Stata Syntax for Sections 5.8.2 to 5.8.4, Chapter 5

Stata Syntax for Section 5.8.2

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

//Chapter 5 Section 5.8.2

cd "D:\psa\_e2\Chapter5\data"

use chpt5\_2\_original, replace

tab kuse puse, cell chi2

//Test Sample Balance on Covariates using Wilcoxon Rank-Sum(Mann-Whitney) test

ranksum mratio96, by(kuse)

ranksum pcged97, by(kuse)

ranksum pcg\_adc, by(kuse)

ranksum black, by(kuse)

ranksum age97, by(kuse)

ranksum male, by(kuse)

ttest mratio96, by(kuse)

ttest pcged97, by(kuse)

ttest pcg\_adc, by(kuse)

ttest black, by(kuse)

ttest age97, by(kuse)

ttest male, by(kuse)

//ttest and regression on outcome without bias control

ttest lwss97, by(kuse)

regress lwss97 male black age97 pcged97 mratio96 pcg\_adc kuse, robust cluster(pcg\_id)

//create propensity scores using boost

program boost\_plugin, plugin using("D:\sage\Chapter 5\Sectionsc 5.8.2 to 5.8.4\Data\boost64.dll")

gen x=uniform()

sort x

set seed 1000

boost kuse pcg\_adc age97 mratio96 pcged97 black, ///

 distribution(logistic) maxiter(1000) trainfraction(0.8) pred(ps) inter(4) ///

 shrink(.0005)influence

//distribution of estimated propensity scores

histogram ps if kuse==1, normal title(boost) xtitle("Users") saving(p1,replace)

histogram ps if kuse==0, normal title(boost) xtitle("Nonusers") saving(p0,replace)

graph combine p1.gph p0.gph, xcommon ycommon saving(hist\_p.gph,replace) title("Histograms of Estimated Propensity Scores")

graph box ps, over(kuse) title("Boxplots of Estimated Propensity Scores")

keep kid kuse ps

saveold chpt5\_2ps, replace

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R Syntax for Section 5.8.2 (Running Optmatch)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#Chapter 5, Section 5.8.2 - Run optmatch in R

#optmatch using pscores created by Stata boost

set.seed(10)

setwd ("D:/psa\_e2/Chapter5/data")

library(foreign)

cds <- read.dta("chpt5\_2ps.dta")

attach(cds)

prank <- rank(ps)

names(prank) <- kid

d1 <- outer(prank[kuse==1], prank[kuse==0], "-")

d1 <- abs(d1)

library(optmatch)

#fullmatch fm

fm <- fullmatch(d1)

(fm.d <- matched.distances(fm,d1,pres=TRUE))

unlist(fm.d,max)

mean(unlist(fm.d))

sum(unlist(fm.d))

stratumStructure(fm)

write.table (fm, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/fm.dat")

#variable match, (vm1 - at least 1 at most 4)

vm1 <- fullmatch(d1,min.controls=1,max.controls=4)

(vm1.d <- matched.distances(vm1,d1,pres=TRUE))

unlist(vm1.d,max)

mean(unlist(vm1.d))

sum(unlist(vm1.d))

stratumStructure(vm1)

write.table (vm1, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/vm1.dat")

#variable match, (vm2 - at least 2 at most 4)

vm2 <- fullmatch(d1,min.controls=2,max.controls=4)

(vm2.d <- matched.distances(vm2,d1,pres=TRUE))

unlist(vm2.d,max)

mean(unlist(vm2.d))

sum(unlist(vm2.d))

stratumStructure(vm2)

write.table (vm2, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/vm2.dat")

#variable match, (vm3 - use Hensen's equaion)

vm3 <- fullmatch(d1,min.controls=1.33,max.controls=5.32)

(vm3.d <- matched.distances(vm3,d1,pres=TRUE))

unlist(vm3.d,max)

mean(unlist(vm3.d))

sum(unlist(vm3.d))

stratumStructure(vm3)

write.table (vm3, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/vm3.dat")

#variable match, (vm4 at least 2 at most 7)

vm4 <- fullmatch(d1,min.controls=2,max.controls=7)

(vm4.d <- matched.distances(vm4,d1,pres=TRUE))

unlist(vm4.d,max)

mean(unlist(vm4.d))

sum(unlist(vm4.d))

stratumStructure(vm4)

write.table (vm4, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/vm4.dat")

