

“Study SHARE”

Tesi Rafa Llorens

`rafallorems@euit.fdsll.cat`

Version 1.0

- Joan Vila -

`joanviladomenech@gmail.com`

December 5, 2021

1 Version History

Version	Effective Date	Changes
1	06-Dec-2021	Exploració de les dades + expressió de dubtes

Contents

1	Version History	1
2	Introduction	3
3	Preguntes generals o no relatives a la taula 1	9
4	Preguntes relatives a la taula 1	13

2 Introduction

Es parteix de la base de dades “Easy567depurada.sav”, que va fer arribar en Rafa el 2/12/2021.

La base de dades 34680 registres, a partir de tres entrevistes realitzades a 11560 individus.

Concretament les entrevistes es van realitzar:

```
##  
## 2013 2015 2017  
## 11560 11560 11560
```

A continuació hi ha una taula amb una descripció de les característiques basals (Ola = 5) segons sexe.

Hi són pràcticament totes les variables de la base de dades, més algunes creades per JVila. Algunes variables no hi són, però es comenten a les seccions “PREGUNTES”.

Table 1: General characteristics at recruitment by sex

	Female N=6289	Male N=5271	[ALL] N=11560	p.overall	N
Country:				0.089	11560
Austria	40 (0.64%)	42 (0.80%)	82 (0.71%)		
Belgium	427 (6.79%)	358 (6.79%)	785 (6.79%)		
Czech Republic	481 (7.65%)	338 (6.41%)	819 (7.08%)		
Denmark	670 (10.7%)	553 (10.5%)	1223 (10.6%)		
Estonia	52 (0.83%)	64 (1.21%)	116 (1.00%)		
France	45 (0.72%)	31 (0.59%)	76 (0.66%)		
Germany	1427 (22.7%)	1253 (23.8%)	2680 (23.2%)		
Italy	549 (8.73%)	465 (8.82%)	1014 (8.77%)		
Luxembourg	436 (6.93%)	353 (6.70%)	789 (6.83%)		
Slovenia	363 (5.77%)	262 (4.97%)	625 (5.41%)		
Spain	963 (15.3%)	806 (15.3%)	1769 (15.3%)		
Sweden	808 (12.8%)	720 (13.7%)	1528 (13.2%)		
Switzerland	28 (0.45%)	26 (0.49%)	54 (0.47%)		
Couple data	3086 (49.1%)	3084 (58.5%)	6170 (53.4%)	<0.001	11560
Homo/Hetero relationship: Homo	2 (0.03%)	0 (0.00%)	2 (0.02%)	0.504	11560
Born country Interv.	5665 (90.2%)	4754 (90.3%)	10419 (90.2%)	0.990	11545
Age	63.1 (10.3)	64.3 (9.38)	63.7 (9.90)	<0.001	11560
Born year Group:				<0.001	11560
1900-1928	189 (3.01%)	116 (2.20%)	305 (2.64%)		
1929-1938	844 (13.4%)	791 (15.0%)	1635 (14.1%)		
1939-1948	1713 (27.2%)	1635 (31.0%)	3348 (29.0%)		
1949-1963	3306 (52.6%)	2696 (51.1%)	6002 (51.9%)		
1964-1999	237 (3.77%)	33 (0.63%)	270 (2.34%)		
Age Group:				<0.001	11560
<50	207 (3.29%)	24 (0.46%)	231 (2.00%)		
>85	206 (3.28%)	141 (2.68%)	347 (3.00%)		
50-64	3238 (51.5%)	2619 (49.7%)	5857 (50.7%)		
65-74	1753 (27.9%)	1668 (31.6%)	3421 (29.6%)		
75-84	885 (14.1%)	819 (15.5%)	1704 (14.7%)		

