

# “МОЗГОВОЙ ШТУРМ” 2017-2018

ПРЕПОДАВАТЕЛИ: Кузнецова  
Мария,  
Москалева Александра

# ПЛАН ЗАНЯТИЯ:

- Виды экспериментов
- Виды данных
- Как исследователь собирает и обрабатывает данные
- Как сформулировать гипотезу, цели и задачи исследования?
- Вопрос о причине и следствии: ложная корреляция, одновременная связь. Примеры эконометрических исследований
- И обязательно подпишись на нашу группу в вк:  
[https://vk.com/mozgovoy\\_shturm2017](https://vk.com/mozgovoy_shturm2017)

# ВИДЫ

# ЭКСПЕРИМЕНТОВ:

Лабораторный	Полевой (естественный)
<ul style="list-style-type: none"><li>• В искусственных условиях</li><li>• Под контролем исследователя</li><li>• Испытуемый знает, что за ним наблюдают, о целях и задачах исследования</li></ul>	<ul style="list-style-type: none"><li>• Привычные условия</li><li>• Ход событий не контролируется исследователем</li><li>• Испытуемый не знает, что за ним наблюдают</li></ul>



# ТИПЫ ДАННЫХ, ИСПОЛЬЗУЕМЫХ В ЭКОНОМЕТРИКЕ:

## Пространственные выборки

- Набор независимых выборочных данных на данный момент

## Панельные данные

- Набор значений нескольких переменных за несколько лет

## Временные ряды

- Выборка, ключевую роль в которой играет значение показателя в тот или иной период

# ТИПЫ ДАННЫХ, ИСПОЛЬЗУЕМЫХ В ЭКОНОМЕТРИКЕ:

## Экспериментальные данные

- Получаются из контролируемых случайных экспериментов

## Наблюдаемые статистические

- Сбор реальных данных



# КАК И ГДЕ СОБРАТЬ ДАННЫЕ?

## Indicators

Featured indicators

All indicators

Quick search

### Agriculture & Rural Development

Agricultural irrigated land (% of total agricultural land)

Agricultural machinery, tractors per 100 sq. km of arable land

Arable land (% of land area)

Cereal yield (kg per hectare)

Employment in agriculture, female (% of female employment)

Fertilizer consumption (kilograms per hectare of arable land)

Forest area (% of land area)

Improved sanitation facilities, rural (% of rural population with access)

Land area (sq. km)

Livestock production index (2004-2006 = 100)

Rural population

Rural poverty gap at national poverty lines (%)

Surface area (sq. km)

Agricultural land (% of land area)

Agriculture, value added (% of GDP)

Arable land (hectares per person)

Crop production index (2004-2006 = 100)

Employment in agriculture, male (% of male employment)

Food production index (2004-2006 = 100)

Forest area (sq. km)

Improved water source, rural (% of rural population with access)

Land under cereal production (hectares)

Permanent cropland (% of land area)

Rural population (% of total population)

Rural poverty headcount ratio at national poverty lines (% of rural population)

Agriculture & Rural Development

Aid Effectiveness

Climate Change

Economy & Growth

Education

Energy & Mining

Environment

External Debt

Financial Sector

Gender

Health

Infrastructure

Poverty

Private Sector

Public Sector

<https://data.worldbank.org/indicator>

Help / Feedback

# КАК И ГДЕ СОБРАТЬ ДАННЫЕ?

→ ↻ www.gks.ru



[ВСЕРОССИЙСКАЯ СЕЛЬСКОХОЗЯЙСТВЕННАЯ ПЕРЕПИСЬ 2016 г.](#)



[СПЛОШНОЕ НАБЛЮДЕНИЕ ЗА ДЕЯТЕЛЬНОСТЬЮ МАЛОГО И СРЕДНЕГО БИЗНЕСА за 2015 год](#)



[ЗАРАБОТНАЯ ПЛАТА ОТДЕЛЬНЫХ КАТЕГОРИЙ РАБОТНИКОВ СОЦИАЛЬНОЙ СФЕРЫ И НАУКИ](#)



[ПРЕДОСТАВЛЕНИЕ ДАННЫХ БУХГАЛТЕРСКОЙ \(ФИНАНСОВОЙ\) ОТЧЕТНОСТИ](#)



[ПРИМЕНЕНИЕ ОКВЭД2 и ОКПД2 ДЛЯ РАЗРАБОТКИ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ](#)

