
Economía Aplicada: Problem Set N°1

Milton Bronstein Felipe García Vassallo Santiago López Franco Riottini

1. Repaso Teórico

1. Verdadero. Bajo los supuestos clásicos se cumple el Teorema de Gauss-Markov que demuestra que Mínimos Cuadrados Ordinarios es el mejor estimador (de mínima varianza) en la clase de estimadores lineales e insesgados.
2. El supuesto necesario para que MCO sea insesgado es el de exogeneidad, es decir que las características observadas no estén correlacionadas con el término del error. Por otro lado, para garantizar la existencia del estimador no se debe violar el supuesto de multicolinealidad perfecta.
3. Al levantar estos supuestos no se podría hacer inferencia causal, salvo que se adopten estrategias para mitigar los problemas que traería esto.
4. En muestras finitas, el estimador tiene una distribución normal ya que la distribución normal pasa a través de funciones lineales. Dado que al modelo lineal se le aplica una regresión lineal, la distribución del error se “transfiere” a la del estimador. Por otro lado, en muestras donde el tamaño tiende a infinito, por el teorema central del límite el estimador tiene una distribución normal, y no es necesario suponer la distribución del término de error.
5. Verdadero. Los estadísticos utilizados para los tests de hipótesis tienen todos una distribución conocida y que debemos suponer.
6. Falso. Las propiedades asintóticas de consistencia y distribución asintótica normal son totalmente compatibles, de hecho el estimador de mínimos cuadrados ordinarios cumple con ambas.
7. Falso. Consistencia es una propiedad que se cumple cuando el tamaño de la muestra tiende a infinito, mientras que insesgadez es una propiedad de muestra finita. Un estimador puede cumplir con ninguna, las dos o alguna de ellas.

2. Primeros Pasos en Stata

2.2. Missings

Una vez limpiada la base, encontramos que las variables *monage*, *htself*, *tincmr*, *obeso* y *totexpr* presentan valores faltantes en más del 5 % de las observaciones.

Realizamos un resumen de la base de datos en detalle (con `summary`, `detail`). Algunas variables con valores irregulares que encontramos son *totexpr*, el gasto real total de los hogares, y *15tincm_r*, el ingreso real de los hogares. En algunas observaciones había valores negativos, lo que es absurdo. Estos valores los reemplazamos con "missing".

2.3. Cleaning de irregulares

Realizando un resumen de la base de datos en detalle (con summary, detail). Algunas variables con valores irregulares que encontramos son *totexpr*, el gasto real total de los hogares, y *15tincm_r*, el ingreso real de los hogares. Si bien en algunas variables, como "Hip Circumference", encontramos valores que parecen bastante elevados, por no conocer que serían "valores adecuados" decidimos dejarlos. Sin embargo, en algunas observaciones monetarias de ingreso y gasto había valores negativos, lo que es absurdo. Estos valores los reemplazamos con "missing", mediante las funciones:

```
replace totexpr3=. if totexpr3<0
replace tincm_r=. if tincm_r<0
```

2.4. Data sort

Utilizando el siguiente código, ordenamos los datos de tal forma que la primera variable que aparezca en la base sea el *id* (id) del individuo, la segunda el *sitio* (*site*) donde se encuentra y la tercera el *sexo* (*female*). Luego, ordenen las filas de mayor a menor según *totexpr*:

```
order id site female totexpr
gsort -totexpr
```

2.5. Estadística Descriptiva

Todas las personas encuestadas a partir de las cuales se elaboró el Cuadro 1 eran mayores de edad (18 años en adelante) y el promedio ha sido de 45,85 años, por lo que la mejor manera de expresarlo es un promedio redondeado de 46 años. En este cuadro, además, se observa que el 58,2% de las personas encuestadas fueron mujeres. Luego, podemos ver que el gasto medio de los hogares de la muestra ha sido de 7656,02.

La satisfacción con la vida es una variable que se obtiene de una encuesta que va de 1 a 5, donde 5 implica que el encuestado está muy satisfecho con su vida, y 1 para nada satisfecho. El valor promedio de respuesta es 2.45, lo que significa que en general los encuestados están más cerca de estar insatisfechos con su vida que satisfechos.

Cuadro 1: Estadística Descriptiva

	Mean	SD	Min	Max
Satisfaction with Life	2.45	1.11	1.00	5.00
Age in years	45.85	17.80	18.00	100.00
Woman	0.58	0.49	0.00	1.00
Waist Circumference	85.64	13.55	37.00	168.00
Hip Circumference	101.48	11.50	40.00	180.00
HH Expenditures Real	7656.02	8914.26	147.83	128868.79
Observations	2361			

