

# Sesion\_5\_STATA

July 29, 2022

## 1 Sesión 5 de Stata

Estoy probando esta nueva interacción de *Jupyter Notebooks* con *Stata* para poder trabajar de manera más cómoda (como un ser humano).

La idea de esta sesión es trabajar dos tipo de transformación que cambian la naturaleza misma de la tabla al reacomodar su estructura: **collapse** y **reshape**. El uso de estos comandos cambia **por completo** su tabla (base de datos).

### 1.1 Setup

Como siempre, haremos un *do file* con su respectivo setup y nuestro (ahora obsoleto) *log*.

```
[1]: *=====

// PROYECTO: Quinta introductoria a STATA
// PROGRAMA: sesion_5.do
// TAREA: De long a wide
// CREADO POR: Manuel Toral
// FECHA DE CREACIÓN: 2022-07-28
// FECHA DE ACTUALIZACIÓN: 2022-07-28

*=====

cd "C:\Users\User\Documents\MUCD_stata_intro\Sesion_5"
log using "logs_notebooks\log_sesion5.smcl", text replace name(sesion5)
```

C:\Users\User\Documents\MUCD\_stata\_intro\Sesion\_5

```
-----
      name:  sesion5
      log:   C:\Users\User\Documents\MUCD_stata_intro\Sesion_5\logs_notebooks\lo
> g_sesion5.smcl
      log type:  text
      opened on: 29 Jul 2022, 10:07:03
```

## 2 Datos

Vamos a usar los datos de incidencia delictiva del Secretariado, para no perder la costumbre.

```
[8]: // 1. Datos
import delimited "data\IDEFC_NM_jun22.csv", clear
```

(19 vars, 25,088 obs)

Ante de continuar, podemos observar cómo la tabla tiene una estructura *horizontal (wide)* que tiene columnas con información *del mismo tipo* repartidas por varias variables. Es decir, si vemos la primer fila, las variables **enero** y **febrero** contienen números que son de la misma naturaleza y son consistentes entre sí: el número de homicidios dolosos con arma de fuego.

```
[3]: list in 1
```

```

+-----+
1. | año | clave_~t | entidad | bienjurídicoafectado |
   | 2015 | 1 | Aguascalientes | La vida y la Integridad corporal |
   |-----+-----+
   | tipodedelito | subtipodedelito | modalidad | enero | febrero |
   | Homicidio | Homicidio doloso | Con arma de fuego | 3 | 0 |
   |-----+-----+
   | marzo | abril | mayo | junio | julio | agosto | septie~e | octubre |
   | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 |
   |-----+-----+
   | noviem~e | diciem~e |
   | 2 | 1 |
   +-----+

```

Y esta estructura está muy bien para cuando necesitamos estudiar un tipo de fenómeno, pero justo una de las ventajas de la tabla de incidencia delictiva es que nos permite hacer **comparaciones en el tiempo**. Por lo que el reto es hacer *una sola variable* que agrupe los números comparables.

### 3 Colapsar

Pero antes, para facilitarnos la vida, vamos a hacer un colapso. Esto nos va a permitir utilizar la información que necesitamos y no más. La idea de “colapsar” es agregar. Es decir, como su nombre lo indica, colapsar la estructura de nuestra tabla en formas más simples a través de operaciones aritméticas.

Vamos a resolver esta pregunta, una que casi nadie se hace en México: *¿Cómo se comportan los homicidios dolosos con arma de fuego en el tiempo?*

Como pueden ver en la documentación, `collapse` tiene una estructura como esta:

```
collapse variable1 variable2 by(grupo)
```

En español podría ser: *colapsa la estructura de esta tabla con base en la “variable 1” y la “variable 2”, pero conserva los grupos de la variable “grupo”.*

Entonces, la instrucción para resolver la pregunta que acabamos de plantear tendría que ser de la siguiente forma (en español): 1. *conserva sólo aquellas observaciones que se tengan como subtipo de delito “Homicidio doloso” y como modalidad “con arma de fuego”.* 2. *colapsa la estructura de*

IDEFC\_NM\_jun22 con base en las variables de “enero” a “diciembre”, pero conserva los grupos de la variable “año”.

```
[4]: // 2. Transformaciones y limpieza}

// 2.1 Conserva sólo el subtipo de delito y modalidad que definen "Homicidio
↳doloso con arema de fuego"

keep if subtipodelito == "Homicidio doloso" & modalidad == "Con arma de fuego"

d
tab modalidad
tab subtipodelito
```

Contains data

```
obs:      256
vars:      19
```

variable name	storage type	display format	value label	variable label
año	int	%8.0g		Año
clave_ent	byte	%8.0g		Clave_Ent
entidad	str31	%31s		Entidad
bienjurídicoaño	str52	%52s		Bien jurídico afectado
tipodelito	str78	%78s		Tipo de delito
subtipodelito	str78	%78s		Subtipo de delito
modalidad	str84	%84s		Modalidad
enero	int	%8.0g		Enero
febrero	int	%8.0g		Febrero
marzo	int	%8.0g		Marzo
abril	int	%8.0g		Abril
mayo	int	%8.0g		Mayo
junio	int	%8.0g		Junio
julio	int	%8.0g		Julio
agosto	int	%8.0g		Agosto
septiembre	int	%8.0g		Septiembre
octubre	int	%8.0g		Octubre
noviembre	int	%8.0g		Noviembre
diciembre	int	%8.0g		Diciembre

Sorted by:

Note: Dataset has changed since last saved.

