

ЭКОНОМЕТРИКА

Семинар 1

ОСНОВНЫЕ МОМЕНТЫ

? Преподаватели:

- Лекции: Зандер Евгения Викторовна
- Семинары: Сырцова Екатерина Александровна

? В конце семестра:

- У групп менеджмента – зачет;
- У групп экономической безопасности – экзамен;

? Система оценки:

- За семестр – **автоматом** зачет/любая оценка
- Экзамен/зачет «с нуля»



ОЦЕНКА В СЕМЕСТРЕ ДЛЯ ГРУПП МЕНЕДЖМЕНТА:

- 50 % - коллоквиумы (10 в течение семестра: 6 (7) теоретических, 4 (3) – практических);
- 20 % - письменная работа по теме: «Модели с бинарной зависимой переменной»;
- 30 % - Итоговый тест (контрольная работа).

- Зачет с 51 %



ОЦЕНКА В СЕМЕСТРЕ ДЛЯ ГРУПП ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ:

- 60 % - коллоквиумы (12 в течение семестра: 7 теоретических, 5 – практических);
 - 20 % - контрольная работа по теме: «Предварительная обработка и регрессионный анализ пространственных данных»;
 - 20 % - контрольная работа по теме: «Модели с бинарной зависимой переменной. Временные ряды».
- ? «3» – с 51 %
- ? «4» – с 67 %
- ? «5» – с 84 %



ОБЩЕНИЕ

- Электронные курсы:
 - ? Для групп экономической безопасности – «Эконометрика для ЭБ»
 - ? Для групп менеджмента – ОММСЭП (эконометрика)
- Электронная почта: ekkoryakova@yandex.ru
- Вконтакте – в случае крайней необходимости (не рассчитывать на ответ после 00.00 и в выходные)



ЛИТЕРАТУРА

- ? Зандер Е. В., Ибрагимов Н. М. Эконометрика: учебное пособие. / Сибирский федеральный университет. Красноярск, 2007.
- ? Елисеева И. И., Юзбашев М. М. Общая теория статистики: учебник / Под ред. И.И. Елисеевой. — 5-е изд., перераб. и доп. — М.: Финансы и статистика, 2004.
- ? Доугерти К. Введение в эконометрику: учебник. 3-е изд./Пер. с англ. М.: ИНФРА-М, 2009. — 465 с.



ЭКОНОМЕТРИКА

- ? Эконометрика — это наука об измерении количественных и качественных экономических взаимосвязей с помощью математических и статистических методов.
- ? В современной эконометрике существуют два основных направления: теоретическое и прикладное.
- ? Целью теоретической эконометрики является развитие методов оценки количественных и качественных экономических взаимосвязей
- ? Прикладная эконометрика занимается применением статистического инструментария для анализа экономических проблем различного уровня.



ПОВТОРЕНИЕ ОСНОВНЫХ ПОНЯТИЙ ТЕОРИИ ВЕРОЯТНОСТЕЙ И СТАТИСТИКИ

- ? Случайная переменная
- ? Математическое ожидание
- ? Дисперсия



СЛУЧАЙНАЯ ПЕРЕМЕННАЯ

- ? Случайная переменная – любая переменная, значение которой не может быть точно предсказано.
 - Дискретная (число очков на кубике)
 - Непрерывная (температура в комнате)

- ? Составляющие:
 - Постоянная
 - Случайная



МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

- Дискретной величины:

$$E(X) = x_1 p_1 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i$$

- Непрерывной величины:

$$E(X) = \int X f(X) dX$$

- Обозначение: $E(X)$ или μ_x
- Математическое ожидание – теоретическое среднее



ДИСПЕРСИЯ

- Дискретной величины:

$$D_X^2 = \sum_{i=1}^n (X - \mu_X)^2 p_i$$

- Непрерывной величины:

$$D_X^2 = \int (X - \mu_X)^2 f(X) dX$$

- Дисперсия – мера разброса
- Обозначение D^2 или σ_X^2



ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА

- ? Генеральная совокупность – все существующие объекты
- ? Выборка – часть генеральной совокупности, по которой проводится исследование



ТЕОРЕТИЧЕСКИЕ ВЕЛИЧИНЫ И ИХ ОЦЕНКИ

- ? Генеральная совокупность
- ? Теоретические величины
- ? Мат. Ожидание
- ? Дисперсия
- ? Выборка
- ? Оценки
- ? Выборочное среднее
- ? Выборочная дисперсия и т.д.

Оценки должны обладать следующими свойствами:

- Несмещенность
- Состоятельность
- Эффективность



ЭТАПЫ ЭКОНОМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ

- ? Определение проблемы/темы
- ? Сбор данных – составление выборки
- ? Предварительная обработка данных:
 - Расчет и анализ описательных статистик
 - Проверка распределения на однородность
 - Проверка распределения на нормальность
 - Корректировка: приведение распределение к однородному и нормальному



ВЫБОРКИ

? Принципы построения выборки:

- Независимость наблюдений
- Случайность
- Достаточный объем

- **Репрезентативность** — соответствие характеристик выборки характеристикам генеральной совокупности в целом. Репрезентативность определяет, насколько возможно обобщать результаты исследования на всю генеральную совокупность, из которой она была собрана.



ОШИБКИ ВЫБОРКИ



ТИПЫ ДАННЫХ

- ? пространственные данные (англ. cross-sectional data) — наборы показателей экономических переменных, полученных в определенный момент времени;
- ? временные ряды (англ. time series) — выборки наблюдений, в которых важны не только сами наблюдаемые значения случайных величин, но и порядок их расположения друг за другом;
- ? пространственно-временные данные (англ. panel data) - прослеженные во времени пространственные выборки.

<https://www.fedstat.ru/indicator/data.do?id=33379>



ВЫБОРКИ

- ? Тип данных
- ? Единица наблюдения
- ? Объем выборки
- ? Показатели (для каждого наименования и единицы измерения)
- ? Период
- ? Источник данных



ВРЕМЯ, ПОТРАЧЕННОЕ НА ЧТЕНИЕ (В НЕДЕЛЮ), ЧАСОВ

Readers round the world

WHERE DO PEOPLE READ MORE?

Hours per week per person, Source: World Culture Score Index



САМЫЕ ЧИТАЮЩИЕ СТРАНЫ МИРА

| | | | |
|-------------------|-------|----------------|------|
| Индия | 10,42 | Австралия | 6,18 |
| Тайланд | 9,24 | Индонезия | 6 |
| Китай | 8 | Аргентина | 5,54 |
| Филиппины | 7,36 | Турция | 5,54 |
| Египет | 7,3 | Испания | 5,48 |
| Чехия | 7,24 | Канада | 5,48 |
| Россия | 7,06 | Германия | 5,42 |
| Швеция | 6,54 | США | 5,42 |
| Франция | 6,54 | Италия | 5,36 |
| Венгрия | 6,48 | Мексика | 5,3 |
| Саудовская Аравия | 6,48 | Великобритания | 5,18 |
| Гонконг | 6,42 | Бразилия | 5,12 |
| Польша | 6,3 | Тайвань | 5 |
| Венесуэла | 6,24 | Япония | 4,06 |
| Южная Африка | 6,18 | Корея | 3,06 |

Показатель – время, потраченное на чтение в неделю, часы (данные за 2014 год)

Пространственные данные

Единица наблюдения – страна

Объем выборки – 30 наблюдений

Источник – World Culture Score Index



СОПОСТАВИМОСТЬ ДАННЫХ

? Список самых кассовых фильмов:

- Аватар \$2,8 млрд
- Титаник \$2,2 млрд
- Звёздные войны: Пробуждение силы \$2,1 млрд
- Мир юрского периода \$1,7 млрд
- Мстители \$1,5 млрд
- Форсаж 7 \$1,5 млрд
- Мстители: Эра Альтрона \$1,4 млрд
- Гарри Поттер и дары Смерти. Часть 2 \$1,3 млрд
- Холодное сердце \$1,3 млрд
- Железный человек 3 \$1,2 млрд

? Список самых кассовых фильмов с учетом инфляции (в ценах 2014 г.):

- Унесённые ветром \$3,4 млрд
- Аватар \$3,0 млрд
- Звёздные войны: Пробуждение силы \$2,8 млрд
- Титаник \$2,5 млрд
- Звуки музыки \$2,4 млрд
- Инопланетянин \$2,3 млрд
- Десять заповедей \$2,2 млрд
- Доктор Живаго \$2,1 млрд
- Челюсти \$2,0 млрд
- Белоснежка и семь гномов \$1,8 млрд



