



АНАЛИЗ ДАННЫХ В КОНТЕКСТЕ ПРОЦЕССА ФОРМИРОВАНИЯ МИРОВОЗЗРЕНИЯ СОВРЕМЕННЫХ СТУДЕНТОВ

(ОПЫТ ПРЕПОДАВАНИЯ
В КУБАНСКОМ ГАУ ИМ. И.Т. ТРУБИЛИНА)

Кацко Игорь Александрович – заведующий кафедрой статистики и
прикладной математики, д.э.н., к.т.н., профессор



1922-1926

Обучаются ученые с мировым именем: П.П. Лукьяненко и В.С. Пустовойт

1942

Эвакуация в Самарканд

1922

Основан Кубанский сельскохозяйственный институт (КСХИ)



1964

Переезд на новую территорию по улице Калинина, 13



1967

За заслуги в подготовке специалистов сельского хозяйства и развитие научных исследований награжден орденом Трудового Красного Знамени

1991

КСХИ преобразован в Кубанский государственный аграрный университет



9001:2008

Подтвержден сертификат соответствия системы менеджмента качества

2012



2017

2005

Визит Президента Российской Федерации В.В. Путина



1500 волонтеров приняли участие в Олимпиаде в Сочи

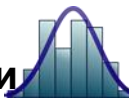
2014



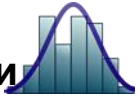
«Образование» 2006

Победитель приоритетного национального проекта «Образование»





С 1922 г. подготовлено более 130 тыс. специалистов. На 22 факультетах обучается более 17 тысяч студентов по очной и заочной формам.

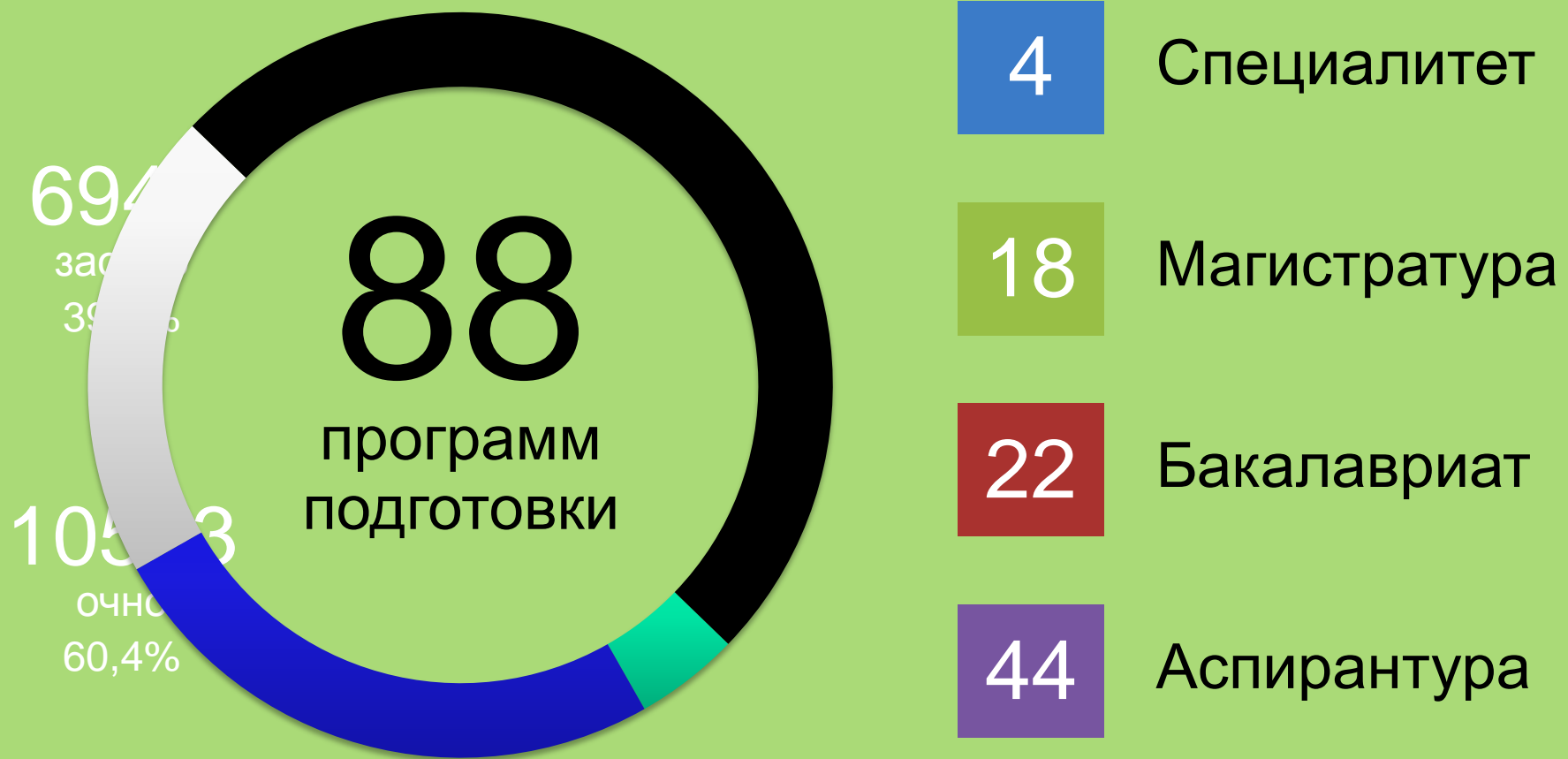


Ботанический сад
им. проф. И. С. Косенко

Спортивные объекты



Программы подготовки



Количество обучающихся

17524

6941

заочно
39,6%

10583

очно
60,4%

1531

специалисты

3047

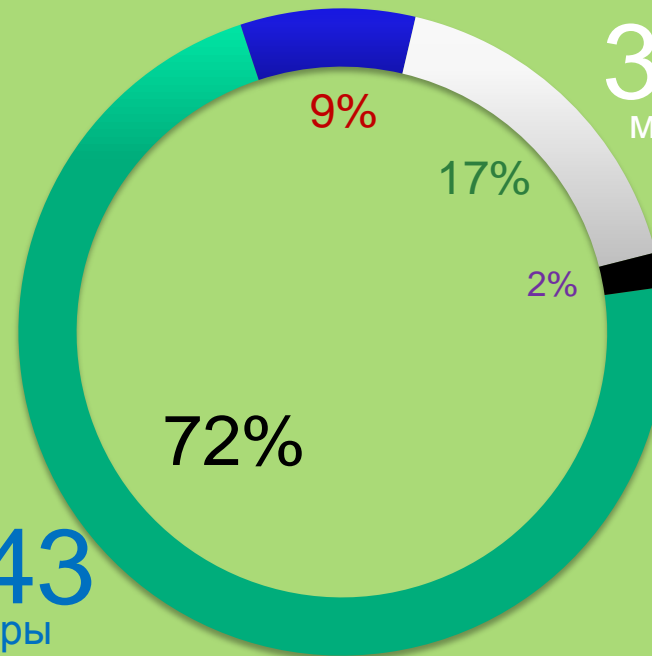
магистры

303

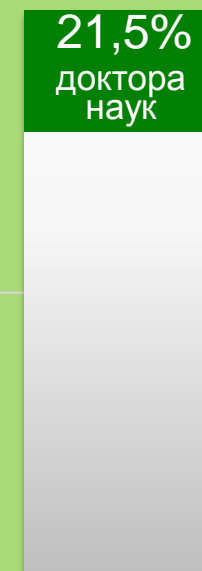
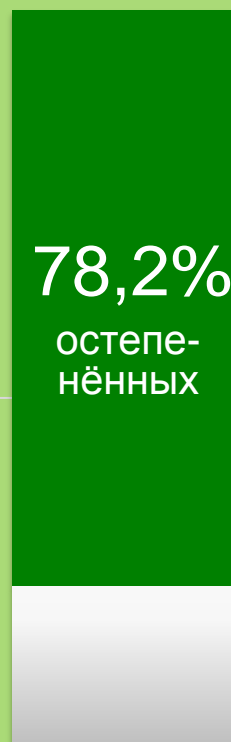
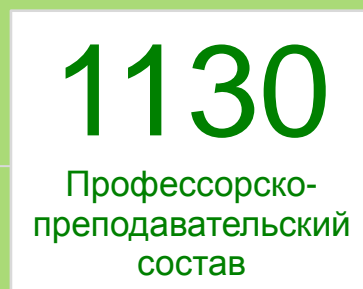
аспиранты

12643

бакалавры



Сотрудники университета

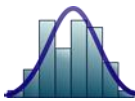




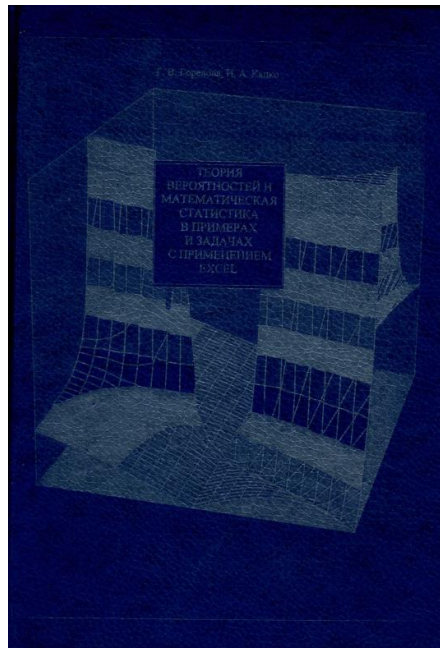
На 174 га разместились 20 учебных и учебно-лабораторных корпусов, 20 студенческих общежитий на 9,5 тысяч мест, 2 научно-исследовательских института (НИИ биотехнологии и сертификации пищевой продукции, в состав которого входит учебно-научно-производственный инновационный комплекс «Технолог»; НИИ прикладной и экспериментальной экологии), ботанический сад, опытная станция, студенческая поликлиника, комбинат студенческого питания, спортивный комплекс с плавательным бассейном, футбольный стадион с площадками для игровых видов спорта, трибунами на 880 мест и искусственным покрытием.

