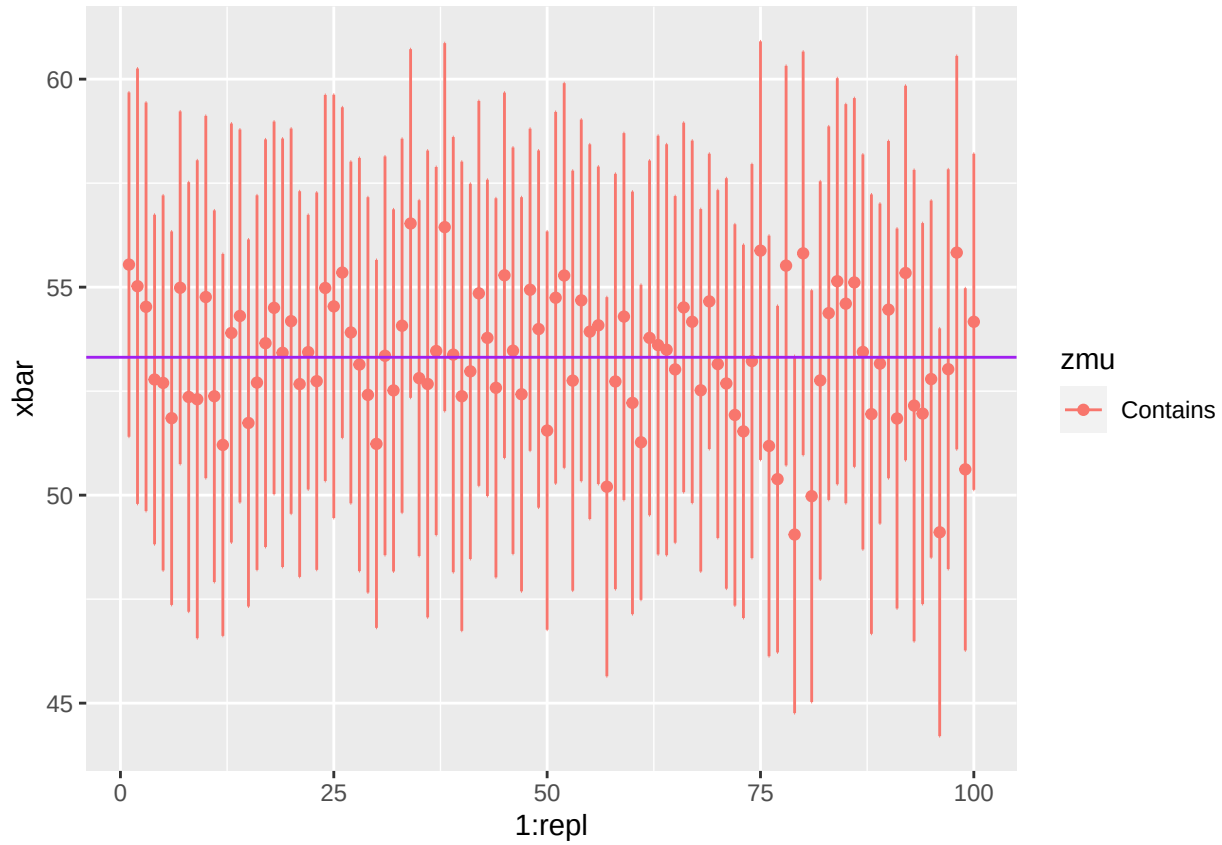
2023-05-01

```
##############################################################################
## SIMULATION
## YOU DO NOT NEED TO LEARN THIS CODE
## IT'S USED ONLY TO ILLUSTRATE THE MEANING OF A CONFIDENCE INTERVAL
##############################################################################

library(tidyverse)
```

