

```

/*****

```

## Problem Set 1

Universidad de San Andrés  
Economía Aplicada  
2024

```

*****/

```

```

* ENVIRONMENT

```

```

*=====

```

```

clear all

```

```

global main ""

```

```

global input "$main/input"

```

```

global output "$main/output"

```

```

use "$input/data_russia.dta", clear

```

```

*=====

```

```

* 1)

```

```

* Limpieza

```

```

*=====

```

```

*****

```

```

// Identificamos los problemas

```

```

// estan OK : marsta1 marsta2 marsta3 marsta4 height htself inwgt site

```

```

// corregir :

```

```

/*

```

```

- pasar texto a numero (eg 'one' -> 1) + destring

```

```

    geo hattac operat evalhl wtchng operat satlif resprk powrnk econrk

```

```

- binaria en texto

```

```

    smokes obese sex

```

```

- binaria como string sin missings

```

```

    ortho work0 work1 work2 operat hprblm

```

```

- binaria como string con missings

```

```

    hhpres alclmo hattac hhpres hosl3m cmedin highsc

```

```

- tab para hacer varias binarias

```

```

    geo

```

```

- continua como string

```

```

    tincm_r waistc monage belief satecc

```

```

- continua con prefijo string

```

```

    totexpr hipsiz

```

```

*/

```

```

***** corregimos los problemas:

```

```

// texto a numero (eg 'one' -> 1)

```

```

foreach var of varlist geo hattac operat evalhl wtchng operat satlif resprk ///