В структуру университета входят также два высокорентабельных учебно-опытных хозяйства «Кубань» и «Краснодарское». Важным показателем качества подготовки в вузе является востребованность его выпускников, их стремительный карьерный рост.

**Более 80% специалистов и руководителей,
работающих в АПК края, — выпускники
аграрного университета**



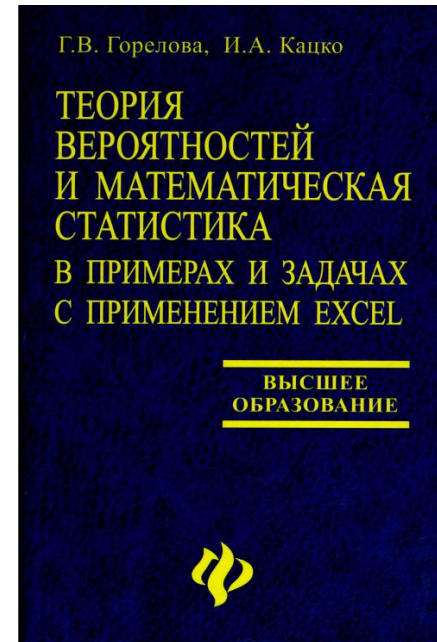
**Кафедра статистики и прикладной математики
образована в 1961 г. под руководством Академика РАСХН Д.Н. Письменной**



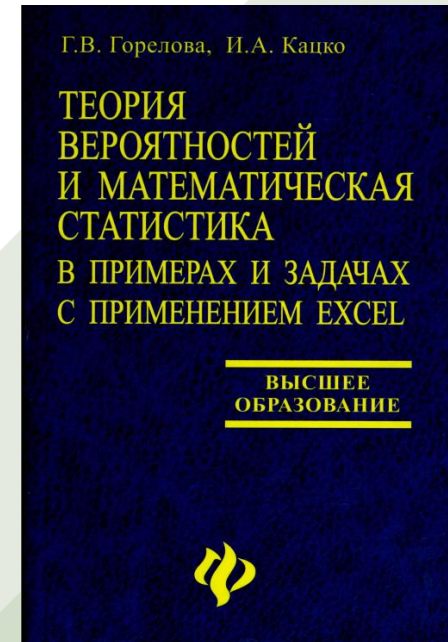
2000
Краснодар



2002

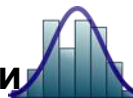


2005
Ростов – на – Дону



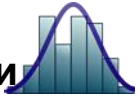
2006

**Анализ данных на ПК для студентов ФПИ стал преподаваться в курсе ТВ и МС,
затем эконометрики и спецкурсов с 1997 г. : Excel 5, 7, 97 далее Statistica 5, 6**



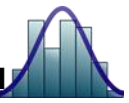
ИСТОРИЯ СОТРУДНИЧЕСТВА

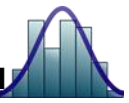
*2006 г.
НАУЧНЫЙ ТУРИЗМ*

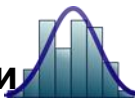


Участники X международной научно-практической конференции по системному анализу («Таймази» - альплагерь ТРТУ, сев. Осетия – Дигория, июль 2006г) – как первопричина и идейный источник появления настоящего практикума (или конструктивный подход к подготовке аналитиков, предлагаемый жителями «нижнего» домика – «верхним» жителям).

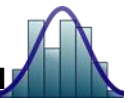








Начальник альплагеря «Таймази» Александр Цымбал и Г.В. Горелова



ИСТОРИЯ СОТРУДНИЧЕСТВА

2006 г. –

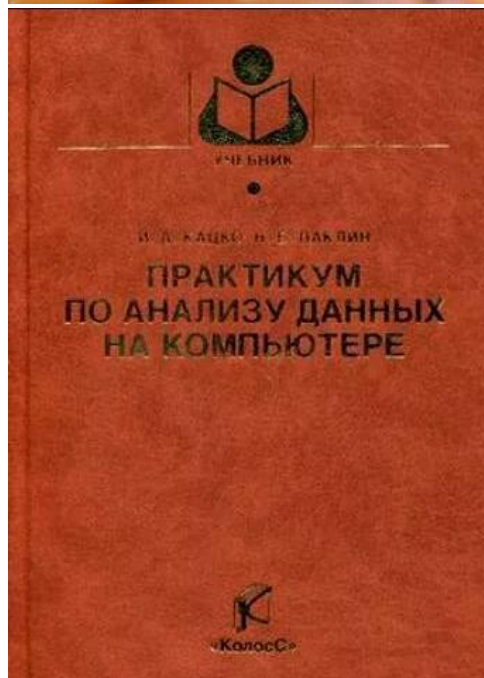
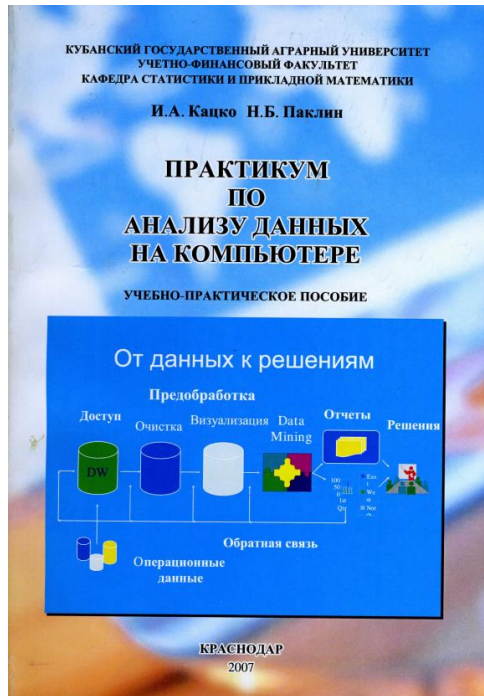
***подписан договор о сотрудничестве
КубГАУ и BaseGroup Labs,***

2007 г. –

***издан практикум по анализу данных
на компьютере (Краснодар),***

2009 г. –

переиздание практикума (Москва)



Часть I Статистический анализ данных

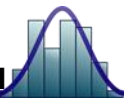
Введение в статистические методы анализа данных

Дисперсионный анализ
Регрессионный анализ
Ковариационный анализ
Кластерный и дискриминантный анализ
Факторный анализ
Анализ временных рядов

Часть II Интеллектуальный анализ данных

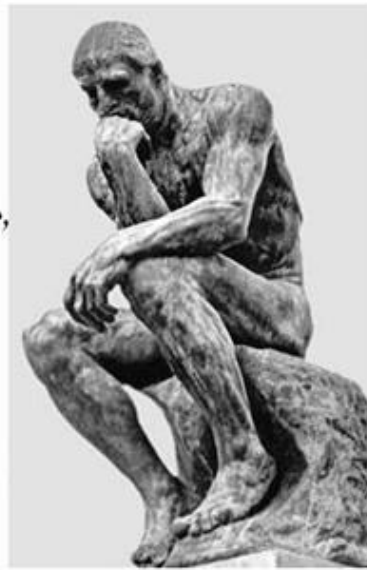
Введение в методы интеллектуального анализа данных

Data Mining система PolyAnalyst
Аналитическая платформа Deductor.
Хранилища данных
Многомерные отчеты и OLAP
Искусственные нейронные сети.
Многослойный персептрон
Логистическая регрессия и деревья решений в задаче кредитного скоринга



DATA

*Нам говорят «безумец» и «фантаст»,
Но, выйдя из зависимости грустной,
С годами мозг мыслителя искусный
Мыслителя искусственно создаст.
И. Гёте*



IT
KNOWLEDGE



Вид КвбГАУ из космоса

...
управление
...
проблемы
...



Укрупнённая схема предприятия: производство, организация и управление

как известно, проблемы принято классифицировать как

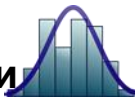
Структурированные (I),

Слабоструктурированные (II),

Неструктурированные (III)

проблемы управления организационными системами относятся к **слабоструктурированным**

проблемам ...



