Maestría en Economía

Economía Aplicada

Prof. Martin A. Rossi

Tutores: Paola Llamas y Tomas Pacheco

Problem Set 2: Multiple hypothesis testing

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

Fecha de entrega: 30 de agosto de 2024

PROBLEM SET 2: TEST DE MÚLTIPLES HIPÓTESIS

GARCIA OJEDA - HAUSVIRTH - HAYDUK - SALVATIERRA

En este problem set se corrrige la inferencia del artículo de Attanasio, Cattan, Fitzsi- mons, Meghir y Rubio-Codina (2020), que investiga los mecanismos por los cuales intervenciones en la niñez tienen efectos en habilidades socioemocionales y cognitivas de los niños.

1. Nuestra Table 1 replica particularmente la Tabla 2 del articulo mencionado, que muestra los resultados de visitas a los hogares1 en cuatro outcomes: desarrollo cognitivo, desarrollo socioemocional, inversión de los padres en materiales de juego e inversión de los padres en actividades de juego. A continuación se presenten las regresiones en una tabla que contiene los cuatro paneles, donde las columnas refieren a los distintos outcomes y la fila a la variable de tratamiento.

TABLE 1.
Treatment Impacts on Raw Measures and Latent Factors

Panel A: Child's cognitive skills at follow-up						
	(1) Cognitive b/se	(2) Receptive language b/se	(3) Expressive language b/se	(4) Fine motor b/se	(5) Words b/se	(6) Complex phrases b/se
Treatment Effect	0.250	0.174	0.029	0.073	0.086	0.057
	(0.063)	(0.063)	(0.062)	(0.059)	(0.064)	(0.056)
Number of Observations	1263	1263	1262	1261	1321	1321
Panel B: Child's socio-emotional skills at follow-up						
	(1) ICQ: Difficult (-) b/se	(2) ICQ: Unsociable (-) b/se	(3) ICQ: Unstoppable (-) b/se	(4) ECBQ: Inhibitory control b/se	(5) ECBQ: Attention b/se	
Treatment Effect	0.073	0.037	0.031	0.000	0.072	
	(0.045)	(0.055)	(0.054)	(0.058)	(0.048)	
Number of Observations	1325	1325	1325	1322	1322	
Panel C: Material investments at follow-up						
	(1) FCI: Number of types of play materials b/se	(2) FCI: Number of coloring and drawing books b/se	(3) FCI: Number of toys to learn movement b/se	(4) FCI: Number of toys to learn shapes b/se	(5) FCI: Number of shop-bought toys b/se	
Treatment Effect	0.217	-0.124	-0.049	0.426	0.020	
	(0.064)	(0.056)	(0.065)	(0.088)	(0.061)	
Number of Observations	1325	1325	1325	1325	1325	
Panel D: Time investment at follow-up						
	FCI: Number of types of play activities in last 3 days b/se	FCI: Number of times told a story to child in last 3 days b/se	FCI: Number of times read to child in last 3 days b/se	FCI: Number of times played with toys in last 3 days b/se	FCI: Number of times named things to child in last 3 days b/se	
Treatment Effect	0.280	0.153	0.403	0.183	0.144	
	(0.051)	(0.064)	(0.068)	(0.060)	(0.048)	
Number of Observations	1325	1325	1325	1325	1325	

2. Enfrentamos un problema de múltiples hipótesis cuando lo que queremos testear contiene múltiples outcomes, tratamientos, o subdimensiones. Es necesario corregir la inferencia, dado que los procedimientos estándar que se utilizan habitualmente, dejan de ser confiables cuando se evalúan múltiples hipótesis simultánemente. El efecto del tratamiento puede dar significativo por azar, y no porque tenga algún efecto real sobre las variables relevantes. Esto es asi dado que la probabilidad de cometer al menos un falso positivo crece exponencialmente con el número de hipotesis. Por ejemplo, de testear simultaneamente K hipotesis, es decir, $h_1, h_2, ...h_K$, con un nivel de significatividad de α. Entonces, en cada test, la probabilidad de cometer un falso positivo es α, mientras que la probabilidad de cometer al menos un falso positivo o Family-Wise Error Rate (FWER) crece exponencialmente con el número de hipótesis:

$$\begin{aligned} FWER &= Prob(h_1 \neq FP \vee ... \vee h_K = FP) \\ &= 1 - Prob(h_1 \neq FP \wedge ... \wedge h_K \neq FP) \\ &= 1 - [Prob(h_1 \neq FP) \times ... \times Prob(h_K \neq FP)] \\ &= 1 - (1-\alpha)^K \end{aligned}$$

De esta manera, como mencionamos antes, comparar individualmente los p-values con α puede llevar a conclusiones erróneas.

3. En Table 2 se corrige los p-values utilizando las tres estrategias vistas: Bonferroni; Holm; y Benjamini, Krieger y Yekutieli, y a partir de allí, se exporta una tabla idéntica a la del punto 1 agregando tres filas con los p-values corregidos, que presenta los impactos estimados de recibir visitas domiciliarias en el desarrollo cognitivo, el desarrollo emocional, la inversión de los padres en materiales de juego y la inversión de los padres en actividades de juego.

