

# Описание данных

Н.В. Артамонов

12 февраля 2023 г.

## 1 Описание наборов данных

В этом разделе приведены описания основных наборов данных, используемых для решения задач.

Для работы с нужным датасетом необходимо загрузить его командой `data(dataset, package)`

Таблица 1: Набор данных `Labour` из пакета `Ecdat` (569 наблюдений) с данными о бельгийских фирмах за 1996 г.

<code>capital</code>	капитал (в млн евро)
<code>labour</code>	число сотрудников
<code>output</code>	выпуск (в млн евро)
<code>wage</code>	зарплата на одного сотрудника (в тыс евро)

Таблица 2: Набор данных **Electricity** из пакета **Ecdat** (158 наблюдений) с данными о производителях электроэнергии в US за 1970 г.

cost	общие издержки за год
q	общий выпуск электроэнергии
pl	уровень зарплата (wage rate)
pk	цена привлечения капитала (capital price index)
pf	цена на топливо (fuel price)

Таблица 3: Набор данных **sleep75** из пакета **wooldridge** (706 наблюдений). Основные переменные. Источник данных [5]

age	возраст (в годах)
educ	уровень образования (в года)
inlf	бинарная, 1 если участник рынка труда
leis1	нерабочее время, sleep-totwrk
smsa	бинарная, 1 если живёт в мегаполисе
male	гендерный фактор (бинарная, 1 если мужчина)
marr	семейный статус (бинарная, 1 если женат/замужем)
prot	бинарная, 1 если протестант
selfe	бинарная, 1 если самозанятый
sleep	продолжительность сна (мин/нед)
south	географический фактор (бинарная, 1 если живёт на юге)
spsepay	доход супруга/супруги
spwrk75	бинарная, 1 если супруг(а) работает
totwrk	занятость (мин/нед)
union	бинарная, 1 если член профсоюза
yngkid	бинарная, 1 если есть дети младше 3 лет
ytsmarr	сколько лет женат/замужем
hrwage	почасовая оплата

Таблица 4: Набор данных `wage2` из пакета `wooldridge` (935 наблюдений).  
Основные переменные. Источник данных [6]

<code>wage</code>	месячная зарплата
<code>hours</code>	недельная занятость в часах
<code>IQ</code>	результаты теста IQ
<code>KWW</code>	результаты теста knowledge of world work
<code>educ</code>	уровень образования (в годах)
<code>exper</code>	опыт работы в годах
<code>tenure</code>	стаж работы на текущем месте
<code>age</code>	возраст (в годах)
<code>married</code>	семейный статус (бинарная, 1 если женат/замужем)
<code>south</code>	географический фактор (бинарная, 1 если живёт на юге)
<code>urban</code>	место жительства (1 если живет в городе)
<code>sibs</code>	число братьев/сестёр
<code>brthord</code>	какой по счёту ребёнок в семье
<code>meduc</code>	уровень образования матери (в годах)
<code>feduc</code>	уровень образования отца (в годах)

Таблица 5: Набор данных `wage1` из пакета `wooldridge` (526 наблюдений).  
Основные переменные.

<code>wage</code>	средняя почасовая оплата
<code>educ</code>	уровень образования (в годах)
<code>female</code>	гендерный фактор
<code>exper</code>	опыт работы
<code>tenure</code>	стаж на текущем месте работы
<code>married</code>	семейный статус
<code>smsa</code>	живёт ли в мегаполисе (бинарная)
<code>south</code>	географический фактор (бинарная)
<code>west</code>	географический фактор (бинарная)
<code>northcen</code>	географический фактор (бинарная)

Таблица 6: Набор данных `loanapp` из пакета `wooldridge` (1989 наблюдений). Основные переменные. Источник данных [9]

<code>approve</code>	бинарная, 1 если кредитная заявка одобрена
<code>appinc</code>	доход заявителя (в \$1000)
<code>mortno</code>	бинарная, 1 если нет ипотечной кредитной истории
<code>unem</code>	уровень безработицы в отрасли в %
<code>dep</code>	количество иждивенцев
<code>male</code>	гендерный фактор
<code>married</code>	семейный статус
<code>yjob</code>	стаж на текущей работе
<code>self</code>	бинарная, 1 если самозанятый

Таблица 7: Набор данных `SwissLabor` о рынке труда Швейцарии из пакета `AER` (872 наблюдений). Источник данных [7]

<code>participation</code>	Является ли участником рынка труда? (фактор, "yes"/"no")
<code>income</code>	логарифм дополнительного дохода (nonlabor income)
<code>age</code>	возраст (в десятилетиях)
<code>education</code>	уровень образования
<code>youngkids</code>	число маленьких детей (младше 7 лет)
<code>oldkids</code>	число старших детей (старше 7 лет)
<code>foreign</code>	является ли иностранцем? (фактор, "yes"/"no")

Таблица 8: Набор данных `Consumption` из пакета `Ecdat` об индивидуальных доходах и расходах в Канаде (квартальные данные с 1947Q1 по 1996Q4).

