STA304_A2_Q3

data

```r
x_1 <- c(204, 143, 82, 256, 275, 198)
y_1 <- c(210, 160, 75, 280, 300, 190)
x_2 <- c(137, 189, 119, 63, 103, 107, 159, 63, 87)
y_2 <- c(150, 200, 125, 60, 110, 100, 180, 75, 90)

n1 <- length(x_1)
n2 <- length(x_2)

N1 <- 120
N2 <- 180
N <- N1+N2

tau_x_1 <- 24500
tau_x_2 <- 21200

mu_x_1 <-  tau_x_1 / N1
mu_x_2 <- tau_x_2 / N2
```

Part 1

(a)

calculate $\hat{\tau}$

```r
y1_mean <- mean(y_1)
y2_mean <- mean(y_2)
tau_hat <- N1*y1_mean + N2*y2_mean
tau_hat
```

```
## [1] 46100
```

calculate $\hat{V(\hat{\tau})}$

```r
s1_square = var(y_1)
s2_square = var(y_2)
V_tau_hat = N1^2 * (1-(n1/N1)) * (s1_square/n1) + N2^2 * (1-(n2/N2)) * (s2_square/n2)
V_tau_hat
```

```
## [1] 23075975
```

(b)

```r
x1_mean <- mean(x_1)
x2_mean <- mean(x_2)
R1_hat <- y1_mean / x1_mean
R2_hat <- y2_mean / x2_mean
```

calculate $S_{R_1}^2$ and $S_{R_2}^2$

```
S_R1_square <- 0
i <- 1
while (i < n1+1){
  residual_square = (y_1[i] - R1_hat * x_1[i])^2
  S_R1_square = S_R1_square + residual_square / (n1-1)
  i = i + 1
}
```

```
S_R2_square <- 0
i <- 1
while (i < n2+1){
  residual_square = (y_2[i] - R2_hat * x_2[i])^2
  S_R2_square = S_R2_square + residual_square / (n2-1)
  i = i + 1
}
```

calculate $\hat{\mu_{y,SR}}$

```
mu_y_SR_hat <- (N1/N)* mu_x_1 * R1_hat + (N2/N)* mu_x_2 * R2_hat
mu_y_SR_hat
```

```
## [1] 160.6882
```

calculate $var(\hat{\mu_{y,SR}})$

```
var_mu_y_sr_hat <- (N1/N)^2 * (1-(n1/N1)) * S_R1_square/n1 + (N2/N)^2 * (1-(n2/N2)) * S_R2_square/n2
```

Eventually, get the estimate of total potential sales and its variance.

```
total_ratio = N*mu_y_SR_hat
var_total_ratio = N^2 * var_mu_y_sr_hat
total_ratio
```

```
## [1] 48206.45
```

```
var_total_ratio
```

```
## [1] 564612.8
```

(c)

```
fit_1 <- lm(y_1~x_1)
fit_2 <- lm(y_2~x_2)
b_1 <- fit_1$coefficients[2]
b_2 <- fit_2$coefficients[2]
```

Calculate mean eastimator for strata 1 and strata 2 separately

```
mu_yL_1_hat <- y1_mean + b_1*(mu_x_1*x1_mean)
mu_yL_2_hat <- y2_mean + b_2*(mu_x_2*x2_mean)
```

Calculate estimated variance of above two variable

```
e1 <- residuals(fit_1)
MSE_1 = sum(e1^2)/(n1-1)
e2 <- residuals(fit_2)
```

```r
MSE_2 = sum(e2^2)/(n2-1)

Var_mu_yL_1_hat <- (1-n1/N1)*(MSE_1/n1)
Var_mu_yL_2_hat <- (1-n2/N2)*(MSE_2/n2)
```

Then get weighted result

```r
mu_y_SR_hat <- (N1/N)*mu_yL_1_hat + (N2/N)*mu_yL_2_hat
Var_mu_y_SR_hat <- (N1/N)^2 * Var_mu_yL_1_hat + (N2/N)^2 * Var_mu_yL_2_hat
```

Thus, get the total potential sales and variance

```r
total_potential_sales_L <- N * mu_y_SR_hat
var_total_potential_sales_L <- (N^2) * Var_mu_y_SR_hat

total_potential_sales_L
```

```
##     x_1
## 8099212
```

```r
var_total_potential_sales_L
```

```
## [1] 462002.2
```

(d)

i)

```r
ER_ratio_to_basic <- V_tau_hat/var_total_ratio
ER_ratio_to_basic
```

```
## [1] 40.87044
```

ii)

```r
ER_ratio_to_reg <- var_total_potential_sales_L / var_total_ratio
ER_ratio_to_reg
```

```
## [1] 0.8182639
```

iii)

```r
ER_reg_to_basic <- V_tau_hat / var_total_potential_sales_L
ER_reg_to_basic
```

```
## [1] 49.94776
```

(e)

I recommend regression method since its variace is smallest.

Part 2

(a)

calculate $\hat{\tau}$

```
y_3 <- c(y_1, y_2)
y3_mean <- mean(y_3)
tau = N*y3_mean
tau
```

```
## [1] 46100
```

calculate $\hat{V(\hat{\tau})}$

```
s3_square = var(y_3)

V_tau_hat = (N^2) * (1-((n1+n2)/N)) * (s3_square/(n1+n2))
V_tau_hat
```

```
## [1] 30769143
```

(b)

```
y_st_bar <- 0.4*y1_mean + 0.6*y2_mean
x_st_bar <- 0.4*x1_mean + 0.6*x2_mean
R_cr <- y_st_bar / x_st_bar
mu_x = (24500+21200)/300
mu_y_hat = mu_x * R_cr
```

caculate $S^2_{R_{CR,1}}$ and $S^2_{R_{CR,2}}$

```
S_R1_square <- 0
i <- 1
while (i < n1+1){
  residual_square = (y_1[i] - R_cr * x_1[i])^2
  S_R1_square = S_R1_square + residual_square / (n1-1)
  i = i + 1
}

S_R2_square <- 0
i <- 1
while (i < n2+1){
  residual_square = (y_2[i] - R_cr * x_2[i])^2
  S_R2_square = S_R2_square + residual_square / (n2-1)
  i = i + 1
}
```

```
var_mu_y_CR_hat <- (N1/N)^2 * (1-(n1/N1))*(S_R1_square/n1) + (N2/N)^2 * (1-(n2/N2))*(S_R2_square/n2)
```

Thus, get the total potential sales

```
total_potential_sales_CR = N*mu_y_hat
var_total_potential_sales_CR = N^2 * var_mu_y_CR_hat

total_potential_sales_CR
```

```
## [1] 48209.84
```

```
var_total_potential_sales_CR
```

```
## [1] 562470.3
```

(c)

```
ER_ratio_to_basic <- V_tau_hat / var_total_potential_sales_CR
ER_ratio_to_basic
```

```
## [1] 54.70359
```

(d)

I recommend ratio method since its variance is smaller