ЧТЕНИЕ КНИГ В РОССИИ



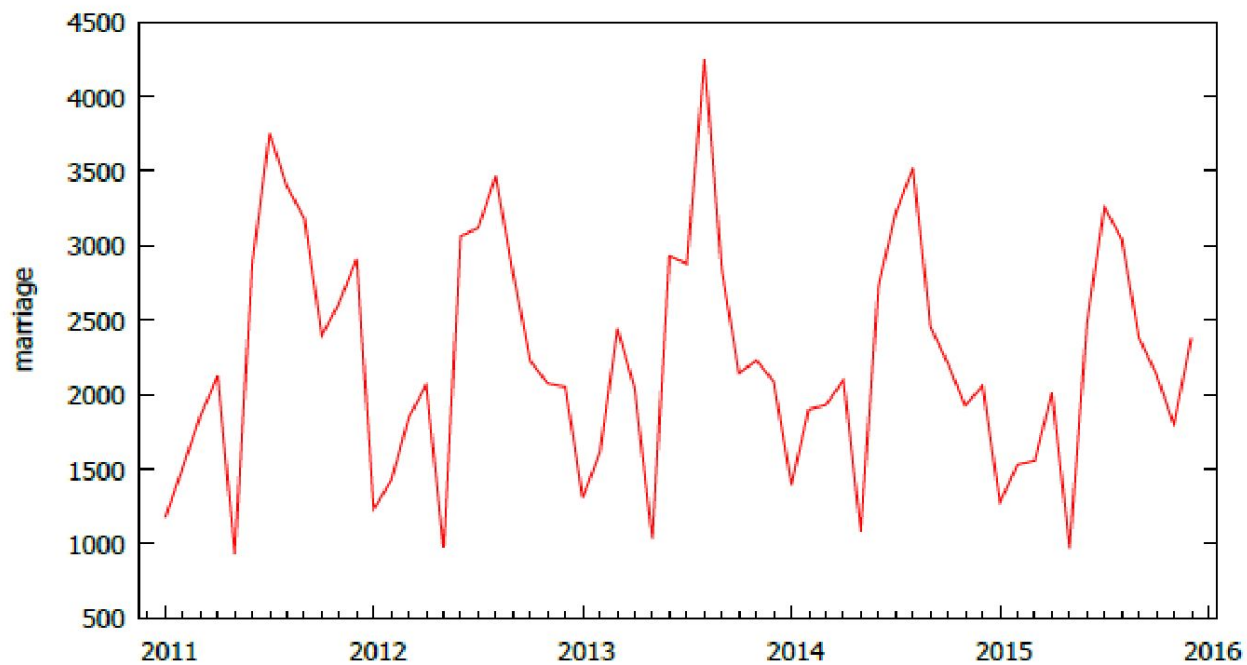
ЧТЕНИЕ КНИГ В РОССИИ

| Год | Доля людей, не читающих книги, % |
|-----------|----------------------------------|
| 1996 | 20 |
| 1998 | 24 |
| 2002 | 28 |
| 2005 | 25 |
| 2008 | 20 |
| 2009 | 35 |
| 2010 | 48 |
| 2012-2013 | 37 |

- ? Доля населения, не читающего книги, %
- ? Временная выборка
- ? Единица наблюдения - год
- ? Период наблюдения 1996-2013
- ? Объем выборки – 8 наблюдений (т.к. данные не за каждый год)
- ? Источник: опросы Левада-центра и ВЦИОМ



КОЛИЧЕСТВО ЗАРЕГИСТРИРОВАННЫХ БРАКОВ В КРАСНОЯРСКОМ КРАЕ



ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ

- ? Среднее
- ? Медиана, квантиль, дециль
- ? Мода
- ? Дисперсия
- ? Стандартное отклонение
- ? Вариация



СРЕДНЕЕ ЗНАЧЕНИЕ

? Пример № 1 (Про принтеры*)

- Руководитель службы качества получил задание: сравнить качество принтеров своей компании и компании конкурентов на основании данных о:
 - ? Количестве проданных принтеров;
 - ? Количестве проблем с качеством в течение гарантийного периода (с одним и тем же принтером могут обращаться несколько раз)
- Руководитель службы качества рассчитал описательную статистику – среднее количество проблем с качеством на каждый проданный принтер:
 - ? Для своей фирмы – 9,1
 - ? Для фирмы-конкурента – 2,8

? Вывод?

*Пример взят из Уилан Ч. Голая статистика. Самая интересная книга о самой скучной науке / пер. с англ. И Веригина. – М.: Манн, Иванов и Фербер, 2016. – 352 с.



СРЕДНЕЕ ЗНАЧЕНИЕ

- ? В баре 10 человек, каждый в год получает 35 000 \$
- ? В баре 10 человек, каждый в год получает 35 000 \$ и Билл Гейтс
- ? Средний годовой доход людей, находящихся в баре – 35 000 \$
- ? Средний годовой доход людей, находящихся в баре – 91 000 000 \$



МЕДИАНА

- ? В баре 10 человек, каждый в год получает 35 000 \$
- ? Средний годовой доход людей, находящихся в баре – 35 000 \$
- ? Медиана – 35 000 \$
- ? В баре 10 человек, каждый в год получает 35 000 \$ и Билл Гейтс
- ? Средний годовой доход людей, находящихся в баре – 91 000 000 \$
- ? Медиана – 35 000 \$



СРЕДНЕЕ ЗНАЧЕНИЕ, МЕДИАНА, МОДА

- ? Медиана – середина вариационного (ранжированного) ряда;
- ? Мода – значение признака, встречающееся в выборке чаще всего;

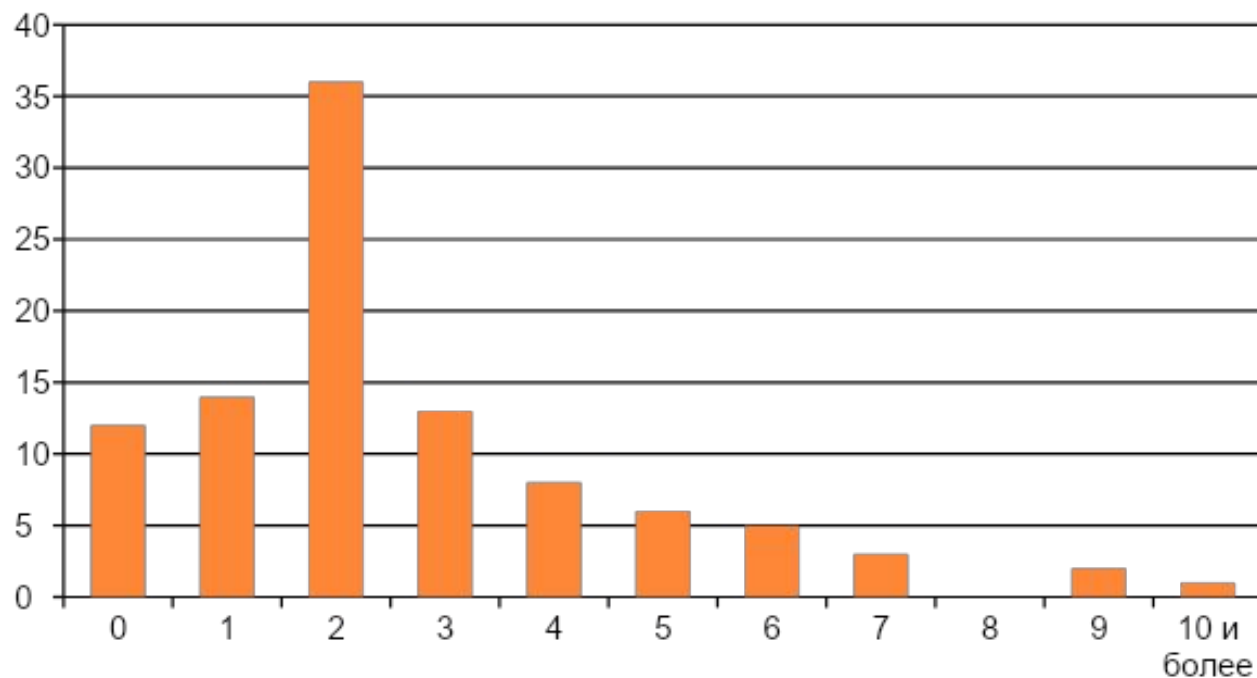


ПРИМЕР ПРО ПРИНТЕРЫ

- Среднее количество проблем с качеством на каждый проданный принтер:
 - ? Для своей фирмы – 9,1
 - ? Для фирмы-конкурента – 2,8
- Медиана для количества проблем с качеством на каждый проданный принтер:
 - ? Для своей фирмы – 1
 - ? Для фирмы-конкурента – 2
- Вывод?