#pairmatch pm

pm <- pairmatch(d1, controls=1)

(pm.d <- matched.distances(pm,d1,pres=TRUE))

unlist(pm.d,max)

mean(unlist(pm.d))

sum(unlist(pm.d))

stratumStructure(pm)

write.table (pm, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/pm.dat")

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Stata Syntax for Section 5.8.2 (After optimal matching, merge matched sets to the original data and perform ***imbalance*** check)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

//Chapter 5 Section 5.8.2

//After matching, we now merge the matched sets to the original data

cd "D:\psa\_e2\Chapter5\data"

set more off

use chpt5\_2\_original,replace

sort kid

save chpt5\_2a,replace

clear

insheet kid fm using "D:\psa\_e2\Chapter5\data\fm.dat"

sort kid

merge kid using chpt5\_2a

save chpt5\_2a, replace

tab fm kuse

use chpt5\_2a,replace

drop \_merge

sort kid

save chpt5\_2a,replace

clear

insheet kid vm1 using "D:\psa\_e2\Chapter5\data\vm1.dat"

sort kid

merge kid using chpt5\_2a

save chpt5\_2a, replace

tab vm1 kuse

use chpt5\_2a,replace

drop \_merge

sort kid

save chpt5\_2a,replace

clear

insheet kid vm2 using "D:\psa\_e2\Chapter5\data\vm2.dat"

sort kid

merge kid using chpt5\_2a

save chpt5\_2a, replace

tab vm2 kuse

use chpt5\_2a,replace

drop \_merge

sort kid

save chpt5\_2a,replace

clear

insheet kid vm3 using "D:\psa\_e2\Chapter5\data\vm3.dat"

sort kid

merge kid using chpt5\_2a

save chpt5\_2a, replace

tab vm3 kuse

use chpt5\_2a,replace

drop \_merge

sort kid

save chpt5\_2a,replace

clear

insheet kid vm4 using "D:\psa\_e2\Chapter5\data\vm4.dat"

sort kid

merge kid using chpt5\_2a

save chpt5\_2a, replace

tab vm4 kuse

use chpt5\_2a,replace

drop \_merge

sort kid

save chpt5\_2a,replace

clear

insheet kid pm using "D:\psa\_e2\Chapter5\data\pm.dat"