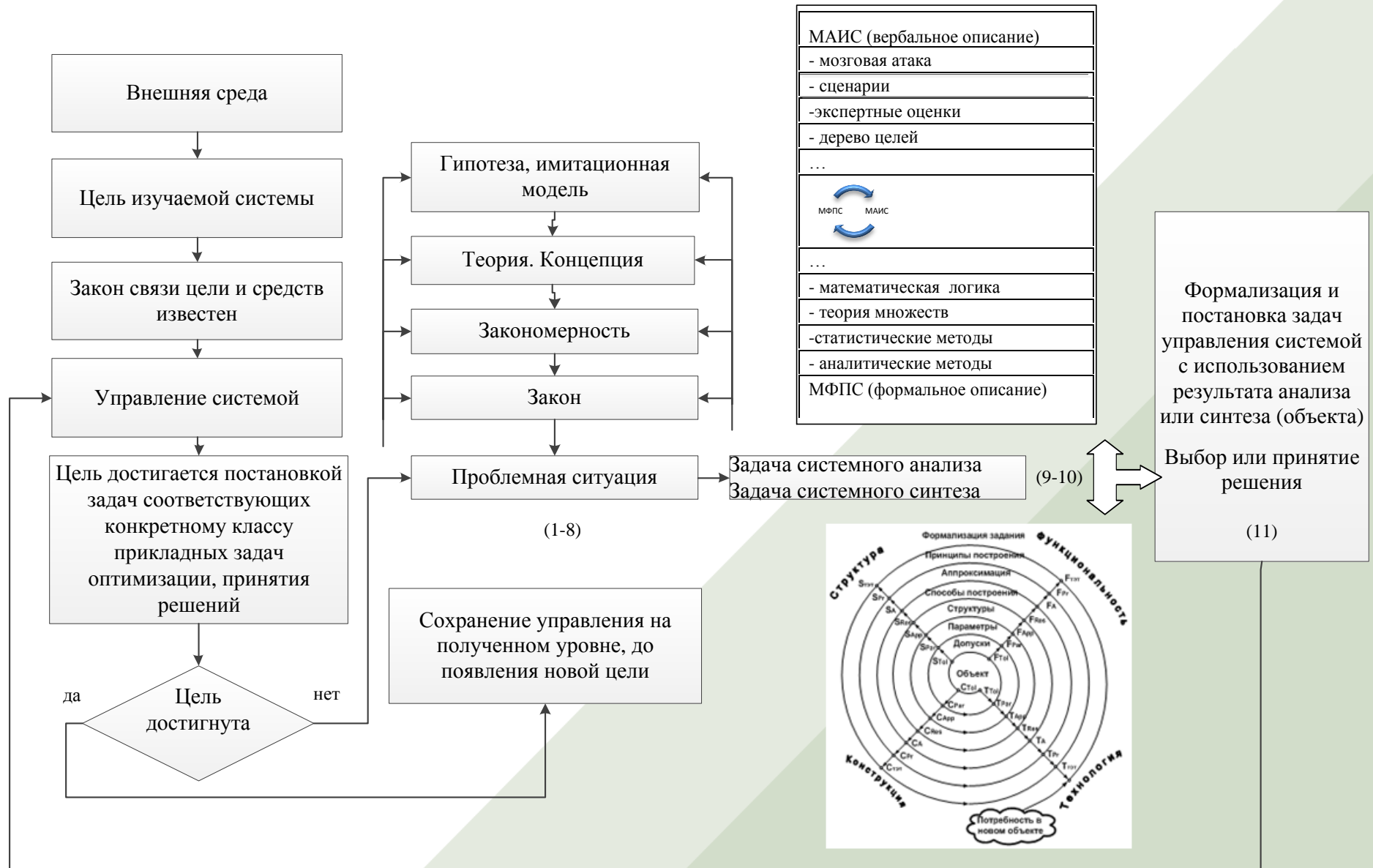
Этапы прикладных системных исследований

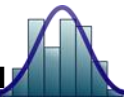
(Тарасенко Ф, 2010 г.)

- 1) **Фиксация проблемы.**
- 2) **Диагностика проблемы.**
- 3) **Составление списка стейкхолдеров.**
- 4) **Выделение проблемного месива (mess) и извлечение сведений путем построения все более содержательных моделей.**
- 5) **Определение конфигуратора (минимального набора профессиональных языков), определяющегося природой проблемы.**
- 6) **Целевыявление .**
- 7) **Определение критериев.**
- 8) **Экспериментальное исследование системы сельского хозяйства.**
- 9) **Построение и усовершенствование моделей.**
- 10) **Генерирование альтернатив.**
- 11) **Выбор, или принятие решения.**
- 12) **Реализация улучшающего вмешательства.**



Общая схема прикладных системных исследований – анализа известных систем и синтеза новых объектов (систем)



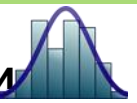


*...недостатком советской системы образования была попытка **формировать человека-творца**, а сейчас задача заключается в том, чтобы **взрастить квалифицированного потребителя**, способного квалифицированно пользоваться результатами творчества других.*

А.А. Фурсенко

**ЗАЧЕМ НУЖЕН АНАЛИЗ ДАННЫХ ? –
основной вопрос студентов (и не только ...)**

***Попытка дать ответ на этот вопрос –
обзор современного состояния науки об обработке
данных и ее роли в формировании мировоззрения
современных студентов***



Наука о том, как обрабатывать данные –

ПРИКЛАДНАЯ СТАТИСТИКА

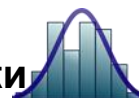
(в Англии, США, России)

или

АНАЛИЗ ДАННЫХ

(во Франции)

Цель: свертка (сжатие) информации,
содержащейся в данных для
получения новых знаний и (или)
поддержки процесса принятия
решений



Все люди от природы стремятся к знанию

Аристотель.
Метафизика

Цель – саморазвитие для управления окружающей реальностью

Власть = сила + богатство + знания

Э.Тоффлер.
Метаморфозы власти, 2009

Знания – неотделимые и кодифицированные

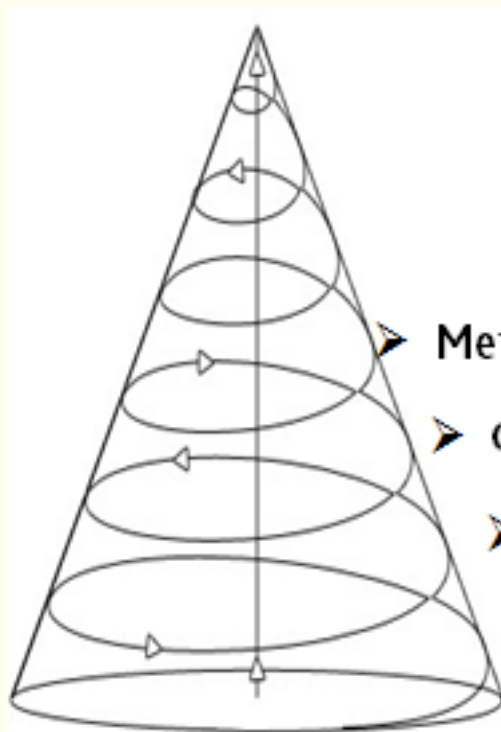
Математика – одна из основных форм представления кодифицированных знаний

Реформы системы образования в 70-80-е гг. под флагом А.Н. Колмогорова для сближения высшей и элементарной математики, привела к ситуации, выраженной словами В.И. Арнольда: ***«... повсеместно наблюдается отвращение к математике и стремление всех правителей отомстить за перенесенные в школе унижения ее уничтожением.»***

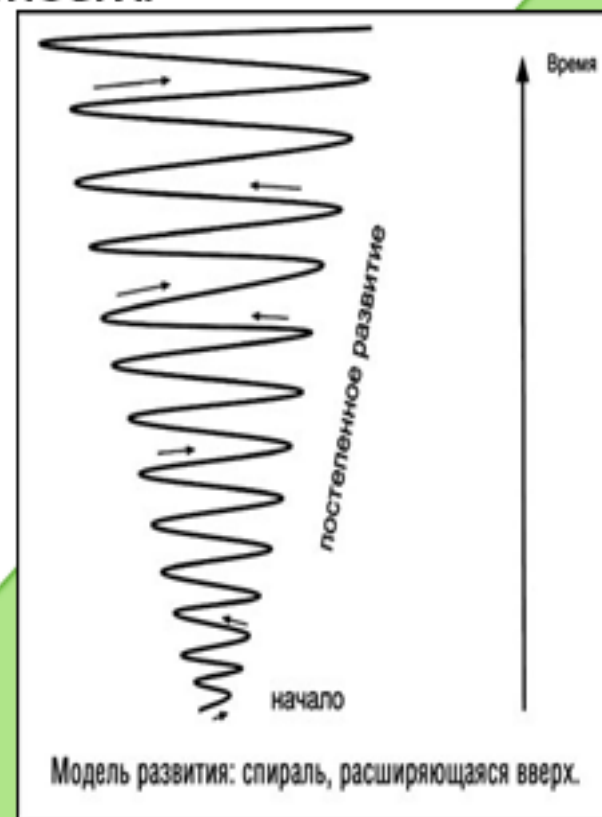


К – классификация А.Н. Колмогорова

Четыре области человеческой (математической) деятельности:

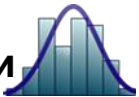


- **Метаматематика**
- **Формализованная математика**
- **Содержательная математика**
- **Изучение реального мира и практическое воздействие на него**



Ψ - "лестница форм души" Аристотеля

УрМ-4	постепенная формализация моделей (метаматематика, системное мышление)	<u>Творчество</u>
УрМ-3	ассоциативное мышление (алгебра, разведочный - РАД и (или) интеллектуальный анализ данных - ИАД)	<u>Нетиповые ситуации</u>
УрМ-2	интуитивное дологическое мышление (арифметика, аналитическая статистика)	<u>Воспроизведение типовых ситуаций</u>
УрМ-1	сенсомоторный (практический) интеллект (геометрия Евклида, описательная статистика)	<u>Узнавание</u>



Серия: вероятность, статистика и прикладные исследования в аграрном университете

Вероятность, статистика и прикладные исследования в аграрном университете

Теория вероятностей и математическая статистика

Вероятность, статистика и прикладные исследования в аграрном университете

Статистика

Вероятность, статистика и прикладные исследования в аграрном университете

Практикум по эконометрике

Вероятность, статистика и прикладные исследования в аграрном университете

Модели и методы прикладных системных исследований (практикум)

От данных к решениям

Преобразование

Данные → Обработка → Визуализация → Data Mining → Отчеты

Operational systems Technology KDD

Обратная связь

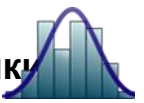
Этапы прикладных исследований для формализации и постановки задач управления в социально-экономических системах

1. Фиксация проблемы (формулирование проблемы и ее документирование).
2. Диагностика проблемы ("писать то, что всегда появляется при неудаче и никогда не появляется в случае успеха").
3. Составление списка стейкхолдеров (непосредственных участников проблемной ситуации: "stakeholders" - держатель ставок).
4. Выделение проблемного ядра (mess) и извлечение сведений путем построения все более содержательных моделей.
5. Определение конфигуриатора (минимального набора профессиональных языков, определяющегося природой проблемы).
6. Цельматрица.
7. Определение критериев.
8. Экспериментальное исследование.
9. Построение и усовершенствование моделей системы на основе чередования МАИС и МФИС или создание нового объекта (подсистемы).
10. Генерирование альтернатив.
11. Выбор, или принятие решения.
12. Реализация улучшающего вмешательства.