Modalidad	Freq.	Percent	Cum.
Con arma de fuego	256	100.00	100.00
Total	256	100.00	

Subtipo de delito	Freq.	Percent	Cum.
Homicidio doloso	256	100.00	100.00
Total	256	100.00	

```
[5]: collapse (sum) enero-diciembre, by(año)
d
list
```

Contains data

```
obs:      8
vars:     13
```

variable name	storage type	display format	value label	variable label
año	int	%8.0g		Año
enero	double	%8.0g		(sum) enero
febrero	double	%8.0g		(sum) febrero
marzo	double	%8.0g		(sum) marzo
abril	double	%8.0g		(sum) abril
mayo	double	%8.0g		(sum) mayo
junio	double	%8.0g		(sum) junio
julio	double	%8.0g		(sum) julio
agosto	double	%8.0g		(sum) agosto
septiembre	double	%8.0g		(sum) septiembre
octubre	double	%8.0g		(sum) octubre
noviembre	double	%8.0g		(sum) noviembre
diciembre	double	%8.0g		(sum) diciembre

Sorted by: año

Note: Dataset has changed since last saved.

+-----+

1.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2015	679	702	684	734	816	751	756	865
	septie~e	octubre	noviem~e	diciem~e					
	776	794	859	793					
2.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2016	884	842	901	853	1001	959	1151	1217
	septie~e	octubre	noviem~e	diciem~e					
	1198	1139	1188	1145					
3.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2017	1296	1220	1278	1275	1437	1474	1355	1394
	septie~e	octubre	noviem~e	diciem~e					
	1468	1647	1504	1543					
4.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2018	1511	1410	1626	1632	1700	1635	1788	1765
	septie~e	octubre	noviem~e	diciem~e					
	1770	1727	1697	1816					
5.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2019	1776	1722	1658	1614	1722	1813	1803	1704
	septie~e	octubre	noviem~e	diciem~e					
	1621	1685	1739	1670					
6.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2020	1736	1643	1881	1789	1720	1668	1748	1693
	septie~e	octubre	noviem~e	diciem~e					
	1592	1633	1558	1556					

7.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2021	1655	1529	1713	1637	1737	1489	1598	1718
	septie~e		octubre	noviem~e		diciem~e			
	1588		1671	1441		1594			
8.	año	enero	febrero	marzo	abril	mayo	junio	julio	agosto
	2022	1468	1366	1543	1479	1621	1542	0	0
	septie~e		octubre	noviem~e		diciem~e			
	0		0	0		0			

```
[6]: // 3. Convertir de wide a long

rename (enero-diciembre) (homicidios(#)) ,addnumber

gen id = _n

reshape long homicidios, i(id) j(mes)

d

reshape wide homicidios, i(id) j(mes)

d
```

(note: j = 1 2 3 4 5 6 7 8 9 10 11 12)

```
-----
Number of obs.                8    ->    96
Number of variables           14    ->     4
j variable (12 values)        ->    mes
xij variables:
```

```
-----

Contains data
  obs:          96
  vars:          4
-----
```

variable name	storage type	display format	value label	variable label
id	float	%9.0g		
mes	byte	%9.0g		
año	int	%8.0g		Año
homicidios	double	%8.0g		

Sorted by: id mes

Note: Dataset has changed since last saved.

(note: j = 1 2 3 4 5 6 7 8 9 10 11 12)

Number of obs.	96	->	8
Number of variables	4	->	14

xij variables:

Contains data

obs:	8
vars:	14

variable name	storage type	display format	value label	variable label
id	float	%9.0g		
homicidios1	double	%8.0g		1 homicidios
homicidios2	double	%8.0g		2 homicidios
homicidios3	double	%8.0g		3 homicidios
homicidios4	double	%8.0g		4 homicidios
homicidios5	double	%8.0g		5 homicidios
homicidios6	double	%8.0g		6 homicidios
homicidios7	double	%8.0g		7 homicidios
homicidios8	double	%8.0g		8 homicidios
homicidios9	double	%8.0g		9 homicidios
homicidios10	double	%8.0g		10 homicidios
homicidios11	double	%8.0g		11 homicidios
homicidios12	double	%8.0g		12 homicidios
año	int	%8.0g		Año

Sorted by: id

Note: Dataset has changed since last saved.

```
[7]: log close session5
```

```
      name:  session5  
      log:   C:\Users\User\Documents\MUCD_stata_intro\Sesion_5\logs_notebooks\lo  
> g_session5.smcl  
      log type:  text  
      closed on: 29 Jul 2022, 10:07:10  
-----
```