sort kid

merge kid using chpt5\_2a

drop \_merge

save chpt5\_2, replace

tab pm kuse

//Run imbalance (Guo, 2008) to generate dx and dxm

imbalance chpt5\_2 mratio96 kuse fm mratio96\_fm

imbalance chpt5\_2 mratio96 kuse vm1 mratio96\_vm1

imbalance chpt5\_2 mratio96 kuse vm2 mratio96\_vm2

imbalance chpt5\_2 mratio96 kuse vm3 mratio96\_vm3

imbalance chpt5\_2 mratio96 kuse vm4 mratio96\_vm4

imbalance chpt5\_2 mratio96 kuse pm mratio96\_pm

imbalance chpt5\_2 pcged97 kuse fm pcged97\_fm

imbalance chpt5\_2 pcged97 kuse vm1 pcged97\_vm1

imbalance chpt5\_2 pcged97 kuse vm2 pcged97\_vm2

imbalance chpt5\_2 pcged97 kuse vm3 pcged97\_vm3

imbalance chpt5\_2 pcged97 kuse vm4 pcged97\_vm4

imbalance chpt5\_2 pcged97 kuse pm pcged97\_pm

imbalance chpt5\_2 pcg\_adc kuse fm pcg\_adc\_fm

imbalance chpt5\_2 pcg\_adc kuse vm1 pcg\_adc\_vm1

imbalance chpt5\_2 pcg\_adc kuse vm2 pcg\_adc\_vm2

imbalance chpt5\_2 pcg\_adc kuse vm3 pcg\_adc\_vm3

imbalance chpt5\_2 pcg\_adc kuse vm4 pcg\_adc\_vm4

imbalance chpt5\_2 pcg\_adc kuse pm pcg\_adc\_pm

imbalance chpt5\_2 black kuse fm black\_fm

imbalance chpt5\_2 black kuse vm1 black\_vm1

imbalance chpt5\_2 black kuse vm2 black\_vm2

imbalance chpt5\_2 black kuse vm3 black\_vm3

imbalance chpt5\_2 black kuse vm4 black\_vm4

imbalance chpt5\_2 black kuse pm black\_pm

imbalance chpt5\_2 age97 kuse fm age97\_fm

imbalance chpt5\_2 age97 kuse vm1 age97\_vm1

imbalance chpt5\_2 age97 kuse vm2 age97\_vm2

imbalance chpt5\_2 age97 kuse vm3 age97\_vm3

imbalance chpt5\_2 age97 kuse vm4 age97\_vm4

imbalance chpt5\_2 age97 kuse pm age97\_pm

imbalance chpt5\_2 male kuse fm male\_fm

imbalance chpt5\_2 male kuse vm1 male\_vm1

imbalance chpt5\_2 male kuse vm2 male\_vm2

imbalance chpt5\_2 male kuse vm3 male\_vm3

imbalance chpt5\_2 male kuse vm4 male\_vm4

imbalance chpt5\_2 male kuse pm male\_pm

use mratio96\_fm

append using mratio96\_vm1

append using mratio96\_vm2

append using mratio96\_vm3

append using mratio96\_vm4

append using mratio96\_pm

append using pcged97\_fm

append using pcged97\_vm1

append using pcged97\_vm2

append using pcged97\_vm3

append using pcged97\_vm4

append using pcged97\_pm

append using pcg\_adc\_fm

append using pcg\_adc\_vm1

append using pcg\_adc\_vm2

append using pcg\_adc\_vm3

append using pcg\_adc\_vm4

append using pcg\_adc\_pm

append using black\_fm

append using black\_vm1

append using black\_vm2

append using black\_vm3

append using black\_vm4

append using black\_pm

append using age97\_fm

append using age97\_vm1

append using age97\_vm2

append using age97\_vm3

append using age97\_vm4

append using age97\_pm

append using male\_fm

append using male\_vm1

append using male\_vm2

append using male\_vm3

append using male\_vm4

append using male\_pm

list name dx dxm

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Stata Syntax for Section 5.8.3 (Post-Full-Matching Analysis Using Hodges-Lehmann Aligned Rank Test)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

//Chapter 5 Section 5.8.3 Post-Full-Matching Analysis of Outcome

//Hodges-Lehamann aligned rank test

cd "D:\psa\_e2\Chapter5\data"

set more off

//Run hodgesl (Guo, 2008) program to conduct Hodges-Lehmann test

// following full-matching

use chpt5\_2, clear

g fm\_s=string(fm)+"s"

save chpt5\_2\_new, replace

hodgesl chpt5\_2\_new lwss97 fm\_s kuse fm\_results

use fm\_results, replace

list if \_n <= 20

//dxm for the outcome variable is Cohen's d (i.e., effect size)

//Run imbalance (Guo, 2008) program to get Cohen's d

imbalance chpt5\_2 lwss97 kuse fm cohen\_d

use lwss97\_fm

rename dxm Cohen\_d

list

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Stata Syntax for Section 5.8.4 (Post-Pair-Matching Analysis Using Regression of Difference Scores)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

//Chapter 5 Section 5.8.4

//Regression of difference-scores based on pair-matched sample

cd "D:\psa\_e2\Chapter5\data"

set more off

use chpt5\_2, replace

drop if kuse==0

rename lwss97 y1

rename male male1

rename black black1

rename age97 age971

rename pcged97 pcged971

rename mratio96 mratio961

sort pm

keep pm kuse y1 male1 black1 age971 pcged971 mratio961 pcg\_id

save "c:\tmp\k1", replace

use chpt5\_2, replace

drop if kuse==1

rename lwss97 y0

rename male male0

rename black black0

rename age97 age970

rename pcged97 pcged970

rename mratio96 mratio960

sort pm

keep pm kuse y0 male0 black0 age970 pcged970 mratio960 pcg\_id

save "c:\tmp\k0", replace

merge pm using "c:\tmp\k1"

gen y=y1-y0

gen male=male1-male0

gen black=black1-black0

gen age97=age971-age970

gen pcged97=pcged971-pcged970

gen mratio96=mratio961-mratio960

regress y male black age97 pcged97 mratio96 , robust cluster(pcg\_id)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_