ГИСТОГРАММА РАСПРЕДЕЛЕНИЯ

- ? По оси X – число проблем с качеством на один принтер
- ? По оси Y – частота, % (т.е. у какого количества принтеров наблюдалось такое количество проблем с качеством)

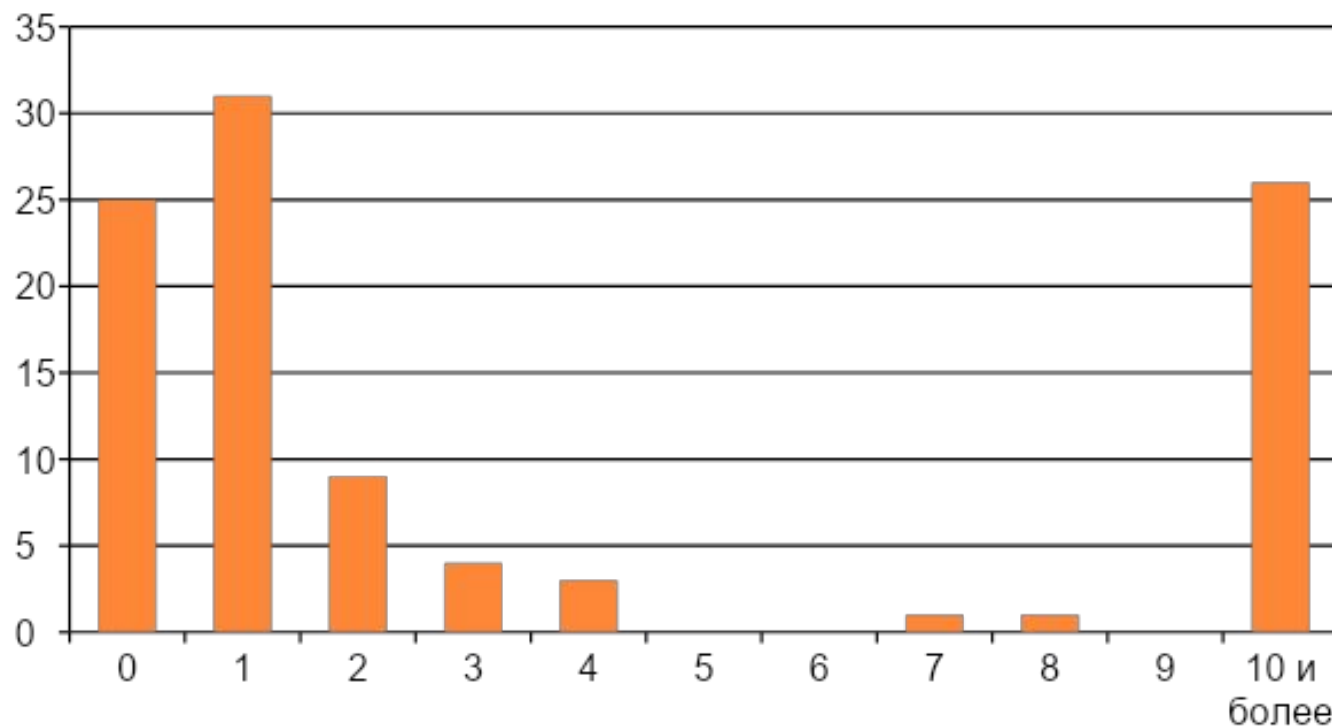


*Пример взят из Уилан Ч. Голая статистика. Самая интересная книга о самой скучной науке / пер. с англ. И Веригина. – М.: Манн, Иванов и Фербер, 2016. – 352 с.



ГИСТОГРАММА РАСПРЕДЕЛЕНИЯ

- ? По оси X – число проблем с качеством на один принтер
- ? По оси Y – частота, % (т.е. у какого количества принтеров наблюдалось такое количество проблем с качеством)



*Пример взят из Уилан Ч. Голая статистика. Самая интересная книга о самой скучной науке / пер. с англ. И Веригина. – М.: Манн, Иванов и Фербер, 2016. – 352 с.

КВАРТИЛИ, ДЕЦИЛИ, ПЕРЦЕНТИЛИ

- ? α -квантиль — это некоторое значение, которое не превышает случайная величина с заданной вероятностью
- ? 0,25-квантиль называется первым (или нижним) квартилем;
- ? 0,5-квантиль называется медианой или вторым квартилем;
- ? 0,75-квантиль называется третьим (или верхним) квартилем.



ПРИМЕР

- ? Рассчитаны описательные статистики для величины «сумма баллов за семестр» для групп ГМУ по эконометрике за весенний семестр, 2016 г.
- ? Среднее 60,7
- ? Медиана 73,4
- ? *Стандартное отклонение 33,6*
- ? *Вариация 0,55*
- ? 10-я перцентиль (1-й дециль) 0
- ? 20-я перцентиль (2-й дециль) 18,2
- ? 30-я перцентиль (3-й дециль) 55,1
- ? 50-я перцентиль (5-й дециль, медиана) 73,4
- ? 90-я перцентиль (9-й дециль) 92,4



СТАНДАРТНОЕ ОТКЛОНЕНИЕ

? 250 пассажиров
самолета

? 250 участников
марафона

? Средний вес – 70,3 кг

? Средний вес – 70,3 кг

? Стандартное
отклонение (средний
разброс) – 12 кг

? Стандартное
отклонение (средний
разброс) – 2 кг



СТАНДАРТНОЕ ОТКЛОНЕНИЕ

? Пример

- «У вас берут на анализ кровь, по результату которого некий показатель = 134.
- Среднее значение этого показателя у человека вашего возраста = 122
- Врач успокаивает: стандартное отклонение для этого показателя равно 18»

? Таким образом, для большинства людей значение этого показателя лежит в диапазоне от 104 ($122 - 18$) до 140 ($122 + 18$), т.е. результат анализа находится в пределах нормы



ВАРИАЦИЯ

- Вариация – различие значений по какому-либо признаку за один и тот же промежуток времени
- Измеряется с помощью коэффициента вариации:

$$V = \frac{S}{\bar{X}}$$

- Чем ближе коэффициент вариации к нулю, тем более однородны значения признака
- Если коэффициент вариации $> 0,33$ выборка считается неоднородной (граница условная)



Что делать в случае, если выборка неоднородна?

- ? Использование относительных показателей вместо абсолютных
- ? Группировка наблюдений
- ? Отсев аномальных наблюдений (выбивающиеся максимальные и минимальные значения, обязательно объяснить, почему это наблюдение оказалось аномальным)

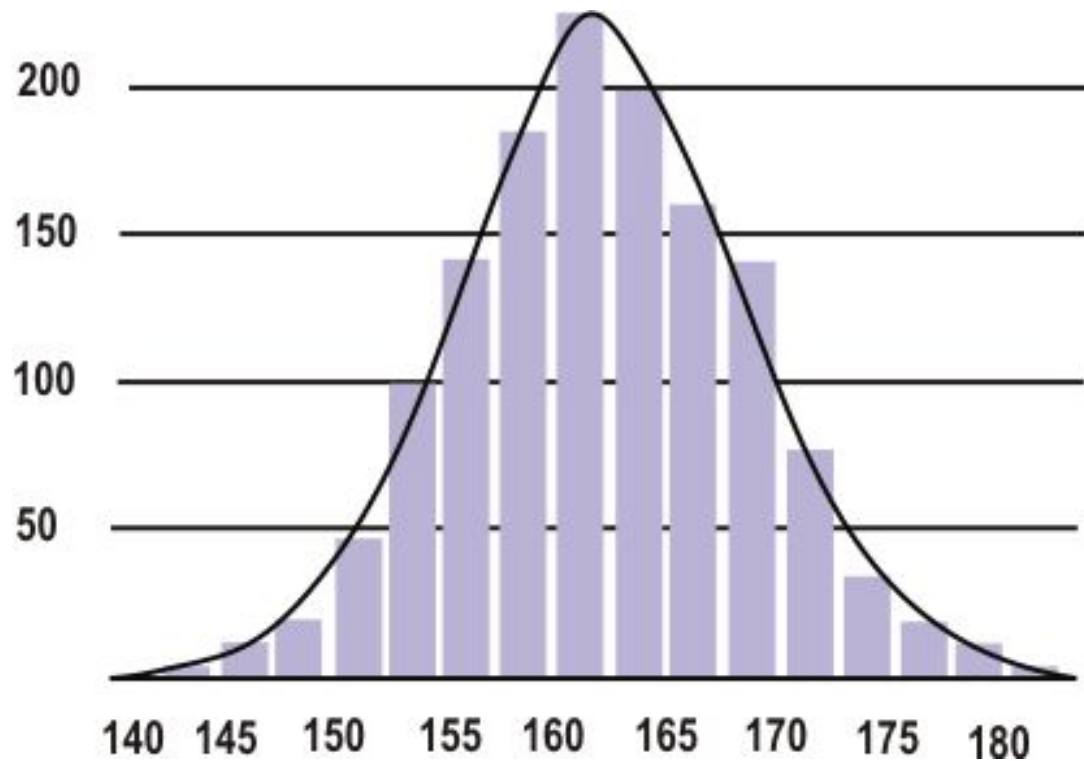


ВОПРОС

- ? По словам остроумного Мишеля Матвеева, в этом семействе был один великий ученый, была парочка редкостных кретинов, но в основном это были нормальные бюргеры.
- ? Назовите упомянутого ученого.



НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ



Распределение роста



НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

- ? Среднее, мода и медиана совпадают
- ? Симметрично относительно среднего
- ? Очень большие и очень маленькие значения маловероятны



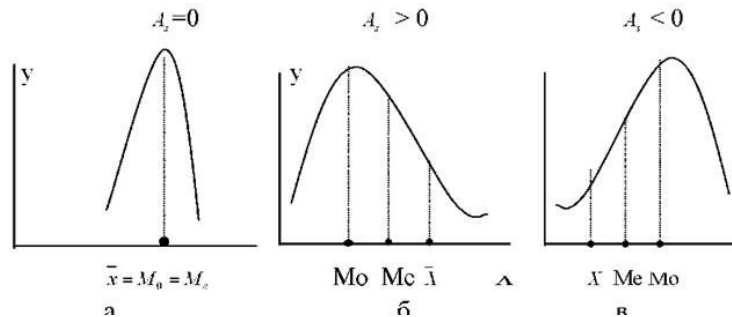
НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

- ? 68,2 % наблюдений находится в пределах одного стандартного отклонения от среднего;
- ? 95,4 % наблюдений – двух стандартных отклонений;
- ? 99,7 % наблюдений – трех стандартных отклонений

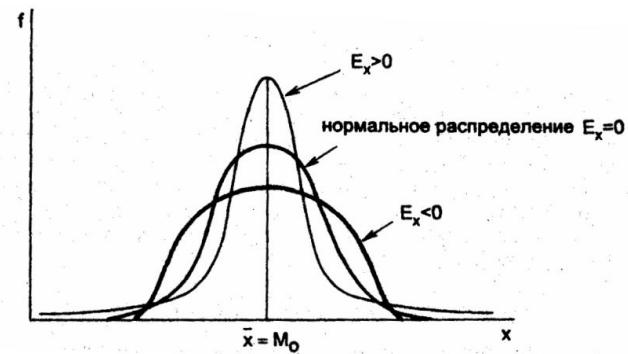


АСИММЕТРИЯ И ЭКСЦЕСС

- ? Для нормального распределения коэффициенты асимметрии и эксцесса равны нулю.
- ? Коэффициент асимметрии:



- ? Коэффициент эксцесса:



ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ. КОЭФФИЦИЕНТЫ АСИММЕТРИИ И ЭКСЦЕССА

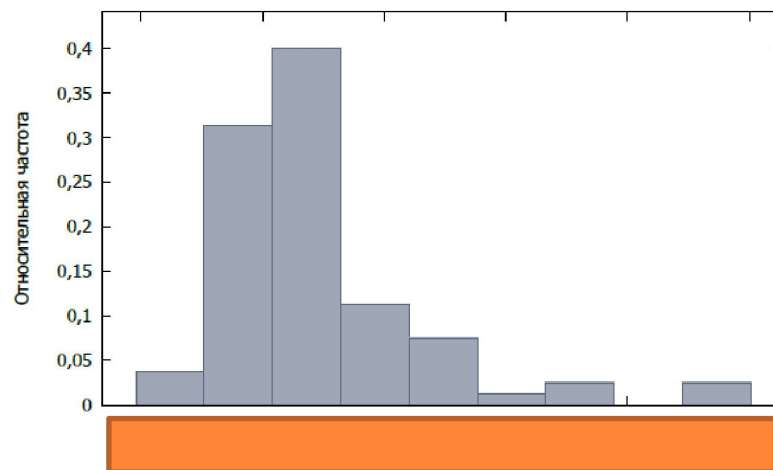
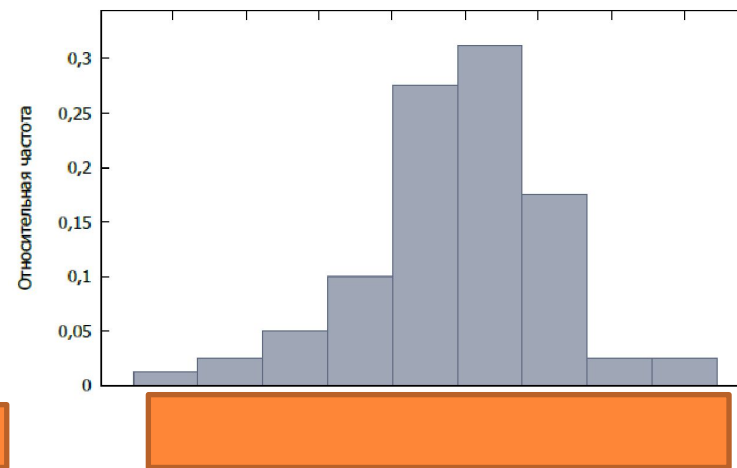
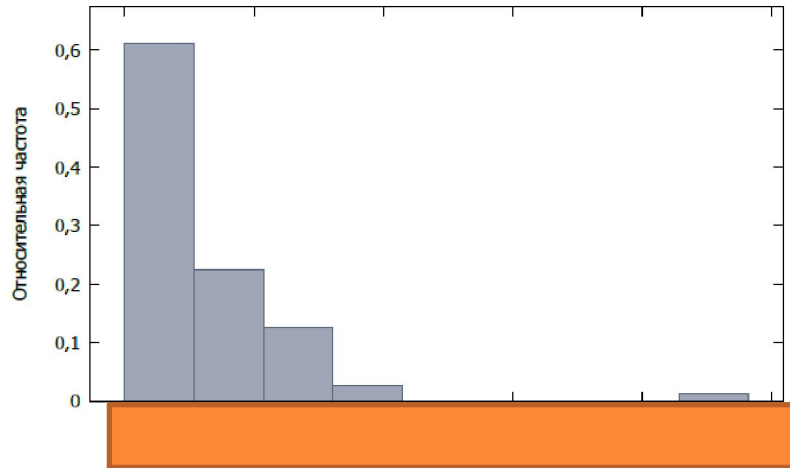
- ? Среднедушевые доходы, руб.:
- ? Асимметрия 1,92; эксцесс 4,51.

- ? Количество преступлений экономической направленности, шт.:
- ? Асимметрия 3,44; эксцесс 17,48.

- ? Доля городского населения, %:
- ? Асимметрия -0,56; эксцесс 0,95.



ГИСТОГРАММА РАСПРЕДЕЛЕНИЯ



ПРОВЕРКА РАСПРЕДЕЛЕНИЯ НА НОРМАЛЬНОСТЬ

- ? Соотношение среднего, моды и медианы
- ? Коэффициенты асимметрии и эксцесса
- ? Гистограмма распределения
- ? Формальные критерии, например, [хи-квадрат](#)



ПРИВЕДЕНИЕ РАСПРЕДЕЛЕНИЯ К НОРМАЛЬНОМУ ЗАКОНУ

Если распределение имеет крутую левую ветвь гистограммы и пологую правую, выполняют преобразование выборки по формулам:

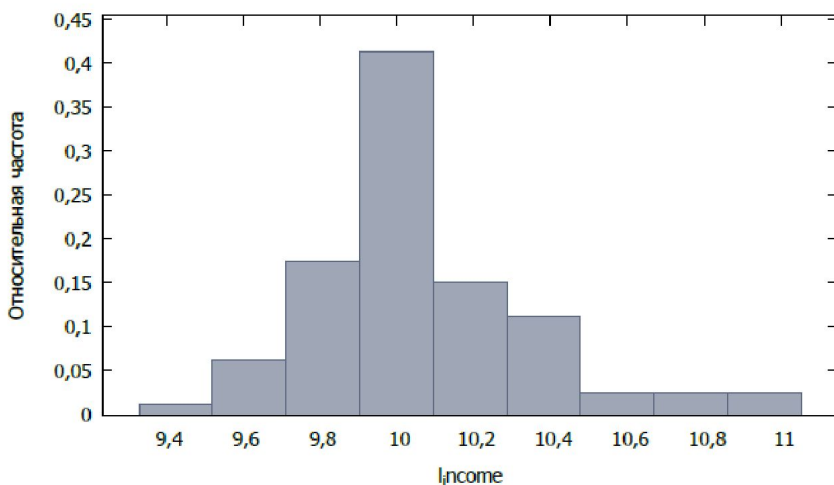
1. $x' = \lg(x \pm a) \cdot 10^b$,
2. $x' = 1/x$,
3. $x' = 1/\sqrt{x}$.

После каждого преобразования вычисляют критерий χ^2 и принимают то преобразование, которое дало наименьший χ^2 . Для распределений, смещенных вправо, исходные данные преобразуют по формуле $x' = x^a$ (при $a = 1,5; 2$) и после каждого преобразования вычисляют χ^2 , выбирают a , при котором χ^2 минимален.



