

Apéndice. Código PS1 (STATA)

2023-08-17

```
* PS1 Economía Aplicada
* Juan Diego Barnes \\ Diego Fasan \\ Francisco Legaspe \\ Rodrigo Martin

*cargar la base
clear all
global main "C:/Users/Usuario/Desktop/MAESTRIA/Economía Aplicada/TPs/TP1"
global input "$main/input"
global output "$main/output"
global programs "$main/programs"
use "$input/data_russia",clear

*****
**PUNTO 1
*****
***Vizualizar la base
*browse

** Corregimos los errores de entrada regulares en las variables
* Creamos una lista de las variables que tienen patrones regulares de error
local varlist econrk powrnk resprk satlif satecc highsc belief monage
               cmedin hprblm hosl3m wtchng evalhl operat hattac smokes
               alclmo waistc hhpres geo work0 work1 work2 ortho marsta1
               marsta2 marsta3 marsta4
* Loop para todas las varlist
quietly foreach var of local varlist {
    replace `var' = "1" if `var' == "one"
    replace `var' = "2" if `var' == "two"
    replace `var' = "3" if `var' == "three"
    replace `var' = "4" if `var' == "four"
    replace `var' = "5" if `var' == "five"
    replace `var' = "" if `var' == "."
    replace `var' = "" if `var' == ".a"
    replace `var' = "" if `var' == ".b"
    replace `var' = "" if `var' == ".c"
    replace `var' = "" if `var' == ".d"
    replace `var' = "1" if `var' == "Smokes"
    destring `var', replace
}

** Otras Variables sin patrones regulares
**** obese
replace obese = "1" if obese == "This person is obese"
replace obese = "0" if obese == "This person is not obese"
```

```

destring obese, replace

**** htself
replace htself =. if htself == .b
replace htself = . if htself == .d
destring htself, replace

*** sex
* Generamos indicadores del sexo
replace sex = "1" if sex == "male"
replace sex = "0" if sex == "female"
destring sex, replace
* Definir etiquetas para los valores
label define sex 0 "female" 1 "male"

* Aca probabementen hubiera sido mucho megor cambiar el nombre de
* la variable a male.

**** geo
* generamos indicadores de region
tabulate geo
tab geo, gen(geo)
rename geo1 Zona1
rename geo2 Zona2
rename geo3 Zona3
drop geo

** Estas son variables numericas por los que las revisaremos mas en cuidado
* height
sum height
* esta creeemos esta bien

**** hipsiz
* Primero rempalzamos las , por . en los numeros para poder extraerlos desp
replace hipsiz = substr(hipsiz, ",", ".", .)
* Extraer los números usando regex
split hipsiz, gen(hipsiz)
replace hipsiz = hipsiz3
drop hipsiz1 hipsiz2 hipsiz3
destring hipsiz,replace

**** totexpr
* exactamente el miso proceso que para hipsiz
replace totexpr = substr(totexpr, ",", ".", .)
split totexpr, gen(totexpr)
replace totexpr = totexpr3
drop totexpr1 totexpr2 totexpr3
destring totexpr,replace

**** tincm_r

```

```

* Pensar si tienen mas sentido remplazamos primero las "," por 0
* replace tincm_r = "0" if tincm_r == ","
* Remplazamos las , de los numeros por .
replace tincm_r = subinstr(tincm_r, ",", ".", .)
destring tincm_r, replace

****Vizualizar la base
* browse
*Miramos que quedo todo bien

*****
*PUNTO 2
*****
* Chequeamos por valor faltantes
* generamos un loop que nos cuente cuantas NA, ., hay en cada variable

*foreach var of varlist _all {
*   local missing_count = 0
*   quietly count if missing(`var')
*   local missing_count = r(N)
*   di "Missing values in `var': " `missing_count'
*}

* Otra alternativa (mas eficiente)
* ssc install mdesc
mdesc

*****
*PUNTO 3
*****

** totexpr (El Gasto Real de los hogares)
summ totexpr // TIENE VALORES NEGATIVOS (IRREGULAR DATA)
replace totexpr =. if totexpr < 0

** tincm_r (Ingreso Real de los hogares)
summ tincm_r // TIENE VALORES NEGATIVOS (IRREGULAR DATA)
replace tincm_r =. if tincm_r < 0

* Estamos agregando mas missings con esto, chequeamos cuantos:
mdesc

*****
*PUNTO 4
*****
*** Ordenamos los datos segun el siguiente criterio:
* La primera variable que aparezca en la base debería ser el id del individuo,
* la segunda el sitio (site) donde se encuentra
* y la tercera el sexo (sex).
order id site sex totexpr
* Luego, ordenen las filas de mayor a menor según totexpr.
gen neg_totexpr = -totexpr

