



PUCP



STATISTICS • VISUALIZATION • DATA MANIPULATION • REPORTING

DEPARTAMENTO DE ECONOMÍA
LABORATORIO DE ECONOMETRÍA: STATA
1ECO31

Sesión 4

Estructura de Bases

Docente: Juan Palomino



PUCP

Índice

1

El comando collapse

2

El comando append

3

El comando reshape

4

El comando preserve y restore

5

Duplicados de Observaciones

6

El comando merge

1. El comando collapse

El comando collapse

	sectores	region	vab	empleados
1	Agricultura	Amazonas	1200	25
2	Agricultura	Lima	1350	90
3	Agricultura	Loreto	4980	30
4	Agricultura	Tacna	850	30
5	Agricultura	Ucayali	320	20
6	Mineria	Amazonas	1100	15
7	Mineria	Lima	2300	35
8	Mineria	Loreto	3250	45
9	Mineria	Tacna	3948	20
10	Mineria	Ucayali	2910	30
11	Pesca	Amazonas	430	10
12	Pesca	Lima	290	15
13	Pesca	Loreto	365	25
14	Pesca	Tacna	950	30
15	Pesca	Ucayali	220	45
16	Electricidad	Amazonas	4392	60
17	Electricidad	Lima	2144	70
18	Electricidad	Loreto	201	80
19	Electricidad	Tacna	9212	90
20	Electricidad	Ucayali	988	100



collapse (mean) vab (sum) empleados, by(sectores)

	sectores	vab	empleados
1	Agricultura	1740	195
2	Electricidad	3387.4	400
3	Mineria	2701.6	145
4	Pesca	451	125

Valor promedio
del VAB

Suma total de
empleados

El comando collapse

	sectores	region	vab	empleados
1	Agricultura	Amazonas	1200	25
2	Agricultura	Lima	1350	90
3	Agricultura	Loreto	4980	30
4	Agricultura	Tacna	850	30
5	Agricultura	Ucayali	320	20
6	Mineria	Amazonas	1100	15
7	Mineria	Lima	2300	35
8	Mineria	Loreto	3250	45
9	Mineria	Tacna	3948	20
10	Mineria	Ucayali	2910	30
11	Pesca	Amazonas	430	10
12	Pesca	Lima	290	15
13	Pesca	Loreto	365	25
14	Pesca	Tacna	950	30
15	Pesca	Ucayali	220	45
16	Electricidad	Amazonas	4392	60
17	Electricidad	Lima	2144	70
18	Electricidad	Loreto	201	80
19	Electricidad	Tacna	9212	90
20	Electricidad	Ucayali	988	100



collapse (sum) vab (max) empleados, by(region)

	region	vab	empleados
1	Amazonas	7122	60
2	Lima	6084	90
3	Loreto	8796	80
4	Tacna	14960	90
5	Ucayali	4438	100

Valor total del
VAB

Valor máximo de
empleados

2. El comando append

El comando append

Este comando permite juntar dos bases verticalmente, es decir, combinar observaciones entre dos bases, siempre y cuando una base contenga la mayoría de variables de la otra base.

Master dataset				
	nombre	edad	...	ciudad
1	Carmen	45	...	Lima
2	Luis	65	...	Lima
3	Manuel	21	...	Lima
...
n	Rosa	89	...	Lima



Using dataset				
	nombre	edad	...	ciudad
1	Esteban	45	...	Cusco
2	Miryam	54	...	Cusco
3	Manuel	15	...	Cusco
...
n	Lucy	8	...	Cusco



Resultado				
	nombre	edad	...	ciudad
1	Carmen	45	...	Lima
2	Luis	65	...	Lima
3	Manuel	21	...	Lima
...
n	Rosa	89	...	Lima
(n+1)	Esteban	45	...	Cusco
(n+2)	Miryam	54	...	Cusco
(n+3)	Manuel	15	...	Cusco
...
(n+m)	Lucy	8	...	Cusco

3. El comando reshape

Base wide

Es aquella donde la temporalidad del identificador (individuo, región, país) se captura mediante la columna de las variables.

id	sex	ingreso2018	ingreso2019	ingreso2020
A	Hombre	520	600	620
B	Mujer	600	580	500
C	Hombre	700	650	600
D	Mujer	550	600	650

Base long

Es aquella donde la temporalidad del identificador (individuo, región, país) se captura mediante la fila de observaciones.

id	sex	año	ingreso
A	Hombre	2018	520
A	Hombre	2019	600
A	Hombre	2020	620
B	Mujer	2018	600
B	Mujer	2019	580
B	Mujer	2020	500
C	Hombre	2018	700
C	Hombre	2019	650
C	Hombre	2020	600
D	Mujer	2018	550
D	Mujer	2019	600
D	Mujer	2020	650

El comando reshape long

Este comando permite pasar de una base wide a una base long.

```
reshape long ingreso, i(id sex) j(año)
```

id	sex	ingreso2018	ingreso2019	ingreso2020
A	Hombre	520	600	620
B	Mujer	600	580	500
C	Hombre	700	650	600
D	Mujer	550	600	650



id	sex	año	ingreso
A	Hombre	2018	520
A	Hombre	2019	600
A	Hombre	2020	620
B	Mujer	2018	600
B	Mujer	2019	580
B	Mujer	2020	500
C	Hombre	2018	700
C	Hombre	2019	650
C	Hombre	2020	600
D	Mujer	2018	550
D	Mujer	2019	600
D	Mujer	2020	650

