



Universidad de
San Andrés

PS1 Economía Aplicada

Elaborado por:

**Casiano, Denys
Quispe, Anzony
Rigirozzi, Gonzalo
Sambrana, Gerónimo**

1. Repaso Teorico

1.1. Pregunta 1

De acuerdo al Teorema de Gauss Markov, el estimador de MCO es MELI si cumple con los supuestos clásicos son:

- Linealidad: La variable dependiente es una función lineal de las variables especificadas en el modelo. La especificación debe ser lineal en sus parámetros.
- Exogeneidad: Me garantiza que el estimador sea insesgado ($E(u|X) = 0$).
- Homocedasticidad: cuando la varianza del error condicional a las variables explicativas es constante a lo largo de las observaciones ($V(u_i|X) = \sigma^2$).
- No correlacion Serial: Los errores de un periodo no guardan relacion con errores de otro periodo ($Cov(u_i, u_j|X) = 0$). Es decir, el termino de error de este periodo no correlaciona con el termino de error del periodo anterior.
- No multicolinealidad: Me garantiza que exista un estimador de MCO ($\rho(X) = K$).

1.2. Pregunta 2

El supuesto que garantiza que el estimador sea insesgado, es el de exogeneidad $E(u|X) = 0$. Por otra parte el supuesto que necesitamos para garantizar la existencia del estimador es el de multicolinealidad. Si hay multicolinealidad perfecta, deja de existir el estimador de MCO.

1.3. Pregunta 3

Cuando no se cumple homocedasticidad la varianza deja de ser constante. Los coeficientes siguen siendo lineales e insesgados pero ya no poseen mínima varianza. No son eficientes. Por otro lado, cuando no se cumple el supuesto de No autocorrelacion, los términos de perturbación aleatoria no son independientes entre si, sino que existe una correlación entre la perturbación de un período y la del período anterior. Los errores estandar del estimador no seran los más eficientes. El estimador seguirá siendo insesgado, pero no sera el mejor posible.

1.4. Pregunta 4

Si el termino de error μ es normal, por la linealidad del modelo Y hereda la linealidad de μ . Y a su vez $\hat{\beta}$ hereda la linealidad de Y .

1.5. Pregunta 5

La forma más común que tienen los economistas para solucionar los problemas a los que se enfrentan es hacer supuestos clave. Un ejemplo de esto puede notarse en inferencia. Resulta que, para hacerla, es necesario suponer como se distribuye el error, condicional en los valores de las variables explicativas X .

Sin embargo esto no quiere decir que estrictamente siempre haga falta el supuesto. De hecho, en el contexto de economía aplicada, la aversión a los supuestos hizo que se buscara una manera de solucionar el problema que no implique realizar supuestos acerca de la distribución del error condicional en las X .

Entonces, haciendo uso correcto de la ley de los grandes numeros así como del teorema central del limite, se tiene la capacidad de conocer la distribución exacta del parametro que se estima mediante OLS. Como esta distribución sigue siendo normal, se puede hacer inferencia sin hacer un supuesto fuerte acerca de la distribución. Todo esto en un contexto en el cual se puede contar con un tamaño muestral considerable.

1.6. Pregunta 6

Falso, un estimador puede ser las dos cosas al mismo tiempo.

Consistencia lo probamos usando Ley de Grandes Numeros y Normalidad asintótica usando Ley de Grandes Numeros y Teorema Central del límite.

1.7. Pregunta 7

Las propiedades de consistencia e insesgadez de los estimadores difieren en un aspecto central. El hecho es que para lograr la insesgadez del estimador es necesario hacer una serie de supuestos, dentro de ellos que la esperanza de u condicional en x sea igual a 0 ($E(u|x) = 0$).

Por otra parte, la consistencia no necesita hacer supuestos sobre la distribución del componente de error, sino que solamente indica que al crecer indefinidamente el tamaño de la muestra, el estimador debería converger en probabilidad al verdadero parametro poblacional, basandose en la Ley de Grandes Numeros.

Hay que tener en cuenta que puede cumplirse la consistencia sin que sea posible la insesgadez. Suponiendo consistencia estoy seguro de que mis resultados son buenos, simplemente teniendo un tamaño muestral lo suficientemente grande y sin

imponer grandes supuestos.

El supuesto para insesgadez es un supuesto mucho más fuerte, que muchas veces ni siquiera es necesario de hacer. Sin embargo, con muestras chicas necesito un supuesto más fuerte para la buena propiedad de mi estimador, aunque eso sea a costa de conclusiones mucho más dudosas y cuestionables. Si tengo acceso a un N grande, siempre conviene hacer el supuesto más débil en esta literatura, en el presente caso, el supuesto para consistencia.

2. Primeros pasos en STATA:

2.1. Ejercicios del 1 al 4:

Todos los ejercicios de este apartado se desarrollaron y contestaron en el dofile de STATA. Revisar al final del documento.

