



```

clear all

// global main "C:\Users\Hp Support\Videos\03 - Cursos\06 - Economía Aplicada\03 - Tra
> bajos prácticos\Applied_Econometrics\PS1"
// global input "$main/input"
// global output "$main/output"

global main "C:\Users\geron\Documents\GitHub\Applied_Econometrics\PS1"
global input "$main/input"
global output "$main/output"

*-----
* 1 Structure of the log file name
*-----

        global dofilename "tp1"
        cap log close
        local td: di %td CY-N-D date("$S_DATE", "DMY")
        local td = trim("`td'")
        local td = subinstr("`td'", " ", " ", .)
        local td = subinstr("`td'", ":", ":", .)
        log using "output/${dofilename}-`td'_1", text replace
        local today "`c(current_time)'"
        local curdir "`c(pwd)'"
        local newn = c(N) + 1

*-----
* 2 Desarrollo de Preguntas
*-----

use "input/data_russia", clear

* Pregunta 1: Utilizando los comandos replace, split, destring y encode, emprolijen la
> base:

* Convirtiendo las variables en formato string a numérico
// Inspeccionando variables alfanumericas
tab econrk
tab powrnk
tab resprk
tab satlif
tab wtchng
tab evalhl
tab operat
tab hattac
tab geo

// Reemplazamos las comas por puntos, separando aquellas variables que unian erroneame
> nte texto y valores numericos.
split hipsiz, parse("") g(hipsiz)
replace hipsiz = hipsiz3
replace hipsiz = subinstr(hipsiz, ",", ".", .)
drop hipsiz1 hipsiz2 hipsiz3

split totexpr, g(totexpr)
replace totexpr = totexpr3
replace totexpr = subinstr(totexpr, ",", ".", .)
drop totexpr1 totexpr2 totexpr3

replace tincm_r = subinstr(tincm_r, ",", ".", .)

```

```
// Reemplazamos los caracteres especificos en cada variable mediante un loop. Las obse
> rvaciones presentadas en texto las volvimos numericas y aquellas variables que por s
> us características eran binarias las pusimos en el formato 0-1.
foreach x of varlist sex econrk obese powrnk resprk satlif wtchng evalhl operat hattac
> smokes tincm_r hipsiz totexpr geo{
    replace `x' = "1" if (`x'== "one" )
    replace `x' = "2" if (`x'== "two" )
    replace `x' = "3" if (`x'== "three" )
    replace `x' = "4" if (`x'== "four" )
    replace `x' = "5" if (`x'== "five" )
    replace `x' = "." if (`x'== ".b" ".c" ".d" )
    replace `x' = "," if (`x'== "," )
    replace `x' = "1" if (`x'== "Smokes" )
    replace `x' = "0" if (`x'== "female" )
    replace `x' = "1" if (`x'== "male" )
    replace `x' = "0" if (`x'== "This person is not obese" )
    replace `x' = "1" if (`x'== "This person is obese" )
}

// Convertimos aquellas variables que se encontraban en formato string a formato numer
> ico via el comando destring.
ds, has(type string)
foreach var in `r(varlist)' {
    destring `var', replace
}

// Finalmente, luego del paso anterior chequeamos que efectivamente no haya quedado al
> guna variable en formato string.
ds, has(type string)

// Chequeamos unicidad de las observaciones, es decir, que no existan valores repetido
> s.
isid id

*Ejercicio 2:
// Via el comando mdesc visto en clase observamos que variables presentaban missing va
> lues y, a partir del uso de un loop, mostramos aquellas que cumplen con el criterio
> establecido en el tp1, de tener missing values que representen el 5% de los datos o
> mas.

// ssc install mdesc // instalamos
ds, has(type numeric)
local var_miss
foreach var in `r(varlist)' {
    mdsc `var'
    if `r(percent)' > 5{
        local var_miss `var_miss' `var'
        display "`var_miss'"
    }
}
mdsc `var_miss'

// Las variables que muestran missing values superiores al 5% son: monage, obese, htse
> lf, totexpr y tincm_r.

* Ejercicio 3:

// Identificamos si existen variables que presenten valores negativos para luego poder
> concluir acerca de la racionalidad de ello.
ds, has(type numeric)
local varegative
foreach var in `r(varlist)' {
    summ `var' if `var'<0
    if `r(N)' > 0 {
        local varegative `varegative' `var'
    }
} // observamos variables numéricas con valores negativos.
