

# Nota técnica 1: Constucción de bases globales dinámica y estática

Centro de Estudios Espinosa Yglesias

Agosto 2021

## Resumen

Éste archivo baja y comprime las bases de datos de la ENOE del 1T 2020 y la ENOE\_N del 4T 2020 para construir dos bases de datos a partir de las tablas de datos *Sociodemográfico, Cuestionario de ocupación y empleo parte I* y *Cuestionario de ocupación y empleo parte II*. La base de datos **estática** se limita a unir todas las bases de datos disponibles, mientras que la base de datos **dinámica** compara los resultados de ciertas variables para personas con entrevistas disponibles a lo largo de un año.

## Descarga y condensación de bases de datos ENOE y ENOE\_N

### 1. Definición de directorios

Definimos el directorio en donde guardaremos las bases de datos.

```
35 gl root = "/Users/miusuario/midirectorio"
36 cap mkdir "$root/Bases ENOE"
37 gl bases = "$root/Bases ENOE"
```

### 2. Descarga y comprensión de datos

El siguiente paso es descargar y comprimir las bases de datos de la Encuesta Nacional de Ocupación y Empleo para los periodos que nos interesan (1T y 3T del 2020).

Encuesta Nacional de Ocupación y Empleo 2020-1T

```
44 /* (2.1): ENOE 1T 2020. */
45 disp "Trabajando para bases año 2020 trimestre 1"
46
47 * Creamos directorios y bajamos archivos:
48 capture mkdir "$bases/2020trim1_dta"
49 cd "$bases/2020trim1_dta"
50
51 * Revisamos si existen bases de datos para no volver a bajar
52 capture confirm file "COE1T120'.dta"
53 if _rc!=0 {
54     copy "https://www.inegi.org.mx/contenidos/programas/enoe/15ymas/microdatos/
55         2020trim1_dta.zip" 2020trim1_dta.zip
56
57     * Revisamos tamaño de archivo para ver si existe el año indicado. Si no existe,
58     paramos, si existe, seguimos:
59 }
```

```

60 qui checksum 2020trim1_dta.zip
61 if r(filelen)/1000000 < 1 {
62     disp "Aún no existen archivos para el año 2020 trimestre 1"
63     break
64 }
65 else {
66     unzipfile 2020trim1_dta.zip
67
68     ***** COMPRESIÓN DE BASES *****
69     * Primero base COE1T Y COE2T
70     use "COE1T120.dta", clear
71     rename *, lower
72     qui compress
73     save, replace
74     use "COE2T120.dta", clear
75     rename *, lower
76     qui compress
77     save, replace
78
79     * Segundo base hogar
80     use "HOGT120.dta", clear
81     rename *, lower
82     qui compress
83     save, replace
84
85     * Tercero base sociodemográficos
86     use "SDEMT120.dta", clear
87     rename *, lower
88     qui compress
89     save, replace
90
91     * Por último base de vivienda
92     use "VIVT120.dta", clear
93     rename *, lower
94     qui compress
95     save, replace
96 }
97 * Por último, se borra el archivo zip:
98 erase 2020trim1_dta.zip
99 }

```

#### Encuesta Nacional de Ocupación y Empleo (nueva edición) 2020-3T

```

99 /* (2.2): ENOE_N 3T 2020. */
100 disp "Trabajando para bases año 2020 trimestre 3"
101
102 * Creamos directorios y bajamos archivos:
103 capture mkdir "$bases/2020trim3_dta"
104 cd "$bases/2020trim3_dta"
105
106 * Revisamos si existen bases de datos para no volver a bajar
107 capture confirm file "COE1T320.dta"
108 if _rc!=0 {

