



MAESTRÍA EN ECONOMÍA

Economía Aplicada

PROF. MARTIN A. ROSSI

TUTORES: PAOLA LLAMAS Y TOMAS
PACHECO

**Problem Set 1: Repaso y
primeros pasos en Stata**

Garcia Ojeda, Juan
Hausvirth, Martina
Hayduk, Gaspar
Salvatierra, Elias Lucas D.

Fecha de entrega: 23 de agosto de 2024

PROBLEM SET 1: REPASO Y PRIMEROS PASOS EN STATA

GARCIA OJEDA, HAUSVIRTH, HAYDUK, SALVATIERRA

1. REPASO TEORICO

1. Según el Teorema de Gauss-Markov, el estimador de Mínimos Cuadrados Ordinarios (MCO) es el “Mejor Estimador Lineal Insesgado”, es decir, es el que tiene la menor varianza entre todos los estimadores lineales e insesgados y, por ende, el más eficiente bajo los siguientes supuestos:

- (1) Linealidad en los parámetros: El modelo debe ser lineal en los parámetros, aunque no necesariamente en las variables.
- (2) No multicolinealidad perfecta: No debe haber una colinealidad perfecta entre las variables explicativas.
- (3) Exogeneidad: El término de error debe tener esperanza condicional cero, es decir, $E(u|X) = 0$. Esto implica que los regresores no están correlacionados con los errores.
- (4) Homoscedasticidad: Los términos de error tienen varianza constante, $Var(u|X) = \sigma^2$.
- (5) No autocorrelación: Los términos de error no están correlacionados entre sí, $Cov(u_i, u_j) = 0$ para $i \neq j$.

Los resultados del teorema de Gauss-Markov solo se cumplen bajo los supuestos mencionados. Si alguno de estos no se cumple, el estimador de MCO podría no ser eficiente, y existen métodos alternativos que podrían ser más apropiados.

2. Para que el estimador de Mínimos Cuadrados Ordinarios (MCO) sea insesgado, es necesario suponer que el error tiene esperanza condicional cero. En otras palabras, $E(u|X) = 0$. Esto implica que el término de error u no está correlacionado con las variables explicativas X . Si se cumple, la estimación por MCO dará un valor promedio que coincide con el verdadero parámetro poblacional β .

Sin embargo, para garantizar la existencia del estimador de MCO el supuesto más importante es la no multicolinealidad perfecta. De esta forma, $X'X$ es invertible. Este supuesto asegura que no hay una relación lineal exacta entre las variables explicativas permitiendo que la matriz mencionada sea invertible. Si hay multicolinealidad perfecta, la matriz no tiene inversa, y por lo tanto, el estimador $\hat{\beta} = (X'X)^{-1}X'Y$ no puede ser calculado.

3. Levantar los supuestos de homocedasticidad o no correlación serial provoca un aumento en la varianza del $\hat{\beta}$. Esto provoca que, aunque el estimador de Mínimos Cuadrados Ordinarios (MCO) siga siendo insesgado y consistente, ya no sea el estimador lineal insesgado más eficiente.

4. Cuando asumimos que los términos de error μ_i siguen una distribución normal, podemos decir que el estimador de Mínimos Cuadrados Ordinarios (MCO), $\hat{\beta}$, también sigue una distribución normal. En un modelo de regresión lineal clásico, el estimador de MCO se define como: $\hat{\beta} = (X'X)^{-1}X'Y$. Notar que $Y = X\beta + u$, entonces, $\hat{\beta} = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$. Simplificando lo que sería la matriz identidad, resulta $\hat{\beta} = \beta + (X'X)^{-1}X'u$. La normalidad de los errores implica que las observaciones, y , siguen una distribución normal. El

estimador de MCO es una combinación lineal de las observaciones y , y cualquier combinación lineal de variables normales es normal.