```

```

sort neg_totexpr
drop neg_totexpr
****Vizualizar la base
browse

*****
*PUNTO 5
*****
* Queremos generar estadísticas descriptivas de sexo, la edad en años,
* la satisfacción con la vida, la circunferencia de la cintura,
* la circunferencia de la cadera y el gasto real

* Primero necesitamos generar la variable de edad de años
gen yearage = floor(monage/12)

summ sex
summ yearage, detail
summ satlif, detail
summ hipsiz, detail
summ totexp, detail

* Generamos labels para hacerla mas facil de leer
label var sex "Hombres"
label var yearage "Edad en Años"
label var satlif "Satisfacción con la vida"
label var hipsiz "Tamaño de caderas"
label var totexp "Gasto Real del Hogar"
* Armamos una unica tabla de descriptivas
estpost summarize sex yearage satlif hipsiz totexp, listwise
esttab using "$output/Table 1.tex", cells("mean sd min max") ///
nomtitle nonumber replace label
* Otra alternativa
* tabstat sex yearage satlif hipsiz totexp, stats(mean range min max)

*****
*PUNTO 6
*****
* Setup general para los graficos:
grstyle init // http://repec.sowi.unibe.ch/stata/grstyle/getting-started.html
grstyle set horizontal
graph set window fontface "Palatino Linotype"
grstyle color background white
grstyle color heading black // Title in black
* Comparemos las distribuciones entre sexo
* Compare the distribution of wage across genders
tway (kdensity hipsiz if sex==1, lwidth(0.7)) ///
      (kdensity hipsiz if sex==0, lwidth(0.7)), ///
legend(order(1 "Hombres" 2 "Mujeres" )) title(" ") ///
yttitle("Densidad") xttitle("Tamaño de caderas")
graph export "$output/hips_histogram_menvswomen.png", replace

```

```

* test de diferencia de medias entre sexo. recordar que 1 es hombre 0 mujer
ttest hipsiz, by(sex)

** Por las dudas realizamos el test para distintas varianzas
ttest hipsiz, by(sex) unequal

* Se rechazan las hipotesis nulas de que las mujeres tienen caderas mas grandes
* Tambien se rechaza que la hipotesis nula de que no existen diferencias
* No se puede rechazar la hipotesis nula de que las caderas de los hombre
* son mas chicas

*****
*PUNTO 7
*****
* Probemos modelos para explicar la felicidad
* Realizamos una primera estimacion, solo un gues no es es un modelo propuesto
* reg satlif econrk totexpr tincm_r satecc height hipsiz highsc yearage cmedin
    wtchng evalhl smokes alclmo waistc hhpres ortho marsta2 marsta3
    marsta4 Zona2 Zona3

* Dividimos el Gasto Real y el ingreso por 100 unidades monetarias
* para poder interpretar mejor el coef

gen tincm_100 = tincm/100
label var tincm_100 "Ingreso Real del Hogar"

gen totexp_100 = totexp/100
label var totexp_100 "Gasto Real del Hogar"

* Ahora nos preguntamos que tiene sentido que se correlacione con la felicidad
* realizamos una matriz de scatter de las variables principales que consideramos
* de interes

* Generamos los labels faltantes para el grafico
label var satecc "Satisfaccion Econ."
label var cmedin "Cobertura Medica"
label var wtchng "Cambio de Peso Ult Año"
label var evalhl "Autoeval. salud"
label var smokes "Fumador"
label var alclmo "Toma Alcohol"

graph matrix satlif tincm_100 yearage, diagonal() title(" ")
graph export "$output/scater_matrix.png", replace

* Modelo 1"
* En el primer modelo queremos ver la relacion de la satisfaccion con la vida
* con el ingreso y el gasto de
reg satlif tincm_100 totexp_100 satecc yearage cmedin evalhl marsta1 marsta2
    marsta3 marsta4 Zona2 Zona3, robust

* La edad puede tener un efecto no lineal

```

```
gen sq_yearag = year^2
label var sq_yearag "Años^2"

* Modelo 2
reg satlif tincm_100 totexp_100 satecc yearage sq_yearag cmedin evalhl smokes
    alclmo marsta1 marsta2 marsta3 marsta4 Zona2 Zona3, robust
```