<code>yd</code>	индивидуальный располагаемый доход в ценах 1986
<code>se</code>	индивидуальные расходы на потребление в ценах 1986

Таблица 9: Набор данных из файла **Mroz** из пакета **Ecdat** содержит данные о рынке труда замужних женщин. Основные переменные. Источник данных [10]

LFP	бинарная, 1 женщина работает
WHRS	уровень занятости (в часах)
KL6	число детей моложе 6 лет в семье
K618	число детей от 6 до 18 лет в семье
WA	возраст
WE	уровень образования (в годах)
WW	средняя почасовая оплата
HHRS	занятость мужа
HA	возраст мужа
HE	уровень образования мужа (а годах)
HW	зарплата мужа
FAMINC	доход домашнего хозяйства
WMED	уровень образования матери
WFED	уровень образования отца
UN	уровень безработицы в стране проживания
CIT	бинарная, 1 если живет в мегаполисе
AX	предыдущий опыт работы (в годах)

Таблица 10: Набор данных **diamonds** из пакета **ggplot2** с данными о бриллиантах (53940 наблюдений). Основные переменные.

price	цена бриллианта
carat	вес бриллианта (в каратах)
cut	качество огранки (упорядоченный фактор с уровнями Fair<Good< Very Good<Premium<Ideal)
color	цвет (упорядоченный фактор с уровнями J<I<H<G<F<E<D)
clarity	прозрачность (упорядоченный фактор с уровнями I1<SI2<SI1<VS2<VS1<VVS2<VVS1<IF)
x	длина (в мм)
y	ширина (в мм)
z	глубина (в мм)

Таблица 11: Набор данных **Diamond** из пакета **Ecdat** с данными о бриллиантах (308 наблюдений). Основные переменные. Источник данных [4]

carat	вес бриллианта (в каратах)
colour	цвет (фактор с уровнями D,E,F,G,H,I)
clarity	прозрачность (фактор с уровнями IF,VVS1,VVS2,VS1,VS2)
certification	орган по сертификации (фактор с уровнями GIA,IGI,HRD)
price	цена в Сингапуре

Таблица 12: Набор данных **Icescream** из пакета **Ecdat** о потреблении мороженого в США (недельные данные с 1951-03-18 по 1953-07-11, всего 30 наблюдений). Источник данных [8]

cons	потребление мороженого (в пинтах)
income	средний недельный доход семьи (в \$)
price	цена мороженого (за пинту)
temp	средняя температура (по Фаренгейту)

Таблица 13: Набор данных **nlsw88** из пакета **Counterfactual** о занятости женщин в США (2246 наблюдений). Основные переменные. Источник данных: сайт Stata <http://www.stata-press.com/data/r10/g.html>

hours	недельная занятость (в часах)
married	семейный статус
ttl_exp	общий стаж работы
smsa	бинарная, 1 если живёт в мегаполисе
south	географический фактор (бинарная, 1 если живёт на юге)
wage	почасовая оплата (в \$)
age	возраст (в годах)
grade	уровень образования (в годах)

Таблица 14: Набор данных из файла `default.csv` с сайта <http://meit.mgimo.ru/node/237> о банкротствах по студенческим займам (6778 наблюдений)

<code>default</code>	бинарная переменная, равная 1 если индивид признал себя банкротом по студенческому займу
<code>age</code>	возраст
<code>adepcnt</code>	количество иждивенцев у индивида плюс 1
<code>acadmos</code>	количество месяцев, которые индивид прожил по текущему адресу
<code>majordrg</code>	количество зарегистрированных серьёзных правонарушений у этого индивида
<code>minordrg</code>	количество зарегистрированных мелких правонарушений у этого индивида
<code>ownrent</code>	1, если индивид живёт в собственном доме, и 0, если снимает
<code>income</code>	месячный доход в \$
<code>spending</code>	среднемесячный расход по кредитной карте
<code>inc_rep</code>	income, делённая на количество иждивенцев
<code>exp_inc</code>	доля месячных расходов по кредитной карте в годовой заработной плате
<code>selfempl</code>	1, если индивид самозанятый, и 0 иначе

Таблица 15: Набор данных `stockton3` из пакета `PoEdata` с данными о стоимости домов (2610 наблюдений). Основные переменные. Источник данных <https://github.com/ccolonescu/PoEdata>