ПРИВЕДЕНИЕ РАСПРЕДЕЛЕНИЯ К НОРМАЛЬНОМУ ЗАКОНУ

- ? При правосторонней асимметрии наиболее распространенным методом приведения к нормальному закону является логарифмирование:
- ? После логарифмирования среднедушевых доходов получаем следующую гистограмму и описательные статистики:

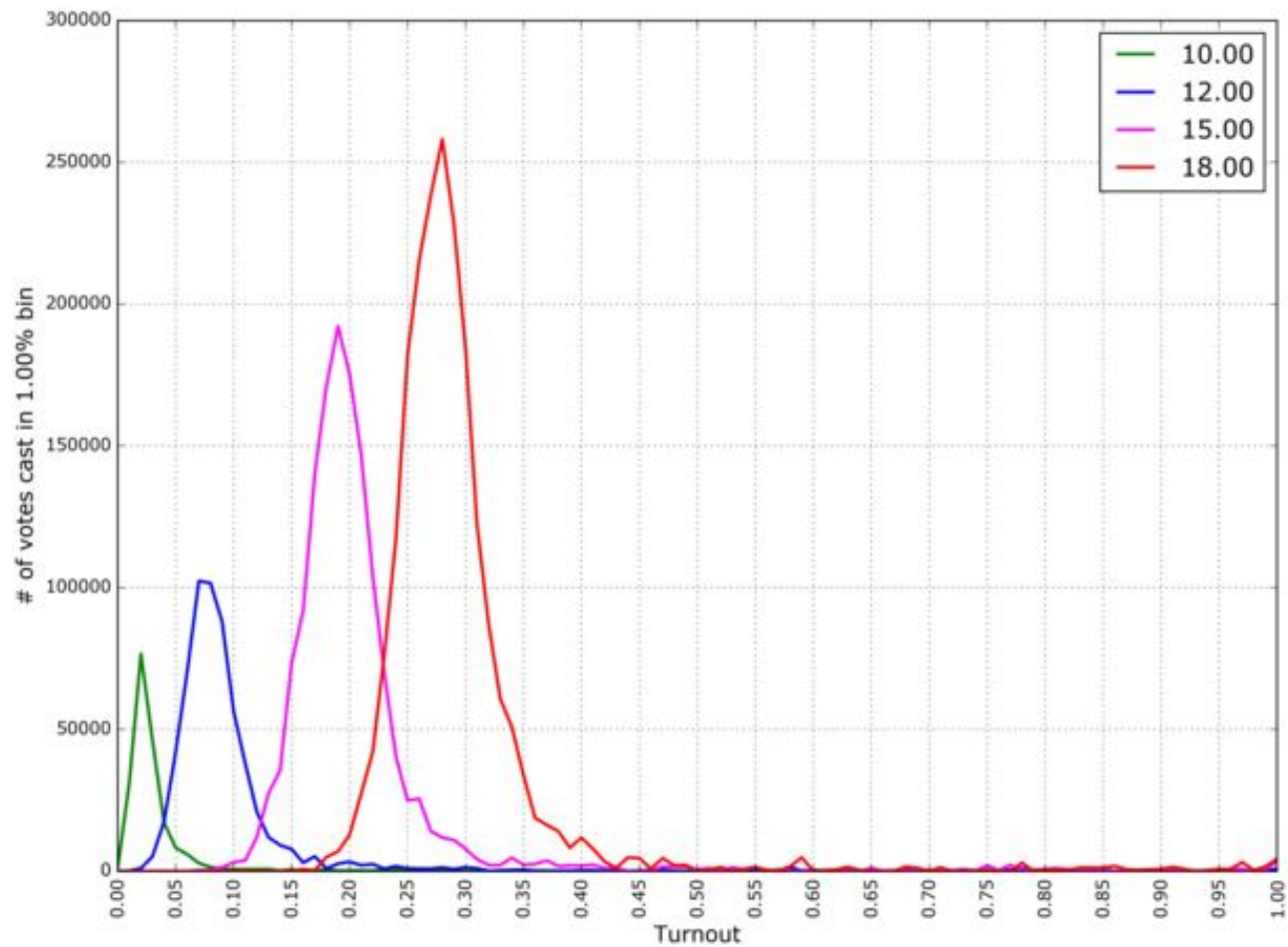


- Асимметрия 0,84; эксцесс 1,23.



