homework4

Homework 4

1

a)

```
data <- data.frame('Stratum' = c(1,2,3,4),
                   'Area'=c(222.81,49.61,50.25,197.81),
                   'Number of tows made'=c(4,6,3,5),
                   'Average number of bushels per tow'=c(0.44,1.17,3.92,1.80),
                   'Sample variance for stratum'=c(0.068,0.042,2.146,0.794))
t_hat <- sum(25 * data$Area * data$Average.number.of.bushels.per.tow) ; t_hat
```

```
## [1] 17727.95
```

```
N_h <- (25 * data$Area)
n_h <- data$Number.of.tows.made
s_h <- data$Sample.variance.for.stratum
SE_t_hat <- sqrt(sum(N_h^2 * (1 - n_h/N_h) * (s_h/n_h))) ; SE_t_hat
```

```
## [1] 2354.492
```

b)

```
data2 <- data.frame('Stratum' = c(1,4),
                   'Area'=c(322.67,197.81),
                   'Number of tows made'=c(8,5),
                   'Average number of bushels per tow'=c(0.63,0.40),
                   'Sample variance for stratum'=c(0.083,0.046))
t_hat <- sum(25 * data2$Area * data2$Average.number.of.bushels.per.tow) ; t_hat
```

```
## [1] 7060.153
```

```
N_h <- (25 * data2$Area)
n_h <- data2$Number.of.tows.made
s_h <- data2$Sample.variance.for.stratum
SE_t_hat <- sqrt(sum(N_h^2 * (1 - n_h/N_h) * (s_h/n_h))) ; SE_t_hat
```

```
## [1] 948.2723
```

2

a)

```
library(SDaA)
N <- 3078
n <- nrow(agsrs)
t_x <- 2087759

y <- agsrs$acres92
ybar <- mean(y)
x <- agsrs$farms87
xbar <- mean(x)

B.hat <- ybar / xbar
t_y.hat <- B.hat * t_x ;t_y.hat
```

```
## [1] 960155061
```

```
e <- y - B.hat * x
SE.B <- 1 / xbar * sqrt(var(e) / n * (1 - n/N))
SE.t_yr <- t_x * SE.B ; SE.t_yr
```

```
## [1] 68446406
```

```
t_y.hat + c(-1,1) * 1.96 * SE.t_yr
```

```
## [1]  826000106 1094310016
```

b)

```
N <- 3078
t_x <- 2087759
n <- nrow(agsrs)
# the sample data
x <- agsrs$farms87
y <- agsrs$acres92

# summary statistics
xbar <- mean(x)
ybar <- mean(y)
s.y <- sd(y)
xbar;ybar;s.y^2
```

```
## [1] 647.7467
```