continued on next page

Table 1 – *continued from previous page*

	Female N=6289	Male N=5271	[ALL] N=11560	p.overall	N
Civil Status:				<0.001	11536
casado/a o pareja registrada	4474 (71.3%)	4240 (80.6%)	8714 (75.5%)		
divorciado/a o separado/a	645 (10.3%)	435 (8.27%)	1080 (9.36%)		
soltero/a	318 (5.07%)	349 (6.64%)	667 (5.78%)		
viudo/a	840 (13.4%)	235 (4.47%)	1075 (9.32%)		
Couple Age	64.7 (9.47)	61.1 (9.64)	62.9 (9.72)	<0.001	9184
Living with a Couple	4717 (75.0%)	4475 (84.9%)	9192 (79.5%)	<0.001	11560
Numero de hijos	2.00 [1.00;3.00]	2.00 [1.00;3.00]	2.00 [1.00;3.00]	0.110	11472
numeronietos	2.00 [0.00;4.00]	1.00 [0.00;3.00]	1.00 [0.00;4.00]	<0.001	11441
personas que conviven en el hogar	2.00 [2.00;2.00]	2.00 [2.00;3.00]	2.00 [2.00;2.00]	<0.001	11560
Algun hijo/a vive cerca:				0.002	10008
no	2717 (49.5%)	2261 (50.0%)	4978 (49.7%)		
si	2255 (41.1%)	1752 (38.7%)	4007 (40.0%)		
sin hijos	512 (9.34%)	511 (11.3%)	1023 (10.2%)		
¿Sigue viva la madre?	1887 (30.0%)	1378 (26.2%)	3265 (28.3%)	<0.001	11539
¿Sigue vivo el padre?	860 (13.8%)	624 (11.9%)	1484 (12.9%)	0.004	11485
Alive brothers	2.00 [1.00;3.00]	2.00 [1.00;3.00]	2.00 [1.00;3.00]	0.375	10308
nivelestudios:				<0.001	11516
Alto	2411 (38.5%)	1644 (31.3%)	4055 (35.2%)		
bajo	1526 (24.4%)	1545 (29.4%)	3071 (26.7%)		
medio	2328 (37.2%)	2062 (39.3%)	4390 (38.1%)		
Años de educación	12.0 [9.00;14.0]	12.0 [9.00;15.0]	12.0 [9.00;14.0]	<0.001	11432
situacionlaboral:				<0.001	11357
Ama de casa	855 (13.9%)	11 (0.21%)	866 (7.63%)		
desempleado	218 (3.54%)	208 (4.00%)	426 (3.75%)		
discapacitado/enfermedad permanente	219 (3.56%)	144 (2.77%)	363 (3.20%)		
empleado/autonomo	2152 (35.0%)	1970 (37.9%)	4122 (36.3%)		
Jubilado/a	2713 (44.1%)	2867 (55.1%)	5580 (49.1%)		
llegarfindemes:				<0.001	11438
bastante facil	1838 (29.6%)	1554 (29.7%)	3392 (29.7%)		
con alguna dificultad	1432 (23.1%)	1048 (20.1%)	2480 (21.7%)		

continued on next page

Table 1 – *continued from previous page*

	Female N=6289	Male N=5271	[ALL] N=11560	p.overall	N
Con gran dificultad	461 (7.42%)	336 (6.43%)	797 (6.97%)		
facilmente	2481 (39.9%)	2288 (43.8%)	4769 (41.7%)		
zonaresidencia:				0.582	11069
area rural	1756 (29.1%)	1520 (30.2%)	3276 (29.6%)		
Gran ciudad	717 (11.9%)	564 (11.2%)	1281 (11.6%)		
Pueblo grande	1092 (18.1%)	906 (18.0%)	1998 (18.1%)		
Pueblo pequeño	1792 (29.7%)	1505 (29.9%)	3297 (29.8%)		
Suburbios/afueras gran ciudad	677 (11.2%)	540 (10.7%)	1217 (11.0%)		
Ingresoshogar	27824 [15985;46260]	31644 [18706;51795]	29701 [17144;49108]	<0.001	11560
Recibió ayuda de otros (fuera de hh)	1181 (18.8%)	834 (15.9%)	2015 (17.5%)	<0.001	11535
Estado de salud infantil:				<0.001	11509
Buena	1740 (27.8%)	1382 (26.4%)	3122 (27.1%)		
Excelente	2004 (32.0%)	1840 (35.1%)	3844 (33.4%)		
Justa	520 (8.29%)	360 (6.87%)	880 (7.65%)		
La salud varía mucho	33 (0.53%)	24 (0.46%)	57 (0.50%)		
Muy buena	1804 (28.8%)	1536 (29.3%)	3340 (29.0%)		
Pobre	168 (2.68%)	98 (1.87%)	266 (2.31%)		
Vacunas durante la infancia	6023 (96.7%)	5048 (96.9%)	11071 (96.8%)	0.504	11435
enfermedades cronicas:				<0.001	11560
entre 0 y 1	4479 (71.2%)	3535 (67.1%)	8014 (69.3%)		
entre 2 y 10	1810 (28.8%)	1736 (32.9%)	3546 (30.7%)		
Acti.Vida Diaria Indice W & H:				0.085	11560
con limitacion	447 (7.11%)	331 (6.28%)	778 (6.73%)		
sin informacion	6 (0.10%)	10 (0.19%)	16 (0.14%)		
sin limitacion	5836 (92.8%)	4930 (93.5%)	10766 (93.1%)		
dificultadmovilidad:				<0.001	11544
Alguna dificultad	1356 (21.6%)	777 (14.8%)	2133 (18.5%)		
Mucha dificultad	274 (4.36%)	165 (3.14%)	439 (3.80%)		
Sin dificultad	4653 (74.1%)	4319 (82.1%)	8972 (77.7%)		
saludautopercibida:				0.001	11546
Buena	2293 (36.5%)	2072 (39.4%)	4365 (37.8%)		