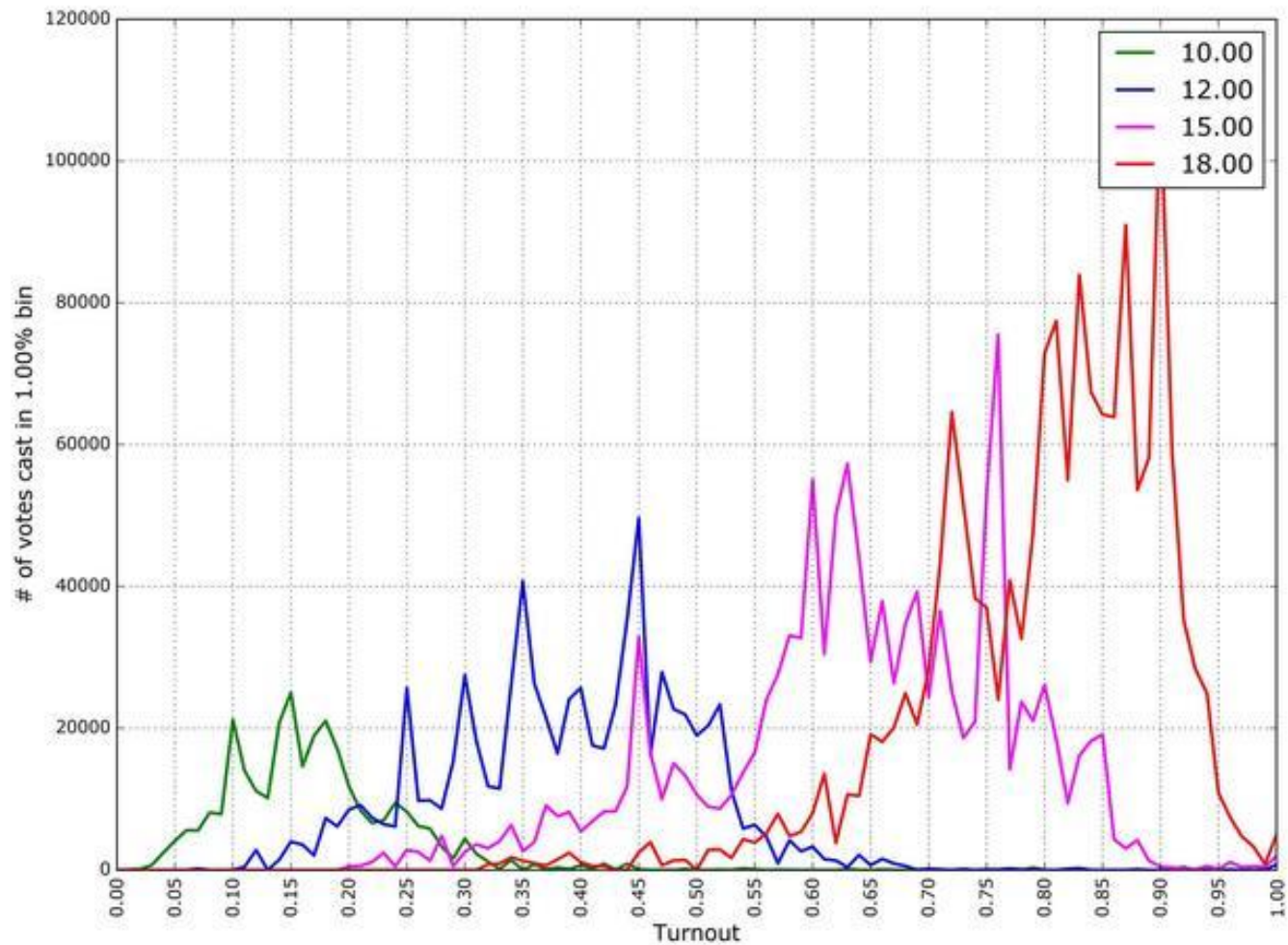
РЕЗУЛЬТАТЫ ВЫБОРОВ

город Москва



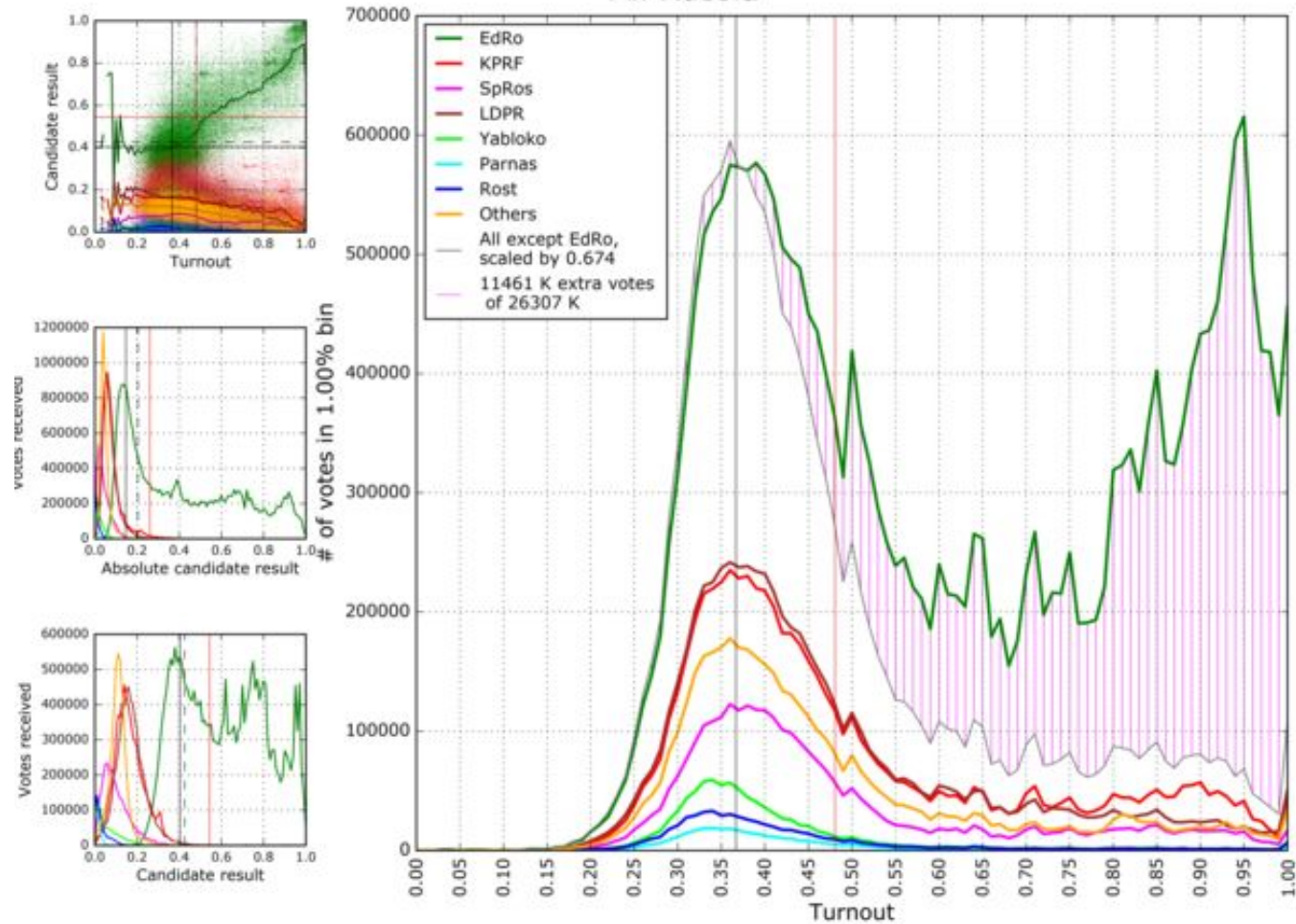
РЕЗУЛЬТАТЫ ВЫБОРОВ

Кемеровская область

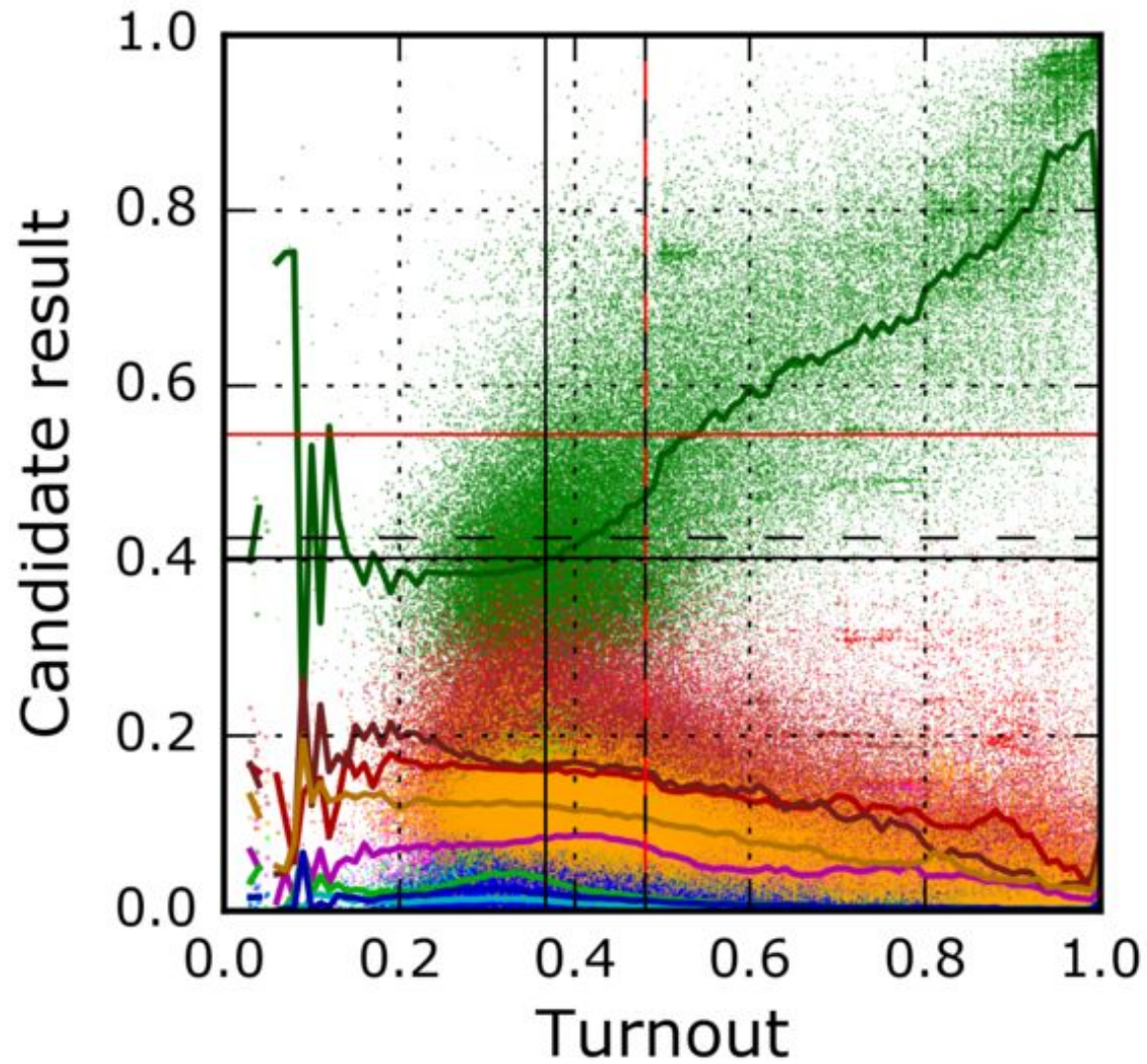


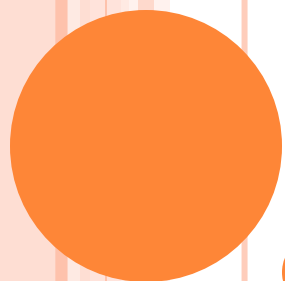
РЕЗУЛЬТАТЫ ВЫБОРОВ

Russia Duma election 2016
All Russia



РЕЗУЛЬТАТЫ ВЫБОРОВ





СТАТИСТИЧЕСКИЕ ВЫВОДЫ

ПРОВЕРКА ГИПОТЕЗ

- ? Нулевая гипотеза H_0
- ? Альтернативная гипотеза H_1



ПРИМЕР 1

- ? Нулевая гипотеза: новый экспериментальный препарат НЕ более эффективен для профилактики малярии, чем плацебо
- ? Альтернативная гипотеза: новый экспериментальный препарат способствует профилактике малярии
- ? Данные: в группе принимающей новый препарат, было зафиксировано меньше случаев заболевания малярией, чем в контрольной группе.
- ? Вывод: нулевая гипотеза отвергается и принимается альтернативная гипотеза



ПРИМЕР 2

- ? Нулевая гипотеза: лечение заключенных от наркозависимости не снижает вероятности их повторного ареста после выходы из тюрьмы
- ? Альтернативная гипотеза: лечение заключенных от наркозависимости снижает вероятность их повторного ареста после выходы из тюрьмы
- ? Данные: через пять лет наблюдений оказалось, что вероятность повторного ареста тех, кого лечили от наркозависимости, и тех, кого не лечили, примерно одинакова.
- ? Вывод: **нулевая гипотеза не отвергается.**



ОШИБКИ ПЕРВОГО И ВТОРОГО РОДА

- ? Ошибка первого рода – отклонение верной нулевой гипотезы
- ? Ошибка второго рода – принятие ложной нулевой гипотезы
- ? Какая хуже?