4. La correción de **Bonferroni** es la forma mas simple, pero también la mas restrictiva de corregir el problema de múltiples hipótesis. Esto ocurre dado que le quita mucho *power* a la estimación, y la hipótesis nula (H_0) se vuelve muy difícil de rechazar. Lo que hace es ajustar el FWER[1]. En una estimación con m hipótesis y un nivel de significatividad α, el ajuste que se hará sobre el nivel de significatividad será α/m.

Por otro lado, el ajuste de **Holm**, si bien es menos restrictiva que Bonferroni, también utiliza el supuesto de independencia para realizar la corrección, lo cual es bastante fuerte, y difícilmente generalizable. Este ajuste lo que hace es ordenar los *p-values* de menor a mayor y comparar, de manera decreciente, cada valor contra el "ranking" del *p-value* $\frac{\alpha}{K+1-k}$. A diferencia del primero, con este método es mas fácil rechazar H_0.

Finalmente, la corrección de **Banjamini, Krieger & Yekuteli**, es la única de las tres que tiene en consideración la dependencia de los outcomes. Al igual que para Holm, se corre la regresión para cada outcome y se guardan los *p-values* originales de forma decreciente. Sin embargo, lo que hace esta estrategia, a diferencia de las otras dos, es hacer una permutación de la data bajo la hipótesis de no significatividad (es decir, se varía el status de cada individuo aleatoriamente), para luego computar los nuevos *p-values* (P_r^{**}), y ordenarlos de forma decreciente.

[1]Family Wise Error Rate

TABLE 2.
Treatment Impacts on Raw Measures and Latent Factors

Panel A: Child's cognitive skills at follow-up						
	(1) Cognitive b/se	(2) Receptive language b/se	(3) Expressive language b/se	(4) Fine motor b/se	(5) Words b/se	(6) Complex phrases b/se
Treatment Effect	0.250	0.174	0.029	0.073	0.086	0.057
	(0.063)	(0.063)	(0.062)	(0.059)	(0.064)	(0.056)
P_value	0.000	0.007	0.644	0.224	0.179	0.315
Bonferroni Corrected P_value	0.003	0.147	1.000	1.000	1.000	1.000
Holm Corrected Significance	0.003	0.004	0.017	0.006	0.056	0.007
BKY Corrected P_value	0.001	0.013	0.407	0.215	0.179	0.267
Number of Observations	1263	1263	1262	1261	1321	1321
Panel B: Child's socio-emotional skills at follow-up						
	(1) ICQ: Difficult (-) b/se	(2) ICQ: Unsociable (-) b/se	(3) ICQ: Unstoppable (-) b/se	(4) ECBQ: Inhibitory control b/se	(5) ECBQ: Attention b/se	
Treatment Effect	0.073	0.037	0.031	0.000	0.072	
	(0.045)	(0.055)	(0.054)	(0.058)	(0.048)	
P_value	0.107	0.506	0.569	1.000	0.136	
Bonferroni Corrected P_value	1.000	1.000	1.000	1.000	1.000	
Holm Corrected Significance	0.005	0.010	0.013	0.050	0.005	
BKY Corrected P_value	0.120	0.366	0.398	0.508	0.143	
Number of Observations	1325	1325	1325	1322	1322	
Panel C: Material investments at follow-up						
	(1) FCI: Number of types of play materials b/se	(2) FCI: Number of coloring and drawing books b/se	(3) FCI: Number of toys to learn movement b/se	(4) FCI: Number of toys to learn shapes b/se	(5) FCI: Number of shop-bought toys b/se	
Treatment Effect	0.217	-0.124	-0.049	0.426	0.020	
	(0.064)	(0.056)	(0.065)	(0.088)	(0.061)	
P_value	0.001	0.028	0.452	0.000	0.750	
Bonferroni Corrected P_value	0.020	0.587	1.000	0.000	1.000	
Holm Corrected Significance	0.003	0.004	0.008	0.003	0.025	
BKY Corrected P_value	0.004	0.038	0.342	0.001	0.429	
Number of Observations	1325	1325	1325	1325	1325	
Panel D: Time investment at follow-up						
	(1) FCI: Number of types of play activities in last 3 days b/se	(2) FCI: Number of times told a story to child in last 3 days b/se	(3) FCI: Number of times read to child in last 3 days b/se	(4) FCI: Number of times played with toys in last 3 days b/se	(5) FCI: Number of times named things to child in last 3 days b/se	
Treatment Effect	0.280	0.153	0.403	0.183	0.144	
	(0.051)	(0.064)	(0.068)	(0.060)	(0.048)	
P_value	0.000	0.019	0.000	0.003	0.004	
Bonferroni Corrected P_value	0.000	0.398	0.000	0.060	0.074	
Holm Corrected Significance	0.003	0.004	0.002	0.003	0.003	
BKY Corrected P_value	0.001	0.029	0.001	0.008	0.009	
Number of Observations	1325	1325	1325	1325	1325	

5. En comparacion a los resultados del articulo, los nuevos resultados incorporan los p values corregidos a partir de la tres estrategias vistas: Bonferroni; Holm; y Benjamini, Krieger y Yekutiele.

