

Tipos de datos y variables

Erika R. Badillo

erika.badilloen@unaula.edu.co

Facultad de Economía

Universidad Autónoma Latinoamericana

- La estructura de los datos económicos
- Tipo de variables económicas

- Wooldridge, Jeffrey (2013). *Introducción a la econometría*. 5a edición, Cengage Learning. [Capítulo 1.3](#)
- Gujarati, Damodar (2010). *Econometría*. 5a edición, Mc Graw Hill. [Capítulo 1.7](#)

- Datos de sección cruzada (*Cross-sectional data*)
- Datos de series de tiempo (*Time series data*)
- Datos panel o longitudinales (*Panel data*)

Datos de sección cruzada

- **Definición:** un conjunto de datos de sección cruzada consiste en una muestra de individuos, hogares, firmas, ciudades, estados, países, o variedad de otras unidades **tomadas en un punto dado en el tiempo**
- Una importante característica de los datos de sección cruzada es que se asume que estos datos han sido obtenidos de una **muestra aleatoria** de una población objetivo. Por ejemplo: salarios, educación, experiencia y otras características de la población
- Los datos de sección cruzada son ampliamente utilizadas en economía y otras ciencias sociales
- En economía el análisis de estos tipos de datos está cercanamente alineado con el campo de la microeconomía aplicada: economía laboral, organización industrial, economía urbana, economía de la educación, economía de la salud, etc

La estructura de los datos económicos

Datos de sección cruzada

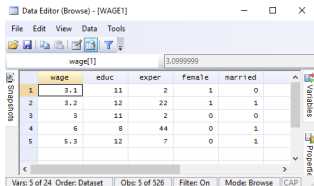
En Stata: datos del *Current Population Survey* de 1976

```
cd "C:\...\Data Wooldridge"
```

```
use WAGE1.dta, clear  
des
```

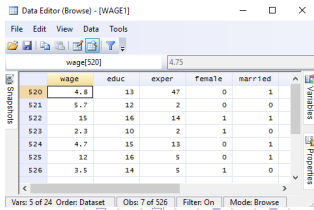
```
contains data from WAGE1.dta  
obs:      526  
vars:     24  
size:    16,032  
16 Sep 1996 15:52  
  
+-----+-----+-----+-----+-----+  
variable name  storage  display  value    variable label  
+-----+-----+-----+-----+-----+  
wage           float    %8.2g    average hourly earnings  
educ           byte     %8.0g    years of education  
exper          byte     %8.0g    years potential experience  
tenure         byte     %8.0g    years with current employer  
nonwhite       byte     %8.0g    =1 if nonwhite  
female         byte     %8.0g    =1 if female  
married        byte     %8.0g    =1 if married  
numdep         byte     %8.0g    number of dependents  
smsa           byte     %8.0g    =1 if live in SMSA  
northern       byte     %8.0g    =1 if live in north central U.S.  
south          byte     %8.0g    =1 if live in southern region  
west           byte     %8.0g    =1 if live in western region  
construc       byte     %8.0g    =1 if work in construc. indus.  
indurman       byte     %8.0g    =1 if in nondur. manuf. indus.  
trcompu        byte     %8.0g    =1 if in trans. commun. pub ut  
trade          byte     %8.0g    =1 if in wholesale or retail  
services       byte     %8.0g    =1 if in services indus.  
profserv       byte     %8.0g    =1 if in prof. serv. indus.  
profocc        byte     %8.0g    =1 if in profess. occupation  
clerocc        byte     %8.0g    =1 if in clerical occupation  
servocc        byte     %8.0g    =1 if in service occupation  
lwage          float    %9.0g    log(wage)  
expersq        int       %9.0g    exper^2  
tenursq        int       %9.0g    tenure^2
```

```
brow wage educ exper female married in 1/5
```



	wage	educ	exper	female	married
1	3.1	11	2	1	0
2	3.2	12	22	1	1
3	3	11	2	0	0
4	6	8	44	0	1
5	5.3	12	7	0	1

```
brow wage educ exper female married in 520/526
```



	wage	educ	exper	female	married
520	4.8	13	47	0	1
521	5.7	12	2	0	0
522	15	16	14	1	1
523	2.3	10	2	1	0
524	4.7	15	13	0	1
525	12	16	5	0	1
526	3.5	14	5	1	0