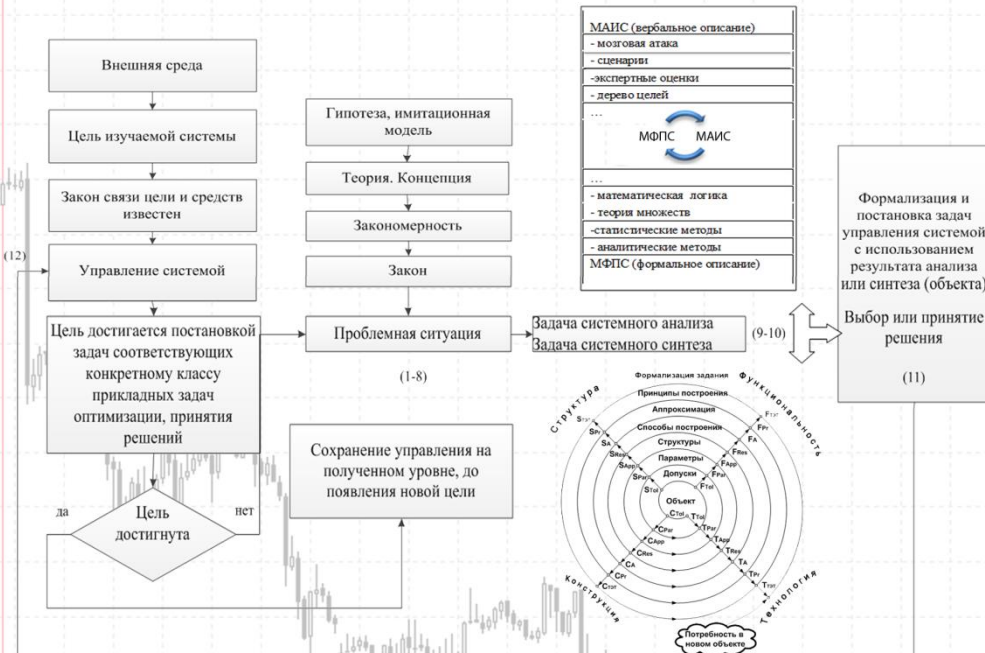
Проблема - это субъективное оценочное суждение к окружающей реальности
Решение проблемы по Р. Ашфуру

- инвариантности (invariance)
- частичное вмешательство (partialism)
- оптимальное (лучшее эффективное) решение (optimism)
- реструктурирование проблемы (restructuring)





Модели и методы прикладных системных исследований

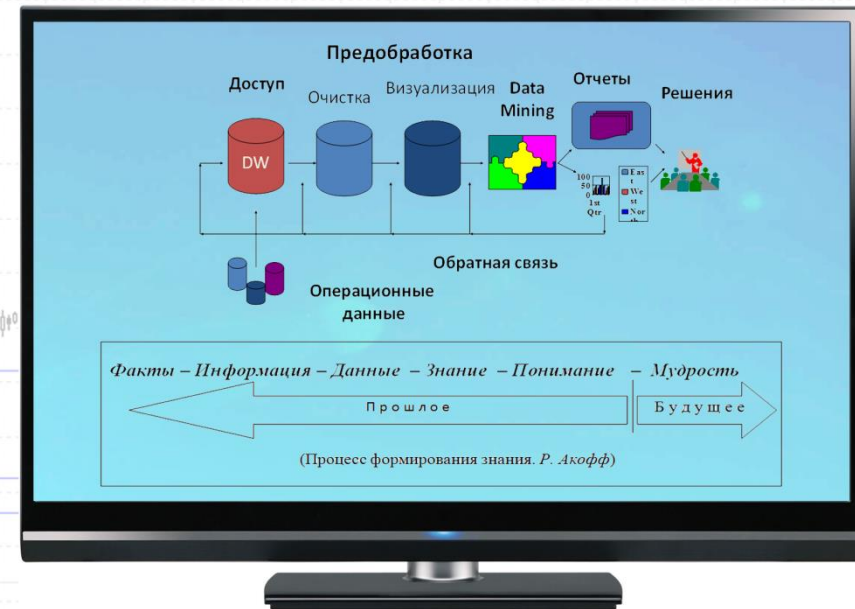


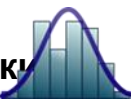
Этапы прикладных исследований для формализации и постановки задач управления в социально-экономических системах

1. Фиксация проблемы (формулирование проблемы и ее документирование).
2. Диагностика проблемы ("искать то, что всегда появляется при неудаче и никогда не появляется в случае успеха").
3. Составление списка стейкхолдеров (непосредственных участников проблемной ситуации; "stakeholders" - держатель ставок)
4. Выделение проблемного месива (mess) и извлечение сведений путем построения все более содержательных моделей.
5. Определение конфигуриатора (минимального набора профессиональных языков), определяющего природу проблемы.
6. Целевыявления.
7. Определение критериев.
8. Экспериментальное исследование.
9. Построение и совершенствование моделей системы на основе чередования МАИС и МФПС или создание нового объекта (подсистемы)
10. Генерирование альтернатив.
11. Выбор, или принятие решения
12. Реализация улучшающего вмешательства.

Проблема - это субъективное отрицательное отношение к окружающей реальности
Решение проблемы по Р. Акоффу

- невмешательство (absolution)
- частичное вмешательство (resolution)
- оптимальное (лучшее эффективное) решение (solution)
- растворение проблемы (dissolution)





РАЗДЕЛ 1. ФОРМАЛИЗАЦИЯ ДАННЫХ – (ПОДХОД СНИЗУ – ВВЕРХ)

ЧАСТЬ I. РАБОТА С ДАННЫМИ В MS Excel

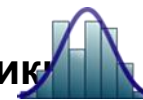
ЧАСТЬ II. АНАЛИЗ ДАННЫХ В СИСТЕМЕ Statistica

ЧАСТЬ III. НЕЛИНЕЙНЫЕ МЕТОДЫ В АНАЛИЗЕ ДАННЫХ (Mat Lab)

ЧАСТЬ IV. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ (Deductor, PolyAnaLyst,
«Статистика поисковых запросов»)

РАЗДЕЛ 2. ФОРМАЛИЗАЦИЯ ЗНАНИЙ – (ПОДХОД СВЕРХУ ВНИЗ)

1. Статистика нечисловых данных в экспертных оценках
2. Математическое представление когнитивных моделей в виде графов
3. Выбор и концептуальное описание предметной области задачи принятия решений
4. Байесовские сети
5. Системно-когнитивный анализ изображений (обобщение, абстрагирование, классификация и идентификация)
6. Системно-структурный синтез



**Воспринимается не то,
что истинно или правильно,
а то, что понятно.**
Народная мудрость

**Многие вещи нам непонятны
не потому, что наши понятия слабы;
но потому, что сии вещи не входят
в круг наших понятий.**
Козьма Прутков

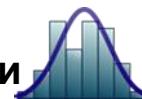
Этапы процесса формирования знания (Аристотель, Р. Акофф):

ФАКТЫ → ИНФОРМАЦИЯ → ДАННЫЕ → ЗНАНИЕ → ПОНИМАНИЕ → МУДРОСТЬ



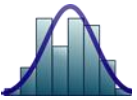
**ДАННЫЕ – ФАКТЫ, ПРЕДСТАВЛЕННЫЕ В ВИДЕ ТАБЛИЦ «ОБЪЕКТ-СВОЙСТВО»
или «свойство-свойство»**

Радость видеть и понимать есть самый прекрасный дар природы.
А.Эйнштейн.