2.2. Ejercicios 5:

El cuadro 1 muestra las estadísticas descriptivas. Se puede ver un promedio de edad en meses de 556 meses (correspondiente a 46.33 años aproximadamente), con un desvío standard de 213.6 meses (17.75 años aproximadamente). La media de satisfacción con la vida tiene una media de 2.45, y un desvío de 1.11. También circunferencia de cintura tiene una media de 85.64 cm, y una desviación standard de 13.54 cm. Mientras que la circunferencia de cadera tiene una media de 101.47 cm y desvío de 11.5 cm.

Cuadro 1: Estadísticas descriptivas

	Mean	SD	Min	Max
Sexo	.4180432	.4933418	0	1
Edad en meses	556.1557	213.6785	216	1207
Satisfacción con la vida	2.449386	1.113896	1	5
Circunferencia de la cintura	85.63537	13.54693	37	168
Circunferencia de la cadera	101.4756	11.5034	40	180
Gasto real	7656.016	8914.258	147.83	128868.8
Observations	2361			

2.3. Ejercicio 6:

La figura 6 muestra 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 aproximadamente en un mismo valor, la distribución para los hombres presenta una menor dispersión.

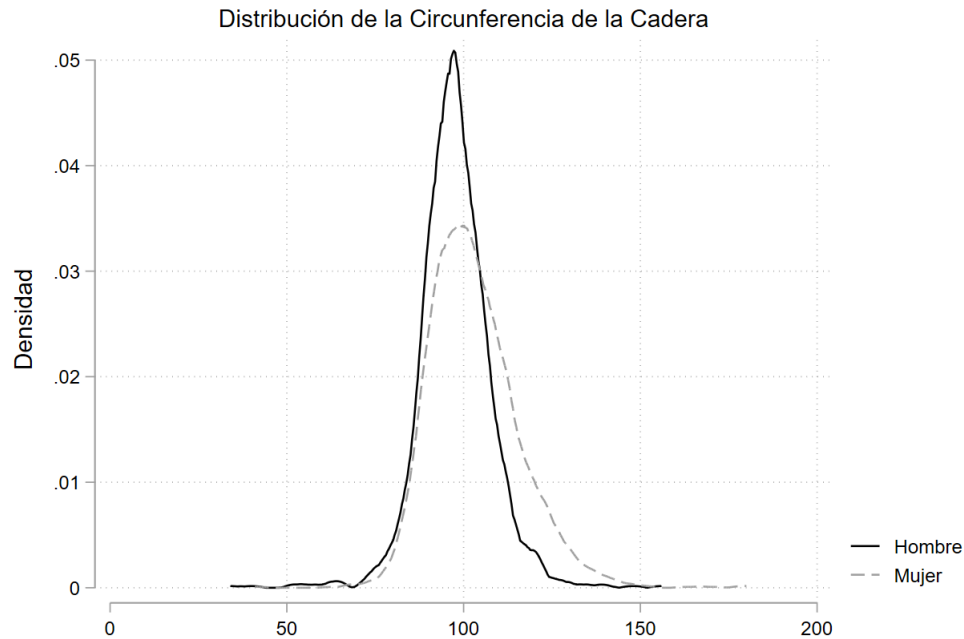


Figura 1: Distribución de la circunferencia de la cadera

EL cuadro 2, muestra los resultados del test de diferencia de medias que tiene como hipótesis nula que la diferencia de medias (en la circunferencia de la cadera) entre mujeres y hombres es cero fue rechazada al 99 % de confianza, enfrentandola con la hipótesis alternativa donde la diferencia de medias fue positiva. Es decir, el resultado del test fue que en promedio las mujeres tienen mayor circunferencia de la cadera que los hombres.

Cuadro 2: Test de Diferencia de Medias

	(1) Sample Completo
Circunferencia de la Cadera	5.3132*** (0.45)
<i>N</i>	2797

Errores standard en parentesis.

* $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$

Note: El P-value responde a la H_a : $\text{diff} > 0$. Siendo diff la diferencia de medias entre mujeres y hombres.

2.4. Ejercicio 7:

La figura 2 presenta gráficos que relacionan la variable dependiente (Satisfacción con la vida) y covariables que consideramos importantes para explicarla. De las figuras presentadas no se puede sacar una conclusión exacta acerca de la relación existente entre el rango económico, la altura, y los gastos con la variable dependiente.

En el cuadro 3 se observan las dos especificaciones sugeridas para el modelo de regresión. Se presentan los valores de los parámetros para las variables de interés, así como sus correspondientes desvíos estándar.

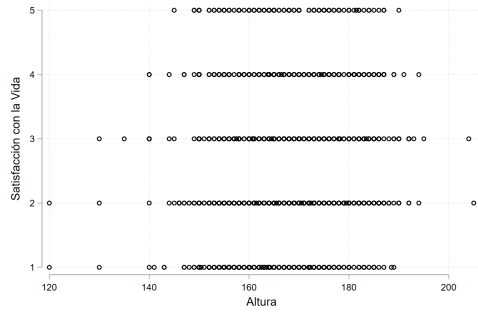
Por un lado, el modelo 1 toma variables de control que consideramos que pueden ser relevantes a la hora de explicar la satisfacción con la vida, como ser la altura, los gastos de la persona, su escala de rango económico, si cuenta con seguro médico y finalmente, si dejó su trabajo.

