Survmeth 895 Homework 5

David (Daiwei) Zhang

March 6, 2017

Problem 1 (a)

```
nSim = 1e4
theta1=rbeta(nSim, 29.5, 9.5)
yexc1=rbinom(nSim,1150,theta1)
theta2=rbeta(nSim, 17.5, 16.5)
yexc2=rbinom(nSim,488,theta2)
theta3=rbeta(nSim, 6.5, 27.5)
yexc3=rbinom(nSim,180, theta3)
theta4=rbeta(nSim, 4.5, 34.5)
yexc4=rbinom(nSim, 58,theta4)
t= 60+yexc1+yexc2+yexc3+yexc4
mean(t)
```

```
## [1] 1221.452
```

```
sd(t)
```

```
## [1] 90.79888
```

```
quantile(t, c(0.025, 0.975))
```

```
##  2.5% 97.5%
##  1035  1389
```

Problem 1(b)

```
theta = rbeta(nSim, 59.5, 89.5)
yexc = rbinom(nSim, 1876, theta)
s = 60 + yexc
mean(s)
```

```
## [1] 807.832
```

```
sd(s)
```

```
## [1] 77.8373
```

```
quantile(s, c(0.025, 0.975))
```

```
##  2.5% 97.5%
##   661   962
```

Problem 1(c)

Here the sizes of the strata differ greatly from each other. While the first two strata should have much more weight than the last two, in the unstratified analysis they are given equal weight and make the prediction much closer to the mean in the last two strata than it is supposed to.

Problem 2(a)

```
require(HDInterval)
```

```
## Loading required package: HDInterval
```

```
m <- 5
strata.size <- c(635, 570, 475, 303, 89)
sample.size <- c(84, 125, 138, 112, 41)
exc.size <- strata.size - sample.size
sample.mean <- c(4.24, 11.63, 15.85, 23.59, 29.61)
sample.var <- c(27.54, 55.84, 71.70, 192.32, 334.93)

nsimul=1000
result <- matrix(0, nsimul, m)
for (j in 1:m){
  sampsize= sample.size[j]
  k = exc.size[j]
  ybar=sample.mean[j]
  ssquare=sample.var[j]
  for (i in 1:nsimul){
    tmp=rnorm(sampsize-1)
    chisq=sum(tmp*tmp)
    sigmasq=(sampsize-1)*ssquare/chisq;
    mu=ybar+sqrt(sigmasq/
                   sampsize)*rnorm(1)
    ybark=mu+sqrt(sigmasq/k)*rnorm(1)
    total = ybark * k + sampsize * ybar
    result[i,j] = total
  }
}
allCows = rowSums(result)
mean(allCows)
```

```
## [1] 26592.16
```

```
sd(allCows)
```

```
## [1] 698.9308
```

```
hdi(allCows)
```

```
##    lower    upper
## 25327.03 28056.61
## attr(,"credMass")
## [1] 0.95
```

Problem 2(b)

```
sampsize = sum(sample.size)
k = sum(exc.size)
ybar = sum(sample.mean * sample.size) / sum(sample.size)
ssquare = 136.19
nsimul=1000
result <- matrix(0, nsimul, 1)
```

```
for (i in 1:nsimul){
  tmp=rnorm(sampsize-1)
  chisq=sum(tmp*tmp)
  sigmasq=(sampsize-1)*ssquare/chisq;
  mu=ybar+sqrt(sigmasq/
                 sampsize)*rnorm(1)
  ybark=mu+sqrt(sigmasq/k)*rnorm(1)
  total = ybark * k + sampsize * ybar
  result[i,1] = total
}
allCows = result
mean(allCows)
```

```
## [1] 32571.05
```

```
sd(allCows)
```

```
## [1] 917.5781
```

```
hdi(allCows)
```

```
##           [,1]
## lower 30777.52
## upper 34415.07
## attr(,"credMass")
## [1] 0.95
```

Problem 2(c)

Since the stratum sizes are very skewed and the sample sizes are quite symmetric, the stratum-blind analysis underweights the small farms and overpredicts the total number of cows.