A partir de estos nuevos resultados, muchos coeficientes ya no son estadisticamente significativos dado que ahora se rechaza muchas mas veces la hipotesis nula de no significatividad.

Ademas, los p-values corregidos mediante el método de Benjamini, Krieger y Yekutile tienden a ser más similares a los reportados en el artículo, por la dependencia existente entre los outcomes que hace

que esta metodologia sea hace menos restrictiva en comparación con el resto de los métodos, que asumen independencia entre las hipótesis testadas.

```
/*******************************************************************************
                          Semana 3: Problem Set 2

                          Universidad de San Andrés
                              Economía Aplicada
                                   2024
Gaspar Hayduk; Juan Gabriel García Ojeda; Elias Lucas Salvatierra; Martina Hausvirth
*******************************************************************************/


* Source: https://www.aeaweb.org/articles?id=10.1257/app.20200204

/*******************************************************************************
Este archivo sigue la siguiente estructura:

0) Set up environment and globals

1) Regressions (panel A, B, C y D)
2) Correccion de Bonferroni
3) Correccion de Holm
4) Correccion de Benjamini, Krieger, and Yekutieli (2006)

*******************************************************************************/



* 0) Set up environment
*==============================================================================*

*gl main "/Users/gasparhayduk/Desktop/Economía Aplicada/aplicadaps2"
gl main "C:\Users\Usuario\OneDrive\Juanga\OneDrive\JUANGA\Maestria\Udesa\Economia
Aplicada\Problem sets\PS2"
gl input "$main/input"
gl output "$main/output"

* Open data set

use "$input/measures.dta", clear

* Global with control variables

global covs_eva "male i.eva_fu"
global covs_ent "male i.ent_fu"

*-- Agregamos etiquetas a la variable de interes y a los outcomes:

* Outcomes panel A
label var b_tot_cog1_st "Cognitive"
label var b_tot_lr1_st "Receptive language"
label var b_tot_le1_st "Expressive language"
label var b_tot_mf1_st "Fine motor"
label var mac_words1_st "Words"
label var mac_phrases1_st "Complex phrases"

*Outcomes panel B
label var bates_difficult1_st  "ICQ: Difficult (-)"
label var bates_unsociable1_st  "ICQ: Unsociable (-) "
label var bates_unstoppable1_st  "ICQ: Unstoppable (-)"
label var roth_inhibit1_st  "ECBQ: Inhibitory control"
label var roth_attention1_st  "ECBQ: Attention"

*Outcomes panel C
label var fci_play_mat_type1_st  "FCI: Number of types of play materials"
label var Npaintbooks1_st  "FCI: Number of coloring and drawing books"
label var Nthingsmove1_st  "FCI: Number of toys to learn movement"
label var Ntoysshape1_st  "FCI: Number of toys to learn shapes"
label var Ntoysbought1_st  "FCI: Number of shop-bought toys"

```

```
*Outcomes panel D
label var fci_play_act1_st  "FCI: Number of types of play activities in last 3 days"
label var home_stories1_st  "FCI: Number of times told a story to child in last 3 days"
label var home_read1_st  "FCI: Number of times read to child in last 3 days"
label var home_toys1_st  "FCI: Number of times played with toys in last 3 days"
label var home_name1_st  "FCI: Number of times named things to child in last 3 days"

*Variable de tratamiento
label var treat "Treatment Effect"



* 1) Regressions
*================================================================================*



*********************************************************************************
* PANEL A (Child's cognitive skills at follow up)
*********************************************************************************


local bayley "b_tot_cog b_tot_lr b_tot_le b_tot_mf"
eststo clear
foreach y of local bayley{
local append append
if "`y'"=="b_tot_cog" local append replace
    cap drop V*
    reg `y'1_st treat `y'0_st $covs_eva , cluster(cod_dane)
}
esttab using "$output/Cuadro1_panelA.tex", se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///


local macarthur "mac_words mac_phrases"
foreach y of local macarthur{
    cap drop V*
    reg `y'1_st treat mac_words0_st $covs_ent , cluster(cod_dane)
}
esttab using "$output/Cuadro1_panelA.tex", se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///


*********************************************************************************
* PANEL B (Child's socio-emotional skills at follow up)
*********************************************************************************

local bates "bates_difficult bates_unsociable bates_unstoppable"
eststo clear
foreach y of local bates{
    cap drop V*
    reg `y'1_st treat `y'0_st $covs_ent, cl(cod_dane)
}
esttab using "$output/Cuadro1_panelB.tex", se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///

local roth "roth_inhibit roth_attention"
foreach y of local roth{
    cap drop V*
    reg `y'1_st treat bates_difficult0_st $covs_ent , cluster(cod_dane)
}
esttab using "$output/Cuadro1_panelB.tex", se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///