5. No es estrictamente necesario, en todos los casos, suponer la distribución del término error para hacer inferencia. Si la muestra es lo suficientemente grande, por el teorema central del límite (TCL) no es necesario realizar supuestos sobre los errores. Además, si se utilizan otro tipo de métodos (no paramétricos) como Bootstrap, no es necesario asumir una distribución específica del término error.

6. Un estimador puede ser tanto consistente como asintóticamente normal al mismo tiempo. La consistencia asegura que el estimador converge al valor verdadero del parámetro, mientras que la normalidad asintótica proporciona información sobre la distribución del estimador a medida que la muestra se hace grande.

7. Consistencia e insesgadez son dos propiedades distintas de un estimador. La propiedad de consistencia refiere al hecho que un estimador converja al verdadero parámetro poblacional en el límite cuando el número de observaciones tiende a infinito. Por su parte, la propiedad de insesgadez indica que el estimador en valor esperado es igual al verdadero parámetro poblacional. Un estimador podría tranquilamente ser insesgado y consistente, solamente uno de las dos o ninguno; razón por la cual no son lo mismo.

2. PRIMEROS PASOS EN STATA

1. Resuelto en el do-file

2. Las variables que resultan tener mas de 5% de missing values son: Age (in months), Self-reported Height, HH Income Real, HH Expenditures Real, Obese.

3-4. Resuelto en el do-file

5.

	Mean	SD	Min	Max
Sexo	0.42	0.49	0.00	1.00
Edad	46.36	17.77	18.00	100.58
Satisfacción con la vida	2.48	1.14	1.00	5.00
Circunferencia de la cadera	84.62	13.97	29.00	168.00
Gasto real	7790.59	9129.68	147.83	128868.79
Observations	2818			

La tabla muestra las características descriptivas de las variables: sexo, edad, satisfacción con la vida, circunferencia de la cadera, y gasto real. En un total de 2.818 observaciones, vemos las medias, el desvío estándar, los mínimos y los máximos para cada variable.

En el caso del sexo, dado que es una variable binaria que toma valores 1 y 0 según se refiera a sexo masculino y femenino respectivamente; presenta una media de 0.42, por lo cual el 42% de la muestra son masculinos

La edad va desde los 18 hasta los 100 años, con una media de 46 años, y un desvío de 17,77; lo cual indica que existe mucha dispersión en los datos.

Para el caso de la variable "satisfacción con la vida" el rango posible es $[1, 6]$, siendo 6 el máximo grado de satisfacción; sin embargo observamos en la muestra que el máximo reportado es 5 y el mínimo es 1. La media de felicidad es 2.48, y el desvío es de 1,14.

El gasto real presenta el grado de dispersión mas alta, de 9,129.68, con un mínimo de 147.83, un

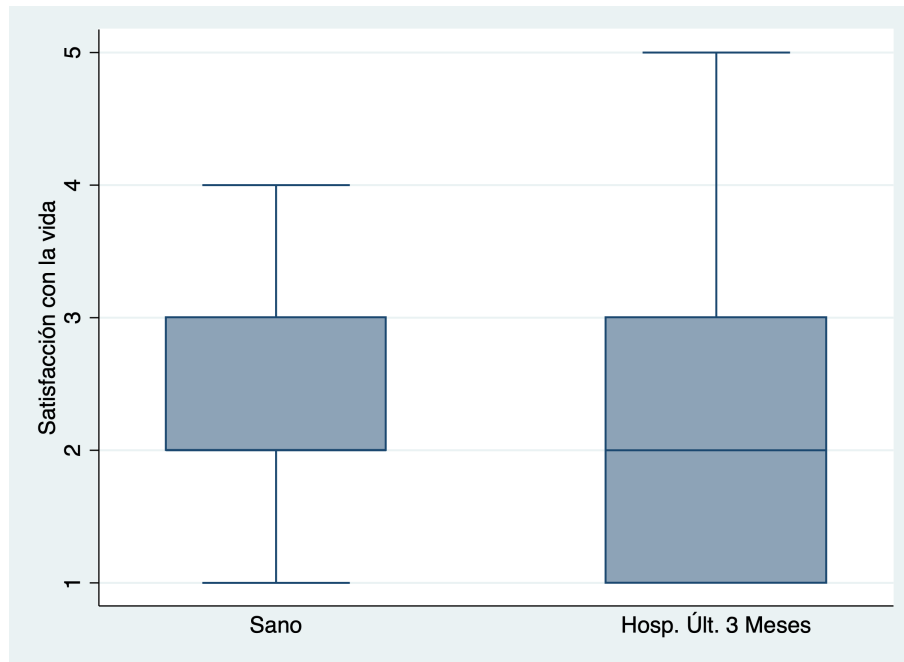
máximo de 128,868.79, y una media de 7,790.59. Finalmente, para el caso de la circunferencia de la cadera, existe un máximo de 168cm, un mínimo de 29cm, una media de 84.62cm, y un desvío estándar de 13.97cm