// Variables con valores negativos
summ `varegative'
```

```
// Summary valores negativos
ds, has(type numeric)
foreach var in `varegative' {
    display "Variable :==> `var'"
    replace `var'=. if(`var'<0)
} // reemplazamos
/*
Las variables cuyas observaciones fueron reemplazdas
por missing son imposibles de tener valores negativos dado la descripción
de la variable.
totexpr : HH Expenditures Real
tinclr : HH Income Real
Cabe especificar que no tenemos variables con valores superlativos
que por lo general suelen representar valores missing (eg. 99999).
*/

* Ejercicio 4:
// utilizamos el comando order visto en clase para ordenar la base de datos de acuerdo
> al criterio solicitado. Ademas con gsort ordenamos los datos de acuerdo al valor de
> la variable totexpr, desde el mayor hasta el menor.
order id site sex
gsort -totexpr

* Ejercicio 5:

foreach var of varlist sex monage satlif waistc hipsiz totexpr {
    summarize `var'
}
//Aquí básicamente le dimos una etiqueta útil a cada una de las variables, para que lu
> ego sea más fácil comprender que representa cada una.
label variable sex "Sexo"
label variable monage "Edad en meses"
label variable satlif "Satisfacción con la vida"
label variable waistc "Circunferencia de la cintura"
label variable hipsiz "Circunferencia de la cadera"
label variable totexpr "Gasto real"

// Exportamos a latex algunas estadísticas descriptivas útiles de las variables.
estpost summarize sex monage satlif waistc hipsiz totexpr, listwise

#delimit ;
    esttab using "output/tables/Table1.tex", replace cells("mean sd min max")
    collabels("Mean" "SD" "Min" "Max" )
    nomtitle nonumber label note("R");
#delimit cr

* Ejercicio 6:
// con el comando kdensity visto en clase y con el uso del twoway, mostramos en un mis
> mo gráfico la distribución de la circunferencia de la cadera, discriminando por sexo
> . En ella se puede ver que, a pesar de que ambas distribuciones están centradas apro
> ximadamente en un mismo valor, la distribución para los hombres presenta una menor d
> ispersión.
#delimit ;
    twoway (kdensity hipsiz if sex==1) ||
    (kdensity hipsiz if sex==0), legend( label(1 "Hombre") label(2 "Mujer") )
    title("Distribución de la Circunferencia de la Cadera" )
    ytitle("Densidad") xtitle( "" ) ;
    graph export "output/figures/hipsiz_histogram_menvswomen.png", replace ;
#delimit cr
// A partir de acá usamos el comando ttest para realizar una diferencia de medias y pr
> obar las hipótesis relevantes. Luego exportamos la tabla, tratando de cambiar el for
> mato para que se vea de una forma más intuitiva.
ttest hipsiz, by(sex)
eststo test1: estpost ttest hipsiz, by( sex )

#delimit ;
```

```

global note_nv
" \item Note: El P-value responde a la Ha: diff > 0. Siendo diff
la diferencia de medias entre mujeres y hombres.";

esttab test1
using "output/tables/ttest1.tex", replace
cell(
    b( pattern( 1 ) star pvalue( p_u ) fmt(4))
    se( pattern(1 ) par fmt(2) ) )
starlevels(\sym{*} 0.10 \sym{**} 0.05 \sym{***} 0.01)
varlabels(
    hipsiz "Circunferencia de la Cadera" )
mtitle("Sample Completo")
collabels( none )
prehead( "\begin{sidewaystable}[htbp] \fontsize{10}{6}\selectfont \centeri
> ng \protect
\captionsetup{justification=centering}
\caption{ Test de Diferencia de Medias}
\label{ttest}
{ \begin{threeparttable}\begin{tabular}{l*{1}{c}}" \to
> prule )
posthead(\hline)
postfoot(\hline \hline "\multicolumn{2}{l}{\footnotesize Errores stand
> ard en parentesis.}\\"
"\multicolumn{2}{l}{\footnotesize \sym{*} \ (p<0.10\),
\sym{**} \ (p<0.05\), \sym{***} \ (p<0.01\)}\\" \end{tab
> ular}
\begin{tablenotes}
\begin{footnotesize}
${note_nv}
\end{footnotesize}
"\end{tablenotes}
\end{threeparttable}
\end{sidewaystable}") ;

#delimit cr

* Ejercicio 7:
//Finalmente, elegimos variables que consideramos relevantes para hacer un analisis de
> regresion y, planteamos dos casos, uno con pocos controles y otro con mayor cantida
> d de ellos.