2.6. Hips don't lie

2.6.1. Refuting the paper

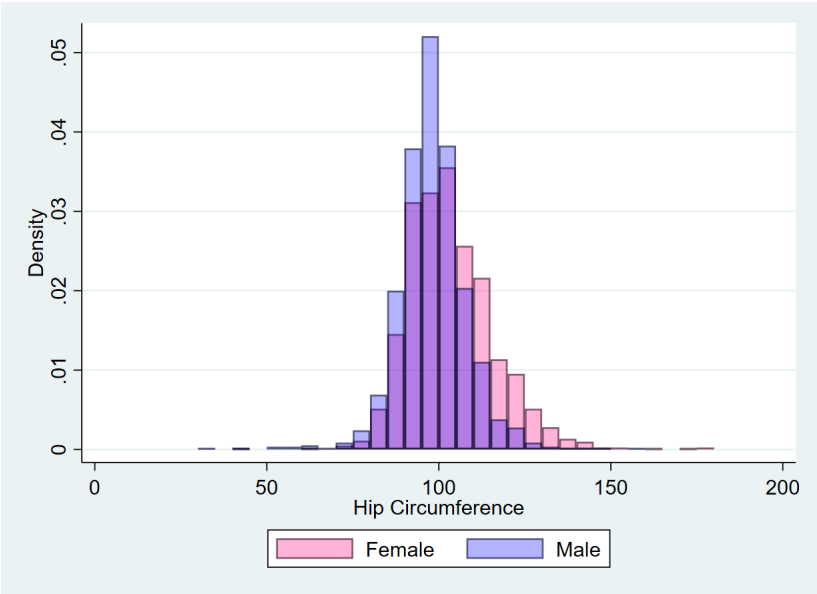


Figura 1: Circunferencia de cadera

En la Figura 1 se observan las medidas de circunferencia de cadera de las personas encuestadas, separando hombres y mujeres. Se observa que esta medida es, en promedio, levemente menor entre los hombres que entre las mujeres. Estos resultados refutan lo dicho en el *paper* de Shakira (2005).

2.6.2. Mean test

En la tabla podemos ver que hay una diferencia de medias estadísticamente significativa en la circunferencia de caderas entre hombres y mujeres. La media de los hombres es de 97.7 cm, mientras que la de las mujeres es de aproximadamente 103 centímetros, por lo que la diferencia de medias es de -5.3 centímetros.

Group	Obs	Mean	Std. Err.	Std. Dev.	[95 % Conf. Interval]	
Hombre	1,16	97.67957	.294948	10.04557	97.10088	98.25826
Mujer	1,637	102.9928	.3139317	12.70163	102.377	103.6085
Combinado	2,797	100.7892	.2261775	11.96177	100.3457	101.2327
Diff		-5.313223	.4480286		-6.191723	-4.434722

Diff = mean(0) - mean(female)

H_o : Diff = 0

H_a : diff < 0

Pr(T < t) = 0.0000

t = -118.591

Degrees of freedom = 2795

H_a : Diff > 0

H_a : Diff != 0

Pr(|T| > |t|) = 0.0000

Pr(T > t) = 1.0000

2.7. Regresiones

2.7.1. Gráficos

En la Figura 2 se observa el nivel de felicidad de las personas encuestadas agrupadas en función del nivel de satisfacción con su situación económica. Para las personas menos satisfechas con su condición económica (1 y 2 puntos) la felicidad mediana es de 2. Este valor asciende a 3 entre las personas que han reportado mayor satisfacción con su condición económica (3 y 4 puntos). Mientras que para las personas más satisfechas con esta condición la felicidad mediana es de 5 (la mayor calificación en esta encuesta). Esto significa que más de la mitad de las personas de este grupo han dado el máximo puntaje para el nivel de felicidad.

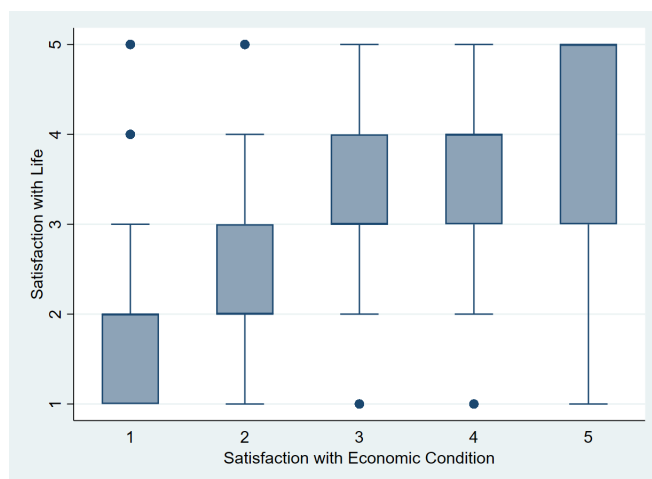


Figura 2: Felicidad según el nivel de satisfacción con la condición económica

En la Figura 3 se presenta la edad de las personas encuestadas en función del nivel de felicidad que han reportado. Se puede observar que la edad mediana decrece cuando aumenta la felicidad. Esto nos indica que en los grupos de personas más felices la proporción de personas jóvenes es mayor.