6.

	Media Mujeres	Media Hombres	Promedio de diferencias	SD	p-value
Circ. de cadera	102.99	97.68	5.31	0.45	0.00
Observations	2797				

Como se puede observar en la presente tabla, en mujeres, la circunferencia media de caderas es de 102.99cm, mientras que en hombres la media es de 97.68. Contrario a la hipótesis establecida en el enunciado, con un desvío estándar de 0.45, podemos decir que en promedio la circunferencia de la cadera de las mujeres es mayor que la de los hombres.

7. Queremos explicar la felicidad de las personas (medido por satlif). Nuestro primer prior es que el estado de salud influye mucho en la felicidad de uno. Consideramos la variable dummy *hosp3m*, que vale 1 si la persona estuvo hospitalizada en los últimos 3 meses y 0 en el caso contrario, el cual consideraremos como Sano. Observamos el siguiente boxplot que nos dirá características de la distribución de la satisfacción con la vida para el grupo que estuvo hospitalizado en los últimos 3 meses y los sanos.



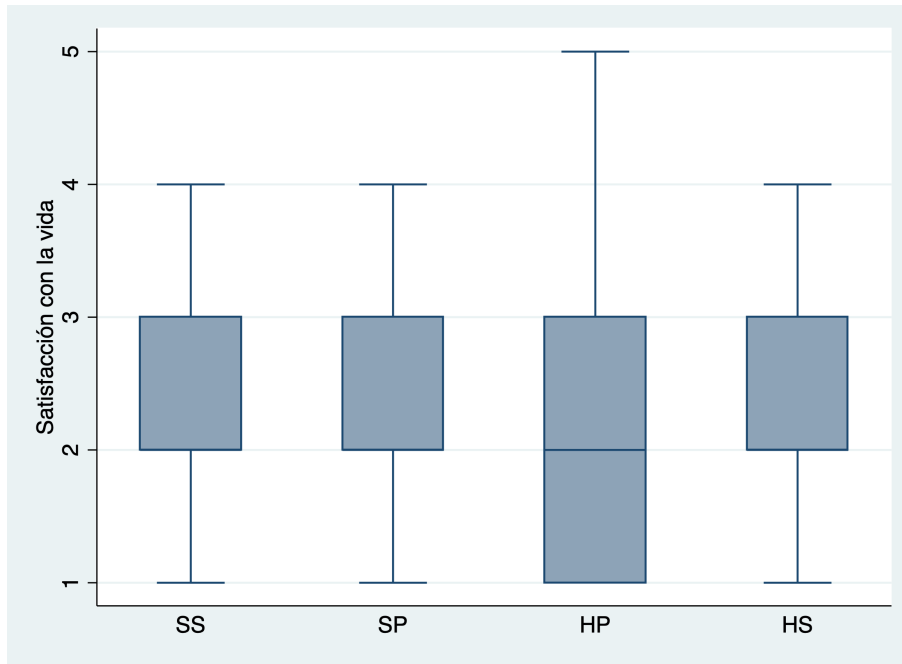
El boxplot nos dice la media, los límites superiores e inferiores y el primer, y tercer cuartil de la distribución de una variable. Podemos observar que la distribución de los sanos se encuentra más centrada en el valor promedio de satisfacción con la vida que en el caso de los hospitalizados en los últimos 3 meses, aunque estos presentan valores máximos por encima de los sanos (podría deberse a un outlier). Exploraremos esta relación a partir de la siguiente ecuación a estimar por OLS:

