part2.410project.R

vicky
Thu Apr 7 22:37:43 2016

```
#*****************************************************
#     STAT 410 Project
#     April 7, 2015
#     Vicky Xu, Qiaochu Zou, Sun
#
#
#*****************************************************
### Data Set Caveat:
# The dataset contains the student's average chemtry score for the 25 differenct class in middel
school.
grade.data=read.csv("410project.csv")
grade.data

##    class.number number.student average.grade Stratum
## 1            45             59       49.2562       1
## 2            41             53       49.5238       1
## 3            50             70       50.3468       1
## 4            21             72       53.3333       1
## 5            19             78       54.0641       1
## 6            39             71       54.2958       1
## 7            30             62       55.1613       1
## 8            28             72       55.3333       1
## 9            33             78       58.2821       1
## 10           24             76       58.6447       1
## 11           35             72       58.9167       1
## 12           48             64       59.1253       1
## 13           49             59       61.3345       2
## 14           42             62       61.4675       2
## 15           36             69       62.1884       2
## 16           46             62       62.5446       2
## 17           22             66       62.7576       2
## 18           34             78       63.5513       2
## 19           26             69       63.8116       2
## 20           29             69       65.5362       2
## 21           52             58       66.4429       2
## 22           51             67       67.2943       2
## 23           16             72       68.3889       2
## 24           44             60       69.2044       2
## 25           55             55       69.5665       2
## 26           23             75       69.6851       2
## 27           53             64       70.4568       2
## 28           15             69       70.4637       2
## 29           43             70       73.0673       3
## 30           37             69       74.3766       3
## 31           40             58       75.4583       3
## 32           32             66       75.7273       3
## 33           25             69       75.7536       3
## 34           54             61       75.8723       3
## 35           20             72       76.5694       3
## 36           17             79       79.3867       3
## 37           27             50       79.4400       3
## 38           47             71       79.8437       3
## 39           38             71       81.8873       3
## 40           31             71       85.0423       3

grade.data$class.number <- as.factor(grade.data$class.number)
# Defining values
y <- grade.data$average.grade
x <- grade.data$number.student
N <- 41
b <- 10000
# Population Parameters
var <- var(grade.data$average.grade)
mu <- mean(grade.data$average.grade)
var
```

```
## [1] 93.0266
```

```
mu
```

```
## [1] 66.08506
```

```
#######  Stratified Sampling #######
# Stratified into 3 stratum
# First order grade in ascending order
# then grouped into three stratum by stratification principles
# exclude the max and min (outliers), then use (second largest - second smallest)/3.
##10.7878333
# Stratum separations: 49.2562~60.31163, 60.31163~71.0994,71.0994~85.0423.
st <- grade.data$Stratum
table(st)
```

```
## st
##  1  2  3
## 12 16 12
```

```
st
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3
## [36] 3 3 3 3 3 3
```

```
y1 <- y[st==1]
y2 <- y[st==2]
y3 <- y[st==3]
# Population statistic
var1 = var(y1)
var2 = var(y2)
var3 = var(y3)
n1 <- 6
n2 <- 8
n3 <- 6
N1 <- 12
N2 <- 16
N3 <- 12
# Sample size
n = 20
# Optimal allocation:
n*N1*var1 / (N1*var1 +N2*var2 + N3*var3)
```

```
## [1] 6.463061
```

```
# n1=6
n*N2*var2 / (N1*var1 +N2*var2 + N3*var3)
```

```
## [1] 7.631549
```

```
# n2=8
n*N3*var3 / (N1*var1 +N2*var2 + N3*var3)
```

```
## [1] 5.905389
```

```
# n3=6
# Stratified sampling uses SRS for each stratum
# Population Parameters
var <- var(grade.data$average.grade)
mu <- mean(grade.data$average.grade)
var
```

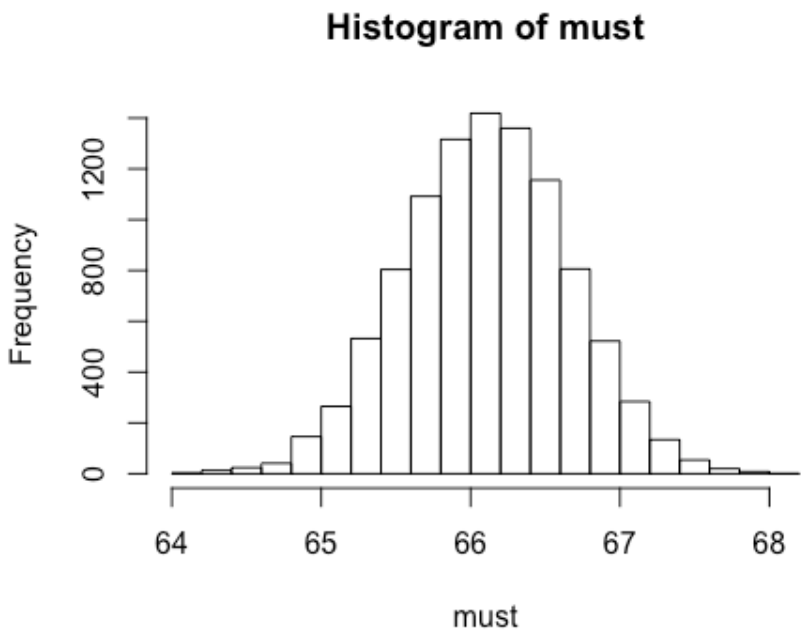
```
## [1] 93.0266
```

```
mu
```

```
## [1] 66.08506
```

```
must <- c()
b = 10000
for (k in 1:b){
s1 <- sample(N1,n1)
must1 <- mean(y1[s1])
s2 <- sample(N2,n2)
must2 <- mean(y2[s2])
s3 <- sample(N3,n3)
must3 <- mean(y3[s3])
must[k] <- (must1+must2+must3)/3
}
```

```
hist(must)
```




```
mean_st <- mean(must)
mse_st <- mean((must-mu)^2)
bias_st <- mean(must)-mu
bias_st
```

```
## [1] 0.01841557
```

```
mse_st
```

```
## [1] 0.3020069
```

```
mean_st
```

```
## [1] 66.10348
```