```

```

powrnk econrk{

```

```

    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' = "6" if `var' == "six"
replace `var' = "7" if `var' == "seven"
replace `var' = "8" if `var' == "eight"
replace `var' = "9" if `var' == "nine"
replace `var' = "10" if `var' == "ten"
}

// variables que no deberian tener textos
replace smokes="1" if smokes == "Smokes"
replace sex="1" if sex == "male"
replace sex="0" if sex == "female"
replace obese="1" if obese == "This person is obese"
replace obese="0" if obese == "This person is not obese"
replace hipsiz = substr(hipsiz, 18, .)
replace totexpr = substr(totexpr, 19, .)

// destringeamos
foreach var of varlist geo hattac operat evalhl wtchng satlif resprk ///
powrnk econrk ortho work0 work1 work2 hprblm hhpres alclmo hosl3m cmedin ///
highsc sex tincm_r waistc monage belief satecc hipsiz obese smokes totexpr ///
marsta1 marsta2 marsta3 marsta4{
    destring `var', gen(`var'_) dpcomma
    drop `var'
    rename `var'_ `var'
}

// esta deberian ser 3 binarias porque son 3 categorias (no especificadas)
tab geo, gen(geo_area)
drop geo
// puede ser que belief tambien pero no especifica
// interpreto que los valores 1-5 son para qué tan creyentes son
// y no que grupo religioso pertenecen
*=====
* 2)
* MISSINGS
*=====
mdesc
// tincm_r, htself y totexpr tienen 6.5% de missings
// monage y obese tienen 7.2% de missings
*=====
* 3)
* IRREGULARES
*=====
/*
Corregir:
    - no deberian ser cero o negativos (gasto, peso, ingresp)
      tincm_r totexpr inwgt
*/

foreach var of varlist tincm_r totexpr inwgt{
    replace `var' = . if `var' <=0
}
```

```
// Gastos mayores a ingresos
replace totexpr = . if totexpr > tincm_r

mdesc // quedaron con muchisimos missings
*=====*
* 4)
* ORDEN
*=====*
order id site sex
sort totexpr
*=====*
* 5)
* ESTADISTICAS DESCRIPTIVAS
*=====*
gen yage = floor(monage/12)
// Resumir variables sex yage satlif waistc totexpr
label var sex "Sexo"
label var yage "Edad en años"
label var satlif "Satisfacción con la vida"
label var waistc "Circunferencia de la cadera"
label var totexpr "Gasto total real"

* Exportar
estpost summarize sex yage satlif waistc totexpr, listwise
esttab using "$output/tables/ej5.tex", cells("mean sd min max") ///
collabels("Mean" "SD" "Min" "Max") nomtitle nonumber replace label
*=====*
* 6)
* HIPS DON'T LIE
*=====*
***** a)
quietly summarize hipsiz if sex == 1, detail
local median_female = r(mean)

quietly summarize hipsiz if sex == 0, detail
local median_male = r(mean)

* achicado para que se vea mejor
gen hs = hipsiz if ( (hipsiz < 140) & (hipsiz > 60) )

twoway (kdensity hs if sex==1, color(red)) ///
(kdensity hs if sex==0, color(blue)), ///
legend(label(1 "Females") label(2 "Males")) ///
title("Distribution of Hip Size") ///
ytitle("Density") xtitle("Hip Size") ///
xline(`median_female`, lcolor(blue) lwidth(medium)) ///
xline(`median_male`, lcolor(red) lwidth(medium)) ///
xscale(range(65 135))
graph export "$output/figures/hipsiz_density_menvswomen_means.png", replace

***** b)
ttest hipsiz, by(sex)
estpost ttest hipsiz, by(sex) listwise esample
esttab using "$output/tables/hips.tex", wide nonumber mtitle(Difference) ///
```

```

cells("b count se t df_t p") ///
collabels("Diff. mean" "Obs" "Diff. Sd" "T-Stat" "df" "p-value") replace label
*=====*
* 7)
* REGRESION
*=====*
***** a)
***** Graficos
graph box yage, over(satlif) ///
    title("Edades para cada grupo de felicidad") ///
    ytitle("Edad") b1title("Satisfaccion") ///
    graphregion(lcolor(black) lwidth(medium))
graph export "$output/figures/box_age_satlif.png", replace

graph bar (count), over(sex, label(angle(45))) over(satlif) ///
    asyvars ///
    bar(1, color("255 205 147")) bar(2, color("71 125 171")) ///
    title("Satisfaccion por sexo") ///
    legend(order(1 "Male" 2 "Female")) ///
    ytitle("") b1title("Satisfaccion") ///
    graphregion(lcolor(black) lwidth(medium))
graph export "$output/figures/bar_sex_satlif.png", replace

***** b)
* Analizamos las variables
/*
***** basicas
monage sex geo_area1 geo_area2           // características básicas
waistc hipsiz height                     // características físicas

***** salud
obese inwgt wtchng                       // Peso
cmedin                                   // Seguro Médico
hprblm hosl3m operat hattac              // Si en ultimo tiempo...
evalhl                                   // Salud autoevaluacion
smokes                                  // Fuma
alclmo                                  // Alcohol

***** economicas
work0 work1 work2                       // Estado Laboral
totexpr                                  // Gasto Total
tincm_r                                  // Ingreso Total

***** personales
satecc
highsc                                  // Satisfaccion con su economia
htself                                  // Altura autoevaluacion
econrk powrnk resprk                    // Rank Ladder
marsta1 marsta2 marsta3 marsta4         // Estado Civil
belief ortho                             // Creencia religiosa
*/

```

\* Vamos a hacer una especificacion

```
label var geo_area1 "Geo 1"  
label var geo_area2 "Geo 2"
```

\*\*\*\*\* Especificacion 1

```
reg satlif yage sex tincm_r geo_area1 geo_area2 ///  
/// las obvias: sexo edad ingreso ciudad  
///  
resprk /// econrk powrnk podrian ser tambien, esta posiblemente esta mejor  
///relacionada con percepcion de satisfaccion  
cmedin /// tener seguro te relaja de preocuparte si tienes un  
///accidente por ahí pero va a estar super correlacionada con ingreso  
///pero no con salud, lo muestra finkelstein  
///  
evalhl /// esta (o podría ser la de hospitalizado/problemas) para ver si el se  
///considera saludable  
///  
work0 work1 /// status laboral -> trabajar te hace mas satisfecho? (mas ingresos)  
///  
marsta1 marsta2 marsta3 // estar casado te puede hacer mas satisfecho,  
//divorciado o viudo(omitida) quizás menos  
  
outreg2 using "$output/tables/Table1.tex", replace label
```

\*\*\*\*\* Especificacion 2

```
reg satlif yage sex tincm_r geo_area1 geo_area2 marsta1 marsta2 marsta3 ///  
work0 work1 ///  
evalhl hprblm hosl3m operat hattac obese inwgt wtchng belief econrk highsc ///  
outreg2 using "$output/tables/Table2.tex", replace label  
*=====*
```