$$\text{satisfacción}_i = \alpha + \beta * \text{Hosp}_i + \mu_i \quad (2.1)$$

Los resultados de la estimación se encuentran en la columna 1 de la tabla de regresiones. Podemos notar que haber sido hospitalizado en los últimos 3 meses se asocia con niveles de satisfacción promedio 0.2 puntos menores. El signo es esperable debido a que una persona que ha sido

hospitalizada puede estar atravesando una situación de estrés, la cual influye negativamente en su satisfacción con la vida. Este resultado es significativo al 5%, por lo que existe una relación negativa entre ambas variables.

Por otro lado, resulta interesante incluir una interacción que contemple el estado de salud en conjunción con el estado civil. Nuestro prior es que estar enfermo afectaría a la felicidad, pero podría ser diferente entre solteros y casados. Consideremos un análisis preliminar de la distribución de la satisfacción con la vida considerando hospitalizados en pareja (HP) y solteros (HS), y sanos en pareja (SP) y solteros (SS). Del gráfico de boxplot se puede observar a primera vista que cuando se contemplan los hospitalizados en pareja la distribución de su satisfacción con la vida es mas dispersa que los hospitalizados solteros. Para el caso de los sanos pareciera no haber importantes diferencias entre casados y solteros respecto a su satisfacción con la vida. De este análisis resulta esperable que los casados hospitalizados sean menos felices (¿castigo doble?).



A continuación, estimamos la siguiente ecuación que incluye como regresoras el estado civil, el status de hospitalizado en los ultimos 3 meses, e interacciones entre hospitalización y estado civil. En particular, la ecuación a estimar por OLS es:

$$satisfacción_i = \alpha + \beta_1 Hosp_i + \beta_2 Married_i + \beta_3 Married_i Hosp_i \quad (2.2)$$

Analizando la salida de la regresión en la tabla 1 observamos que los signos van en linea con lo esperado: tanto los casados como los hospitalizados son menos felices en relación a sus respectivas categorías base, y de la interacción entre ambos tambien. Sin embargo, ningun coeficiente resultó ser significativamente distinto de cero, por lo que no es correcto interpretar los valores que resultan de la estimación ya que se consideran azarosos.

TABLE 1. Tabla de Regresiones

	(1) Satisfaction with Life	(2) Satisfaction with Life
Married		-0.0355 (0.0446)
Hospitalized last 3 months	-0.206** (0.0962)	-0.185 (0.138)
Married&Hosp		-0.0433 (0.192)
Constant	2.495*** (0.0222)	2.514*** (0.0326)
Observations	2,788	2,788
R-squared	0.002	0.002

