
Economía Aplicada: Problem Set N^o2

Milton Bronstein Felipe García Vassallo Santiago López Franco Riottini

1. Los problemas de inferir con hipótesis múltiples

Al evaluar múltiples hipótesis de manera simultánea, los procedimientos estándar para hacer inferencia no son confiables. En este sentido, nos referimos a evaluar múltiples hipótesis simultáneamente a cuando realizamos algunas de estas tres cosas:

- Realizamos múltiples regresiones al mismo tiempo con la misma variable independiente y múltiples variables dependientes.
- Realizamos una regresión con un mismo *outcome* y distintas variables independientes.
- Realizamos regresiones con una variable independiente y una variable dependiente pero sobre distintas submuestras.

Cuando esto sucede, aumenta nuestra probabilidad de obtener al menos un falso positivo, es decir, de rechazar la hipótesis nula cuando esta es verdadera. De esta forma podría estar sacando conclusiones erróneas si no tengo en cuenta que estoy evaluando múltiples hipótesis de manera simultánea. Por ende, para realizar un proceso estadístico correcto, debemos corregir o los p-valores o los niveles de significatividad para no obtener conclusiones equivocadas.

En este problem set atacaremos este problema utilizando distintas estrategias con la intención de corregir los resultados del *paper* “Estimating the production function for human capital: results from a randomized controlled trial in Colombia” de Attanasio, Cattán, Fitzsimmons, Meghir & Rubio-Codina (2020).

2. Las estrategias

Para corregir los p-valores usamos las tres estrategias que conocemos: Bonferroni, Holm y Benjamini, Krieger y Yekutieli.

- Bonferroni: Esta estrategia busca disminuir la *Family Wise Error Rate* (FWER), que es la probabilidad de cometer al menos un falso positivo en una de las hipótesis evaluadas. Para eso se toma el nivel de significatividad α con el que queremos trabajar, y se lo corrige por la cantidad de hipótesis que estamos testeando. La fórmula es α/m , donde “m” es la cantidad de hipótesis. Esta estrategia tiene la desventaja de que es muy restrictiva, y si bien tiene éxito en disminuir la probabilidad de ocurrencia de falsos positivos, también aumentan las probabilidades de que ningún coeficiente sea estadísticamente significativo. Por otro lado, esta estrategia utiliza un supuesto fuerte que es el de **independencia entre las hipótesis**.
- Holm: Esta estrategia toma todos los *p-values* (p) de cada una de las K hipótesis testeadas y los ordena de menor a mayor. Luego se calcula:

$$\frac{\alpha}{K + 1 - k}$$

donde K es el número de hipótesis a testear y k es la posición que cada p -value obtiene en el ordenamiento mencionado previamente. Luego, se debe comparar cada p -value ordenado contra su correspondiente valor que se obtiene de esta fórmula. Este proceso se realiza hasta encontrar el primer p -value tal que:

$$p \geq \frac{\alpha}{K + 1 - k}$$

Una vez que esto se cumple, éste y todos los p_k siguientes en la sucesión dejan de ser estadísticamente significativos. Sólo deben tomarse los anteriores como tales para rechazar las hipótesis nulas correspondientes.

- Benjamini, Krieger y Yekutieli: Esta estrategia consiste en ordenar de forma decreciente, según su nivel de significatividad, los p-valores de las K hipótesis que estemos testeando. Luego se permutan las observaciones que fueron parte del tratamiento con las que son control y se ‘‘simulan’’ regresiones con estas permutaciones. De ahí obtenemos nuevos p-valores. Estos p-valores también son ordenados de forma decreciente según su nivel de significatividad, y se busca un $p_r^* = \min\{p_r^*, p_{r+1}^*, \dots, p_k^*\}$, donde r denota el ‘‘ranking’’ original. Estas dos operaciones deben ser repetidas una gran cantidad de veces (L) y para cada *outcome* que estudiemos contamos la cantidad de veces que $p_r^* < p_r$, que llamaremos S_r . Luego calculamos $p_r^{fwer*} = S_r/L$. Con esto podemos ver en que proporción de las simulaciones permutadas el p_r^* fue menor al original. Por último se computa un p-valor final que es el máximo de los p_r^{fwer*} .

Estas estrategias tienen la similitud principal de buscar resolver los problemas de realizar inferencia cuando estamos evaluando múltiples hipótesis de manera simultánea. En general, las tres toman como insumo la cantidad de hipótesis que estamos testeando. La estrategias de Holm y la de Benjamini, Krieger y Yekutieli tienen la similitud de que trabajan principalmente con los p-valores de las hipótesis y los ordenan de cierta forma para luego quedarse con aquellos que cumplan ciertas condiciones restrictivas y que permitan hacer inferencia. Las estrategias de Holm y de Bonferroni también comparten la similitud de que trabajan con *alpha*, el nivel de significatividad que es tomado para rechazar o no las hipótesis nulas. Asimismo, las estrategias de Benjamini, Krieger y Yekutieli y la de Bonferroni usan para resolver los problemas de inferencia la *Family Wise Error Rate*.

