Example of Data Management in STATA

Carlos Noton *

January 18, 2012

```
clear
clear matrix
set mem 5G
set linesize 80
set matsize 500
capture log close
capture program drop _all
* THIS LINE CHANGE IN EVERY COMPUTER!!!   cd "C:\Users\Carlos Noton\Dropbox\mysel

log using "LOG_COFFEE_ZERO", replace

use "bc_260.dta",clear
*****************************
*FIXING THE INCOMPLETE STRINGS
*****************************
replace bcode="2264400264600"    if bcode=="22644002646"
replace bcode="2550080031800"    if bcode=="25500800318"
replace bcode="2550080288600"    if bcode=="25500802886"
replace bcode="2800031429300"    if bcode=="28000314293"
replace bcode="4300079830000"    if bcode=="43000798300"
replace bcode="4300079450000"    if bcode=="43000794500"
replace bcode="4300079460900"    if bcode=="43000794609"
replace bcode="5500074219100"    if bcode=="55000742191"
replace bcode="5500021003400"    if bcode=="55000210034"
replace bcode="5500021005800"    if bcode=="55000210058"
replace bcode="4300079450000"    if bcode=="43000794500"
replace bcode="7515707429700"    if bcode=="75157074297"
replace bcode="7515707451800"    if bcode=="75157074518"
replace bcode="7724370033200"    if bcode=="77243700332"
```

*C.Noton@Warwick.ac.uk

```
replace bcode="7323791031110"    if bcode=="732379103111"
replace bcode="8001172900000"    if bcode=="80011729"
replace bcode="8280520000270"    if bcode=="828052000027"
tab bcode
sort bcode
save "bc_260.dta", replace


outsheet using bc_260_example, replace
clear

insheet using "bc_260_example.out"

*INVARIANT CHARACTERISTICS OF THE COFFEE: grams, supplier, flavour, etc


use "fake_coffee_data.dta", clear
drop year week
*****************************
*FIXING THE INCOMPLETE STRINGS
*****************************

replace bcode="2264400264600"    if bcode=="22644002646"
replace bcode="2550080031800"    if bcode=="25500800318"
replace bcode="2550080288600"    if bcode=="25500802886"
replace bcode="2800031429300"    if bcode=="28000314293"
replace bcode="4300079830000"    if bcode=="43000798300"
replace bcode="4300079450000"    if bcode=="43000794500"
replace bcode="4300079460900"    if bcode=="43000794609"
replace bcode="5500074219100"    if bcode=="55000742191"
replace bcode="5500021003400"    if bcode=="55000210034"
replace bcode="5500021005800"    if bcode=="55000210058"
replace bcode="4300079450000"    if bcode=="43000794500"
replace bcode="7515707429700"    if bcode=="75157074297"
replace bcode="7515707451800"    if bcode=="75157074518"
replace bcode="7724370033200"    if bcode=="77243700332"
replace bcode="7323791031110"    if bcode=="732379103111"
replace bcode="8001172900000"    if bcode=="80011729"
replace bcode="8280520000270"    if bcode=="828052000027"
sort bcode
*tab bcode
```

```
merge bcode using bc_260

************************************************
* PROBLEMATIC bcodes *
* NOT IN bc260.dta BUT IN THE OTHER
tab bcode if _merge==1

* NOT IN fake_coffee_data.dta BUT IN THE OTHER
tab bcode if _merge==2
tabulate _merge

************************************************

label var bcode "Unique Product Code (string)"
label var Name  "Product Name"
label var Supplier  "Supplier"
label var Grams  "Grams"
label var q "Units"
label var p  "Retail Price"
label var date  "Date Code"
label var D_flavour "Flavoured"
label var D_decaf "Decaffeinated"
label var D_bean "Bean"
label var D_inst "Instant"
label var D_ground "Ground"
label var D_prom "Promotion"
label var population  "Local Population"
label var chain_id  "Chain Name"
label var store_id  "Store Code"
label var store_name  "Store Address"