Классификация шкал

	Шкала	Допустимые операции	Характеристики центральной тенденции	Примеры
качественные	1. Номинальная (шкала наименований или категориальная)	$X=Y$, $X \neq Y$ (сравнения)	Мода	Пол (м, ж), Национальность (...)
	2. Порядковая (ординальная)	$X=Y$, $X < Y$, $X > Y$ (сравнения и порядка)	Медиана (а так же мода)	Состояние здоровья, Красота, Мнение
количественные	3. Интервальная	$X=Y$, $X < Y$, $X > Y$, $X - Y$, $X + Y$	Средняя арифметическая (а так же мода, медиана)	Дата, Широта, Температура ($^{\circ}\text{C}$)
	4. Отношений	$X=Y$, $X < Y$, $X > Y$, $X - Y$, $X + Y$ $X \cdot Y$, X / Y	Средняя геометрическая (а так же средняя арифметическая, мода, медиана)	Возраст (от 0 до 99), Температура ($^{\circ}\text{K}$)

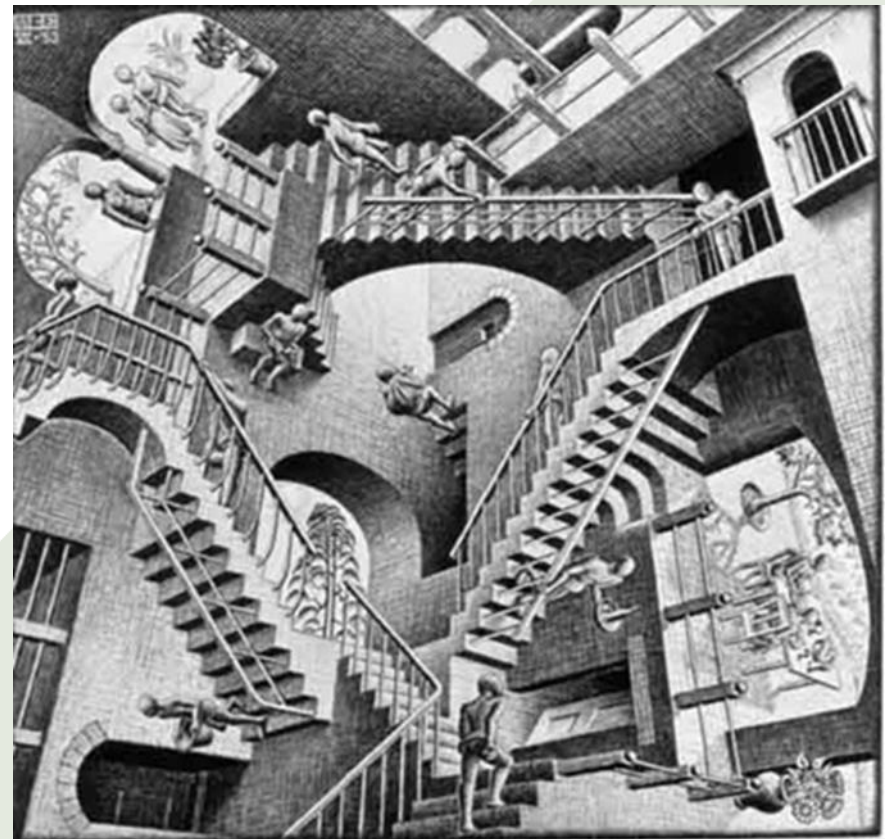


Проблемы измерения

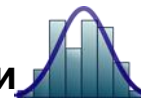
1. НЕОДНОРОДНОСТЬ НАБЛЮДЕНИЙ
2. КАЧЕСТВО ДАННЫХ (артефакты, пропуски, дубликаты, противоречия и пр.)
3. **ОГРАНИЧЕННОСТЬ ВОСПРИЯТИЯ**. Человеку удобнее всего думать об однородных вещах, если их число не превышает семи (7 ± 2 – число Миллера)
4. СУБЪЕКТИВНОСТЬ ВОСПРИЯТИЯ И МЫШЛЕНИЯ



My wife and my mother-in-law



Относительность



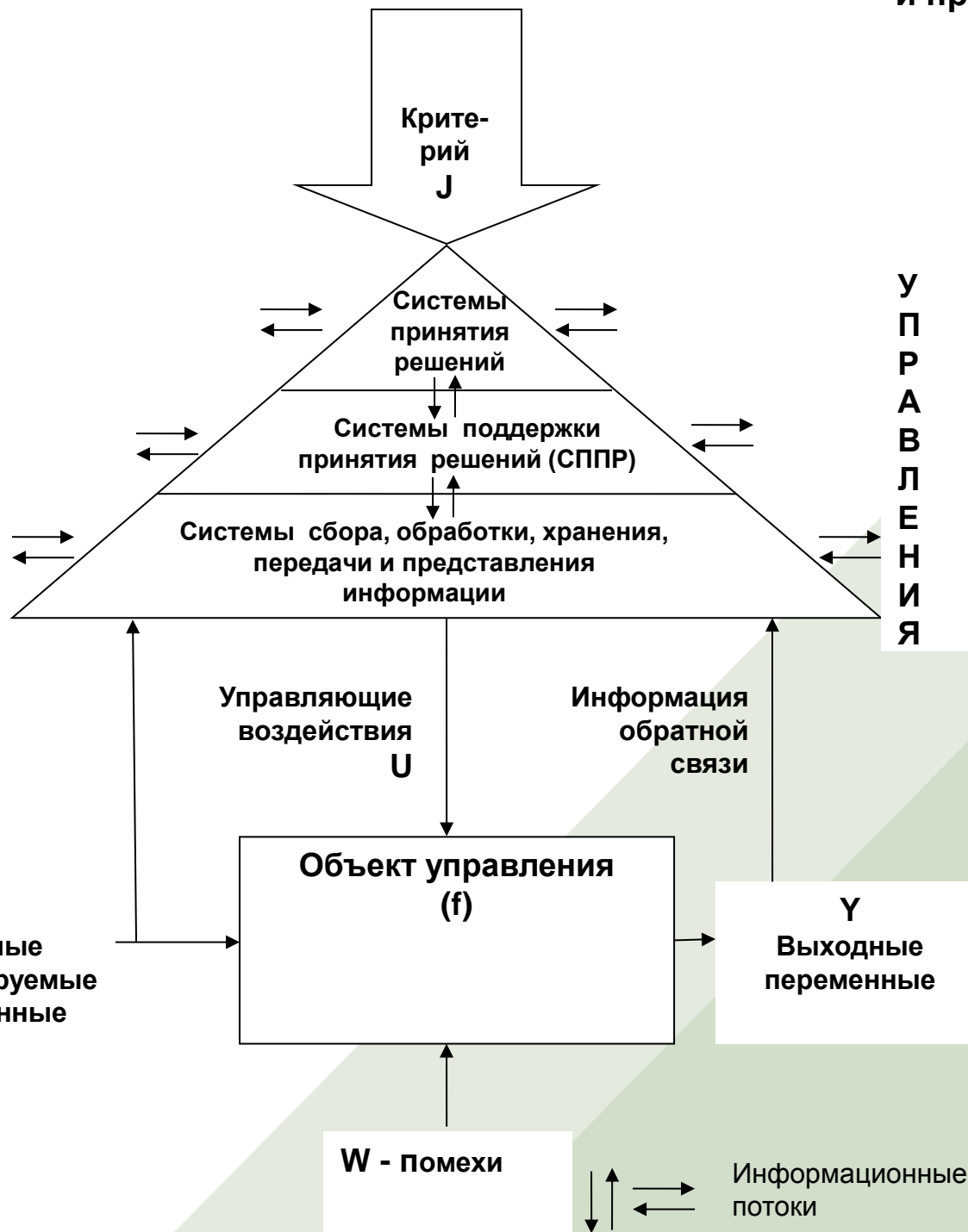
7 ± 2



Чёрный лебедь, птица из рода лебедей (Cygnus) семейства утиных (Anatidae). До XVIII века была неизвестна науке.

СИСТЕМА

УПРАВЛЕНИЯ



СППР :

статистические пакеты, системы класса Data Mining и KDD, экспертные системы

...

X
Входные контролируемые переменные

Управляющие воздействия
U

Информация обратной связи

Объект управления (f)