Las diferencias surgen principalmente en el grado de complejidad y sofisticación de cada una de las estrategias, en el orden en el que las presentamos es claro que la complejidad va de menor a mayor. Además, las ventajas y desventajas propias de cada estrategia son las expresiones fundamentales de sus diferencias. Mientras que la estrategia de Bonferroni es muy restrictiva, la de Benjamini, Krieger y Yekutieli no lo es tanto y nos da una mayor probabilidad de que podamos rechazar hipótesis nulas. En primer lugar, la corrección de Bonferroni permite igualar la FWER al nivel de significatividad y tiene como contra que supone independencia entre las hipótesis. Ésta es la principal restricción de esta corrección. En segundo lugar, la corrección de Holm es menos restrictiva en ese sentido. En tercer lugar, la corrección de Benjamini, Krieger y Yekutieli cuenta con la ventaja de que considera la dependencia entre los distintos outcomes que pudiere haber. Esto supera la restricción que planteaba la corrección de Bonferroni.

3. Comparación de resultados

Primero replicamos en el Cuadro 1 la Tabla 2 del paper, sin corregir la inferencia causal, pero organizando los resultados presentados de otra manera, con los outcomes en columnas y los tratamientos en las filas. Sin embargo, en este cuadro todavía no corregimos los p-valores atendiendo los problemas del testeo de hipótesis múltiples, que será el foco de esta sección. En el Cuadro 2 nuevamente replicamos la tabla, pero incluimos los p-valores que surgen de las tres estrategias que hemos ido comentando.

Con respecto a los resultados del Panel A del Cuadro 2, como podemos ver los p-valores originales y de las 3 estrategias correctoras, difieren mucho. Para dos variables que resultaron significativas en algún grado para el test original (*Cognitive* y *Receptive language*) las correcciones aplicadas arrojan que siguen siendo significativas (inclusive, aumentan su significatividad). En las otras cuatro, que no son significativas, solo la corrección de Holm arroja significatividad al 5 % ahora, realizándolo en 3 de las 4 variables (la única por fuera es *Expressive language* la cual está en el límite).

En el Panel B del Cuadro 2 tenemos todas variables que no son significativas en el test original. Ahora, con las correcciones realizadas, el p-valor de la estrategia de Holm es significativo en todas excepto en una variable (*ECBQ: Inhibitory control*) la cual se encuentra en el límite del 5 % de significatividad. Sin embargo, para las otras dos correcciones aplicadas los estimadores siguen siendo no significativos.

Para el Panel de *Material investments* en el Cuadro 2 la significancia al 5 % en el test original se logra solamente en dos variables (columna 1 y 4). Para la primera columna, cualquiera sea la estrategia aplicada sigue siendo significativa, y lo mismo sucede para la cuarta. Para las otras tres columnas los resultados son diversos. Bajo las tres estrategias, la variable *FCI: N° of coloring and drawing books* es significativa al 5 % mientras antes no lo era. Sin embargo, hay una regularidad que atraviesa a todas las variables. Las variables de la columna 4 y 5 son ahora significativas bajo la estrategia de Holm mientras que bajo las otras siguen siendo no significativas.

Por último, al observar el Panel D del Cuadro 2, se ve que antes de las correcciones las estimaciones de los coeficientes de los *outcomes* resultaban significativas al 5 %. Luego de haber realizado las tres correcciones de los p-valores, se observa que para las variables explicadas en las columnas 1 y 3 las estimaciones siguen siendo significativas al 5 %. En cambio, en la columna 2 no lo es según la corrección de Bonferroni mientras que sí lo es luego de haber realizado la corrección de BKY. En las estimaciones de las columnas 4 y 5 se puede ver que las estimaciones siguen siendo significativas al 5 % incluso luego de todas las correcciones de los p-valores.

Cuadro 1: Tabla 2 de Attanasio *et al* (2020)

