

GUIA DE DATOS DE PANEL

Este documento contiene una descripción de los diferentes datos de panel que se usan en el libro de “Econometric analysis of cross section and panel data” (2002) de J. Wooldridge. Para ello es necesario distinguir qué bases son datos de panel y cuáles son solo de corte transversal.

Libro	Datos de Panel	Corte Transversal
NO	AIRFARE	401KSUBS
SI	CORNWELL	APPLE
NO	EZUNEM	ATTEND
SI	GPA	BWGT
SI	JTRAIN1	CARD
SI	KEANE	CPS78_85
SI	LOWBIRTH	CPS91
NO	MATHPNL	CRIME
SI	MURDER	FERTIL1
NO	NLS81_87	FERTIL2
SI	NORWAY	FRINGE
NO	PATENT	HPRICE
SI	PRISON	INJURY
NO	Q	JTRAIN2
SI	WAGEPAN	LABSUP
		LOANAPP
		MROZ
		NBASAL
		NLS80
		OPENNESS
		PENSION
		RECID
		SMOKE
		VOTE

Es importante tener en cuenta que todas las bases de datos de panel se pueden utilizar también como datos de corte transversal, pero no al revés

La columna “libro” de la tabla anterior solo hace referencia a las bases de datos de panel, y un “si” indica que hay ejemplos o ejercicios disponibles en el libro de Wooldridge para esa base en concreto. Las páginas donde se encuentran se detallan en la explicación de cada base.

Para cualquier consulta sobre esta guía pueden ponerse en contacto a través de email en palo1980@hotmail.com y songorka@mundofree.com

AIRFARE

Datos de panel de vuelos comerciales desde 1997 hasta 2000.

Nº de variables: 16

Nº de observaciones: 4.596

Nº de individuos: 1149 vuelos

Nº de períodos: 4 años, desde 1997 hasta 2000

Definición de algunas de las variables:

- *origin* → lugar de salida del avión
- *destin* → lugar de destino del vuelo
- *dist* → distancia recorrida
- *fare* → precio del billete
- *id* → identificador de ruta
- *dist* → distancia en millas
- *passen* → media de pasajeros por día.

CORNWELL

Mide el nivel de delitos por condado

Nº de variables: 59

Nº de observaciones: 630

Nº de individuos: 90 tipos de delitos

Nº de períodos: 7, de 1981 hasta 1987

Definición de algunas de las variables:

- *county* → identificador del estado.
- *crm rte* → crímenes cometidos per capita
- *prbconv* → probabilidad de ser condenado.
- *prbpris* → probabilidad de ir a prisión.
- *avgsen* → media de la sentencia (en días).
- *polpc* → policía per capita.
- *wser* → salario semanal en sector de servicios industriales.
- *wmfc* → salario semanal en el sector manufacturero.

Posibles modelos:

- **Página 181, ejercicio 7.11 y página 294 , ejercicio 10.9**

$E(\text{crm rte} | \text{prbarr}, \text{prbconv}, \text{prbpris}, \text{avgsen}, \text{polpc})$

- **Página 305, ejemplo 11.3**

$$\Delta \log(\text{crrmrte}) = 0,65 + 0,0212 \Delta \log(\text{crrmrte})_{-1}$$

EZUNEM

Datos de panel sobre el número de peticiones de subsidio de desempleo y de la presencia de parques empresariales.

Nº de variables: 198

Nº de observaciones: 41

Nº de individuos: 22 ciudades

Nº de períodos: 9 períodos, desde 1980 hasta 1988

Definición de algunas de las variables:

- *uclms* → solicitudes de subsidio de desempleo
- *ez* → toma valor 1 si tiene parque empresarial.
- *cX* → variable ficticia de ciudad (x toma distintos valores dependiendo de la ciudad).
- *dY* → variable ficticia de año (Y toma valores 80, 81, 82...88).

GPA

Contiene los datos de notas medias de estudiantes deportistas (GPA: Grade Point Average). Los datos están divididos en dos semestres (otoño y primavera) para cada estudiante. GPA es la nota media del expediente de los estudiantes en Estados Unidos.

