

# Homework6

## homework 6

Name: Yi Chen, UNI: YC3356

### Problem 1

(a) one-stage cluster sampling with equal cluster size.

```
N = 100 # number of clusters
n = 5 # number of cluster to sample
M = 8 # number of item in ith cluster and same as m in one stage cluster
M0 = sum(rep(M,N))
yi_bar = rep(c(3,2,0),c(2,1,2)) / M
ti_hat = M * yi_bar
tunb_hat = (N/n) * sum(ti_hat)
s_t = sqrt((1/(n-1)) * sum((ti_hat - tunb_hat/N) ^ 2))
se_t = N * s_t / sqrt(n) * sqrt(1 - n/N)
tunb_hat ; se_t
```

```
## [1] 160
```

```
## [1] 66.10598
```

(b) simple random sampling

```
N = 800
n = 40
y = rep(c(1,0),c(8,n-8))
ys_bar = sum(y) / n
t_hat = N * ys_bar
ss_2 = (1/n) * sum((y - ys_bar)^2)
se_t = sqrt(N^2 * (ss_2/n) * (1-n/N))
t_hat; se_t
```

```
## [1] 160
```

```
## [1] 49.31531
```

### Problem 2

(a) two stage cluster sampling with different cluster size

## i. unbiased estimation

```

N = 20 # number of cluseter
n = 4  # number of cluseter be sampled
M0 = 940 # total population size
Mi = c(17,56,23,64) # cluster size
mi = c(10,25,12,30) # sample size in step two
yi_bar = c(32.7,36.1,30.3,33.4) # cluster sample mean
si_2 = c(26.3,21.4,23.6,29.1) # cluster sample variance
ti_hat = Mi * yi_bar

tunb_hat = (N/n) * sum(ti_hat)

yunb_hat = tunb_hat / M0

st_2 = (1/(n-1)) * sum((ti_hat - tunb_hat/N)^2)

se_tunb_hat = sqrt(N^2*(st_2/n)*(1-n/N)+sum(((si_2*M0^2)/mi)*(1 - mi/Mi)))

se_yunb_hat = se_tunb_hat / M0

yunb_hat;se_yunb_hat

```

```
## [1] 28.78723
```

```
## [1] 8.015045
```

## ii. ratio estimation

```

yr_bar_hat = ((N/n) * sum(ti_hat))/((N/n) * sum(Mi))

sr_2 = (1/(n-1)) * sum(Mi^2 * (yi_bar - yr_bar_hat)^2)

se_yr_bar_hat = sqrt((sr_2/(n * mean(Mi)^2)) * (1 - n/N) + (1/(n*N*mean(Mi)^2)) * sum(si_2*(Mi^2/mi) * (1-mi/Mi)))

yr_bar_hat;se_yr_bar_hat

```

```
## [1] 33.825
```

```
## [1] 1.015215
```

Even the ration estimation while biased gives a small standard error.

# Problem 3

Unequal probability sampling one-stage cluster sampling

(a)

```

Mi = c(110,265,360,1265)
Sales = c(44,82,112,362)
M0 = sum(Mi)
phi = Mi / M0
N = 4
n = 2
# every draw is independent
E_tphi_hat = 0
for ( i in 1:N){
  for ( j in 1:N){
    tphi_hat = (1/n) * sum((Sales[i]/phi[i]),(Sales[j]/phi[j]))
    E_tphi_hat = E_tphi_hat + phi[i] * phi[j] * tphi_hat
  }
}
E_tphi2_hat = 0
for ( i in 1:N){
  for ( j in 1:N){
    tphi2_hat = ((1/n) * sum((Sales[i]/phi[i]),(Sales[j]/phi[j])))^2
    E_tphi2_hat = E_tphi2_hat + phi[i] * phi[j] * tphi2_hat
  }
}
v_tphi_hat = E_tphi2_hat - E_tphi_hat^2
E_tphi_hat;v_tphi_hat

```

```
## [1] 600
```

```
## [1] 1410.124
```

(b)

The srs is unbiased estimation. Thus the mean is still 600.

```

N=4
n=2
y.i<-c(44,82,112,362)
S.2<-var(y.i)
t.srs.var<-N^2*(1-n/N)*S.2/n
t.srs.var

```

```
## [1] 82997.33
```