```

```
136
137 ********************************************************************************
138 * PANEL C (Material investments)
139 ********************************************************************************
140
141 local fcimat "fci_play_mat_type Npaintbooks Nthingsmove Ntoysshape Ntoysbought"
142 eststo clear
143 foreach y of local fcimat{
144     cap drop V*
145     reg `y'1_st treat fci_play_mat_type0_st $covs_ent , cluster(cod_dane)
146 }
147 esttab using "$output/Cuadro1_panelC.tex", se replace label ///
148 keep(treat) ///
149 cells(b(fmt(3)) se(par fmt(3))) ///
150
151
152 ********************************************************************************
153 * PANEL D (Time investments)
154 ********************************************************************************
155 local fcitime "fci_play_act home_stories home_read home_toys home_name"
156 eststo clear
157 foreach y of local fcitime{
158     cap drop V*
159     reg `y'1_st treat fci_play_act0_st $covs_ent , cluster(cod_dane)
160 }
161 esttab using "$output/Cuadro1_panelD.tex", se replace label ///
162 keep(treat) ///
163 cells(b(fmt(3)) se(par fmt(3))) ///
164
165
166
167
168 * 2) Correccion de Bonferroni
169 *================================================================================*
170
171 * Tenemos 21 outcomes de interes, por lo que hay que testear 21 hipotesis. Guardamos ese numero
    en un escalar:
172 scalar hyp = 21
173
174 * si tengo m hipotesis uso alpha/m como nivel de significancia.
175 * por lo que debemos ver si los p-valores < alpha/m
176 * por comodidad, veremos si m*p-valores < alpha (el codigo hace esto).
177
178 * Hay que correr las regresiones otra vez, almacenar el p-valor y generar el nuevo p-valor
    corregido. Hacemos esto para cada panel.
179
180 ********************************************************************************
181 * PANEL A (Child's cognitive skills at follow up)
182 ********************************************************************************
183
184
185 local bayley "b_tot_cog b_tot_lr b_tot_le b_tot_mf"
186 eststo clear
187 foreach y of local bayley{
188 local append append
189 if "`y'"=="b_tot_cog" local append replace
190     cap drop V*
191     reg `y'1_st treat `y'0_st $covs_eva , cluster(cod_dane)
192     eststo: test treat = 0
193     estadd scalar p_value = r(p)
194     estadd scalar corr_p_value = min(1,r(p)*hyp)
195 }
196 esttab using "$output/Cuadro2_panelA.tex", se replace label ///
197 keep(treat) ///
198 cells(b(fmt(3)) se(par fmt(3))) ///
199 stats(p_value corr_p_value blank N, fmt(3 3 0) labels("P_value" "Bonferroni Corrected P_value" " "
     "Number of Observations"))
200
```

```
201
202 local macarthur "mac_words mac_phrases"
203 foreach y of local macarthur{
204     cap drop V*
205     reg `y'1_st treat mac_words0_st $covs_ent , cluster(cod_dane)
206     eststo: test treat = 0
207     estadd scalar p_value = r(p)
208     estadd scalar corr_p_value = min(1,r(p)*hyp)
209 }
210 esttab using "$output/Cuadro2_panelA.tex", se replace label ///
211 keep(treat) ///
212 cells(b(fmt(3)) se(par fmt(3))) ///
213 stats(p_value corr_p_value blank N, fmt(3 3 0) labels("P_value" "Bonferroni Corrected P_value" " "
     "Number of Observations"))
214
215 ********************************************************************************
216 * PANEL B (Child's socio-emotional skills at follow up)
217 ********************************************************************************
218
219 local bates "bates_difficult bates_unsociable bates_unstoppable"
220 eststo clear
221 foreach y of local bates{
222     cap drop V*
223     reg `y'1_st treat `y'0_st $covs_ent, cl(cod_dane)
224     eststo: test treat = 0
225     estadd scalar p_value = r(p)
226     estadd scalar corr_p_value = min(1,r(p)*hyp)
227 }
228 esttab using "$output/Cuadro2_panelB.tex", se replace label ///
229 keep(treat) ///
230 cells(b(fmt(3)) se(par fmt(3))) ///
231 stats(p_value corr_p_value blank N, fmt(3 3 0) labels("P_value" "Bonferroni Corrected P_value" " "
     "Number of Observations"))
232
233 local roth "roth_inhibit roth_attention"
234 foreach y of local roth{
235     cap drop V*
236     reg `y'1_st treat bates_difficult0_st $covs_ent , cluster(cod_dane)
237     eststo: test treat = 0
238     estadd scalar p_value = r(p)
239     estadd scalar corr_p_value = min(1,r(p)*hyp)
240 }
241 esttab using "$output/Cuadro2_panelB.tex", se replace label ///
242 keep(treat) ///
243 cells(b(fmt(3)) se(par fmt(3))) ///
244 stats(p_value corr_p_value blank N, fmt(3 3 0) labels("P_value" "Bonferroni Corrected P_value" " "
     "Number of Observations"))
245
246 ********************************************************************************
247 * PANEL C (Material investments)
248 ********************************************************************************