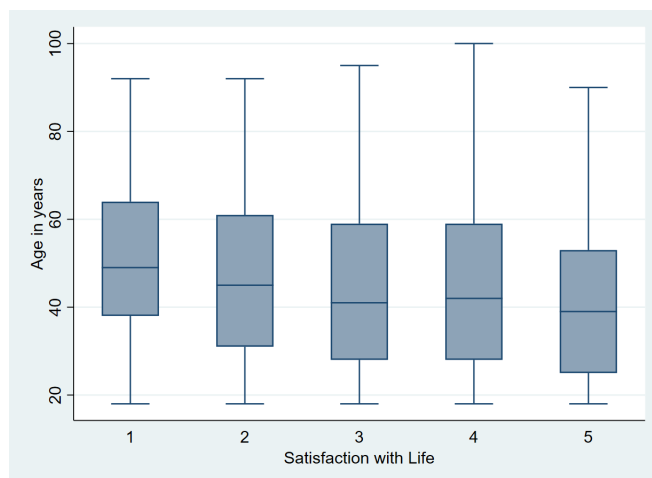


Figura 3: Edad para cada nivel de felicidad

2.7.2. Interpretaciones

En la primera especificación regresamos la satisfacción con la vida sobre el sexo del individuo, su edad, su altura, si tuvo un paro cardíaco, si está casado, si es obeso, si es creyente, si terminó la secundaria, su satisfacción con su condición económica, si toma alcohol, su situación laboral y el gasto real de su hogar. La mayor parte de estas variables no son estadísticamente significativas, las que sí son la edad (al 1 %), si tuvo un paro cardíaco (al 5 %), la altura (al 10 %), su satisfacción con su condición económica (1 %) y el gasto real (al 1 %). La edad y haber tenido un paro cardíaco tienen un efecto negativo sobre la satisfacción con la vida. El coeficiente de la edad nos dice que 10 años más de vida están correlacionados con una caída de 0.05 puntos en el score de satisfacción con la vida, lo que implica un efecto marginal. Por otro lado, haber tenido un paro cardíaco tiene un impacto aproximadamente 5 veces mayor. Dado el estrés y los problemas que causan los paros cardíacos este resultado es esperable. Las otras tres variables tienen un efecto positivo sobre la satisfacción con la vida. De todas formas, solamente la magnitud de la satisfacción con la condición económica es considerable: mayor satisfacción con la situación económica está relacionada con un aumento de 0.543 puntos sobre responder si el individuo está satisfecho con su vida en general. El impacto del gasto real es muy cercano a 0, y totalmente despreciable, mientras que un centímetro más de altura está asociado a un incremento de aproximadamente 0.005 puntos en las respuestas sobre la satisfacción. Un metro más de altura, poco razonable dado que las alturas están fijas y no hay mucha dispersión dentro de la población, recién tendría un impacto de 0.5 sobre las respuestas de los encuestados.

En la segunda columna observamos la regresión de la satisfacción con la vida sobre las mismas variables, pero a las que sumamos la edad al cuadrado, si el individuo está divorciado o si es viudo. En esta nueva especificación son significativas estadísticamente las mismas variables (todas al mismo nivel, excepto la altura que ahora es significativa al 5 %), a las que se suma también si el encuestado trabaja (al 10 %); de las nuevas solamente la edad al cuadrado es estadísticamente significativa (al 1 %). Nuevamente, el gasto real está asociado a un incremento en la satisfacción con la vida, pero el impacto es despreciable. El signo y la magnitud de si el encuestado tuvo un paro cardíaco se mantienen, así como la satisfacción con la condición económica y la altura. En esta segunda regresión, la edad tiene un impacto negativo casi diez veces mayor que en la primera, lo que la hace más significativa en términos empíricos. La edad al cuadrado, por otro lado, está asociada a un incremento, aunque muy marginal, en la satisfacción con la vida. Esta última variable implica que a mayor edad se contrarresta el efecto negativo, hasta cierto punto, cuando llega a un vértice. Por último, estar trabajando está asociado a un incremento de aproximadamente 0.1 puntos en la satisfacción con la vida.

VARIABLE	(1) Satisfaction with Life	(2) Satisfaction with Life
CONSTANT	0.611 (0.516)	0.920* (0.522)
Woman	-0.0424 (0.0543)	-0.0180 (0.0548)
Age in years	-0.00446*** (0.00136)	-0.0277*** (0.00695)
<i>Ageinyears</i> ²		0.000261*** (7.28e-05)
Height	0.00491* (0.00284)	0.00570** (0.00284)
Ever Had Heart Attack	-0.254** (0.112)	-0.270** (0.112)
Married	0.0112 (0.0413)	-0.00389 (0.0562)
Obese	0.0243 (0.0518)	0.0519 (0.0522)
Believer	0.00934 (0.0189)	0.00614 (0.0190)
Finished High School	-0.00112 (0.0471)	0.0215 (0.0474)
Satisfaction with Economic Condition	0.543*** (0.0196)	0.530*** (0.0199)
Drinks Alcohol	0.0446 (0.0420)	0.0526 (0.0420)
Working	0.0385 (0.0441)	0.0887* (0.0460)
HH Expenditures Real	1.25e-05*** (2.29e-06)	1.24e-05*** (2.29e-06)
Divorced		-0.0779 (0.0843)
Widowed		-0.125 (0.0839)
Observations	2,310	2,310
R-squared	0.286	0.291

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1