## Официальная статистика

Национальные счета

Население

Рынок труда, занятость и заработная плата

Предпринимательство

Эффективность экономики России

Показатели, характеризующие импортозамещение в России

Технологическое развитие отраслей экономики

Наука, инновации и информационное общество

Государство, общественные организации

Цены

Финансы

Внешняя торговля

Окружающая среда

Международная статистика

Опережающие индикаторы по видам экономической деятельности

Региональная статистика

Публикации

Базы данных

ССРД МВФ

## Интерактивные статистические сервисы

<http://www.gks.ru/>

поиск по сайту

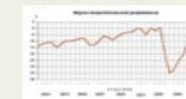
## Новостная лента

05.10.2017



[Об индексе потребительских цен в сентябре 2017 года](#)

05.10.2017



[Потребительские ожидания в России в III квартале 2017 г.](#)

05.10.2017



[Об объеме производства нефтепродуктов и потребительских ценах на них с 25 сентября по 1 октября 2017г.](#)

05.10.2017

[5 октября в рамках 19-ой Российской агропромышленной выставки «Золотая осень-2017» состоялся круглый стол «Предварительные итоги Всероссийской сельскохозяйственной переписи 2016г.»](#)

04.10.2017



[Об оценке индекса потребительских цен с 26 сентября по 2 октября 2017 года](#)

03.10.2017

[3 октября 2017 года в МИА «Россия сегодня» состоялась пресс-конференция заместителя руководителя Росстата К.Э. Лайкама: «Итоги Всероссийской сельскохозяйственной переписи 2016 года. изменения в сельском хозяйстве России за 10 лет»](#)

все новости



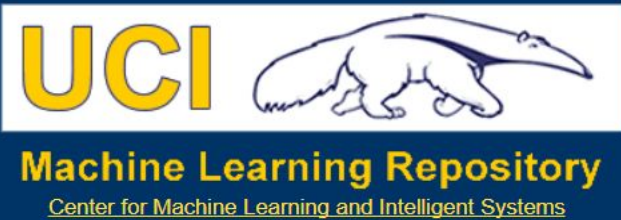
Часто задаваемые вопросы

Федеральная служба государственной статистики



# КАК И ГДЕ СОБРАТЬ ДАННЫЕ?

← → ↻ archive.ics.uci.edu/ml/index.php



[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

  
 Repository  Web 

<http://archive.ics.uci.edu/ml/index.php>

[View ALL Data Sets](#)

## Welcome to the UC Irvine Machine Learning Repository!

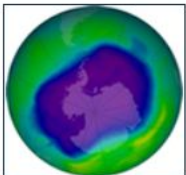
We currently maintain 394 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. Our [old web site](#) is still available, for those who prefer the old format. For a general overview of the Repository, please visit our [About page](#). For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to [contact the Repository librarians](#). We have also set up a [mirror site](#) for the Repository.



### Latest News:

- 04-04-2013:** Welcome to the new Repository admins Kevin Bache and Moshe Lichman!
- 03-01-2010:** [Note](#) from donor regarding Netflix data
- 10-16-2009:** Two new data sets have been added.
- 09-14-2009:** Several data sets have been added.
- 07-23-2008:** [Repository mirror](#) has been set up.
- 03-24-2008:** New data sets have been added!
- 06-25-2007:** Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

### Featured Data Set: [Ozone Level Detection](#)



**Task:** Classification  
**Data Type:** Multivariate, Sequential, Time-Series  
**# Attributes:** 73  
**# Instances:** 2536

Two ground ozone level data sets are included in this collection. One is the eight

### Newest Data Sets:

- 08-28-2017:** [Burst Header Packet \(BHP\) flooding attack on Optical Burst Switching \(OBS\) Network](#)
- 07-23-2017:** [Eco-hotel](#)
- 07-23-2017:** [Las Vegas Strip](#)
- 07-20-2017:** [Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet](#)
- 07-18-2017:** [PM2.5 Data of Five Chinese Cities](#)
- 07-16-2017:** [Sales\\_Transactions\\_Dataset\\_Weekly](#)
- 06-29-2017:** [Data for Software Engineering Teamwork Assessment in Education Setting](#)

### Most Popular Data Sets (hits since 2007):

- 1487462:** [Iris](#)
- 980877:** [Adult](#)
- 735464:** [Wine](#)
- 639868:** [Car Evaluation](#)
- 574824:** [Breast Cancer Wisconsin \(Diagnostic\)](#)
- 570018:** [Forest Fires](#)
- 537839:** [Human Activity Recognition Using Smartphones](#)



# КАК ОБРАБОТАТЬ ДАННЫЕ?

The screenshot displays the Microsoft Excel 2016 interface. The title bar shows the file name "Sales\_Transactions\_Dataset\_Weekly - Excel" and the user name "Маруся Кузнецова". The ribbon includes "Файл", "Главная", "Вставка", "Разметка страницы", "Формулы", "Данные", "Рецензирование", "Вид", and "Что вы хотите сделать?". The ribbon is set to the "Главная" (Home) tab, showing options for text, alignment, numbers, styles, cells, and editing. The formula bar contains the text "Product\_Code,W0,W1,W2,W3,W4,W5,W6,W7,W8,W9,W10,W11,W12,W13,W14,W15,W16,W17,W18,W19,W20,W21,W22,W23,W24,W25,W26,W27,W28,W29,W30,W31,W32,W33,W34,W35,W36,W37,W38,W39,W40,W41,W42,W43,W44,W45,W46,W47,W48,W49,W50". The main grid shows a table with columns A through W and rows 1 through 29. The data in the cells consists of long strings of numbers separated by commas, representing a dataset of sales transactions.