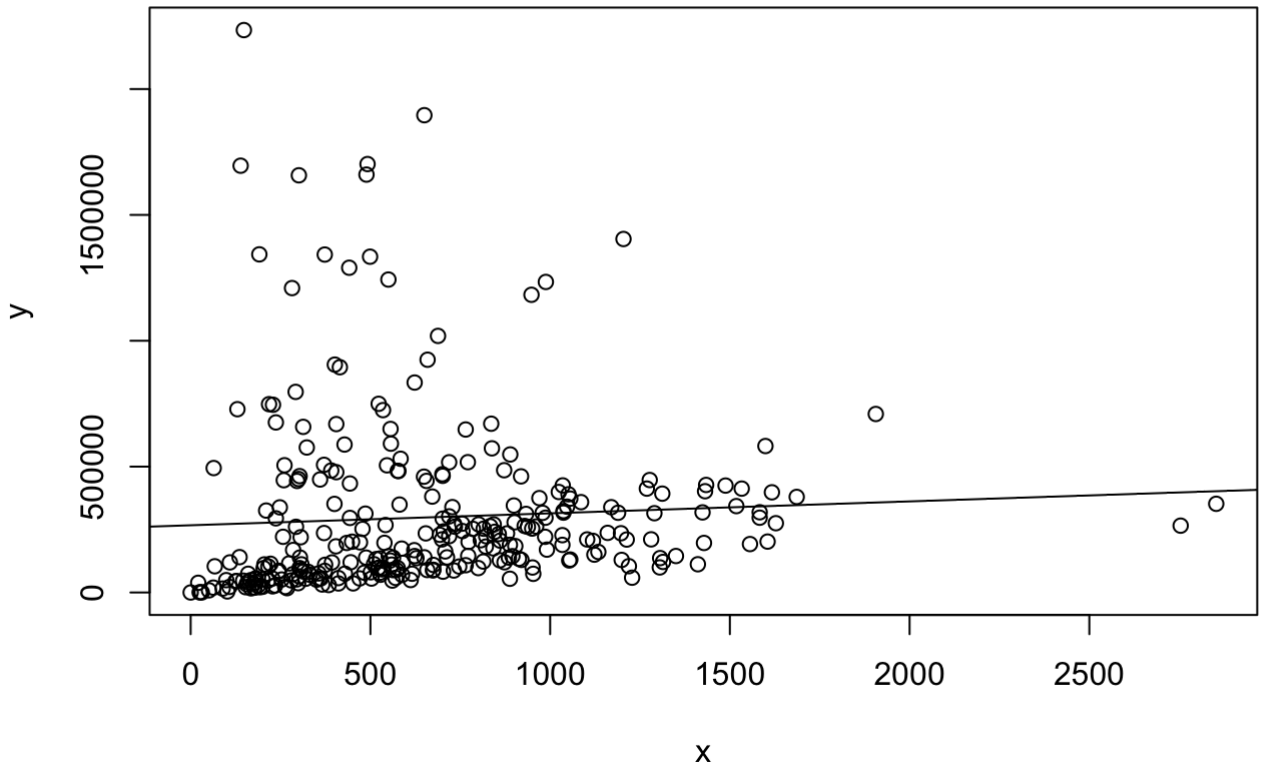
```
## [1] 297897
```

```
## [1] 1.18716e+11
```

```
lsfit(x,y)$coefficients
```

```
##    Intercept            X
## 267029.81421     47.65325
```

```
B1.hat <- cor(x,y) * sd(y) / sd(x)
plot(x,y)
abline(ybar-B1.hat * xbar,B1.hat)
```




```
#regression estimate
x_bar.U <- t_x / N
ybar.hat.reg <- ybar + B1.hat * (x_bar.U - xbar)
ybar.hat.reg
```

```
## [1] 299352.3
```

```
## standard error
B0.hat <- ybar - B1.hat * xbar
e <- y - B0.hat - B1.hat * x
var(e) ; var(y) * (1- cor(x,y)^2) # should be the same
```

```
## [1] 118293647832
```

```
## [1] 118293647832
```

```
SE.ybar.reg <- sd(e) / sqrt(n) * sqrt(1 - n/N)
SE.ybar.reg
```

```
## [1] 18864.79
```

```
# estimate and standard error for population total
N * ybar.hat.reg
```

```
## [1] 921406265
```

```
N* SE.ybar.reg
```

```
## [1] 58065813
```

```
# 95% CI should be
N * (ybar.hat.reg + c(-1,1) * 1.96 * SE.ybar.reg)
```

```
## [1]  807597271 1035215259
```

C

In this problem the N_d (domain size) is unknown

```
N <- 3078 # as we can know from problem 2 : total number of countries in the United Stat
es
y.samp <- agsrs$acres92
domain.samp <- (agsrs$farms87 >= 600)

domain.total <- function(y.samp, domain.samp, d, N){
  n <- length(y.samp)
  x <- (domain.samp==d)
  y <- y.samp; u <- x*y
  u.bar <- mean(u)
  s2.u <- var(u)
  t.yd.hat <- N * u.bar
  V.hat <- N^2 * s2.u/n * (1 - n/N)
  SE <- sqrt(V.hat)
  answer <- c(point.est=t.yd.hat, std.error=SE)
  return(answer)
}
```

For the domain of countries fewer than 600 farms

```
domain.total(y.samp, domain.samp, d=0, N)
```

```
## point.est std.error
## 473559072  55528141
```

For the domain of countries with 600 or more farms

```
domain.total(y.samp, domain.samp, d=1, N)
```

```
## point.est std.error
## 443368037  39595965
```