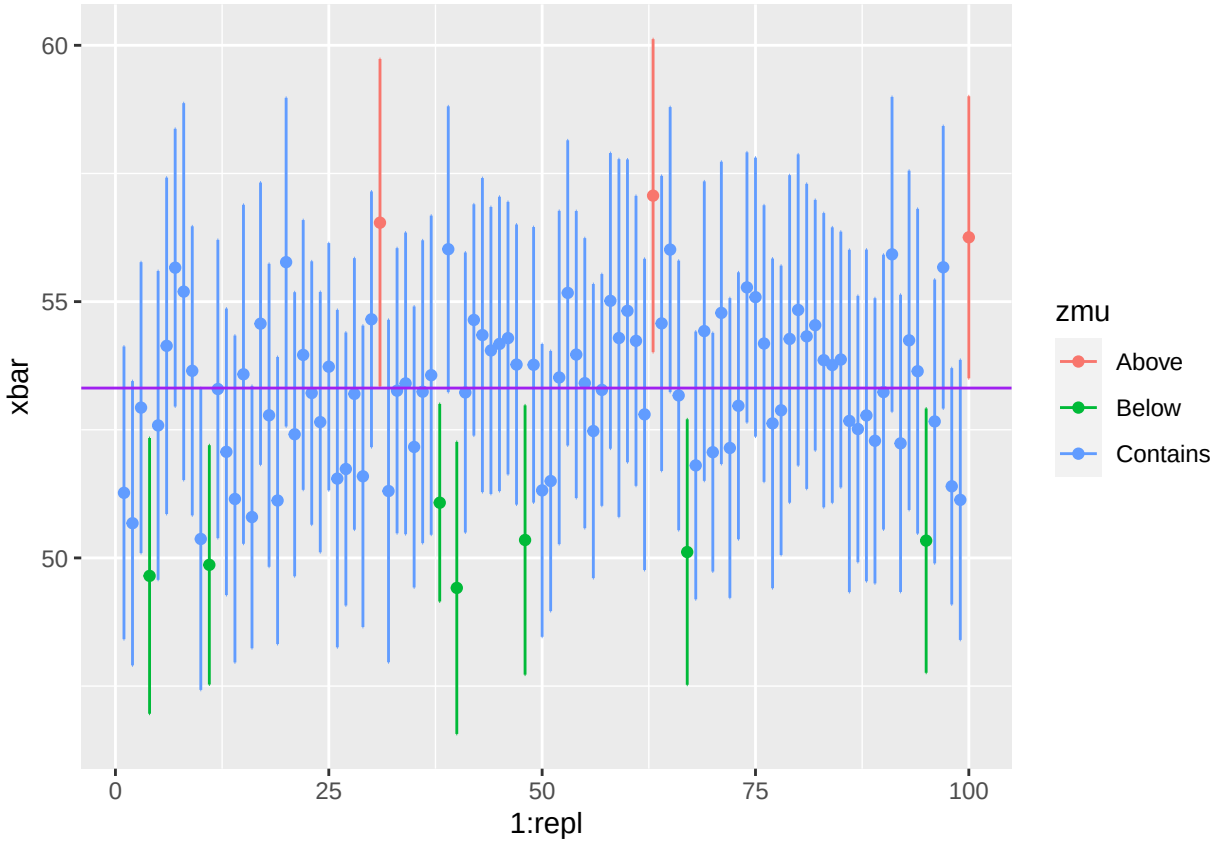
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.2.1     v stringr 1.5.0
## v readr   2.1.3     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## this function construct CI when sigma is known
ci_mu1 <- function(x, n, CI=.95) {
  xbar = mean(x)
  se = sigma/sqrt(n) # if we know sigma
  c(xbar, xbar+c(-1,1)*qnorm(CI+(1-CI)/2)*se)
}

## this function construct CI using s (when sigma is unknown)
ci_mu2 <- function(x, n, CI=.95) {
  xbar = mean(x)
  se = sd(x)/sqrt(n) # if we do not know sigma we use and estimate of sigma, the sample std. deviation
  c(xbar, xbar+c(-1,1)*qt(CI+(1-CI)/2, n-1)*se)
}

## A function for changing colors when intervals do not contain the population mean
colorbeta0 = function(bounds, marker) {
  if (marker < bounds[2]) "Above"
  else if (marker > bounds[3]) "Below"
  else "Contains"
}