El comando reshape wide

Este comando permite pasar de una base long a una base wide.

id	sex	año	ingreso
A	Hombre	2018	520
A	Hombre	2019	600
A	Hombre	2020	620
B	Mujer	2018	600
B	Mujer	2019	580
B	Mujer	2020	500
C	Hombre	2018	700
C	Hombre	2019	650
C	Hombre	2020	600
D	Mujer	2018	550
D	Mujer	2019	600
D	Mujer	2020	650



```
reshape wide ingreso, i(id sex) j(año)
```

id	sex	ingreso2018	ingreso2019	ingreso2020
A	Hombre	520	600	620
B	Mujer	600	580	500
C	Hombre	700	650	600
D	Mujer	550	600	650

4. El comando preserve y restore

El comando preserve y restore

Los comandos “**preserve**” y “**restore**” son dos comandos que se deben usar en una misma ejecución de programación.

Ejemplo

- El comando “**preserve**” toma una foto de la base de datos “base_laboral_2020”.
- A partir de ahí, se pueden realizar modificaciones en la base como, por ejemplo, filtrar y guardar bases con las observaciones de Lima y Amazonas.
- Con el comando “**restore**” se revierte todo a la base inicial.

```
use "$works/base_laboral_2020.dta", clear

preserve
keep if dpto==15
save "$works/base_laboral_lima.dta", replace
restore

preserve
keep if dpto==01
save "$works/base_laboral_amazonas.dta", replace
restore
```

5. Duplicados de Observaciones

El comando duplicates

Existe un paquete de comandos que permite hallar, analizar, tabular o eliminar los duplicados de las observaciones de una variable.

- **duplicates list**: enlista todos los duplicados de la base.
- **duplicates report**: reporta duplicados de una variable o identificador.
- **duplicates tag**: crea una variable que identifica la cantidad de duplicados por observación.
- **duplicates drop**: borra los duplicados de alguna variable en particular.

6. El comando merge

El comando merge 1:1

El merge 1:1 (one-to-one) es aquel donde una observación del Master se corresponde con una observación del Using.

Master dataset			
	nombre	edad	sexo
1	Carmen	45	femenino
2	Luis	65	masculino
2	Manuel	21	masculino
3	Rosa	89	femenino



Using dataset			
	nombre	ingresos	profesión
1	Carmen	480	economista
2	Luis	950	abogado
2	Manuel	620	biólogo
3	Rosa	780	antropólogo

Resultado						
	nombre	edad	sexo	ingresos		profesión
1	Carmen	45	femenino	480		economista
2	Luis	65	masculino	950		abogado
3	Manuel	21	masculino	620		biólogo
4	Rosa	89	femenino	780		antropólogo

El comando merge m:1

El merge m:1 (many-to-one) es aquel donde un grupo de observaciones del Master corresponde con observaciones del Using.

Master dataset					
	nombre	edad	sexo	...	ciudad
1	Carmen	45	femenino	...	Lima
2	Luis	65	masculino	...	Lima
3	Manual	21	masculino	...	Lima
4	Rafael	27	masculino	...	Cusco
5	Efrain	65	masculino	...	Cusco
6	Zoila	82	femenino	...	Huaraz
7	Rosa	89	femenino	...	Lima



Using dataset				
	ciudad	población	...	altitud
1	Lima	1200	...	161
2	Cusco	780	...	3399
3	Huaraz	920	...	3052

Resultado								
	nombre	edad	sexo	...	ciudad	población	...	altitud
1	Carmen	45	femenino	...	Lima	1200		161
2	Luis	65	masculino	...	Lima	1200		161
3	Manual	21	masculino	...	Lima	1200		161
4	Rafael	27	masculino	...	Cusco	780		3399
5	Efrain	65	masculino	...	Cusco	780		3399
6	Zoila	82	femenino	...	Huaraz	920		3052
7	Rosa	89	femenino	...	Lima	1200		161

El comando merge 1:m

El merge 1:m (one-to-many) es aquel donde una observación del Master se corresponde con un grupo de observaciones del Master.

Master dataset				
	ciudad	población	...	altitud
1	Lima	1200	...	161
2	Cusco	780	...	3399
3	Huaraz	920	...	3052



Using dataset					
	nombre	edad	sexo	...	ciudad
1	Carmen	45	femenino	...	Lima
2	Luis	65	masculino	...	Lima
3	Manual	21	masculino	...	Lima
4	Rafael	27	masculino	...	Cusco
5	Efrain	65	masculino	...	Cusco
6	Zoila	82	femenino	...	Huaraz
7	Rosa	89	femenino	...	Lima

Resultado								
	nombre	edad	sexo	...	ciudad	población	...	altitud
1	Carmen	45	femenino	...	Lima	1200		161
2	Luis	65	masculino	...	Lima	1200		161
3	Manual	21	masculino	...	Lima	1200		161
4	Rafael	27	masculino	...	Cusco	780		3399
5	Efrain	65	masculino	...	Cusco	780		3399
6	Zoila	82	femenino	...	Huaraz	920		3052
7	Rosa	89	femenino	...	Lima	1200		161



PUCP