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

```

1  /*****
2
3
4          Problem Set 1
5
6          Universidad de San Andrés
7          Economía Aplicada
8          2024
9
10 Gaspar Hayduk; Juan Gabriel García Ojeda; Elias Lucas Salvatierra; Martina Hausvirth
11 *****/
12
13
14
15 * 0) Set up environment
16 *=====*
17 clear all // borro cualquiera cosa que esté abierta
18
19
20
21 * Save path in local or global:
22 global main "/Users/gasparhayduk/Desktop/Economía Aplicada/PS1" //cambiar según directorio de
23 cada uno
24 global input "$main/input"
25 global output "$main/output"
26
27
28 * Base de datos:
29 use "$input/data_russia.dta", clear
30 *=====*
31
32
33
34 * 1) Cleaning data: incisos 1, 2, 3 y 4.
35 *=====*
36
37 *Visualizing data
38 *browse
39
40 *Si corremos 'codebook' nos dirá qué tipo de data es cada variable. El objetivo de esta sección
41 es transformar todas las variables a numericas.
42
43
44 ***---- Las variables econrk, powrnk, evalhl, geo, resprk, satlif, wtchng, hhpres, work0, work1,
45 work2 (y algunas mas) tienen string cuando deberían tener numeros. Por ejemplo, tienen 'four' en
46 lugar de 4, y cuando es 4 es "4".
47 * Estas variables son CATEGORICAS (creo), cada valor representa una categoria.
48 * Primero paso todo al mismo tipo de string, es difícil trabajar con "two" y "2", prefiero tener
49 uno de los dos.
50 * despues uso encode para pasarlo a numerica.
51
52 * Tabuleo para ver los valores posibles:
53 foreach var of varlist econrk powrnk evalhl geo resprk satlif wtchng marsta1 marsta2 marsta3
54 hhpres work0 work1 work2 satecc highsc belief cmedin hprblm hosl3m operat hattac alclmo ortho
55 hhpres {
56     tab `var'
57 }
58
59
60
61 * Paso todo a un mismo tipo de string, donde dice "two" lo convertimos en "2". y convertimos
62 todos los missing a una misma categoria.
63 foreach var of varlist econrk powrnk evalhl geo resprk satlif wtchng marsta1 marsta2 marsta3
64 hhpres work0 work1 work2 satecc highsc belief cmedin hprblm hosl3m operat hattac alclmo ortho
65 hhpres {
66     replace `var' = "1" if `var' == "one"
67     replace `var' = "2" if `var' == "two"
68     replace `var' = "3" if `var' == "three"
69     replace `var' = "4" if `var' == "four"
70     replace `var' = "5" if `var' == "five"
71     replace `var' = "6" if `var' == "six"
72     replace `var' = "7" if `var' == "seven"
73     replace `var' = "8" if `var' == "eight"
74     replace `var' = "9" if `var' == "nine"
75     replace `var' = "." if `var' == ".b"
76     replace `var' = "." if `var' == ".c"
77     replace `var' = "." if `var' == ".d"

```

```

59     replace `var' = "0" if `var' == "zero"
60 }
61
62
63 * Defino el label para encode:
64
65 label define mylabel 0 "0" 1 "1" 2 "2" 3 "3" 4 "4" 5 "5" 6 "6" 7 "7" 8 "8" 9 "9" 50 "."
    // 50 será mi valor de missing, luego lo cambio
66
67
68
69 * Aplico encode:
70
71 foreach var of varlist econrk powrnk evalhl geo resprk satlif wtchnr marsta1 marsta2 marsta3
hhpres work0 work1 work2 satecc highsc belief cmedin hprblm hosl3m operat hattac alclmo ortho {
72     encode `var', generate (`var'_enc) label (mylabel)
73     drop `var'
74     rename `var'_enc `var'
75 }
76
77 * Seteo el valor de 50 como missing.
78 foreach var of varlist econrk powrnk evalhl geo resprk satlif wtchnr marsta1 marsta2 marsta3
hhpres work0 work1 work2 satecc highsc belief cmedin hprblm hosl3m operat hattac alclmo ortho {
79     replace `var'=. if `var'==50
80 }
81
82
83
84
85 ***--- La variable waistc está guardada como string pero es un float (un continuo), con monage
pasa lo mismo. Las paso a float con destring.
86 destring waistc, replace
87 destring monage, replace
88
89
90
91 ***--- Ahora queda trabajar con las variables hipsiz y totexpr. Debemos extraer el numero del
texto (que seguira siendo un string pero despues usamos destring) .
92 **Para esto usamos split:
93
94 * Spliteo hipsiz:
95 split hipsiz, gen (h) // Esto genera 3 variables: h1, h2 y h3. h1 dice 'hip', h2 dice
'circunsference' y h3 la medida. Quiero esto ultimo.
96 * Elimino las primeras dos variables y a la ultima le cambio el nombre:
97 drop h1
98 drop h2
99 drop hipsiz
100 rename h3 hipsiz
101
102 * Spliteo totexpr
103 split totexpr, gen (h)
104 drop h1
105 drop h2
106 drop totexpr
107 rename h3 totexpr
108
109
110 * QUEDA CONVERTIR LOS "," EN MISSING Y DESPUES PASAR A FLOAT CON destring.REPETIR PARA 'totexpr'.
TAMBIEN HAY QUE HACER ESTO CON LA VARIABLE 'tincm_r'
111
112 ***--- Pasamos "," a "." en hip_size, tot_expr y tincm_r
113
114 replace hipsiz = subinstr(hipsiz, ",", ".", .)
115 replace totexpr = subinstr(totexpr, ",", ".", .)
116 replace tincm_r = subinstr(tincm_r, ",", ".", .)
117
118 ***--- Pasamos hip_size, tot_expr y tincm_r a float
119 destring hipsiz, replace force