Y
Выходные переменные

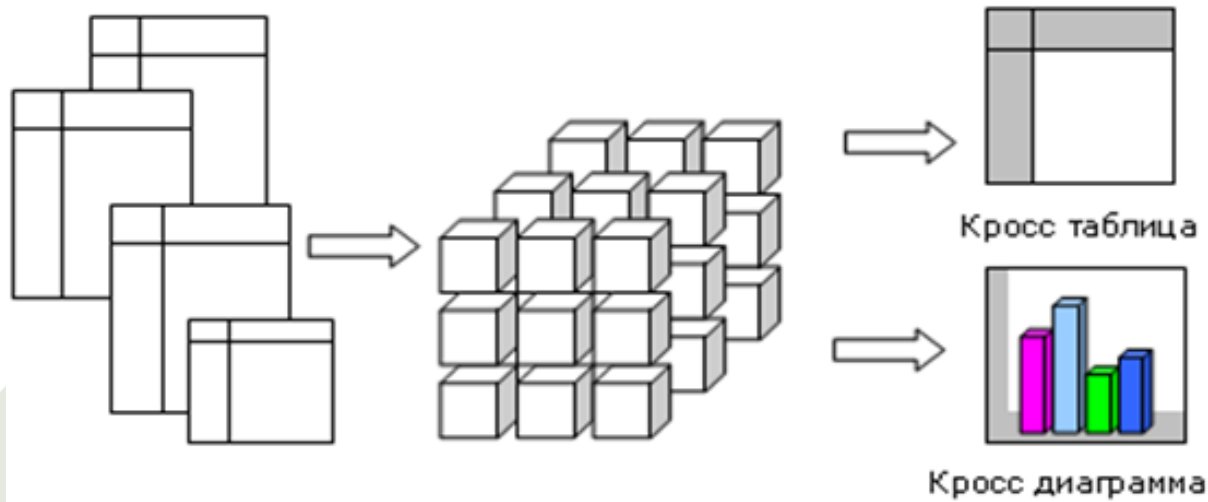
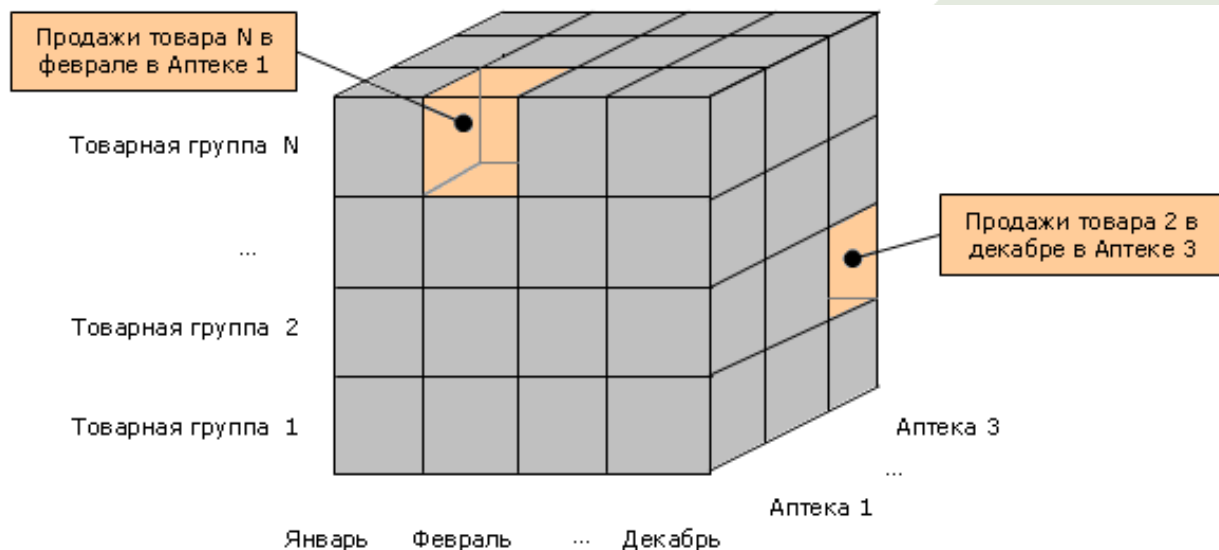
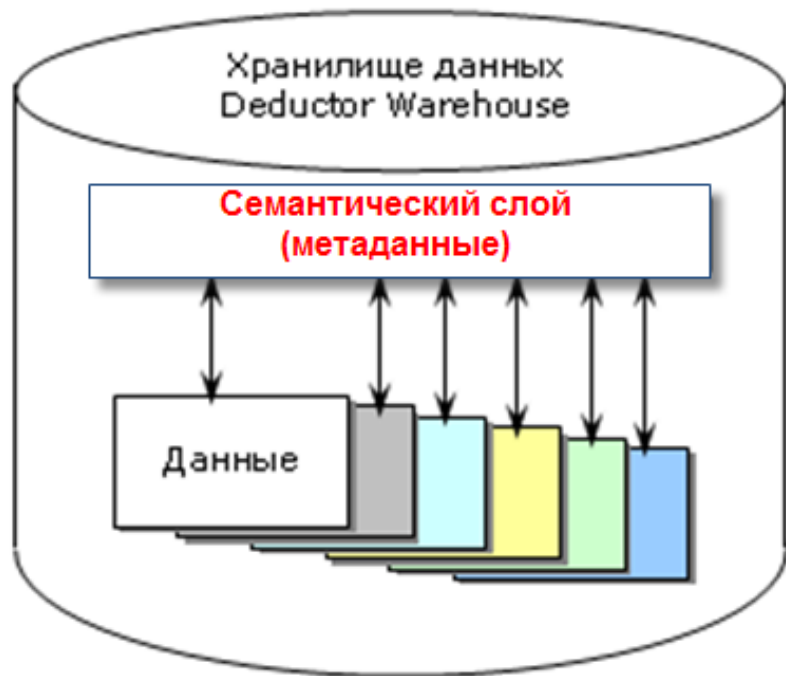
W - помехи

Информационные потоки



7 ± 2

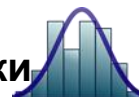
оценка текущего состояния системы
(описательные статистики, отчеты OLAP)



Плоские таблицы

N-мерный куб

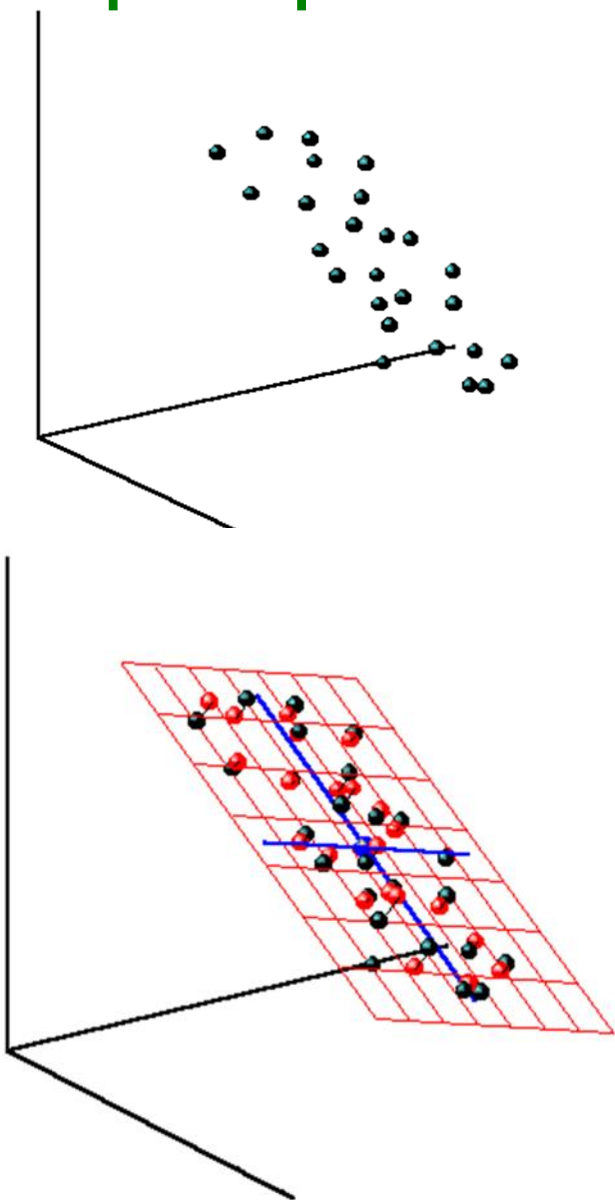
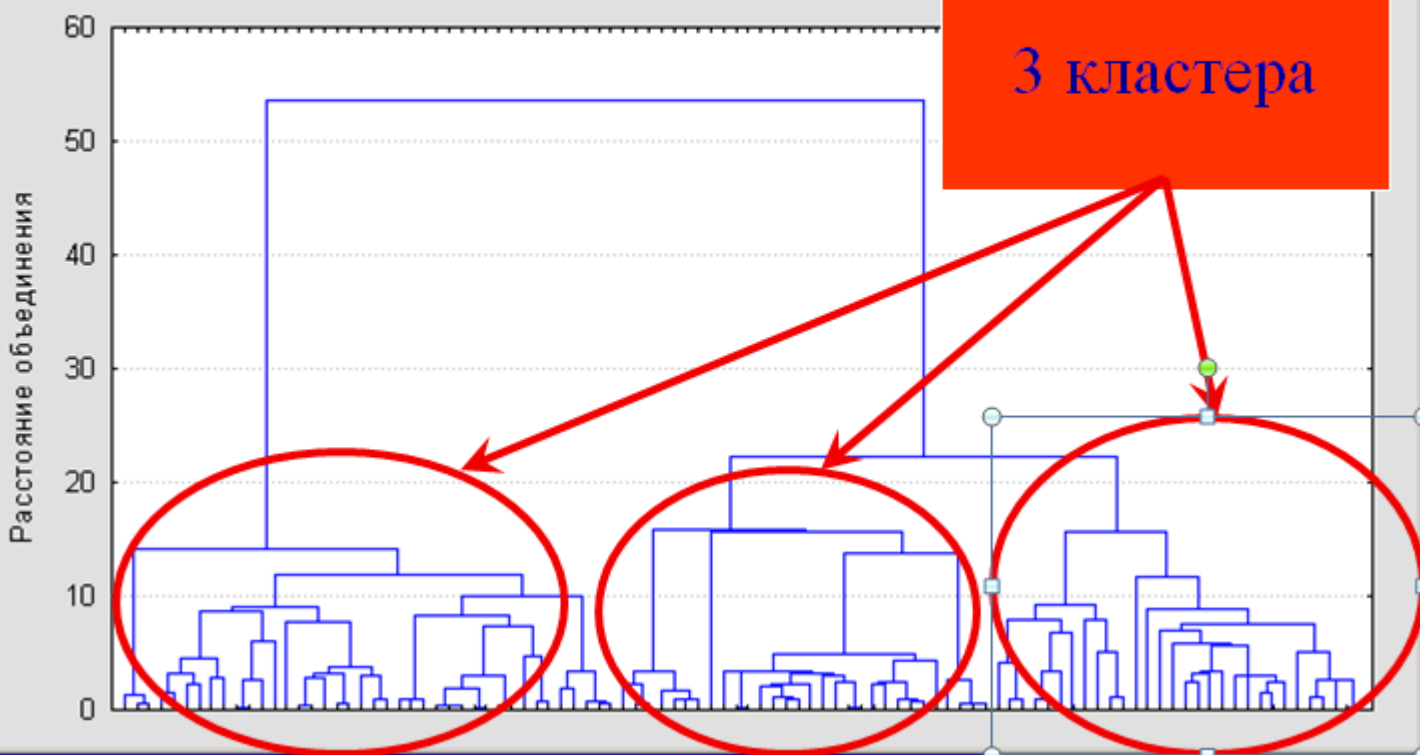
Отображения среза