Nº de variables: 26

Nº de observaciones: 732

Nº de individuos: 366 estudiantes deportistas

Nº de períodos: 2 semestres

Definición de algunas de las variables:

- *term* → se distingue entre trimestre académico de otoño de 1988 y trimestre académico de primavera de 1989
- *sat* → indica los resultados del test de aptitudes escolares.
- *tohor* → total horas por semestre.
- *black* → toma valor 1 en el caso en que sea de raza negra
- *white* → toma valor 1 en el caso en que sea de raza blanca
- *spring* → toma valor 1 en el caso de que sea primavera
- *season* → toma valor 1 si se encuentra en temporada de competición.
- *verbmath* → ratio entre la nota de lengua y matemáticas del test SAT
- *id* → código de estudiante
- *sport* → tipo de deporte

Posibles modelos:

- **Página 173, el ejemplo 7.8**

$$trmgpa_{it} = \mathbf{b}_0 + \mathbf{b}_1 spring_t + \mathbf{b}_2 cumgpa_{it} + \mathbf{b}_3 crsgpa_{it} + \mathbf{b}_4 firstsem_{it} + \mathbf{b}_5 season_{it} + \mathbf{b}_6 SAT_i + \mathbf{b}_7 verbmth_i + \mathbf{b}_8 hsperc_i + \mathbf{b}_9 hssize_i + \mathbf{b}_{10} black_i + \mathbf{b}_{11} female + u_{it}$$

JTRAIN1

Base de Datos acerca de la cantidad de chatarra producida por 150 empresas de Michigan.

Nº de variables: 30

Nº de observaciones: 471

Nº de individuos: 150 empresas

Nº de períodos: 3 años (1987, 1988 y 1989)

Definición de algunas de las variables:

- *scrap* → mide el ratio de cantidad de chatarra por cada 100 artículos.
- *fcode* → identificador de las empresas.
- *sales* → número de ventas.
- *avgsal* → media del salario de los empleados.
- *employ* → número de empleados por planta
- *grant* → toma valor 1 si se recibe subvención
- *totrain* → total de empleados entrenados
- *hrsemp* → mide el ratio número de horas trabajadas / número de horas de entrenamiento por empleado.

Posibles modelos:

- **Página 172, ejemplo 7.7; Página 272 + 276, ejemplo 10.5**

$$\log(scrap) = \mathbf{b}_1 d88 + \mathbf{b}_2 d89 + \mathbf{b}_3 grant + \mathbf{b}_4 grant_{-1}$$

d88 y *d89* son variables que toman valor 1 en el caso que año sea 1988 ó 1989 respectivamente

- **Página 337, ejercicio 11.15**

$$\log(scrap_{it}) = \mathbf{q} + \mathbf{b}_1 hrsemp_{it} + c_i + u_{it}$$

KEANE

Datos de panel sobre escolarización y empleo para una muestra de hombres (2231) desde 1981 hasta 1987.

Nº de variables: 25

Nº de observaciones: 12.723

Nº de individuos: 2.231

Nº de períodos: 7 años, desde 1981 hasta 1987

Definición de algunas de las variables:

- *id* → variable identificador de los hombres.
- *numyr* → número de años en la muestra
- *educ* → años de escolarización.
- *wage* → salario medio anual.
- *educ* → años de educación
- *expwc* → experiencia en trabajos de “cuello blanco” (de oficina)
- *expbc* → experiencia en trabajos de “cuello azul”
- *expser* → experiencia en el sector servicios.
- *black* → toma valor 1 si es de raza negra.

Posibles modelos:

- **Página 515, ejercicio 15.19**

$$P(\text{employ}_t = 1 | \text{employ}_{i,t-1}) = \Phi(\mathbf{d}_0 + p\text{employ}_{i,t-1})$$

LOWBIRTH

Base de datos acerca del porcentaje de bebés con bajo peso al nacer medido en los años 1987 y 1990.