249
250 local fcimat "fci_play_mat_type Npaintbooks Nthingsmove Ntoysshape Ntoysbought"
251 eststo clear
252 foreach y of local fcimat{
253     cap drop V*
254     reg `y'1_st treat fci_play_mat_type0_st $covs_ent , cluster(cod_dane)
255     eststo: test treat = 0
256     estadd scalar p_value = r(p)
257     estadd scalar corr_p_value = min(1,r(p)*hyp)
258 }
259 esttab using "$output/Cuadro2_panelC.tex", se replace label ///
260 keep(treat) ///
261 cells(b(fmt(3)) se(par fmt(3))) ///
262 stats(p_value corr_p_value blank N, fmt(3 3 0) labels("P_value" "Bonferroni Corrected P_value" " "
     "Number of Observations"))
263
264 ********************************************************************************
```

```
* PANEL D (Time investments)
*******************************************************************************
local fcitime "fci_play_act home_stories home_read home_toys home_name"
eststo clear
foreach y of local fcitime{
    cap drop V*
    reg `y'1_st treat fci_play_act0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    estadd scalar p_value = r(p)
    estadd scalar corr_p_value = min(1,r(p)*hyp)
}
esttab using "$output/Cuadro2_panelD.tex", se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///
stats(p_value corr_p_value blank N, fmt(3 3 0) labels("P_value" "Bonferroni Corrected P_value" " "
 "Number of Observations"))



* 3) Correccion de Holm
*================================================================================*

* Defino el nivel de significancia:
scalar signif = 0.05
* Defino la cantidad de hipotesis:
scalar hyp = 21

* Creo una matriz de 21 filas y 1 columna donde almaceno los p-valores:
mat p_values = J(21,1,.)

* Creo la variable para iterar:
scalar i = 1

*---------Ahora debo rellenar la matriz de p-valores. Debo correr todas las regresiones:

*--- PANEL A (Child's cognitive skills at follow up)
local bayley "b_tot_cog b_tot_lr b_tot_le b_tot_mf"
foreach y of local bayley{
local append append
if "`y'"=="b_tot_cog" local append replace
    cap drop V*
    reg `y'1_st treat `y'0_st $covs_eva , cluster(cod_dane)
    eststo: test treat = 0
    mat p_values[i,1] = r(p)
    scalar i = i +1
}

local macarthur "mac_words mac_phrases"
foreach y of local macarthur{
    cap drop V*
    reg `y'1_st treat mac_words0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    mat p_values[i,1] = r(p)
    scalar i = i +1
}

*--- PANEL B (Child's socio-emotional skills at follow up)
local bates "bates_difficult bates_unsociable bates_unstoppable"
foreach y of local bates{
    cap drop V*
    reg `y'1_st treat `y'0_st $covs_ent, cl(cod_dane)
    eststo: test treat = 0
    mat p_values[i,1] = r(p)
    scalar i = i +1
}

local roth "roth_inhibit roth_attention"
```

```
foreach y of local roth{
    cap drop V*
    reg `y'1_st treat bates_difficult0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    mat p_values[i,1] = r(p)
    scalar i = i +1
}

*--- PANEL C (Material investments)
local fcimat "fci_play_mat_type Npaintbooks Nthingsmove Ntoysshape Ntoysbought"
foreach y of local fcimat{
    cap drop V*
    reg `y'1_st treat fci_play_mat_type0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    mat p_values[i,1] = r(p)
    scalar i = i +1
}

*--- PANEL D (Time investments)
local fcitime "fci_play_act home_stories home_read home_toys home_name"
foreach y of local fcitime{
    cap drop V*
    reg `y'1_st treat fci_play_act0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    mat p_values[i,1] = r(p)
    scalar i = i +1
}


*--- Ahora trabajo con la matriz.

preserve
clear // borro la base que tengo abierta
svmat p_values // abro la matriz que guarde antes
gen var = _n // identificador de a qué outcome corresponde el pvalor.
sort p_values1 // ordeno los p-valores de menor a mayor.

gen alpha_corr = signif/(hyp+1-_n) // genero el nivel de significancia ajustado segun formula

gen significant = (p_values1<alpha_corr) // marco con 1 a aquellos p-valores que son menor que el
nivel de significancia ajustado.

replace significant = 0 if significant[_n-1]==0 // le doy un cero a los que no son significativos.


*odeno:
sort var


*restauro
restore


* Almacenos los niveles de significatividad corregidos.
* Rechazamos la hipotesis nula de treat = 0 si el p-valor original es menor que el nivel de
significancia ajustado
mat alpha_corr_h = [0.0027778, 0.0035714, 0.016667, 0.00625, 0.055556, 0.0071429, 0.0045455, 0.01,
 0.0125, 0.05, 0.005, 0.0029412, 0.0041667, 0.0083333, 0.0026316, 0.025, 0.0025, 0.0038462,
0.002381, 0.003125, 0.0033333]


*---Queda agregar los niveles de significancia ajustados a la tabla. Hay que hacerlo para cada
panel.

*definimos un iterados para poder ir accediendo a los elementos de la matriz con los niveles de
significancia ajustados.
scalar i = 1