ПРИМЕРЫ

- ? Нулевая гипотеза: письмо не спам
- ? Альтернативная гипотеза: письмо — спам

- ? Нулевая гипотеза: пациент не болен раком
- ? Альтернативная гипотеза: пациент болен раком



ПРИМЕРЫ

- ? Спам-фильтр
- ? Ошибка первого рода: отбраковка письма, которое на самом деле не является спамом
- ? Ошибка второго рода: пропуск письма, являющегося спамом

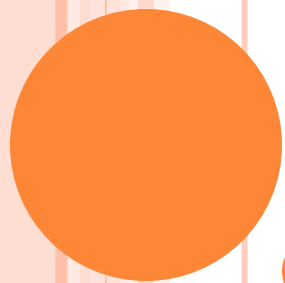
- ? Выявление рака
- ? Ошибка первого рода: отправили на дополнительные обследования того, кто не болен раком
- ? Ошибка второго рода: ошибочно не диагностировали раковое заболевание



УРОВЕНЬ ЗНАЧИМОСТИ

- ? Уровень значимости – вероятность отклонения нулевой гипотезы при условии, что она истинная. Другими словами, уровень значимости – вероятность ошибки I рода.
- ? Самый часто используемый уровень значимости – 5 %
- ? Также используются 1% и 10 %
- ? Ужесточить или ослабить?





ОЦЕНКА ВЗАИМОСВЯЗИ

Корреляция

КОВАРИАЦИЯ

- ? Ковариация является базовой мерой линейной связи между двумя случайными величинами:

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

- ? Ковариация является неудобной мерой связи, т.к. зависит от масштаба (единиц измерения), поэтому на практике не используется.



КОРРЕЛЯЦИЯ

Мерой линейной статистической связи двух случайных величин, имеющих нормальное распределение, является коэффициент парной корреляции:

$$\rho_{XY} = \frac{E[(X - E(X))(Y - E(Y))]}{\sigma(X) \cdot \sigma(Y)} = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Оценивают коэффициент парной корреляции, как правило, по формуле:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{S(X) \cdot S(Y)}.$$



СВОЙСТВА КОЭФФИЦИЕНТА ПАРНОЙ КОРРЕЛЯЦИИ

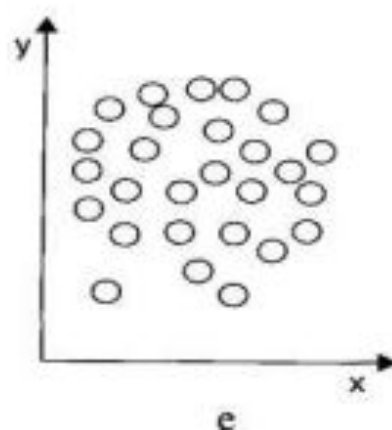
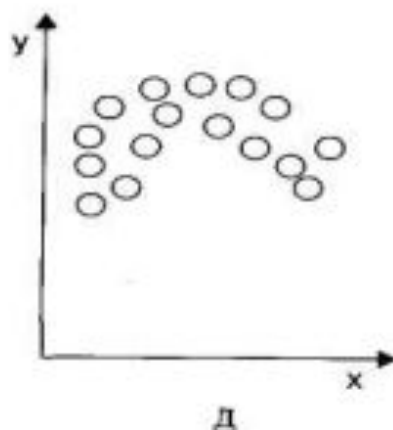
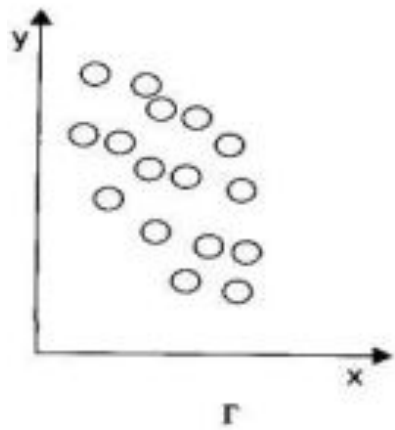
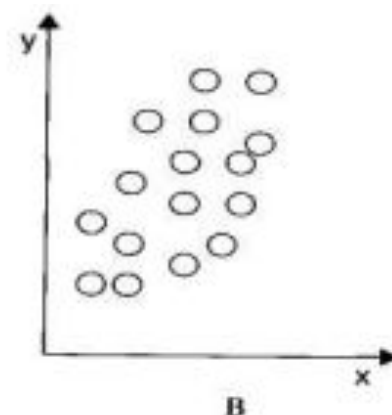
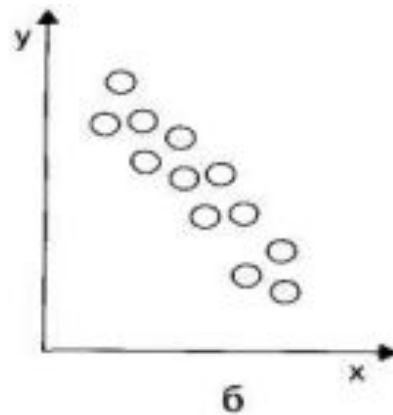
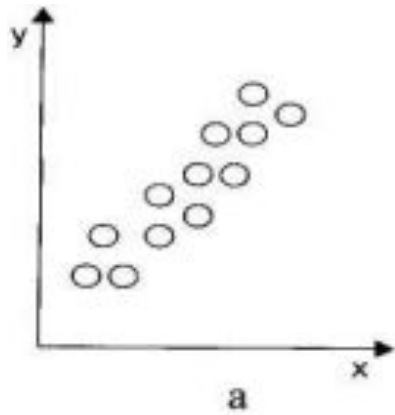
- ? Коэффициент корреляции показывает направление и тесноту связи

Свойства коэффициента парной корреляции:

1. $r_{xy} = r_{yx}$ (свойство симметричности: невозможно установить причину и следствие).
2. $|r_{xy}| \leq 1$.
3. $|r_{xy}| = 1 \Rightarrow$ функциональная линейная связь.
4. $r_{xy} = 0 \Rightarrow$ связь отсутствует.
5. $r_{xy} > 0 \Rightarrow$ прямая связь , $r_{xy} < 0 \Rightarrow$ обратная связь.

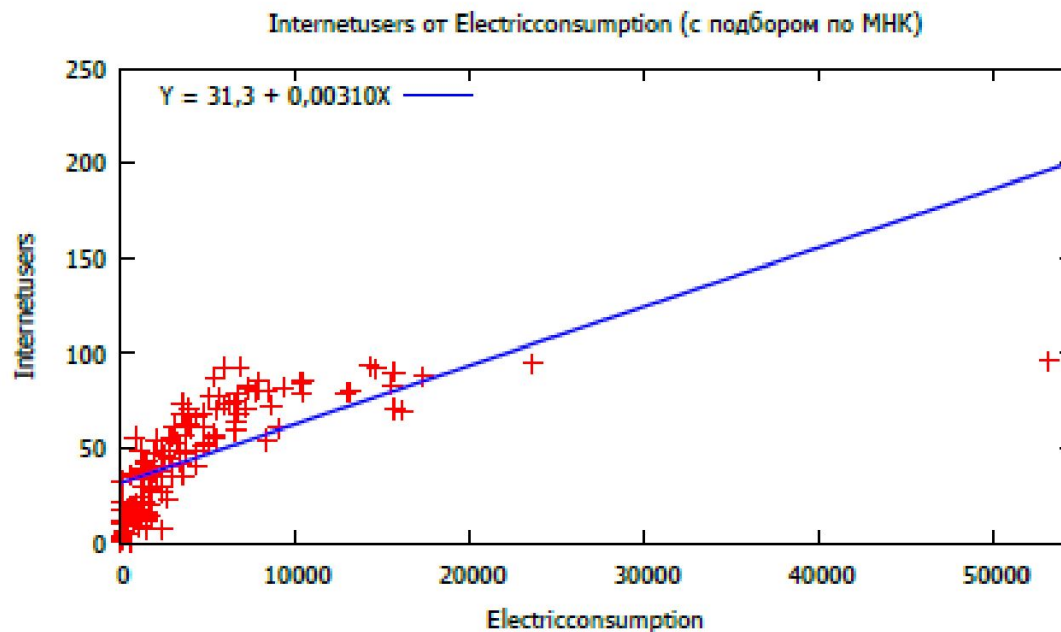


ПОСТРОЕНИЕ ГРАФИКА РАССЕЙВАНИЯ



ПРИМЕР

- ? По выборке из 139 стран рассмотрены два показателя:
- ? 1. Electric consumption – потребление электроэнергии, квт/чел, за 2012 год;
- ? 2. Internet users – количество пользователей интернетом на 100 чел, за 2012 год;



ПРИМЕР

- ? По выборке из 139 стран рассмотрены два показателя:
- ? 1. Electric consumption – потребление электроэнергии, квт/чел, за 2012 год;
- ? 2. Internet users – количество пользователей интернетом на 100 чел, за 2012 год;

- ? Парный коэффициент корреляции между ними равен $r = 0,67$



ПРОВЕРКА ЗНАЧИМОСТИ

- ? Чтобы сделать вывод о наличии/отсутствии корреляционной связи, необходимо проверить значимость коэффициента корреляции.
- ? Алгоритм:
- ? 1. Формулируются две гипотезы: H_0 : связь отсутствует, H_1 : связь существует.
- ? 2. Рассчитывается статистика Стьюдента:

$$t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}.$$

- ? 3. Расчетное значение сравнивается с табличным при уровне значимости (1 %, 5 %, 10 %) и степенях свободы $n - 2$. Если расчетное значение меньше табличного, не отвергается гипотеза H_0 об отсутствии корреляционной связи, если расчетное больше табличного – гипотеза H_1 .



