PS2 code

R Markdown

This is the code part of PS2

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.2
```

```
data <- read_excel("ddk2011test.xlsx")
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV2296 / R2296C48: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV2480 / R2480C48: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV2492 / R2492C48: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV3592 / R3592C48: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV3901 / R3901C48: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA4904 / R4904C27: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5107 / R5107C27: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5116 / R5116C27: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5121 / R5121C27: got '.'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5122 / R5122C27: got '.'
```

```
#data clean
data<-data.frame(data$schoolid,data$totalscore,data$tracking,data$etpteacher,data$girl,data$agetest,data
data<-data[data$data.girl!='.',]
data<-data[data$data.agetest!='.',]
data$data.girl<-sapply(data$data.girl,as.numeric)-2
data$data.agetest<-sapply(data$data.agetest,as.numeric)
data$data.percentile<-sapply(data$data.percentile,as.numeric)
#cluster based
y <- scale(as.matrix(data$data.totalscore))
n <- nrow(y)
x<-cbind(data$data.tracking,data$data.agetest,data$data.girl,data$data.etpteacher,data$data.percentile,
schoolid <- as.matrix(data$data.schoolid)
k <- ncol(x)
```

```r
xx <- t(x)%*%x
invx <- solve(xx)
beta <- solve(xx,t(x)%*%y)
xe <- x*rep(y-x%*%beta,times=k)
# Clustered robust standard error
xe_sum <- rowsum(xe,schoolid)
G <- nrow(xe_sum)
omega <- t(xe_sum)%*%xe_sum
scalee <- G/(G-1)*(n-1)/(n-k)
V_clustered <- scalee*invx%*%omega%*%invx
se_clustered <- sqrt(diag(V_clustered))
print(beta,digits = 5)
```

```
##           [,1]
## [1,]  0.173580
## [2,] -0.041056
## [3,]  0.081706
## [4,]  0.180989
## [5,]  0.017424
## [6,] -0.576944
```

```r
print(se_clustered,digits = 5)
```

```
## [1] 0.07665351 0.01339398 0.02867527 0.03770844 0.00072473 0.13318931
```

```r
#conventional robust
e <- y-x%*%beta
leverage <- rowSums(x*(x%*%invx))
a <- n/(n-k)
sig2 <- (t(e) %*% e)/(n-k)
u1 <- x*(e%*%matrix(1,1,k))
u2 <- x*((e/sqrt(1-leverage))%*%matrix(1,1,k))
u3 <- x*((e/(1-leverage))%*%matrix(1,1,k))

v0 <- invx*as.numeric(sig2)
v1 <- invx %*% (t(u1)%*%u1) %*% invx
v1a <- a * invx %*% (t(u1)%*%u1) %*% invx
v2 <- invx %*% (t(u2)%*%u2) %*% invx
v3 <- invx %*% (t(u3)%*%u3) %*% invx
s0 <- sqrt(diag(v0)) # Homoskedastic formula
s1 <- sqrt(diag(v1)) # HC0
s1a <- sqrt(diag(v1a)) # HC1
s2 <- sqrt(diag(v2)) # HC2
s3 <- sqrt(diag(v3)) # HC3
print(s0)
```

```
## [1] 0.0241654219 0.0084575472 0.0241484133 0.0239139818 0.0004292976
## [6] 0.0834363938
```

```r
print(s1)
```

```
## [1] 0.0241571564 0.0085405393 0.0242239304 0.0238385277 0.0004269618
## [6] 0.0827288533
```

```r
print(s1a)
```

```
## [1] 0.0241709225 0.0085454062 0.0242377345 0.0238521122 0.0004272051
```

```
## [6] 0.0827759967
```

```
print(s2)
```

```
## [1] 0.0241710799 0.0085487064 0.0242377785 0.0238521380 0.0004272418
## [6] 0.0827994159
```

```
print(s3)
```

```
## [1] 0.0241850131 0.0085568932 0.0242516361 0.0238657582 0.0004275219
## [6] 0.0828701442
```