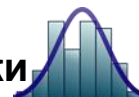


7 ± 2

Задача СППР: классификация и снижение размерности

Древовидная диаграмма
Метод одиночной связи (простое объединение)
Евклидово расстояние





Задача СППР:

7 ± 2

поиск зависимостей, прогнозирование

Итоги регрессии для зависимой переменной: PRICE (flat.sta)

R= .73496181 R2= .54016887 скоррек. R2= .53965615
F(7,6278)=1053.5 p<0.0000 станд. ошибка оценки : 5.1571

N=6286	БЕТА	Ст. ош. БЕТА	B	Ст. ош. B	t(6278)	p-уров.
Св. член			-12.6656	.573034	-22.1027	0.000000
TOTSP	.562726	.008715	.8659	.013412	64.5667	0.000000
DIST_1	.343219	.009037	21.3231	.561427	37.9803	0.000000
WALK	.085491	.008754	1.3084	.133971	9.7664	.000000
BRICK	.170653	.009058	2.9583	.157019	18.8406	0.000000
FLOOR	.069809	.008648	1.4186	.175750	8.0719	.000000
M_1	.092121	.008746	4.5358	.430618	10.5331	.000000
TEL	.046048	.008644	1.3452	.251522	5.3271	.000000

Лучшее по значимости правило:

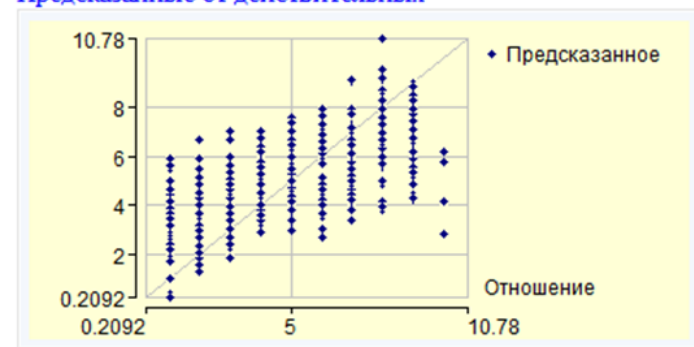
Отношение = (63.2333 * Q5 * if(-5.97024 <= 1 * Q24 * Q5 and 1 * Q24 * Q5 < -5.97024 + 29.8382, 1, 1.02459) + 0.0338207 * Q5 * Q29 * Q29 - 65.8828 * Q5 + 0.427707 * Q5 * Q2 + 0.758525 * Q5 * Q5 + 1.75088 * Q24) / (Q5 * if(-5.97024 <= 1 * Q24 * Q5 and 1 * Q24 * Q5 < -5.97024 + 29.8382, 1, 1.02459) * if(-5.97024 <= 1 * Q24 * Q5 and 1 * Q24 * Q5 < -5.97024 + 29.8382, 1, 1.02459) + 0.924907)

Лучшее по точности правило:

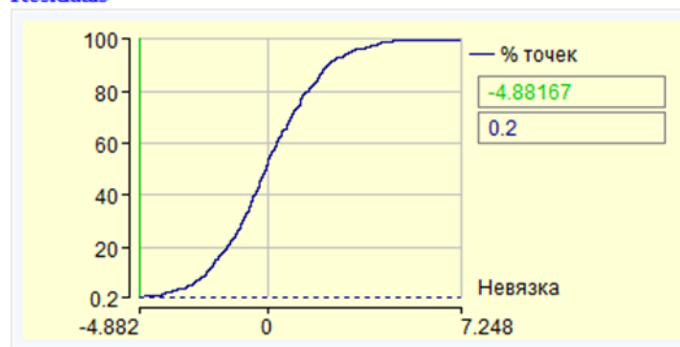
Отношение = (52.3168 * Q5 * if(-5.97024 <= 1 * Q24 * Q5 and 1 * Q24 * Q5 < -5.97024 + 29.8382, 1, 1.02459) + 0.0278703 * Q5 * Q29 * Q29 - 53.7143 * Q5 + 0.356049 * Q5 * Q2 + 0.657115 * Q5 * Q5 + 0.688304 * Q24) / (Q5 * if(-5.97024 <= 1 * Q24 * Q5 and 1 * Q24 * Q5 < -5.97024 + 29.8382, 1, 1.02459) * if(-5.97024 <= 1 * Q24 * Q5 and 1 * Q24 * Q5 < -5.97024 + 29.8382, 1, 1.02459) + 0.924907 - 0.0307857 * Q16 * Q24)

Уровень	Стд.ош.	Стд.откл.	Значим..	R-squared
наиб. знач.	0.6964	1.814	> 100	0.515
наиб. точн.	0.6873	1.791	> 100	0.5277

Предсказанные от действительных



Residuals

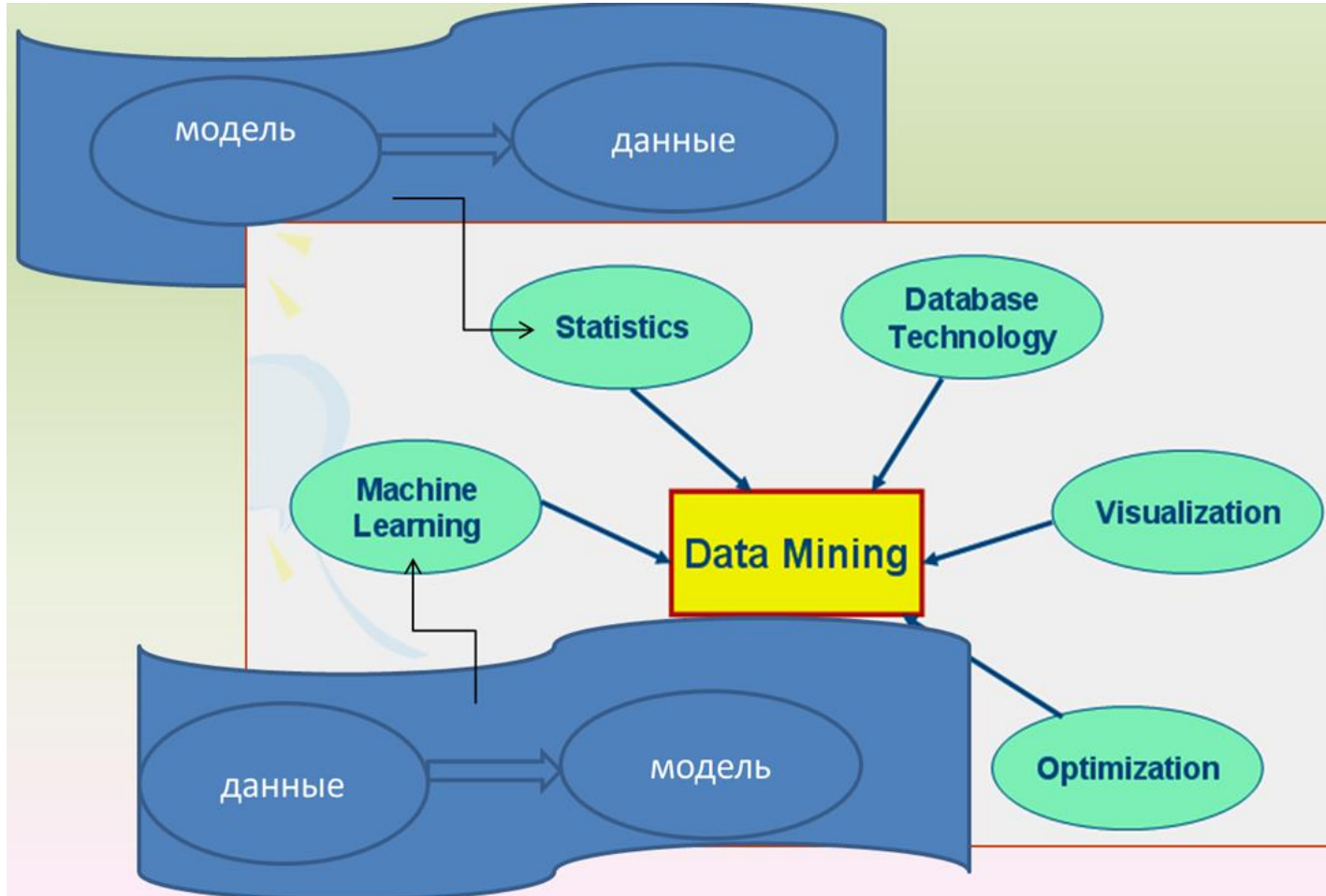


Средняя цена 1 м² недвижимости равна \$ 866 (регрессия в пакете Statistica)

Зависимость переменной «Отношение» от других факторов социологического опроса (Data Mining система – PolyAnalyst, отчёт Поиска законов)