```

```
*--- PANEL A (Child's cognitive skills at follow up)

local bayley "b_tot_cog b_tot_lr b_tot_le b_tot_mf"
eststo clear
foreach y of local bayley{
local append append
if "`y'"=="b_tot_cog" local append replace
    cap drop V*
    reg `y'1_st treat `y'0_st $covs_eva , cluster(cod_dane)
    eststo: test treat = 0
    estadd scalar p_value = r(p)
    estadd scalar corr_p_value = min(1,r(p)*hyp)
    estadd scalar alpha_corr = alpha_corr_h[1,i]
    scalar i = i + 1
}
esttab using "$output/Cuadro2_panelA.tex", p se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///
stats(p_value corr_p_value alpha_corr blank N, fmt(3 3 3 0) labels("P_value" "Bonferroni
Corrected P_value" "Holm Corrected Significance" " " "Number of Observations"))


local macarthur "mac_words mac_phrases"
foreach y of local macarthur{
    cap drop V*
    reg `y'1_st treat mac_words0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    estadd scalar p_value = r(p)
    estadd scalar corr_p_value = min(1,r(p)*hyp)
    estadd scalar alpha_corr = alpha_corr_h[1,i]
    scalar i = i + 1
}
esttab using "$output/Cuadro2_panelA.tex", p se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///
stats(p_value corr_p_value alpha_corr blank N, fmt(3 3 3 0) labels("P_value" "Bonferroni
Corrected P_value" "Holm Corrected Significance" " " "Number of Observations"))


*--- PANEL B (Child's socio-emotional skills at follow up)

local bates "bates_difficult bates_unsociable bates_unstoppable"
eststo clear
foreach y of local bates{
    cap drop V*
    reg `y'1_st treat `y'0_st $covs_ent, cl(cod_dane)
    eststo: test treat = 0
    estadd scalar p_value = r(p)
    estadd scalar corr_p_value = min(1,r(p)*hyp)
    estadd scalar alpha_corr = alpha_corr_h[1,i]
    scalar i = i + 1
}
esttab using "$output/Cuadro2_panelB.tex", p se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///
stats(p_value corr_p_value alpha_corr blank N, fmt(3 3 3 0) labels("P_value" "Bonferroni
Corrected P_value" "Holm Corrected Significance" " " "Number of Observations"))

local roth "roth_inhibit roth_attention"
foreach y of local roth{
    cap drop V*
    reg `y'1_st treat bates_difficult0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    estadd scalar p_value = r(p)
    estadd scalar corr_p_value = min(1,r(p)*hyp)
    estadd scalar alpha_corr = alpha_corr_h[1,i]
    scalar i = i + 1
}
```

```
esttab using "$output/Cuadro2_panelB.tex", p se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///
stats(p_value corr_p_value alpha_corr blank N, fmt(3 3 3 0) labels("P_value" "Bonferroni Corrected P_value" "Holm Corrected Significance" " " "Number of Observations"))


*--- PANEL C (Material investments)


local fcimat "fci_play_mat_type Npaintbooks Nthingsmove Ntoysshape Ntoysbought"
eststo clear
foreach y of local fcimat{
    cap drop V*
    reg `y'1_st treat fci_play_mat_type0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    estadd scalar p_value = r(p)
    estadd scalar corr_p_value = min(1,r(p)*hyp)
    estadd scalar alpha_corr = alpha_corr_h[1,i]
    scalar i = i + 1

}
esttab using "$output/Cuadro2_panelC.tex", p se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///
stats(p_value corr_p_value alpha_corr blank N, fmt(3 3 3 0) labels("P_value" "Bonferroni Corrected P_value" "Holm Corrected Significance" " " "Number of Observations"))


*--- PANEL D (Time investments)

local fcitime "fci_play_act home_stories home_read home_toys home_name"
eststo clear
foreach y of local fcitime{
    cap drop V*
    reg `y'1_st treat fci_play_act0_st $covs_ent , cluster(cod_dane)
    eststo: test treat = 0
    estadd scalar p_value = r(p)
    estadd scalar corr_p_value = min(1,r(p)*hyp)
    estadd scalar alpha_corr = alpha_corr_h[1,i]
    scalar i = i + 1
}
esttab using "$output/Cuadro2_panelD.tex",  se replace label ///
keep(treat) ///
cells(b(fmt(3)) se(par fmt(3))) ///
stats(p_value corr_p_value alpha_corr blank N, fmt(3 3 3 0) labels("P_value" "Bonferroni Corrected P_value" "Holm Corrected Significance" " " "Number of Observations"))







* 4) Correccion de Benjamini, Krieger, and Yekutieli (2006)
*================================================================================*

*---Ya tenemos los p_valores guardados en la matriz p_values. La abro y la guardo como un .dta:
preserve
clear
* abro la matriz para tratarla como una base de datos.
svmat p_values
gen outcome = _n
rename p_values1 pval
save "$output/pvals.dta", replace
restore


*--- Ahora toca usar Michael Anderson's code for sharpened q-values.
* el input es la base de datos con los p-valores originales (que debo corregirlos)
```

```
524  * el output seran los p-valores corregidos