label var location  "Location Name"
label var loc_id  "Location Code"
label define super  1 "BANDERA AZUL" ///
                    2 "ECONOMAX" ///
                    3 "EKONO" ///
                    4 "JUMBO" ///
                    5 "LAS BRISAS" ///
                    6 "LIDER" ///
                    7 "MAICAO" ///
                    8 "MONTECARLO" ///
                    9 "MONTSERRAT" ///
```

```
                    10 "OKMARKET" ///
                    11 "PUERTO CRISTO" ///
                    12 "RIBEIRO" ///
                    13 "STA ISABEL" ///
                    14 "UNIMARC"
label values chain_id super

*************************************************
*SPLITING THE BCODE NUMBER INTO 2 STRINGS then translating into 2 real numbers (t
*************************************************
gen v1= substr(bcode,1,6)
gen v2= substr(bcode,7,7)
gen a1= real(v1)
gen a2= real(v2 )
drop v1 v2
label var a1 "Code First Half"
label var a2 "Code Second Half"

*********************************************************************
*     LEAVING THE USEFUL SAMPLE
*********************************************************************
keep if _merge==3 & Grams!=. & q!=. & loc_id!=. & population!=. & p!=. & date!=.

*****************************************************
* Generating Dummies per Product
*****************************************************
tab  bcode, generate(prod_)

*****************************************************
* Generating Dummies per Supermarket
*****************************************************
tab chain_id, gen(sup_)

*****************************************************
* Generating Dummies per County
*****************************************************
tab location, gen(loc_)

*****************************************************
* Generating Dummies per Week
*****************************************************
```

```
tab date, gen(week_)

gen week = date-2372

*****************************************************
* Generating Dummies per Upstream Supplier
*****************************************************
tab Supplier, gen(ups_)

*********PRICE************
gen lp=log(p)
label var lp   "Log of Price"
sum
compress
local i = 0
while `i' <=4 {
gen dsize_`i' =0
local i = `i' + 1
}
replace dsize_0 =1 if Grams<=99 & Grams!=.
replace dsize_1 =1 if Grams>99  & Grams<=150 & Grams!=.
replace dsize_2 =1 if Grams>150 & Grams<=200 & Grams!=.
replace dsize_3 =1 if Grams>200 & Grams<=250 & Grams!=.
replace dsize_4 =1 if Grams>250 & Grams!=.
*Notar todos excluyentes: Checked!
tab dsize_0 dsize_1
tab dsize_1 dsize_2
tab dsize_1 dsize_3
tab dsize_2 dsize_3
tab dsize_3 dsize_4
sort bcode
drop _merge
save "coffee.dta", replace

use "iv_old.dta", clear
drop week
sort date
save "IV.dta", replace

use "coffee.dta", clear
sort date
```

```
merge date using IV.dta
tab _merge
keep if _merge==3
drop _merge

*******************************************
*  Constructing the prices in Chilean pesos *
*******************************************
local i1 = 1
while `i1' <=15 {
      gen Fiv_`i1'= ner*iv`i1'
  local i1 = `i1' + 1
}


++++++++++++DEMAND DATA SET++++++++++++
save "coffee_D.dta", replace
++++++++++++++++++++++++++++++++++++++

**Dummy DyS*******
gen     DyS=0
replace DyS=1 if chain_id==6 | chain_id==3

******************
*   CENCOSUD     *
******************
* Jumbo+Sta Isabel+Las Brisas, Montecarlo (acquired before 2005)
* Dummy CS *
gen     CS=0
replace CS=1 if chain_id==4 | chain_id==5 | chain_id==8 | chain_id==13
***********************************
*Economax (after 2nd half of 2006= 2006w26=2418.)