## The key function to obtain the simulations results (plots)
```

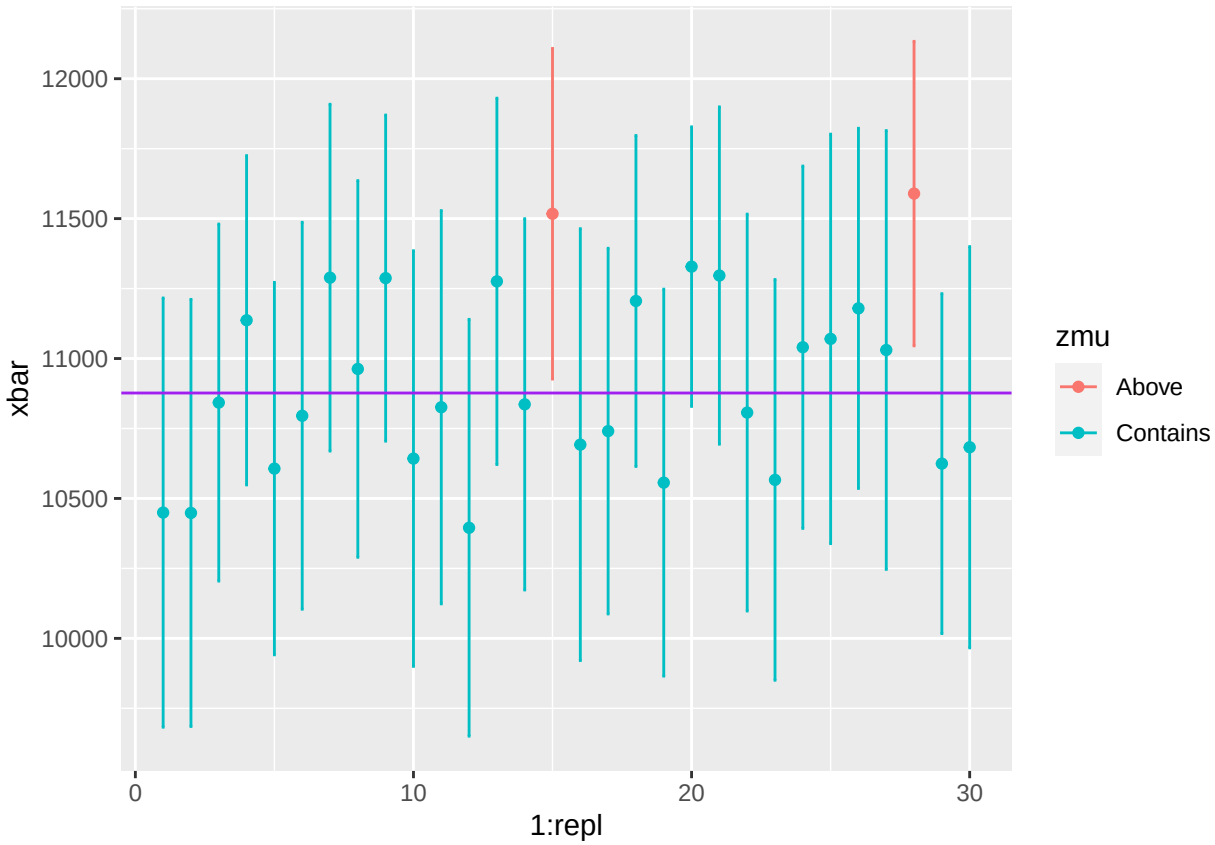
```r
sim.CI.mu = function(x, n, repl = 20, CI = 0.95){
  mu = mean(x)
  sigma = sqrt(mean(x^2) - mean(x)^2)
  x.vec = replicate(repl, sample(x, n, replace = T))
  cimu = apply(x.vec, 2, ci_mu2, n, CI) #change to ci_mu1 if you want to obtain CI with sigma known
  zmu <- apply(cimu, 2, colorbeta0, mu)
  df1 <- data.frame(t(cimu))
  colnames(df1) <- c("xbar", "lb", "ub")
  ggplot(df1, aes(x = 1:repl, y= xbar, colour = zmu)) +
    geom_errorbar(aes(ymin=lb, ymax=ub), width=.1) +
    geom_point() +
    geom_hline(yintercept = mu, colour = "purple")
}

### Observe the arguments you can change are
#### x is your population vector,
#### n is the sample size,
#### repl is the number of replications to be used
#### CL is the confidence level
#### seed is the seed number

#### Let's try simulations with different values
library(fivethirtyeight)
age = congress_age$age
sim.CI.mu(x = age, n=40, repl = 100, CI=.99)
```



```
sim.CI.mu(x = age, n=40, repl = 100, CI=.90)
```




```
sim.CI.mu(x = US_births_1994_2003$births, n = 30, repl = 30, CI = .95)
```


12000
11500
11000
10500
10000
xbar
0
10
20
30
1:repl
zmu
Above
Contains