<i>Panel A: Child's cognitive skills at follow up</i>							
	(1) Bayley: Cognitive	(2) Bayley: Receptive language	(3) Bayley: Expressive language	(4) Bayley: Fine motor	(5) MacArthur: Words the child can say	(6) MacArthur: Complex phrases the child can say	(7) Cognitive factor
treat	0.250*** (0.063)	0.174** (0.063)	0.029 (0.062)	0.073 (0.059)	0.086 (0.064)	0.057 (0.056)	0.116** (0.041)
N	1263	1263	1262	1261	1321	1321	1256
R-Squared	0.21	0.15	0.20	0.17	0.22	0.17	0.38
<i>Panel B: Child's socio-emotional skills at follow up</i>							
	(1) ICQ: Difficult (-)	(2) ICQ: Unsociable (-)	(3) ICQ: Unstoppable (-)	(4) ECBQ: Inhibitory control	(5) ECBQ: Attentional focusing	(6) Socio-emotional factor	
treat	0.073 (0.045)	0.037 (0.055)	0.031 (0.054)	0.000 (0.058)	0.072 (0.048)	0.036 (0.033)	
N	1325	1325	1325	1322	1322	1322	
R-Squared	0.16	0.11	0.13	0.10	0.22	0.21	
<i>Panel C: Material investments</i>							
	(1) FCI: N° of types of play materials	(2) FCI: N° of coloring and drawing books	(3) FCI: N° of toys to learn movement	(4) FCI: N° of toys to learn shapes	(5) FCI: N° of shop-bought toys	(6) Material investment factor	
treat	0.217*** (0.064)	-0.124* (0.056)	-0.049 (0.065)	0.426*** (0.088)	0.020 (0.061)	0.098 (0.049)	
N	1325	1325	1325	1325	1325	1325	
R-Squared	0.27	0.10	0.20	0.13	0.12	0.24	
<i>Panel D: Time investments</i>							
	(1) FCI: N° of types of play activities in last 3 days	(2) FCI: N° of times told a story to child in last 3 days	(3) FCI: N° of times read to child in last 3 days	(4) N° of times played with toys in last 3 days	(5) N° of times named things to child in last 3 days	(6) Time investment factor	
treat	0.280*** (0.051)	0.153* (0.064)	0.403*** (0.068)	0.183** (0.060)	0.144** (0.048)	0.233*** (0.038)	
N	1325	1325	1325	1325	1325	1325	
R-Squared	0.31	0.13	0.12	0.13	0.14	0.28	

Cuadro 2: Tabla 2 de Attanasio *et al* (2020) con P-values corregidos con estrategias Bonferroni, Holm y BKY

Panel A: Child's cognitive skills at follow up						
	(1) Bayley: Cognitive	(2) Bayley: Receptive language	(3) Bayley: Expressive language	(4) Bayley: Fine motor	(5) MacArthur: Words the child can say	(6) MacArthur: Complex phrases the child can say
treat	0.250*** (0.063)	0.174** (0.063)	0.029 (0.062)	0.073 (0.059)	0.086 (0.064)	0.057 (0.056)
P-value	0.000	0.007	0.644	0.224	0.179	0.315
Bonf p-value	0.001	0.042	1.000	1.000	1.000	1.000
Holm p-value	0.008	0.010	0.050	0.017	0.013	0.025
BKY p-value	0.001	0.018	0.508	0.290	0.290	0.338
Number of Observations	1263.00	1263.00	1262.00	1261.00	1321.00	1321.00
R-Squared	0.21	0.15	0.20	0.17	0.22	0.17
Panel B: Child's socio-emotional skills at follow up						
	(1) ICQ: Difficult (-)	(2) ICQ: Unsociable (-)	(3) ICQ: Unstoppable (-)	(4) ECBQ: Inhibitory control	(5) ECBQ: Attentional focusing	
treat	0.073 (0.045)	0.037 (0.055)	0.031 (0.054)	0.000 (0.058)	0.072 (0.048)	
P-value	0.107	0.506	0.569	1.000	0.136	
Bonf p-value	0.533	1.000	1.000	1.000	0.681	
Holm p-value	0.010	0.017	0.025	0.050	0.013	
BKY p-value	0.517	0.744	0.744	1.000	0.517	
Number of Observations	1325.00	1325.00	1325.00	1322.00	1322.00	
R-Squared	0.16	0.11	0.13	0.10	0.22	
Panel C: Material investments						
	(1) FCI: N° of types of play materials	(2) FCI: N° of coloring and drawing books	(3) FCI: N° of toys to learn movement	(4) FCI: N° of toys to learn shapes	(5) FCI: N° of shop-bought toys	
treat	0.217*** (0.064)	-0.124* (0.056)	-0.049 (0.065)	0.426*** (0.088)	0.020 (0.061)	
P-value	0.001	0.028	0.452	0.000	0.750	
Bonf p-value	0.005	0.140	1.000	0.000	1.000	
Holm p-value	0.013	0.017	0.025	0.010	0.050	
BKY p-value	0.002	0.029	0.293	0.001	0.429	
Number of Observations	1325.00	1325.00	1325.00	1325.00	1325.00	
R-Squared	0.27	0.10	0.20	0.13	0.12	
Panel D: Time investments						
	(1) FCI: N° of types of play activities in last 3 days	(2) FCI: N° of times told a story to child in last 3 days	(3) FCI: N° of times read to child in last 3 days	(4) N° of times played with toys in last 3 days	(5) N° of times named things to child in last 3 days	
treat	0.280*** (0.051)	0.153* (0.064)	0.403*** (0.068)	0.183** (0.060)	0.144** (0.048)	
P-value	0.000	0.019	0.000	0.003	0.004	
Bonf p-value	0.000	0.095	0.000	0.014	0.018	
Holm p-value	0.013	0.050	0.010	0.017	0.025	
BKY p-value	0.001	0.005	0.001	0.003	0.003	
Number of Observations	1325.00	1325.00	1325.00	1325.00	1325.00	
R-Squared	0.31	0.13	0.12	0.13	0.14	