Nº de variables: 36 variables

Nº de observaciones: 100

Nº de individuos: 50

Nº de períodos: 2 años, 1987 y 1990

Definición de algunas de las variables:

- *infmort* → mortalidad infantil
- *pcinc* → ingresos per capita
- *state* → nombre del estado al que pertenece
- *bedspc* → número de camas de hospital per capita
- *d90* → variable ficticia que vale 1 en 1990
- *afdcprc* → porcentaje de la población que pertenece al programa “Aid to families with dependent children” (AFDC)
- *phypc* → logaritmo de la proporción de médicos per capita
- *popul* → población.
- *stateabb* → código postal de cada estado.

Posibles modelos:

- **Página 295 el ejercicio 10.11**

$$lowbrth_{it} = \alpha_1 + \alpha_2 d90 + \beta_1 afdcprc_{it} + \beta_2 \log(phypc_{it}) + \beta_3 \log(bedspc_{it}) + \beta_4 \log(pcinc_{it}) + \beta_5 \log(popul_{it}) + c_i + u_{it}$$

MATHPNL

Datos de panel con el porcentaje de alumnos de cuarto y séptimo grado que avanzan satisfactoriamente junto con información sobre el distrito escolar.

Nº de variables: 54 variables

Nº de observaciones: 3.850

Nº de individuos: 550

Nº de períodos: 7 años, desde 1992 hasta 1998

Definición de algunas de las variables:

- *distid* → identificador del distrito
- *enrol* → número de alumnos matriculados en cada colegio
- *expp* → gastos por alumno.
- *drop* → porcentaje de alumnos que abandonan el instituto.
- *grad* → porcentaje de alumnos graduados.
- *avgben* → salario medio de profesores.
- *math4* → porcentaje de alumnos de cuarto grado aprobados en matemáticas.
- *math7* → porcentaje de alumnos de séptimo grado aprobados en matemáticas.
- *staff* → equipamiento por cada 1000 estudiantes.
- *cpi* → índice de precios al consumo.

MURDER:

Datos de Panel acerca del porcentaje de asesinatos de EEUU durante los años 1987, 1990 y 1993.

Nº de variables: 13

Nº de observaciones: 153

Nº de individuos: 51 Estados

Nº de períodos: 3 años, 1987, 1990 o 1993

Definición de algunas de las variables:

- *id* → identificador del estado.
- *state* → código postal de cada estado.
- *mrdrte* → ratio de asesinatos.
- *exec* → total de las ejecuciones en los últimos 3 años.
- *unem* → tasa de desempleo anual.
- *d90* y *d93* → toman valor 1 si están el año 1990 ó 1993 respectivamente.

Posibles modelos:

• **Página 295, ejercicio 10.10**

$$mrdрте_{it} = \mathbf{q}_t + \mathbf{b}_1 exec_{it} + \mathbf{b}_2 unem_{it} + c_i + u_{it}$$

NLS81_87

Datos de panel que reflejan la información sobre temas relacionados con la educación, los salarios e información demográfica, obtenidas a través de encuestas.

Nº de variables: 20

Nº de observaciones: 3.710

Nº de individuos: 530 personas

Nº de períodos: 7 años, desde 1981 hasta 1987

Definición de algunas de las variables:

- *wage* → representa el salario anual medio
- *educ* → años de educación del individuo encuestado
- *exper* → se refiere a la experiencia profesional
- *manuf* → toma el valor 1 si el individuo trabaja en la industria
- *black* → toma el valor 1 si el individuo es de raza negra
- *id* → identificador del individuo

NORWAY

Índice de crímenes en Noruega

Nº de variables: 12

Nº de observaciones: 106

Nº de individuos: 53 distritos de Noruega

Nº de períodos: 2 años, 1972 y 1978

Definición de algunas de las variables:

- *distric* → Identificación de cada uno de los distritos.
- *crime* → Índice de crímenes por cada 1000 personas.
- *clrprcX* → Cantidad de crímenes resueltos en los X años anteriores.
- *avgclr* → media de los crímenes resueltos.