Se puede observar que todas las covariables incorporadas en el primer modelo son significativas al 99% de confianza. Con respecto a los signos de los coeficientes estimados se puede observar que estos cumplen con lo esperado. Cabe resaltar el poco impacto de la variable gasto sobre la variable dependiente. Por otro lado, era de esperar que a medida que la escala de rango económico aumentaba su impacto también lo haría.

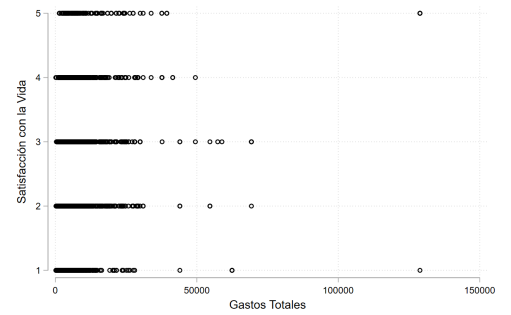
El modelo 2 agrega dos controles extra, como ser divorciado y orthodoxo, ambas variables de tipo Dummy. Como se puede observar la variable divorciado no es significativa, mientras que la variable orthodoxo lo es sólo al 90% de confianza.

Figura 2: Graficos de Correlación

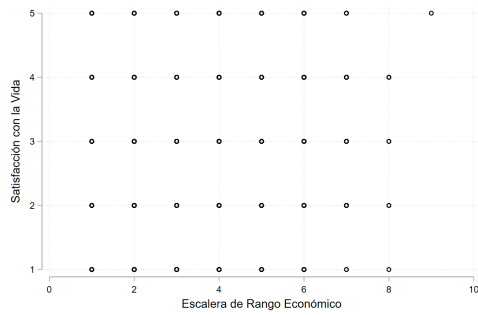
(a) Satisfacción con la Vida y Altura



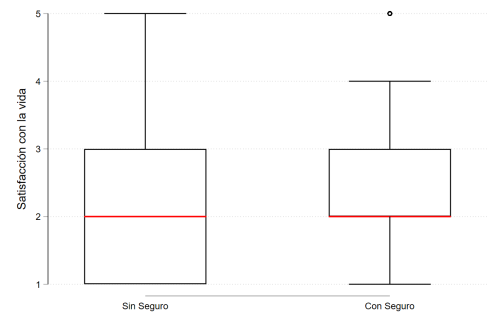
(b) Satisfacción con la Vida y Gastos Totales



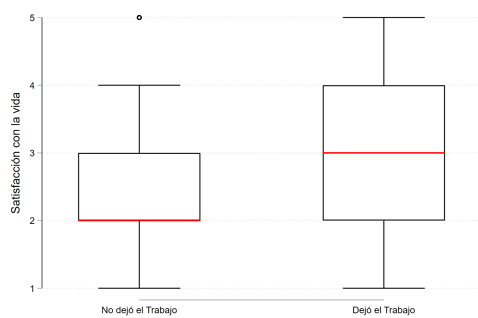
(c) Satisfacción con la Vida y Escala de Rango Económico



(d) Satisfacción con la Vida y Seguro



(e) Satisfacción con la Vida y Dejar el Trabajo



Fuente: Elaboración Propia

Cuadro 3: Especificaciones del Modelo

	Modelo 1	Modelo 2
Altura Reportada	0.009*** (0.002)	0.008*** (0.002)
Gastos Totales	0.000*** (0.000)	0.000*** (0.000)
<i><u>Escala de Rango Económico</u></i>		
Escala 2	0.308*** (0.084)	0.306*** (0.084)
Escala 3	0.476*** (0.079)	0.473*** (0.079)
Escala 4	0.814*** (0.080)	0.807*** (0.080)
Escala 5	1.032*** (0.078)	1.025*** (0.078)
Escala 6	1.344*** (0.112)	1.338*** (0.112)
Escala 7	1.211*** (0.200)	1.212*** (0.199)
Escala 8	0.924** (0.399)	0.936** (0.399)
Escala 9	2.932*** (1.032)	2.839*** (1.033)
Con Seguro Medico	0.201*** (0.068)	0.203*** (0.068)
Dejó el Trabajo	0.316** (0.148)	0.309** (0.147)
Divorciado		-0.102 (0.077)
Orthodoxo		-0.103* (0.059)
N	2,359	2,359
R ²	0.164	0.166



```

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 reemplazadas
por missing son imposibles de tener valores negativos dado la descripción
de la variable.
totexpr : HH Expenditures Real
tincm_r : 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

```