```

```

109 copy "https://www.inegi.org.mx/contenidos/programas/enoe/15ymas/microdatos/
110 enoe_n_2020_trim3_dta.zip" enoe_n_2020_trim3_dta.zip
111
112 * Revisamos tamaño de archivo para ver si existe el año indicado. Si no existe,
113 paramos, si existe, seguimos:
114
115 qui checksum enoe_n_2020_trim3_dta.zip
116 if r(filelen)/1000000 < 1 {
117     disp "Aún no existen archivos para el año 2020 trimestre 3"
118     break
119 }
120 else {
121     unzipfile enoe_n_2020_trim3_dta.zip
122
123     ***** COMPRESIÓN DE BASES *****
124     * Primero base COE1T Y COE2T
125     use "enoen_coe1t320.dta", clear
126     rename *, lower
127     qui compress
128     save "COE1T320.dta", replace
129     erase enoen_coe1t320.dta
130     use "enoen_coe2t320.dta", clear
131     rename *, lower
132     qui compress
133     save "COE2T320.dta", replace
134     erase enoen_coe2t320.dta
135
136     * Segundo base hogar
137     use "enoen_hogt320.dta", clear
138     rename *, lower
139     qui compress
140     save "HOGT320.dta", replace
141     erase enoen_hogt320.dta
142
143     * Tercero base sociodemográficos
144     use "enoen_sdemt320.dta", clear
145     rename *, lower
146     qui compress
147     save "SDEMT320.dta", replace
148     erase enoen_sdemt320.dta
149
150     * Por último base de vivienda
151     use "enoen_vivt320.dta", clear
152     rename *, lower
153     qui compress
154     save "VIVT320.dta", replace
155     erase enoen_vivt320.dta
156 }
157 * Por último, se borra el archivo zip:
158 erase enoe_n_2020_trim3_dta.zip
159 }

```

## Construcción de bases globales

### 3. Base global estática

Cargamos la tabla Sociodemográfica del 1T-2020 para comenzar a unir las demás tablas por medio de sus llaves

```
163 cd "$root"
164 use "$bases/2020trim1_dta/SDEMT120.dta", clear
```

Fusionamos la tabla Sociodemográfica con el Cuestionario de ocupación y empleo parte 1. Después nos quedamos unicamente con aquellas variables que después usaremos.

```
167 qui merge 1:1 cd_a ent con v_sel n_hog h_mud n_ren using "$bases/2020trim1_dta/COE1T120.dta", force
168 keep if _merge==3
169 keep r_def loc mun est est_d ageb t_loc cd_a ent con upm d_sem n_pro_viv v_sel n_hog
170     h_mud n_ent per n_ren c_res par_c sex eda nac_dia nac_mes nac_anio l_nac_c ingocup
171     per fac cs_p13_1 cs_p13_2 cs_p17 clase* imssissste p1 p1c salario
```

Ahora fusionamos con el Cuestionario de ocupación y empleo parte 2. Nuevamente nos quedamos con aquellas variables que nos interesan.

```
162 qui merge 1:1 cd_a ent con v_sel n_hog h_mud n_ren using "$bases/2005trim1_dta/COE2T120.dta", force
163 keep if _merge==3
164 keep r_def loc mun est est_d ageb t_loc cd_a ent con upm d_sem n_pro_viv v_sel n_hog h_mud
165     n_ent per n_ren c_res par_c sex eda nac_dia nac_mes nac_anio l_nac_c ingocup per fac
166     cs_p13_1 cs_p13_2 cs_p17 clase* imssissste p1 p1c p11_1 salario p6c p6b2 p6_9 p6a3
```

Guardamos nuestra base unificada

```
179 save "$root/primertrim.dta", replace
```

Después hacemos lo propio para las bases del tercer trimestre.

Cargamos la tabla Sociodemográfica del 3T-2020 para comenzar a unir las demás tablas por medio de sus llaves

```
182 use "$bases/2020trim3_dta/SDEMT320.dta", clear
```

Renombramos variables trimestrales

```
185 rename est_d_tri est_d
186 rename t_loc_tri t_loc
187 rename fac_tri fac
```

Fusionamos la tabla Sociodemográfica con el Cuestionario de ocupación y empleo parte 1. Después nos quedamos unicamente con aquellas variables que después usaremos.

```
190 qui merge 1:1 cd_a ent con v_sel tipo mes_cal ca n_hog h_mud n_ren using "$bases/
191                                     2020trim3_dta/COE1T320.dta"
192 keep if _merge==3
193 keep r_def loc mun est est_d ageb t_loc cd_a ent con upm d_sem n_pro_viv v_sel tipo
194     mes_cal ca n_hog h_mud n_ent per n_ren c_res par_c sex eda nac_dia nac_mes nac_anio
195     l_nac_c ingocup per fac cs_p13_1 cs_p13_2 cs_p17 clase* imssissste p1 p1c salario
```