Datos de sección cruzada

- Diferentes variables algunas veces corresponden a diferentes períodos de tiempo en datos de sección cruzada
- Por ejemplo, para determinar los efectos de las políticas sobre el crecimiento de largo plazo, se ha estudiado la relación entre el crecimiento en el PIB per capita real sobre un cierto período (1960-1985) y variables determinadas en parte por políticas en 1960, como el consumo del gobierno como porcentaje del PIB y las tasas de adultos con secundaria

TABLE 1.2 A Data Set on Economic Growth Rates and Country Characteristics

obsno	country	gpcrgdp	govcons60	second60
1	Argentina	0.89	9	32
2	Austria	3.32	16	50
3	Belgium	2.56	13	69
4	Bolivia	1.24	18	12
.
.
.
61	Zimbabwe	2.30	17	6

Datos de series de tiempo

- **Definición:** los datos de series de tiempo consiste de observaciones sobre una o más variables a lo largo del tiempo
- Ejemplos: los precios, la oferta monetaria, IPC, PIB, tasas de homicidios, ventas de automóviles, etc
- El tiempo es importante: a diferencia de los datos de sección cruzada, el orden cronológico de las observaciones en los datos de series de tiempo conlleva a potencialmente importante información
- Dos características importantes en los datos de series de tiempo son la **dependencia temporal** (los eventos pasados pueden influir sobre los eventos futuros) y la presencia de **tendencias sobre el tiempo**. Estos factores deben ser tenidos en cuenta antes aplicar técnicas econométricas estándar
- Otra característica a tener en cuenta en este tipo de datos es la **frecuencia de los datos**: datos diarios, semanales, mensuales, trimestrales o anuales. Este tipo de datos presentan fuertes patrones estacionales, lo cual debe ser tenido en cuenta en los modelos de regresión

La estructura de los datos económicos

Datos de series de tiempo

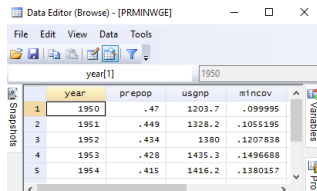
En Stata: datos anuales de Puerto Rico entre 1950 y 1987 sobre la tasa de empleo, salario mínimo y otras variables usadas en Castillo-Freeman y Freeman (1992) para estudiar los efectos del salario mínimo sobre el empleo

```
cd "C:\...\Data Wooldridge"
```

```
use PRMINWGE.DTA.dta, clear  
des
```

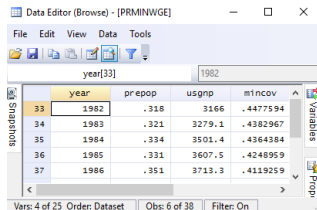
```
Contains data from D:\Dropbox\EAFIT\Cursos EAFIT\2016-II\Econometria I\Data Wooldridge\PRMINWGE.DTA
obs:      38  
vars:     25      17 Aug 1999 21:41  
size:    3,496  
  
+-----+-----+-----+-----+-----+  
variable name   storage   display   value   variable label  
                type      format                label  
+-----+-----+-----+-----+-----+  
year            int       %8.0g          1950-1987  
avgmin          float     %8.0g          weighted avg min wage, 44 indust  
avgwage         float     %8.0g          weighted avg hrly wage, 44 indust  
kaltz           float     %8.0g          Kaitz min wage index  
avgcov         float     %8.0g          weighted avg coverage, 8 indust  
covt           float     %8.0g          economy-wide coverage of min wg  
efwage         float     %8.0g          avg manuf. wage  
pindex         float     %8.0g          Puerto Rican price deflator  
prepop         float     %8.0g          PR employ/popul ratio  
prepopf        float     %8.0g          PR employ/popul ratio, alter.  
prgnp          float     %8.0g          PR GNP  
prunemp        float     %8.0g          PR unemployment rate  
usgnp          float     %8.0g          US GNP  
t              byte       %8.0g          time trend: 1 to 38  
post74         byte       %8.0g          time trend: starts in 1974  
lprunemp       float     %8.0g          log(prunemp)  
lprgnp         float     %8.0g          log(prgnp)  
lavgwage       float     %8.0g          log(avgwage)  
lkaltz         float     %8.0g          log(kaltz)  
lprun_1        float     %8.0g          lprunemp[_n-1]  
lprepop        float     %8.0g          log(prepop)  
lprepop_1      float     %8.0g          lprepop[_n-1]  
lmincov        float     %8.0g          (avgmin/avgwage)*avgcov  
lalnccov       float     %8.0g          log(lnccov)  
lavgmin        float     %8.0g          log(avgmin)
```

```
brow prepop usgnp mincov in 1/5
```