Вывод: Изображение представляет собой скриншот интерфейса программы Microsoft Excel 2016. В центре внимания находится панель «Главная» (Home), на которой расположены различные инструменты для форматирования и редактирования данных. Видно меню лент «Файл», «Главная», «Вставка», «Разметка страницы», «Формулы», «Данные», «Рецензирование», «Вид» и «Что вы хотите сделать?». Основное рабочее пространство занято таблицей данных с колонками, обозначенными буквами от А до W, и строками с номерами от 1 до 29. В ячейках таблицы содержится большое количество длинных строк чисел, разделенных запятыми, что указывает на то, что это массив данных. В строке формул вверху таблицы введена формула: Product\_Code,W0,W1,W2,W3,W4,W5,W6,W7,W8,W9,W10,W11,W12,W13,W14,W15,W16,W17,W18,W19,W20,W21,W22,W23,W24,W25,W26,W27,W28,W29,W30,W31,W32,W33,W34,W35,W36,W37,W38,W39,W40,W41,W42,W43,W44,W45,W46,W47,W48,W49,W50. В панели задач в нижнем левом углу отображены значки различных приложений, а в нижнем правом углу — часы и дата: 22:24 09.10.2017.



# КАК ОБРАБОТАТЬ ДАННЫЕ?

Autosave Sales\_Transactions\_Dataset\_Weekly - режим совместности - Excel Маруся Кузнецова

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Что вы хотите сделать? Поделиться

Вставить Шрифт Выравнивание Число Стили Ячейки Редактирование

Calibri 11 Перенести текст Обобщить

Условное форматирование Форматировать как таблицу Стили Вставить Удалить Формат

Автосумма Заполнить Очистить Сортировка и фильтр Найти и выделить

Product\_Code

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
	Product_C	W0	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18	W19	W20	W21	W22	W23	W24	W25
2	P1	11	12	10	8	13	12	14	21	6	14	11	14	16	9	9	9	14	9	3	12	5	11	7	12	5	
3	P2	7	6	3	2	7	1	6	3	3	3	2	2	6	2	0	6	2	7	7	9	4	7	2	4	5	
4	P3	7	11	8	9	10	8	7	13	12	6	14	9	4	7	12	8	7	11	10	7	7	13	11	8	10	
5	P4	12	8	13	5	9	6	9	13	13	11	8	4	5	4	15	7	11	9	15	4	6	7	11	7	9	
6	P5	8	5	13	11	6	7	9	14	9	9	11	18	8	4	13	8	10	15	6	13	11	6	10	9	8	
7	P6	3	3	2	7	6	3	8	6	6	3	1	1	5	4	3	5	3	5	10	8	4	9	7	5	4	
8	P7	4	8	3	7	8	7	2	3	10	3	5	2	3	4	5	3	7	10	0	3	7	5	1	5	7	
9	P8	8	6	10	9	6	8	7	5	10	10	8	8	15	9	5	11	10	7	13	9	12	11	5	11	11	
10	P9	14	9	10	7	11	15	12	7	13	12	15	15	16	10	9	9	13	8	10	18	18	17	10	16	14	
11	P10	22	19	19	29	20	16	26	20	24	20	31	22	23	19	15	19	22	23	20	33	16	23	23	16	25	
12	P11	15	7	15	14	17	7	10	16	11	8	8	10	10	12	10	16	13	10	13	13	18	12	15	9	8	
13	P12	3	4	1	6	4	3	7	3	5	3	5	6	5	0	4	0	7	1	5	5	3	7	4	4	5	
14	P13	12	10	9	6	10	11	18	8	10	17	11	12	11	13	10	8	9	10	9	8	7	5	12	10	10	
15	P14	14	12	9	11	13	12	8	12	13	10	10	17	14	14	25	18	13	22	12	21	15	17	11	15	12	
16	P15	19	45	47	42	29	44	43	36	25	52	39	42	43	42	43	51	40	44	30	39	39	36	32	39	30	
17	P16	30	27	27	43	29	32	49	41	49	38	42	30	43	43	54	48	34	36	44	43	44	28	35	29	43	
18	P17	49	40	40	28	40	47	44	45	39	33	39	37	33	52	29	45	34	43	40	50	29	36	32	28	35	
19	P18	40	38	39	38	39	33	28	44	36	36	23	38	38	41	43	27	38	31	43	32	35	38	48	38	30	
20	P19	26	31	45	36	31	28	28	34	42	40	43	35	30	33	40	45	48	35	30	29	39	41	30	34	28	
21	P20	13	17	11	10	7	11	17	8	12	10	8	9	8	10	10	13	11	10	7	8	8	8	9	7	10	
22	P21	12	5	9	8	9	6	10	13	6	12	13	10	15	12	8	12	12	9	12	8	9	13	10	7	9	
23	P22	8	14	8	9	17	6	17	15	11	13	9	18	5	8	12	6	17	9	10	6	9	17	10	9	9	
24	P23	3	5	4	3	3	2	5	4	5	5	7	7	5	4	6	6	6	6	8	4	4	3	3	4	9	
25	P24	36	42	27	33	40	48	38	39	41	39	44	35	53	52	43	45	41	42	43	26	35	29	48	24	33	
26	P25	26	28	33	32	20	33	42	29	24	32	45	41	35	39	32	36	31	32	48	26	34	30	36	31	34	
27	P26	14	14	9	8	9	7	9	11	10	14	16	7	15	8	18	8	11	12	9	15	16	20	7	10	12	
28	P27	44	34	33	39	34	30	47	27	45	39	47	39	35	47	29	45	52	38	51	38	41	21	33	33	25	
29	P28	34	32	36	41	31	31	32	29	43	33	37	44	37	27	36	38	24	36	30	32	29	40	41	36	33	
30	P29	13	10	12	17	17	11	15	16	10	12	18	6	14	12	11	10	19	25	9	17	12	15	16	11	14	
31	P30	46	36	45	34	35	36	43	28	26	33	46	42	41	42	37	38	40	35	29	31	43	29	39	35	35	
32	P31	7	17	6	7	9	7	11	10	13	9	4	15	10	7	10	9	11	8	14	9	10	3	3	1	8	
33	P32	15	13	10	4	13	7	10	12	11	15	5	15	8	10	7	16	11	10	7	11	6	8	5	3	12	
34	P33	15	13	11	17	10	18	11	16	8	6	11	10	18	10	14	12	12	12	14	10	10	11	14	11	10	