525  preserve
526
527  use "$output/pvals.dta", clear
528  version 10
529  set more off
530
531  * Collect the total number of p-values tested
532  quietly sum pval
533  local totalpvals = r(N)
534
535  * Sort the p-values in ascending order and generate a variable that codes each p-value's rank
536  quietly gen int original_sorting_order = _n
537  quietly sort pval
538  quietly gen int rank = _n if pval~=.
539
540  * Set the initial counter to 1
541  local qval = 1
542
543  * Generate the variable that will contain the BKY (2006) sharpened q-values
544  gen bky06_qval = 1 if pval~=.
545
546  * Set up a loop that begins by checking which hypotheses are rejected at q = 1.000, then checks
     which hypotheses are rejected at q = 0.999, then checks which hypotheses are rejected at q =
     0.998, etc.  The loop ends by checking which hypotheses are rejected at q = 0.001.
547
548  while `qval' > 0 {
549      * First Stage
550      * Generate the adjusted first stage q level we are testing: q' = q/1+q
551      local qval_adj = `qval'/(1+`qval')
552      * Generate value q'*r/M
553      gen fdr_temp1 = `qval_adj'*rank/`totalpvals'
554      * Generate binary variable checking condition p(r) <= q'*r/M
555      gen reject_temp1 = (fdr_temp1>=pval) if pval~=.
556      * Generate variable containing p-value ranks for all p-values that meet above condition
557      gen reject_rank1 = reject_temp1*rank
558      * Record the rank of the largest p-value that meets above condition
559      egen total_rejected1 = max(reject_rank1)
560
561      * Second Stage
562      * Generate the second stage q level that accounts for hypotheses rejected in first stage:
     q_2st = q'*(M/m0)
563      local qval_2st = `qval_adj'*(`totalpvals'/(`totalpvals'-total_rejected1[1]))
564      * Generate value q_2st*r/M
565      gen fdr_temp2 = `qval_2st'*rank/`totalpvals'
566      * Generate binary variable checking condition p(r) <= q_2st*r/M
567      gen reject_temp2 = (fdr_temp2>=pval) if pval~=.
568      * Generate variable containing p-value ranks for all p-values that meet above condition
569      gen reject_rank2 = reject_temp2*rank
570      * Record the rank of the largest p-value that meets above condition
571      egen total_rejected2 = max(reject_rank2)
572
573      * A p-value has been rejected at level q if its rank is less than or equal to the rank of the
     max p-value that meets the above condition
574      replace bky06_qval = `qval' if rank <= total_rejected2 & rank~=.
575      * Reduce q by 0.001 and repeat loop
576      drop fdr_temp* reject_temp* reject_rank* total_rejected*
577      local qval = `qval' - .001
578  }
579
580
581  quietly sort original_sorting_order
582  pause off
583  set more on
584
585  display "Code has completed."
586  display "Benjamini Krieger Yekutieli (2006) sharpened q-vals are in variable 'bky06_qval'"
587  display "Sorting order is the same as the original vector of p-values"
```

```
588
589 keep outcome pval bky06_qval
590 * ACA ESTÁ EL OUTPUT. es una base de datos con los p-valores corregidos.
591 save "$output/sharpenedqvals.dta", replace
592
593 * hay que notar que esta correccion puede reducir los p-valores.
594 * esto puede ocurrir cuando los outcomes estan muy correlacionados entre sí.
595
596 restore
597
598 *-- estas lineas son para ver el output
599 preserve
600 use "$output/sharpenedqvals.dta", clear
601
602 restore
603 *--
604
605 *-- Almaceno los resultados en una matriz:
606
607 mat bky_pval = [0.001, 0.013, 0.407, 0.215, 0.179, 0.267, 0.12, 0.366, 0.398, 0.508, 0.143, 0.004,
     0.038, 0.342, 0.001, 0.429, 0.001, 0.029, 0.001, 0.008, 0.009]
608
609
610 *--Queda agregar estos p-valores ajustados a la tabla. Como antes, corro las regresiones y agrego
    por panel.
611
612 *definimos un iterados para poder ir accediendo a los elementos de la matriz con los niveles de
    significancia ajustados.
613 scalar i = 1
614
615 *--- PANEL A (Child's cognitive skills at follow up)
616
617 local bayley "b_tot_cog b_tot_lr b_tot_le b_tot_mf"
618 eststo clear
619 foreach y of local bayley{
620 local append append
621 if "`y'"=="b_tot_cog" local append replace
622     cap drop V*
623     reg `y'1_st treat `y'0_st $covs_eva , cluster(cod_dane)
624     eststo: test treat = 0
625     estadd scalar p_value = r(p)
626     estadd scalar corr_p_value = min(1,r(p)*hyp)
627     estadd scalar alpha_corr = alpha_corr_h[1,i]
628     estadd scalar bky_pval = bky_pval[1,i]
629     scalar i = i + 1
630 }
631 esttab using "$output/Cuadro2_panelA.tex", p se replace label ///
632 keep(treat) ///
633 cells(b(fmt(3)) se(par fmt(3))) ///