Posibles modelos:

• **Página 194, ejercicio 10.8**

$$\log(\text{crime}_{it}) = \mathbf{q} + \mathbf{q}_1 d78_t + \mathbf{b}_1 clrprc_{i,t-1} + \mathbf{b}_2 clrprc_{i,t-2} + c_i + u_{it}$$

PATENT

Datos de panel acerca del numero de patentes buscadas y obtenidas en una muestra de 226 empresas a lo largo de información privilegiada de empresas.

Nº de variables: 48

Nº de observaciones: 2.269

Nº de individuos: 226 empresas

Nº de períodos: 10 años, desde 1972 hasta 1981

Definición de algunas de las variables:

- *cusip* → identificador de las empresas.
- *patents* → patentes aplicada.
- *patentsg* → patentes concedidas.
- *sales* → ventas.
- *patentsg* → patentes subvencionadas.
- *employ* → trabajadores.

PRISON

Datos de panel acerca de los ratios de crímenes estatales y las poblaciones reclusas.

Nº de variables: 45

Nº de observaciones: 714

Nº de individuos: 51 estados

Nº de períodos: 3 años, desde 1980 hasta 1983

Definición de algunas de las variables:

- *state* → estados numerados y ordenados por orden alfabético.
- *govelec* → toma valor 1 si se encuentra en elecciones.
- *black* → proporción de individuos de raza negra.
- *unem* → proporción de desempleados.
- *crip* → proporción de crímenes por 100.000 habitantes.
- *incpc* → ingreso per capita nominales.
- *plpc* → ratio de policías por cada 100.000 habitantes.
- *pris* → población de las prisiones por cada 100.000 habitantes.

Posibles modelos:

- **Pagina 309, en el ejemplo 11.6**

$$\log(\text{crime}_{it}) = \theta_t + \beta_1 \log(\text{pris}_{it}) + x_{it} \gamma + c_i + u_{it}$$

Que se puede diferenciar:

$$\Delta \log(\text{crime}_{it}) = \zeta_t + \beta_1 \Delta \log(\text{pris}_{it}) + \Delta x_{it} \gamma + \Delta u_{it}$$

Q

Contiene datos acerca de la Q de Tobin para varias empresas.

Nº de variables: 28

Nº de observaciones: 2068

Nº de individuos: 188 empresas

Nº de períodos: 11 años, desde 1975 hasta 1985

Definición de algunas de las variables:

- *ikn* → ratio inversión/capital con medida en sentido amplio.
- *ikb* → ratio inversión/capital con medida en sentido estrecho.
- *cusip* → identificador de las empresas.
- *qb* → q de Tobin con medida amplia.
- *qn* → q de Tobin con medida estricta.

WAGEPAN

Base de datos de panel sobre hombres que trabajaron todos los años desde 1980 hasta 1987.

Nº de variables: 44

Nº de observaciones: 4360

Nº de individuos: 545

Nº de períodos: 7 años, desde 1980 hasta 1987

Definición de algunas de las variables:

- *nr* → variable que identifica a de los individuos.
- *hours* → horas trabajadas.
- *agric* → toma valor 1 si se dedica al sector de la agricultura.
- *black* → toma valor 1 si es de raza negra.
- *construc* → toma valor 1 si se dedica al sector de la construcción.
- *fin* → toma valor 1 si se dedica al sector financiera.
- *hisp* → toma valor 1 si es hispano.
- *married* → toma valor 1 si está casado.
- *rur* → toma valor 1 si vive en un área rural.
- *union* → toma valor 1 si está afiliado a un sindicato.
- *exper* → experiencia en el mercado de trabajo.
- *educ* → años de educación.

Posibles modelos:

- **Página 21, ejemplo 2.3**

$E(\text{wage} | \text{educ}, \text{exper})$

- **Página 296, Ejercicio 10.45**

$$\log(\text{wage}_{it}) = \mathbf{q}_t + \mathbf{b}_1 \text{educ}_i + \mathbf{b}_2 \text{black}_i + \mathbf{b}_3 \text{hispan}_i + \mathbf{b}_4 \exp \text{er}_{it} \\ + \mathbf{b}_5 \exp \text{er}^2_{it} + \mathbf{b}_6 \text{married}_{it} + \mathbf{b}_7 \text{union}_{it} + c_i + u_{it}$$