replace CS=1 if chain_id==2 & chain_id!=. & date>=2418
tab chain_id CS, miss

** KEEPING THE INTERSECTION WEEKS ONLY
keep if date>=2373  & date <= 2466
tab date
*** 633.148
```

```
* KEEPING THE MEDIUM SIZE ONLY
keep if Grams>99 & Grams<251
sum dsize*

*****************************************************************
* Constructing the Total Quantities per market (big players only)
*****************************************************************
gen qM=.
qui log off
local i1 = 1
while `i1' <=2 {
local i2 = 2444
      while `i2' <=2473 {
egen    X  =  sum(q) if loc_`i1'==1 & date==`i2'
replace qM =  X      if loc_`i1'==1 & date==`i2'
drop X
local i2 = `i2' + 1
}
local i1 = `i1' + 1
}
qui log on


*************************************************
* Population Analysis
*************************************************
sum qM population
gen coffee_perc=qM/population
sum coffee_perc, det
sort location
by location: sum coffee_perc qM pop

tab date location if coffee_perc>.10
tab date location if coffee_perc>.09
tab date location if coffee_perc>.08


**** Total Amount of Gram per consumer per week:
gen sh=q/(0.5*population)
gen sz=1- qM/(0.5*population)
sum qM pop sh sz,d
```

```
sort location
by location: sum sh sz pop

*****************************************
* Name of big players
*****************************************
hist sh
graph export "hist_sh_full.eps", as(eps) preview(on) replace

hist sz
graph export "hist_sz_full.eps", as(eps) preview(on) replace

tab Name if sh>.03
************************************************
*  Nominal Prices
************************************************
sum p, detail
*Histogram Price
hist p
graph export "hist_p_full.eps", as(eps) preview(on) replace

gen lsz=log(sz)
gen lsh=log(sh)
gen LS=log(sh/sz)

label var qM  "Total Demand per Market"
label var sh  "Market Share"
label var sz  "Outside Good Market Share"

label var lsh  "Log of Market Share"
label var lsz  "Log Outside Good Market Share"
label var LS   "Log(sh/sz)"


*Coffee Category
sum D_*

* Several Tabs

tab location
tab store_id
```

```
tab Supplier
tab chain_id
tab Supplier chain_id
tab Supplier CS
tab Supplier DyS
save "coffee_Example.dta", replace

use "coffee_Example.dta", clear
keep if q!=.


*****************************************************************
* Now constructing WEEKLY Quantities per the entire CITY, location no longer impo
*****************************************************************

gen qW_nat=.
qui log off
local i2 = 1
      while `i2' <=94 {
egen    X      =  sum(q) if  week==`i2'
replace qW_nat =  X      if  week==`i2'
drop X
local i2 = `i2' + 1
}
qui log on


gen sh_W_nat=q/qW_nat
gen Mks_W_nat=0
gen Hf_W_nat=0
qui log off
local i2 = 1
      while `i2' <=23 {
local i3 = 1
while `i3' <=4 {
egen    X   =  sum(sh_W_nat*week_`i2'*sup_`i3')
replace Mks_W_nat =  X  if week_`i2'==1 & sup_`i3'==1
drop X
local i3 = `i3' + 1
}
egen    X1   =  max(Mks_W_nat*Mks_W_nat*week_`i2'*CS)
```

```
egen    X2   =  max(Mks_W_nat*Mks_W_nat*week_‘i2’*DyS)
egen    X3   =  max(Mks_W_nat*Mks_W_nat*week_‘i2’*sup_1)
egen    X4   =  max(Mks_W_nat*Mks_W_nat*week_‘i2’*(1-CS)*sup_2)

replace Hf_W_nat  = X1+X2+X4   if week_‘i2’==1
drop X1 X2 X3 X4
  local i2 = ‘i2’ + 1
}
qui log on
dotplot Hf_W_nat, over(date)
graph export "Hf_W_nat.eps", as(eps) preview(on) replace

twoway scatter  p q
twoway (scatter p q) (lfit p q)
twoway (scatter p q, mlabsize(small)) (qfit p q )
twoway (scatter p q) (qfit p q)
graph matrix p q Grams
twoway (scatter Hf_W_nat date, xtitle(date vs HHI) title(Some Very Impertant Grap
log close
```