```



```

120  destring totexpr, replace force
121  destring tincm_r, replace force
122
123
124  ***---- Trabajamos ahora con las variables sex, smokes y obese.
125
126  *Smokes y Sex no tienen missings.
127
128  // Sex
129  replace sex="1" if sex=="male"
130  replace sex="0" if sex=="female"
131  label define label_sex 1 "1" 0 "0" // 1 si es male y 0 si es female
132  encode sex, generate (sex_enc) label (label_sex)
133
134  drop sex
135  rename sex_enc sex
136
137  // Smokes
138  replace smokes="1" if smokes=="Smokes"
139  label define label_smokes 1 "1" 0 "0" // 1 si fuma y 0 si no
140  encode smokes, generate (smokes_enc) label (label_smokes)
141
142  drop smokes
143  rename smokes_enc smokes
144
145  // Obese
146  replace obese="1" if obese=="This person is obese"
147  replace obese="0" if obese=="This person is not obese"
148  label define label_obese 1 "This person is obese" 0 "This person is not obese" 50 "." //1 si es
obeso, 0 si no es obeso y 50 si hay "."
149  encode obese, generate (obese_enc) label (label_obese)
150
151  drop obese
152  rename obese_enc obese
153
154  replace obese=. if obese==50 //transformo los 50 en missing.
155
156
157  ***--- INCISO 2: cuento los missings de cada variable---***
158
159  *Buscamos obtener aquellas variables que presenten mas de 5% de missing values
160
161  mdesc
162
163  // Las variables que resultan tener mas de 5% de missing values son: Age (in months),
Self-reported Height, HH Income Real, HH Expenditures Real, Obese.
164
165  ***--- INCISO 3: veo datos irregulares ---***
166
167  *Chequeamos valores particulares, en general las variables float
168
169  foreach var of varlist inwgt monage htself height waistc tincm_r hipsiz totexpr {
170
171  count if `var' < 0
172  }
173
174
175  *De lo anterior surge que tincm_r y totexpr presentan valores negativos. Ambas variables indican
Ingresos y Gastos respectivamente, por lo que no tiene sentido que tengan valores negativos.
Reemplazamos los valores por "."
176
177  replace tincm_r=. if tincm_r < 0
178  replace totexpr=. if totexpr < 0
179
180  ** Cuento las observaciones donde los gastos sean mayores a los ingresos:
181  count if totexpr > tincm_r
182  * Hay 1,556 observaciones donde los gastos son mayores a los ingresos
183