//Tambien exportamos la tabla con los resultados de la regresion, modificando su forma
> to para que la presentacion sea mucho mas clara.
reg satlif htself totexpr i.econrk i.cmedin i.work1
est store reg1
reg satlif htself totexpr i.marsta3 i.econrk i.cmedin i.work1 i.ortho
est store reg2

#delimit ;
esttab reg1 reg2 using "output/tables/first_model.tex", replace
eqlabels(" " ) ///
style(tab) order( ) mlabel(,none) ///
cells(b(label(coef.) star fmt(%8.3f) ) se(label((z)) par fmt(%6.3f))) ///
starlevels(* 0.10 ** 0.05 *** 0.01) ///
s(N r2, label( "N" "R^2") fmt(%9.0gc %6.3f) ) ///
collabels(none) /// No column names within model
delim("&") /// Type of column delimiter
noobs /// Do not show number of observation used in model
nomtitle ///
label ///
drop( _cons 1.econrk 0.cmedin 0.work1 0.marsta3 0.ortho) ///
width(1.5\hsize) ///
nogaps /// No gaps between rows
booktabs /// Style
nonote /// No notes
varlabels(htself "Altura Reportada"
totexpr "Gastos Totales"
1.cmedin "Con Seguro Medico"
1.work1 "Dejó el Trabajo"
2.econrk "Escala 2"
3.econrk "Escala 3"
4.econrk "Escala 4"
5.econrk "Escala 5"

```

```

        6.econrk "Escala 6"
        7.econrk "Escala 7"
        8.econrk "Escala 8"
        9.econrk "Escala 9"
        1.marsta3 "Divorciado"
        1.ortho "Orthodoxo"
    ) ///
    mgroups( "Modelo 1" "Modelo 2" , pattern( 1 1) ) ///
    nonumbers ///
    refcat( 2.econrk "\Gape[0.25cm][0.25cm]{ \underline{ \textbf{ \textit{ Escala
> de Rango Económico } } }" )" /// Subtitles
    , nolabel) /// Subtitles
    prehead("\begin{table} \small \centering \protect \captionsetup{justification=
> centering} \caption{\label{tab:table1} Especificaciones del Modelo }" "\noindent\res
> izebox{\textwidth}{!}{ \begin{threeparttable}" "\begin{tabular}{lcc}" \toprule) ///
    postfoot(\hline \end{tabular} ///
    \begin{tablenotes} ///
    \begin{footnotesize} ///
    ${note} ///
    \end{footnotesize} ///
    "\end{tablenotes} \end{threeparttable} } \end{table}") ;

#delimit cr

// Por ultimo, presentamos un diagrama de dispersion del regresando contra algunos reg
> resores, para de esa manera poder observar de que manera se relacionan entre si.
#delimit ;
    twoway (scatter satlif htself ),
    ytitle("Satisfacción con la Vida") xtitle( "Altura" ) ;
    graph export "output/figures/Altura.png", replace ;

    twoway (scatter satlif totexpr ),
    ytitle("Satisfacción con la Vida") xtitle( "Gastos Totales" ) ;
    graph export "output/figures/gastos.png", replace ;

    twoway (scatter satlif econrk ),
    ytitle("Satisfacción con la Vida") xtitle( "Escala de Rango Económico" ) ;
    graph export "output/figures/rango_economica.png", replace ;

    graph box satlif, over(cmedin, relabel( 1 "Sin Seguro" 2 "Con Seguro"))
    medtype(ccline) medline( lcolor("red"))
    box( 1, fcolor( "white") lcolor(black) ) ;
    graph export "output/figures/boxplot_cmedin.png", replace ;

    graph box satlif, over(work1, relabel( 1 "No dejó el Trabajo" 2 "Dejó el Traba
> jo"))
    medtype(ccline) medline( lcolor("red")) box( 1, fcolor( "white") lcolor(black)
> ) ;
    graph export "output/figures/boxplot_work1.png", replace ;

#delimit cr

*****
*** PART 3: Log
*****/
log close

```