continued on next page

Table 1 – *continued from previous page*

	Female N=6289	Male N=5271	[ALL] N=11560	p.overall	N
Excelente	714 (11.4%)	609 (11.6%)	1323 (11.5%)		
Muy buena	1278 (20.3%)	1087 (20.7%)	2365 (20.5%)		
Pesima	492 (7.83%)	355 (6.75%)	847 (7.34%)		
Regular	1507 (24.0%)	1139 (21.6%)	2646 (22.9%)		
visitasmedico12meses:				<0.001	10185
de 1 a 5	3407 (60.6%)	2973 (65.1%)	6380 (62.6%)		
de 11 a 20	718 (12.8%)	518 (11.3%)	1236 (12.1%)		
de 21 a 30	162 (2.88%)	132 (2.89%)	294 (2.89%)		
de 31 a 100	144 (2.56%)	105 (2.30%)	249 (2.44%)		
de 6 a 10	1187 (21.1%)	839 (18.4%)	2026 (19.9%)		
ingresohospital12meses	806 (12.8%)	719 (13.7%)	1525 (13.2%)	0.192	11544
Índice de masa corporal	26.3 (5.17)	27.2 (4.04)	26.7 (4.70)	<0.001	11364
Categorías de índice de masa corporal:				<0.001	11364
normal	2705 (44.0%)	1585 (30.4%)	4290 (37.8%)		
obesidad	1216 (19.8%)	1044 (20.0%)	2260 (19.9%)		
peso insuficiente	101 (1.64%)	23 (0.44%)	124 (1.09%)		
sobrepeso	2122 (34.5%)	2568 (49.2%)	4690 (41.3%)		
Fuma a diario	2317 (36.9%)	2939 (55.9%)	5256 (45.6%)	<0.001	11527
bebe alcohol a diario:				<0.001	11541
bebe casi todos los días	717 (11.4%)	1520 (28.9%)	2237 (19.4%)		
bebe entre 1 y 4 días a la semana	1731 (27.6%)	1788 (34.0%)	3519 (30.5%)		
no bebe o menos de 1-2 veces al mes	3832 (61.0%)	1953 (37.1%)	5785 (50.1%)		
Actividad fisica:				<0.001	11546
activo/a	3162 (50.3%)	3035 (57.7%)	6197 (53.7%)		
no activo/a	3122 (49.7%)	2227 (42.3%)	5349 (46.3%)		
Escala de depresion EURO-D:				<0.001	11324
Alta depresión	234 (3.79%)	83 (1.61%)	317 (2.80%)		
Baja depresión	4316 (70.0%)	4291 (83.2%)	8607 (76.0%)		
Moderada depresión	1617 (26.2%)	783 (15.2%)	2400 (21.2%)		
CASP Indice calidad de vida y bienenestar:				<0.001	11071
alta calidad de vida	746 (12.3%)	609 (12.1%)	1355 (12.2%)		

continued on next page

Table 1 – *continued from previous page*

	Female N=6289	Male N=5271	[ALL] N=11560	p.overall	N
baja calidad de vida	1818 (30.1%)	1316 (26.2%)	3134 (28.3%)		
moderada calidad de vida	602 (9.96%)	530 (10.5%)	1132 (10.2%)		
muy alta calidad de vida	2876 (47.6%)	2574 (51.2%)	5450 (49.2%)		