Worksheet

# План исследования

- Определить проблемный вопрос
- Сформулировать цели и задачи исследования
- Сформулировать гипотезу и определить какие данные нужны для её подтверждения/опровержения
- Собрать необходимые данные
- Построить модель
- Сделать выводы с опорой на цели, задачи и гипотезу



# КАК СФОРМУЛИРОВАТЬ ЦЕЛЬ?

- **Цель исследования** — конечный результат, которого хотел бы достичь исследователь при завершении своей работы.

## **Можно поставить целью:**

- выявить...
- установить...
- обосновать...
- уточнить...
- разработать...



# КАК СФОРМУЛИРОВАТЬ ЗАДАЧИ?

- **Задачи исследования** — последовательные действия, которые обеспечивают достижение данной цели.
- **Методы исследования** – способы сбора, обработки и анализа данных.

На нашем курсе – количественные 😊



# КАК ВЫДВИГАТЬ ГИПОТЕЗУ?

- **Гипотеза** — научное предположение, допущение, истинное значение которого неопределенно.

**Гипотеза должна:**

- быть проверяемой
- содержать предположение
- быть логически непротиворечивой





# Вопрос о причине и следствии: ложная корреляция, одновременная связь

Не всегда очевидно как взаимосвязаны два явления:

- **Ложная взаимосвязь.** Связь событий не всегда равно причинно-следственной связи.

<http://tylervigen.com/spurious-correlations>

- **Одновременная связь.** Два события являются и причиной и следствием друг для друга: пример – преступность и продажа оружия; доход и потребление.

# ОДНОВРЕМЕННАЯ ПРИЧИННО-СЛЕДСТВЕННАЯ СВЯЗЬ

- С одной стороны, разрешение на хранение оружия в доме привело к росту преступности
- С другой стороны, рост преступности привел к необходимости разрешения хранения оружия



# ОДНОВРЕМЕННАЯ ПРИЧИННО-СЛЕДСТВЕННАЯ СВЯЗЬ

С одной стороны:

$$\bullet Y = C + I + G$$

Где C- потребление, I- инвестиции, а G- гос.расходы

С другой стороны:

$$\bullet C = F(T; Y)$$



# Домашнее задание – 10 баллов

- Придумать или найти проблемную ситуацию
- Сформулировать к ней гипотезу, цели и задачи
- Понять какие данные вам кажется, что нужно найти для проверки гипотезы
- Описать как вы будете их собирать и обрабатывать