	year	prepop	usgnp	mincov
1	1950	.47	1203.7	.099995
2	1951	.449	1328.2	.1055195
3	1952	.434	1380	.1207838
4	1953	.428	1435.3	.1496688
5	1954	.415	1416.2	.1380157

```
brow prepop usgnp mincov in 33/38
```



	year	prepop	usgnp	mincov
33	1982	.318	3166	.4477594
34	1983	.321	3279.1	.4382967
35	1984	.334	3501.4	.4364384
36	1985	.331	3607.5	.4248959
37	1986	.351	3713.3	.4119259

Datos panel o longitudinales

- **Definición:** los datos panel o longitudinales consisten en series de tiempo para cada unidad de sección cruzada. En un intervalo de tiempo se vigilan las mismas unidades.
- Ejemplo: información sobre salarios, educación y empleo para un conjunto de individuos en un período de 10 años. O datos de crimen de dos años para 150 ciudades

TABLE 1.5 A Two-Year Panel Data Set on City Crime Statistics

obsno	city	year	murders	population	unem	police
1	1	1986	5	350000	8.7	440
2	1	1990	8	359200	7.2	471
3	2	1986	2	64300	5.4	75
4	2	1990	1	65100	5.5	75
.
.
.
297	149	1986	10	260700	9.6	286
298	149	1990	6	245000	9.8	334
299	150	1986	25	543000	4.3	520
300	150	1990	32	546200	5.2	493

Datos panel o longitudinales

Los datos panel presentan importantes ventajas sobre los datos de sección cruzada:

- la existencia de múltiples observaciones sobre la misma unidad permite controlar por **características inobservadas** de los individuos, firmas, etc
- permiten estudiar la importancia de los efectos de **los rezagos**. Esto es de particular importancia teniendo en cuenta que muchas medidas de política puede ser esperadas a tener un impacto después de un tiempo

Variables discretas

- **Definición:** es una variable que puede tomar sólo un finito número de valores
- Ejemplo: el número de niños en un hogar: 0, 1, 2,...
el número de visitas a un mall particular por semana: 0, 1, 2,...
- Las variables discretas son usadas comúnmente en economía para caracterizar cualidades, o no numéricas características de un agente económico \implies en este rol estas variables son llamadas **discretas**, **dummy** o **variables cualitativas**
- Por ejemplo, si se esta estudiando el comportamiento de los hogares, se puede querer indagar sobre el género del cabeza del hogar. Esto se puede hacer con la siguiente variable D :

$$D = \begin{cases} 1 & \text{si el ind. cabeza de hogar es mujer} \\ 0 & \text{si el ind. cabeza de hogar es hombre} \end{cases}$$

- Cualquier característica cualitativa con dos estados (si o no) puede ser caracterizada por una variable discreta (binaria)

Variables continuas

- **Definición:** es una variable que puede tomar cualquier valor real mayor o igual a cero
- Ejemplo: el PIB, la oferta monetaria, la tasa de interés, los ingresos de los hogares...