ПРИМЕР

- ? Проверим значимость найденного коэффициента корреляции = 0,67.
- ? $n - 2 = 139 - 2 = 137$
- ? $t = 10,57$

- ? Вывод: коэффициент значим/не значим на уровне значимости _____



КОРРЕЛЯЦИОННАЯ МАТРИЦА

- ? Способ представления парных коэффициентов корреляции;
- ? Матрица, которая состоит из парных коэффициентов корреляции.

- ? Свойства корреляционной матрицы:
 - Квадратная
 - Симметрична относительно главной диагонали
 - На главной диагонали - единицы



ПРИМЕР

- ? Electric consumption – потребление электроэнергии, квт/чел, за 2012 год;
- ? Internet users – количество пользователей интернетом на 100 чел, за 2012 год;
- ? Population 15-64 – доля населения в возрасте от 15 до 64 лет, %, за 2012 год;
- ? Rural population – доля сельского населения, %, за 2012 год.

| Electric consumption | Internet users | Population15-64 | Rural population | |
|----------------------|----------------|-----------------|------------------|----------------------|
| 1,0000 | 0,6733 | 0,3733 | -0,5159 | Electric consumption |
| | 1,0000 | 0,5772 | -0,6846 | Internet users |
| | | 1,0000 | -0,4401 | Population15-64 |
| | | | 1,0000 | Rural population |

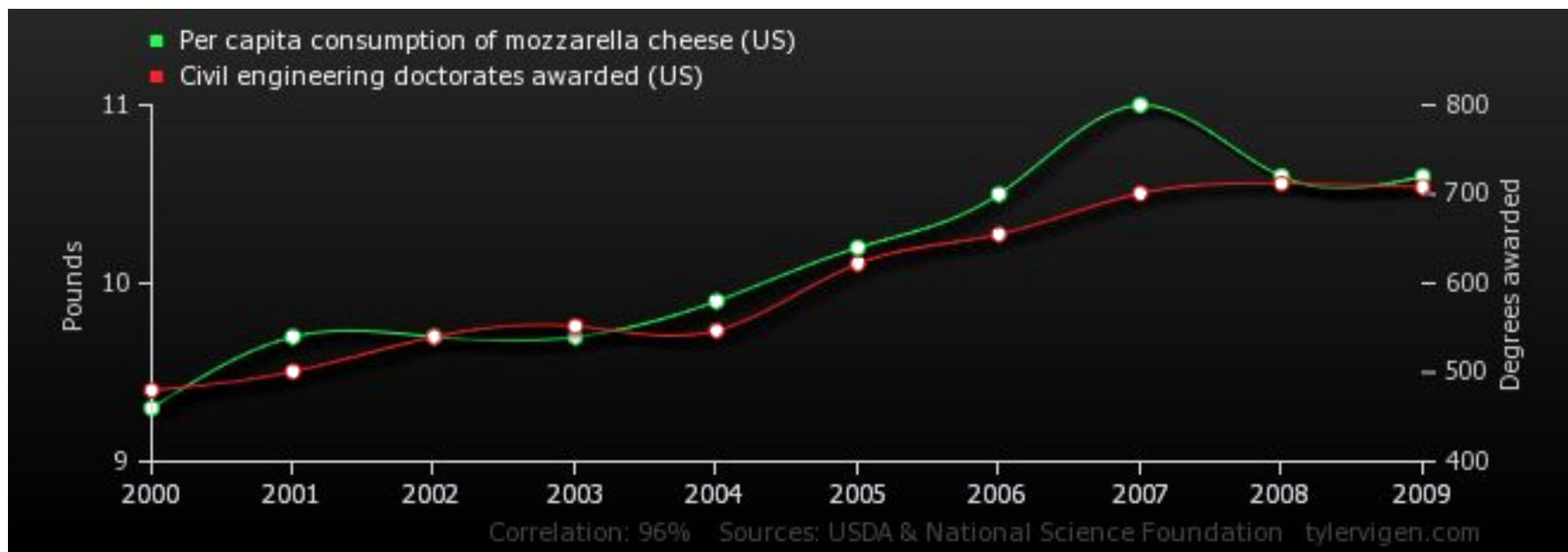


ПРИМЕРЫ КОРРЕЛЯЦИЙ

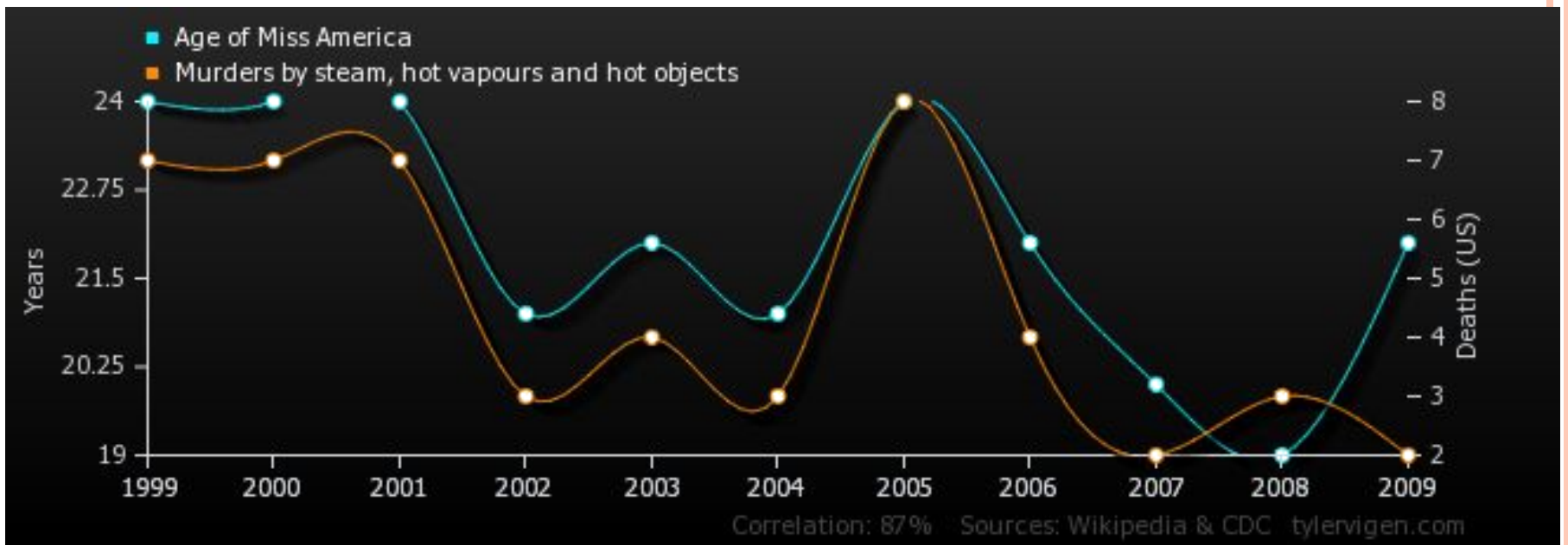
- ? Наблюдается тесная прямая корреляционная связь между количеством аистов и количеством новорожденных;
- ? Наблюдается тесная корреляционная связь между объемом потребления мороженого и количеством утонувших людей;



ЕЩЕ ПРИМЕРЫ



И ЕЩЕ



ПРИМЕР ИСПОЛЬЗОВАНИЯ КОРРЕЛЯЦИИ

- ? Американская компания Netflix (поставщик фильмов и сериалов на основе потокового мультимедиа) дает пользователю рекомендации посмотреть тот или иной фильм на основе корреляционного анализа.
- ? Пользователь выставляет оценки ряду фильмов
- ? Эти оценки сравниваются с рейтингами других кинозрителей, чтобы выявить тех, чьи оценки высоко коррелированы с оценками данного пользователя
- ? Netflix рекомендует пользователю фильмы, понравившиеся его единомышленникам