```

```

184
185 ***--- INCISO 4: ordeno los datos ---***
186
187 order id site sex
188 gsort -totexpr
189
190
191 ***--- Guardo base de datos limpia
192
193 save "$input/russian_clean", replace
194 use "$input/russian_clean.dta", clear
195
196 *=====*
197
198
199
200 * 2) Descriptive statistics
201 *=====*
202
203
204 *----Inciso 5: Tablas Descriptivas ----*
205
206
207 *Modificamos la variable monage (edad) para que figure en años
208
209 gen age = monage/12
210
211 *Agregamos los labels
212 label var sex "Sexo"
213 label var age "Edad"
214 label var satlif "Satisfacción con la vida"
215 label var waistc "Circunferencia de la cadera"
216 label var totexpr "Gasto real"
217
218 *Generamos la tabla y la exportamos
219 estpost summarize sex age satlif waistc totexpr
220 esttab using "$output/tables/Table 1.tex", cells("mean(fmt(2)) sd(fmt(2)) min(fmt(2)) max(fmt(2))"
221 ) collabels("Mean" "SD" "Min" "Max") nomtitle nonumber replace label
222
223 ***--- INCISO 6: Distribución de hipsiz para los hombres y para las mujeres. ---***
224
225 *ssc install grstyle
226
227 *a)
228
229 *Comparamos la distribución de hipsiz entre hombres y mujeres
230 twoway (kdensity hipsiz if sex==0) ///
231 (kdensity hipsiz if sex==1), ///
232 legend(order(1 "Mujeres" 2 "Hombres")) title("Distribución de la circunsferencia de la cadera") ///
233 ytitle("Densidad") xtitle("Circunferencia de la cadera")
234 graph export "$output/figures/hipsiz_histogram_menvswomen.png", replace
235
236 *b) realizamos un test de medias entre sex y hpsiz y exportamos los resultados
237
238 ttest hipsiz, by (sex)
239
240 *Generamos los labels
241 label var hipsiz "Circ. de cadera"
242
243 estpost ttest hipsiz, by (sex)
244 esttab using "$output/tables/Table 2.tex", cells("mu_1(fmt(2)) mu_2(fmt(2)) b(fmt(2)) se(fmt(2))
245 p(fmt(2))") collabels("Media Mujeres" "Media Hombres" "Promedio de diferencias" "SD" "p-value")
246 nomtitle nonumber replace label
247
248 *-----*
249 *=====*
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

249 *=====*
250
251
252 ***--- INCISO 7: Regresando la felicidad de las personas ---***
253
254
255 *** Gráficos Introdutorios:
256
257 **Prior 1: las personas con salud son mas felices:
258 graph box satlif, over(hosl3m, relabel(1 "Sano" 2 "Hosp. Últ. 3 Meses")) ytitle("Satisfacción con
la vida") nooutsides title("") note("")
259 graph export "$output/figures/satlif_salud.png", replace
260
261
262 **Prior 2: las personas sin salud son menos felices, pero seran mas felices si tienen pareja.
Vamos a ver si esta hipotesis se cumple!!
263
264 ** Generamos la interaccion:
265
266 gen interac=.
267
268
269 replace interac = 0 if hosl3m == 0 & marsta1 == 0 // sanos solos
270 replace interac = 1 if hosl3m == 0 & marsta1 == 1 // sanos en pareja
271 replace interac = 2 if hosl3m == 1 & marsta1 == 1 // hospitalizados en pareja
272 replace interac = 3 if hosl3m == 1 & marsta1 == 0 // hospitalizados solos
273
274
275
276 *Grafico 2:
277 graph box satlif, over(interac, relabel(1 "SS" 2 "SP" 3 "HP" 4 "HS")) nooutsides note("") ytitle(
"Satisfacción con la vida")
278 graph export "$output/figures/pareja_salud2.png", replace
279
280
281 *----- Hacemos las regresiones:
282
283 *1: satisfaccion contra salud:
284 reg satlif hosl3m
285 outreg2 using "$output/tables/Table 3.tex", replace label tex(fragment)
286
287 *2: satisfaccion contra salud y pareja:
288 reg satlif i.marsta1##i.hosl3m
289 outreg2 using "$output/tables/Table 3.tex", append label tex(fragment)
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311

```