634 stats(p_value corr_p_value alpha_corr bky_pval blank N, fmt(3 3 3 3 0) labels("P_value"
    "Bonferroni Corrected P_value" "Holm Corrected Significance" "BKY Corrected P_value" " " "Number
    of Observations"))
635
636
637 local macarthur "mac_words mac_phrases"
638 foreach y of local macarthur{
639     cap drop V*
640     reg `y'1_st treat mac_words0_st $covs_ent , cluster(cod_dane)
641     eststo: test treat = 0
642     estadd scalar p_value = r(p)
643     estadd scalar corr_p_value = min(1,r(p)*hyp)
644     estadd scalar alpha_corr = alpha_corr_h[1,i]
645     estadd scalar bky_pval = bky_pval[1,i]
646     scalar i = i + 1
647 }
648 esttab using "$output/Cuadro2_panelA.tex", p se replace label ///
649 keep(treat) ///
650 cells(b(fmt(3)) se(par fmt(3))) ///
```

```
651  stats(p_value corr_p_value alpha_corr bky_pval blank N, fmt(3 3 3 3 0) labels("P_value"
     "Bonferroni Corrected P_value" "Holm Corrected Significance" "BKY Corrected P_value" " " "Number
     of Observations"))
652
653
654  *--- PANEL B (Child's socio-emotional skills at follow up)
655
656  local bates "bates_difficult bates_unsociable bates_unstoppable"
657  eststo clear
658  foreach y of local bates{
659      cap drop V*
660      reg `y'1_st treat `y'0_st $covs_ent, cl(cod_dane)
661      eststo: test treat = 0
662      estadd scalar p_value = r(p)
663      estadd scalar corr_p_value = min(1,r(p)*hyp)
664      estadd scalar alpha_corr = alpha_corr_h[1,i]
665      estadd scalar bky_pval = bky_pval[1,i]
666      scalar i = i + 1
667  }
668  esttab using "$output/Cuadro2_panelB.tex", p se replace label ///
669  keep(treat) ///
670  cells(b(fmt(3)) se(par fmt(3))) ///
671  stats(p_value corr_p_value alpha_corr bky_pval blank N, fmt(3 3 3 3 0) labels("P_value"
     "Bonferroni Corrected P_value" "Holm Corrected Significance" "BKY Corrected P_value" " " "Number
     of Observations"))
672
673  local roth "roth_inhibit roth_attention"
674  foreach y of local roth{
675      cap drop V*
676      reg `y'1_st treat bates_difficult0_st $covs_ent , cluster(cod_dane)
677      eststo: test treat = 0
678      estadd scalar p_value = r(p)
679      estadd scalar corr_p_value = min(1,r(p)*hyp)
680      estadd scalar alpha_corr = alpha_corr_h[1,i]
681      estadd scalar bky_pval = bky_pval[1,i]
682      scalar i = i + 1
683  }
684  esttab using "$output/Cuadro2_panelB.tex", p se replace label ///
685  keep(treat) ///
686  cells(b(fmt(3)) se(par fmt(3))) ///
687  stats(p_value corr_p_value alpha_corr bky_pval blank N, fmt(3 3 3 3 0) labels("P_value"
     "Bonferroni Corrected P_value" "Holm Corrected Significance" "BKY Corrected P_value" " " "Number
     of Observations"))
688
689
690  *--- PANEL C (Material investments)
691
692
693  local fcimat "fci_play_mat_type Npaintbooks Nthingsmove Ntoysshape Ntoysbought"
694  eststo clear
695  foreach y of local fcimat{
696      cap drop V*
697      reg `y'1_st treat fci_play_mat_type0_st $covs_ent , cluster(cod_dane)
698      eststo: test treat = 0
699      estadd scalar p_value = r(p)
700      estadd scalar corr_p_value = min(1,r(p)*hyp)
701      estadd scalar alpha_corr = alpha_corr_h[1,i]
702      estadd scalar bky_pval = bky_pval[1,i]
703      scalar i = i + 1
704
705  }
706  esttab using "$output/Cuadro2_panelC.tex", p se replace label ///
707  keep(treat) ///
708  cells(b(fmt(3)) se(par fmt(3))) ///
709  stats(p_value corr_p_value alpha_corr bky_pval blank N, fmt(3 3 3 3 0) labels("P_value"
     "Bonferroni Corrected P_value" "Holm Corrected Significance" "BKY Corrected P_value" " " "Number
     of Observations"))
710
```

```
711
712   *--- PANEL D (Time investments)
713
714   local fcitime "fci_play_act home_stories home_read home_toys home_name"
715   eststo clear
716   foreach y of local fcitime{
717       cap drop V*
718       reg `y'1_st treat fci_play_act0_st $covs_ent , cluster(cod_dane)
719       eststo: test treat = 0
720       estadd scalar p_value = r(p)
721       estadd scalar corr_p_value = min(1,r(p)*hyp)
722       estadd scalar alpha_corr = alpha_corr_h[1,i]
723       estadd scalar bky_pval = bky_pval[1,i]
724       scalar i = i + 1
725   }
726   esttab using "$output/Cuadro2_panelD.tex", p se replace label ///
727   keep(treat) ///
728   cells(b(fmt(3)) se(par fmt(3))) ///
729   stats(p_value corr_p_value alpha_corr bky_pval blank N, fmt(3 3 3 3 0) labels("P_value"
      "Bonferroni Corrected P_value" "Holm Corrected Significance" "BKY Corrected P_value" " " "Number
      of Observations"))
730
731
732
733
734
735
736
737
738   **********************************************-----FIN-----*****************************************
      ******
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
```