Practical 7

1940223, Pranav Gopalkrishna

2021-02-26

FIND OUT THE GAIN IN EFFICIENCY IN USING STRATIFIED SAMPLING WITH PROPORTIONAL ALLOCATION OVER SRS

Efficiency of an estimator means that the variance of the estimator, measured accross multiple samples, is relatively small.

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# The population, strata, and other key values

popl = iris  
head(popl)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

stra1 = filter(popl, popl$Petal.Width > 0 & popl$Petal.Width <= 0.5)  
stra2 = filter(popl, popl$Petal.Width > 0.5 & popl$Petal.Width <= 1)  
stra3 = filter(popl, popl$Petal.Width > 1 & popl$Petal.Width <= 1.5)  
stra4 = filter(popl, popl$Petal.Width > 1.5 & popl$Petal.Width <= 2)  
stra5 = filter(popl, popl$Petal.Width > 2 & popl$Petal.Width <= 2.5)  
  
n = 60  
N = length(popl$Petal.Width)

#### The following apply only for stratified sampling with proportional allocation…

The sizes of the various strata, denoted by Nh, h = 1, 2, 3, 4, 5…

N1 = length(stra1$Petal.Width)  
N2 = length(stra2$Petal.Width)  
N3 = length(stra3$Petal.Width)  
N4 = length(stra4$Petal.Width)  
N5 = length(stra5$Petal.Width)  
Nh = c(N1, N2, N3, N4, N5)  
print(Nh)

## [1] 49 8 41 29 23

The appropriate sample size is given by n \* Nh / N

n1 = round(n \* N1 / N)  
n2 = round(n \* N2 / N)  
n3 = round(n \* N3 / N)  
n4 = round(n \* N4 / N)  
n5 = round(n \* N5 / N)  
nh = c(n1, n2, n3, n4, n5)  
print(nh)

## [1] 20 3 16 12 9

# Calculating variance of sample mean for stratified sampling with proportional allocation

The value of ybarst is measured for 10000 samples.

ybarstList = c()  
max = 10000  
for(i in c(1:max))  
{  
 sample1 = sample(stra1$Petal.Width, n1, replace = FALSE)  
 sample2 = sample(stra2$Petal.Width, n2, replace = FALSE)  
 sample3 = sample(stra3$Petal.Width, n3, replace = FALSE)  
 sample4 = sample(stra4$Petal.Width, n4, replace = FALSE)  
 sample5 = sample(stra5$Petal.Width, n5, replace = FALSE)  
 samples = c(sample1, sample2, sample3, sample4, sample5)  
   
 ybar1 = mean(sample1)  
 ybar2 = mean(sample2)  
 ybar3 = mean(sample3)  
 ybar4 = mean(sample4)  
 ybar5 = mean(sample5)  
 ybarh = c(ybar1, ybar2, ybar3, ybar4, ybar5)  
   
 ybarst = sum(Nh \* ybarh) / N  
 ybarstList = c(ybarstList, ybarst)  
}

#### Finding the variance of the 10000 ybarst values

prop = var(ybarstList)

# Calculating variance for sample means for simple random sampling (without replacement)

The value of ybar is measured for 10000 samples.

srswor = var(replicate(max, mean(sample(popl$Petal.Width, n, replace = FALSE))))

Comparing the efficiency of the sample means taken using the stratified samples, and the sample means taken using SRS WOR…

if(prop > srswor) {  
 print("SRS is more efficient than proportional allocation.")  
 print(paste("Gain in efficiency is", 1/srswor - 1/prop))  
} else if(prop < srswor) {  
 print("Proportional allocation is more efficient than SRS.")  
 print(paste("Gain in efficiency is", 1/prop - 1/srswor))  
} else {  
 print("Proportional allocation and SRS are equally efficient.")  
}

## [1] "Proportional allocation is more efficient than SRS."  
## [1] "Gain in efficiency is 6849.33752357286"

# Conclusion

From this practical verification, we see that the variance of sample mean measured across a large number of samples is smaller when we use stratified sampling using proportional allocation, as compared to simple random sampling without replacement.

In other terms, stratified sampling using proportional allocation is a more efficient method of estimating population mean, as compared to simple random sampling without replacement.