<code>sprice</code>	цена продажи дома (в \$)
<code>livarea</code>	жилая площадь (кв.футы)
<code>pool</code>	наличие бассейна (бинарная)
<code>lgelot</code>	размер участка (бинарный фактор, 1 если участок больше 5 акров)
<code>age</code>	возраст (в годах)
<code>beds</code>	число спален

Таблица 16: Набор данных из файла `applications.csv` о поступивших на магистерские и PhD-программы (400 наблюдений)

admit	бинарный фактор, 1 если заявка одобрена
GPA	средняя оценка за время обучения
GRE	балл за экзамен graduate record exam
rank	категориальная переменная, обозначающая престиж университета (1 – высший престиж, 4 – низший престиж)

Таблица 17: Панель **Guns** из пакета **AER** с данными по 51 штату США с 1977 по 1999 гг. (всего 1173 наблюдения). Основные переменные. Источник данных [\[1\]](#)

state	factor indicating state
year	factor indicating year
violent	violent crime rate (incidents per 100,000 members of the population)
murder	murder rate (incidents per 100,000).
robbery	robbery rate (incidents per 100,000)
prisoners	incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents; value for the previous year)
afam	percent of state population that is African-American, ages 10 to 64
cauc	percent of state population that is Caucasian, ages 10 to 64
male	percent of state population that is male, ages 10 to 29
population	state population, in millions of people.
income	real per capita personal income in the state (US dollars)
density	population per square mile of land area, divided by 1,000
law	factor. Does the state have a shall carry law in effect in that year?



Таблица 18: Панель **LaborSupply** из пакета **plm**, **Ecdat** с данными по 532 индивидуумам с 1979 по 1988 гг. (всего 5320 наблюдений). Основные переменные. Источник данных [11]

lnhr	логарифм годовой занятости в часах
lnwg	логарифм почасовой оплаты
kids	число детей
age	возраст
disab	бинарная, 1 если плохое здоровье

Таблица 19: Панель **Cigar** из пакета **plm** с данными по 46 штатам США с 1963 по 1992 гг. (всего 1380 наблюдений). Основные переменные. Источник данных [2]

state	state abbreviation
year	the year
price	price per pack of cigarettes
pop	population
pop16	population above the age of 16
cpi	consumer price index (1983=100)
ndi	per capita disposable income
sales	cigarette sales in packs per capita
pimin	minimum price in adjoining states per pack of cigarettes

Таблица 20: Панель **Gasoline** из пакета **plm**, **Ecdat** с данными о потреблении бензина по 18 странам OECD с 1960 по 1978 гг. (всего 342 наблюдений). Основные переменные. Источник данных [3]

lgaspcar	логарифм потребления бензина
lincomer	логарифм реального дохода на душу населения
lrpmg	логарифм реальной цены на бензин
lcarpcar	логарифм объёма рынка машин

## Список литературы

- [1] Ayres, I., and Donohue, J.J. (2003). Shooting Down the ‘More Guns Less Crime’ Hypothesis. *Stanford Law Review*, 55, 1193–1312
- [2] Baltagi B, Levin D (1992). Cigarette taxation: Raising revenues and reducing consumption. *Structural Change and Economic Dynamics*, 3(2), 321-335
- [3] Baltagi, B.H. and Y.J. Griggin (1983) “Gasoline demand in the OECD: an application of pooling and testing procedures”, *European Economic Review*, 22.
- [4] Chu, Singfat (2001) “Pricing the C’s of Diamond Stones”, *Journal of Statistics Education*, 9(2).
- [5] J.E. Biddle and D.S. Hamermesh (1990), “Sleep and the Allocation of Time,” *Journal of Political Economy* 98, 922-943.
- [6] M. Blackburn and D. Neumark (1992), “Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials”, *Quarterly Journal of Economics* 107, 1421-1436.
- [7] Gerfin, M. (1996). Parametric and Semi-Parametric Estimation of the Binary Response Model of Labour Market Participation. *Journal of Applied Econometrics*, 11, 321–339.
- [8] Hildreth, C. and J. Lu (1960) Demand relations with autocorrelated disturbances, Technical Bul- letin No 2765, Michigan State University.
- [9] W.C. Hunter and M.B. Walker (1996), “The Cultural Affinity Hypothesis and Mortgage Lending Decisions”, *Journal of Real Estate Finance and Economics* 13, 57-70
- [10] Mroz, T. (1987) “The sensitivity of an empirical model of married women’s hours of work to economic and statistical assumptions”, *Econometrica*, 55, 765-799.
- [11] Ziliak, Jim (1997) “Efficient Estimation With Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators”, *Journal of Business and Economic Statistics*, 419-431