3 Preguntes generals o no relatives a la taula 1

1. Segons el document Projecte tesis doctoral Rafa Llorens v3.docx, a l'estudi SHARE es van entrevistar 65.653 individus. Perquè la base de dades sòls en té 11560?
2. Com és que tots el 11560 tenen visites a les tres onades (anys 2013, 2015 i 2017). Per exemple, entre el 2013 i el 2017, no es va morir ningú? Tots van sobreviure per poder fer la "Ola" 5, 6 i 7?
3. A les dades que dispo, la primera entrevista es va fer a l'any 2013. Són correctes els següents anys de naixement?

```
##      [,1]
## 1912     1
## 1915     1
## 1916     1
## 1918     1
## 1919     2
## 1920    12
## 1921    10
## 1922    11
## 1923    12
## 1924    30
## 1925    39
## 1926    43
## 1927    49
## 1928    93
## 1929    90
## 1930   100
## 1931   105
## 1932   140
## 1933   159
## 1934   168
## 1935   199
## 1936   213
## 1937   211
## 1938   250
## 1939   227
## 1940   315
## 1941   272
## 1942   290
## 1943   302
## 1944   382
## 1945   330
## 1946   413
## 1947   391
## 1948   426
## 1949   377
## 1950   400
## 1951   411
## 1952   411
## 1953   404
## 1954   385
## 1955   423
## 1956   399
```

```
## 1957 391
## 1958 394
## 1959 393
## 1960 359
## 1961 624
## 1962 546
## 1963 85
## 1964 73
## 1965 44
## 1966 34
## 1967 31
## 1968 22
## 1969 14
## 1970 13
## 1971 9
## 1972 6
## 1973 3
## 1974 1
## 1975 3
## 1976 3
## 1977 2
## 1978 4
## 1979 2
## 1981 1
## 1982 4
## 1983 1
```

4. L'estudi SHARE es va fer amb persones de 50 anys a més. Com és que hi ha aquestes edats?

```
##      [,1]
## 29      1
## 30      2
## 31      3
## 33      1
## 34      1
## 35      5
## 36      3
## 37      1
## 38      4
## 39      1
## 40      3
## 41      7
## 42     17
## 43      7
## 44     19
## 45     32
## 46     31
## 47     35
## 48     55
## 49     77
## 50    304
## 51    601
```

```
## 52 453
## 53 382
## 54 400
## 55 369
## 56 425
## 57 397
## 58 408
## 59 405
## 60 385
## 61 420
## 62 410
## 63 407
## 64 385
## 65 417
## 66 384
## 67 379
## 68 353
## 69 353
## 70 318
## 71 257
## 72 311
## 73 264
## 74 240
## 75 228
## 76 224
## 77 193
## 78 193
## 79 146
## 80 152
## 81 136
## 82 108
## 83 97
## 84 89
## 85 74
## 86 53
## 87 40
## 88 26
## 89 22
## 90 13
## 91 8
## 92 13
## 93 8
## 94 1
## 95 1
## 96 1
## 97 1
## 100 1
```

5. Per què es diu s'ha entrevistat a la parella i no hi ha dades d'algunes parelles?

```
##      N      Y
## 1 1675 6166
## 2 2374      4
```

Hi ha 1675 persones que a la variable “entrevistapareja” tenen un 1 (Entrevista de pareja disponible) i no hi ha ningú que tingui un ID com ell.

Hi ha 4 persones (2 parelles) que tenen el mateix ID de parella, però tenen un 2 a “entrevistapareja”:

```
##      Identificadorpersona Identificadorhogar Identificadorpareja
## 2036          SE-285504-01          SE-285504-A          SE-285504-01-02
## 34593          SE-285504-02          SE-285504-A          SE-285504-01-02
## 25885          SI-548294-01          SI-548294-A
## 25285          SI-548294-02          SI-548294-A
##      entrevistapareja
## 2036                2
## 34593                2
## 25885                2
## 25285                2
```

6. És important treballar amb el nombre de malalties cròniques, o amb la categòrica n’hi ha prou?

```
##      entre 0 y 1  entre 2 y 10
## 0             4376             0
## 1             3638             0
## 2              0            2055
## 3              0            958
## 4              0            367
## 5              0            123
## 6              0             32
## 7              0             10
## 8              0              1
```

7. **IMPORTANT:** la variable *outcome* del teu estudi, és CASP en 4 categories (alta, baixa, moderada i molt alta).

Veuràs s’han de fer models multivariats bastant complicats. Si la variable *outcome* és dicotòmica o numèrica, tot i que ja és prou difícil, no és massa complexa analitzar i sobretot explicar el resultat. Si té, com la teva 4 categories, tot es complica molt.

La pregunta és: no teniu la variable CASP de forma numèrica enlloc de categories?

4 Preguntes relatives a la taula 1

1. Fixa't amb la columna de la dreta del tot “N”. Totes les variables haurien de tenir 11560 individus. Si no els tenen és perquè hi ha *missing*. Per què hi ha *missing*?
Si es fan anàlisis multivariades, sols que tingui missing en una de les variables, és perd tot l'individu!
2. Hi ha dades de 6170 persones que són parella (3085 parelles). Sòls 2 persones (una parella) són homosexuals?
3. Com és que hi ha *missings* a la variable CASP, que és l'*outcome* de l'estudi?