Ahora fusionamos con el Cuestionario de ocupación y empleo parte 2. Nuevamente nos quedamos con aquellas variables que nos interesan.

```
196 qui merge 1:1 cd_a ent con v_sel tipo mes_cal ca n_hog h_mud n_ren using "$bases/  
197                                     2020trim3_dta/COE2T320.dta"  
198 keep if _merge==3  
199 keep r_def loc mun est est_d ageb t_loc cd_a ent con upm d_sem n_pro_viv v_sel tipo  
200     n_hog h_mud n_ent per n_ren c_res par_c sex eda nac_dia nac_mes nac_anio l_nac_c  
201     ingocup per fac cs_p13_1 cs_p13_2 cs_p17 clase* imssissste p1 p1c p9_1 salario  
202     p6c p6b2 p6_9 p6a3
```

Guardamos nuestra base unificada

```
202 save "tercertrim.dta", replace
```

Contruimos la base global estática uniendo ambas bases trimestrales.

```
205 use "primertrim.dta", clear  
206 append using "tercertrim.dta", force
```

Generamos identificador único

```
210 egen foliop = concat(cd_a ent con v_sel n_hog h_mud n_ren sex nac_dia nac_anio nac_mes)  
211 egen folioh = concat(cd_a ent con v_sel n_hog h_mud n_ren)
```

Clasificador del tipo de localidad

```
214 gen rururb = cond(t_loc>=1 & t_loc<=3,0,1)  
215 label define ru 0 "Urbano" 1 "Rural"  
216 label values rururb ru
```

Variable año-trimestre y la misma con lag de un año

```
219 gen year = substr(string(per),2,2)  
220 gen trim = substr(string(per),1,1)  
221 egen yeartrim = concat(year trim)  
222 destring yeartrim, replace  
223 gen int yeartrim_lag = .  
224 replace yeartrim_lag = yeartrim - 9  
225 replace yeartrim_lag = 192 if yeartrim == 201  
226 replace yeartrim_lag = 201 if yeartrim == 204  
227 egen base = group(yeartrim)
```

Variable carácter de año, mes y fecha

```
229 gen anio = "20" + year  
230 destring trim, replace  
231 gen mes = string(trim*3)  
232 replace mes = "0" + mes if strlen(mes)==1  
233 generate str fecha = anio + "-" + mes + "-01"  
234 compress
```

Por último, guardamos base global estática

```
237 save "ENOE_Base Global_Estatica.dta", replace
```

#### 4. Base global dinámica

Tiramos las entrevistas entermedias

```
245 drop if n_ent != 1 & n_ent != 4
```

Mantenemos solo a aquellos que tienen entrevista en el primer y cuarto trimestre

```
248 qui duplicates tag foliop, gen(dup)
249 qui keep if dup == 1
```

Guardamos base temporal

```
252 save "temp.dta", replace
```

Nos quedamos solo con las observaciones de cuatro trimestres antes y renombramos las variables de interés

```
256 qui keep if yeartrim == 201
257 rename ingocup ingocup1
258 rename imssissste imssissste1
259 rename clase1 clase1ini
260 rename clase2 clase2ini
261 rename clase3 clase3ini
```

Guardamos base temporal a

```
264 save "tempa.dta", replace
```

Cargamos nuestra base temporal para realizar el mismo tratamiento al cuarto trimestre.

```
267 use "temp.dta", clear
```

Nos quedamos solo con las observaciones del trimestre actual y renombramos variables.

```
270 qui keep if yeartrim == 204
271 rename ingocup ingocup2
272 rename imssissste imssissste2
273 rename clase1 clase1fin
274 rename clase2 clase2fin
275 rename clase3 clase3fin
276 rename fac factor
277
278 capture drop _merge
```

Ahora unimos las bases efectivas de cada trimestre para general el panel

```

281 qui merge m:m foliop using "tempa.dta"
282 qui keep if _merge == 3
283 qui drop dup

```

Tiramos a ausentes definitivos, nos quedamos con rango de edad de PEA, entrevistas completas y PEA.

```

286 drop if r_def != 0
287 drop if c_res == 2
288 drop if eda < 12 | eda == 99
289 keep if clase1ini == 1
290 compress

```

Guardamos la base global dinámica.

```

293 save "ENOE_Base Global_Dinamica.dta", replace

```

Por último, eliminamos todas las bases de datos temporales.

```

296 erase "primertrim.dta"
297 erase "tercertrim.dta"
298 erase "temp.dta"
299 erase "tempa.dta"
300 shell rmdir "Bases ENOE" /s /q

```