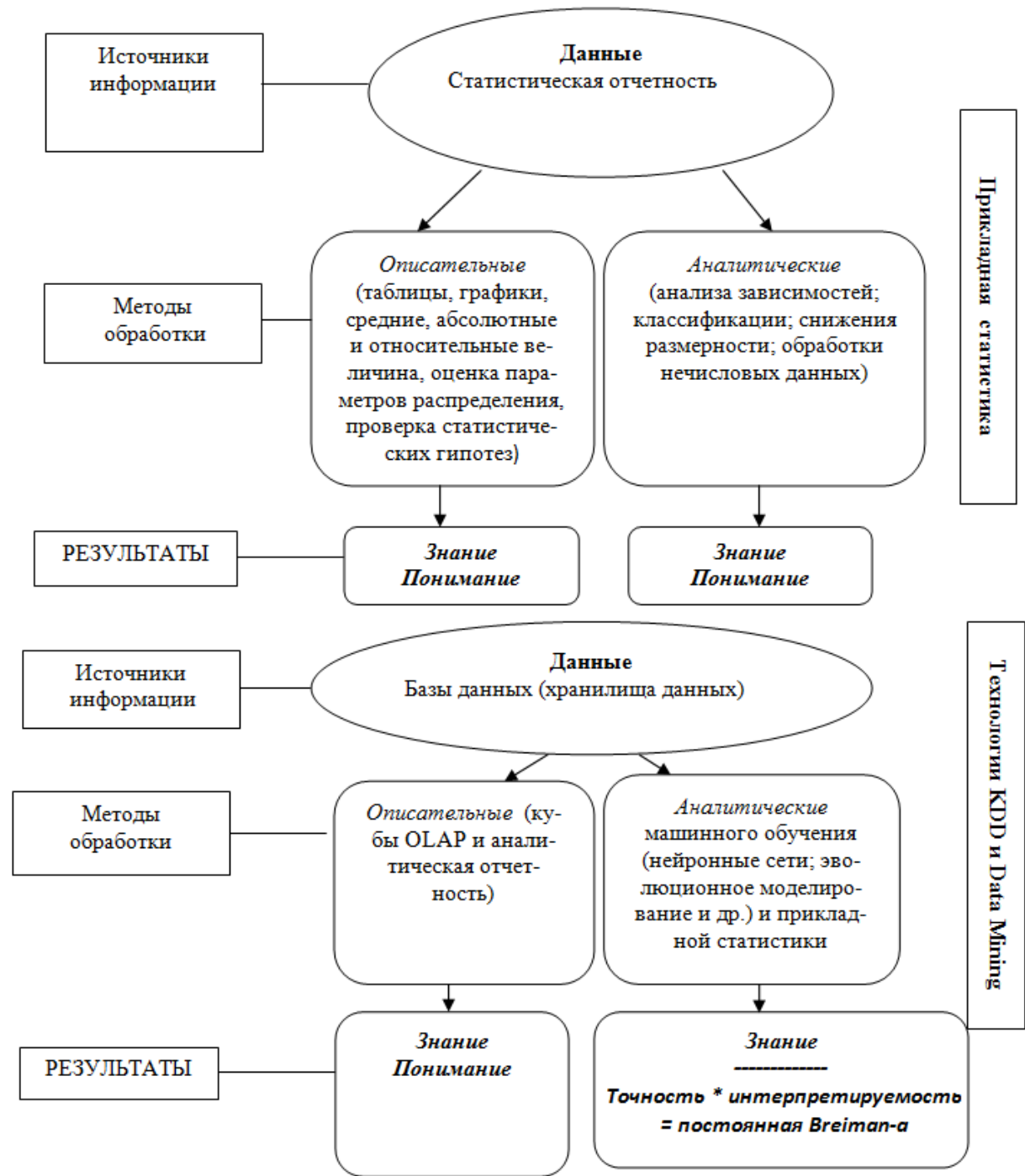
Современные подходы к анализу структурированных данных: Статистика и *Data Mining*

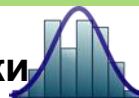


Цели

Разведочный анализ данных (1960-е гг. Дж. Тьюки) – выявление природы данных,
Data Mining (1990-е гг. Г. Пятецкий- Шапиро) – практическое приложение результатов
(ПРИБЫЛЬ!)

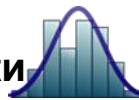
Этапы: выборка, преобразование данных, моделирование, проверка адекватности на других данных, возврат на предыдущие этапы (при необходимости)





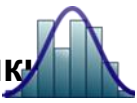
Особенности применения методов статистики и *Data Mining*

<i>Особенности</i>	<i>Статистика</i>	<i>Data Mining</i>
Тип проблем	Хорошо структурированные	Неструктурированные/ Слабоструктурированные
Роль выводов	Большая роль явных выводов	Нет явных выводов
Цель сбора данных и их анализа	Целеориентированный подход к сбору данных	Первичные данные не собираются целенаправленно
Размер совокупности	Небольшая совокупность однородных данных	Большая совокупность неоднородных данных
Интерпретация	<i>Хорошо интерпретируется количественно</i>	<i>Точность × интерпретируемость = = постоянная Бреймана</i>
Парадигма/Подход	Дедуктивная статистическая теория	Дедуктивная теория и индуктивный эвристический подход
Сигнал/Шум, <i>Signal-to- Noise Ratio (STNR)</i>	$STNR > 3$	$0 < STNR \leq 3$
Тип анализа	Подтверждающий	Разведочный
Число переменных	Небольшое (до 10)	Большое (свыше 30-50)

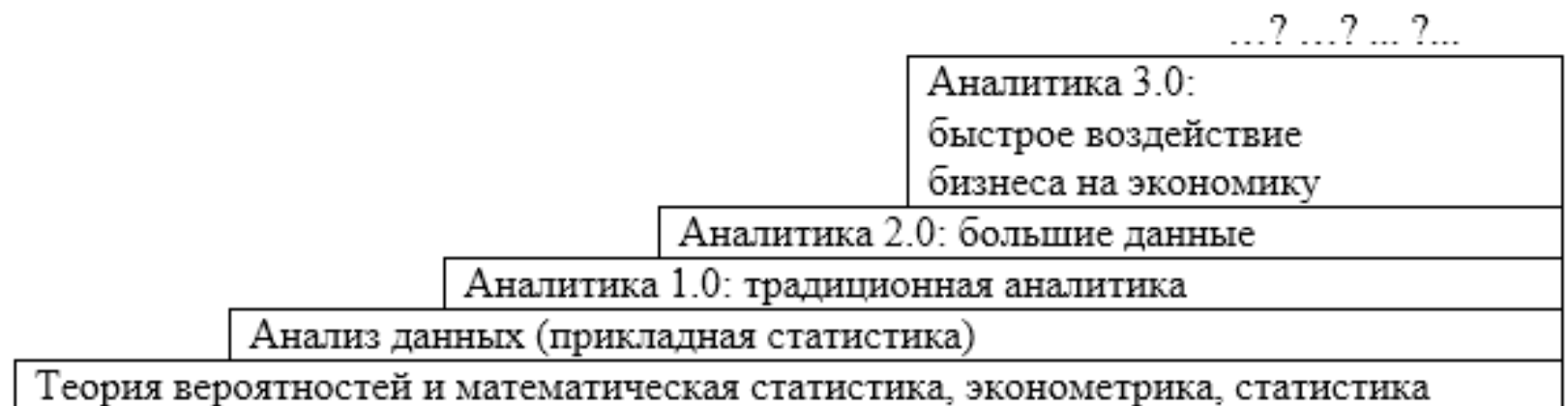


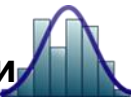
Эволюция методов анализа данных

Общество	Общественное производство	Ограничивающий фактор	Характер изменчивости требований рынка	Требуемый тип мышления	Методы анализа ретроспективных данных
Аграрное	Продукты питания	Земля	Стабильность	Консервативный	Описательные, Аналитические (начало)
Индустриальное	Товары	Капитал	Достаточная стабильность	Рыночный, способный реагировать на изменения	Описательные, Аналитические, Методы машинного обучения (начало)
Постиндустриальное	Информация	Знания	Резкие дискретные скачкообразные изменения	Стратегический, способный увидеть и сменить цели и стратегии фирмы	Описательные, Аналитические, Оперативная отчетность, <i>Data Mining, KDD</i>
Информационное	Знания	Социальное время	Резкие изменения под воздействием информационных факторов	Стратегический, способный реагировать на информационные воздействия	<i>...Text Mining Web Mining Visual Mining Social Mining Big Data Интернет вещей</i>



Проблемы	Данные	Представление данных
Структурированные, хорошо выраженные количественно	Структурированные (табличные)	Таблицы «объект – свойство», «свойство-свойство», представленные в электронных таблицах и реляционных базах данных
Слабоструктурированные, выраженные как количественно, так и качественно	Слабоструктурированные (блоги, сенсорные данные и т.д.)	Сложные форматы, использующие нереляционные платформы, например, mongodb, hadoop – реализующие концепцию распределённых (параллельных) вычислений MapReduce
Неструктурированные, выраженные в основном качественно	Неструктурированные (тексты, видео- и аудиофайлы)	





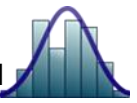
ОБЩЕСТВА
(как и личности)
МОГУТ ЗАНИМАТЬСЯ ЛИШЬ ОГРАНИЧЕННЫМ
ЧИСЛОМ ВОПРОСОВ

7 ± 2

«**Мегатренды**» – система координат в которой
можно объяснить явления этого мира на любом
уровне (личности, организации, общества)

[PCMovies95.xls](#)

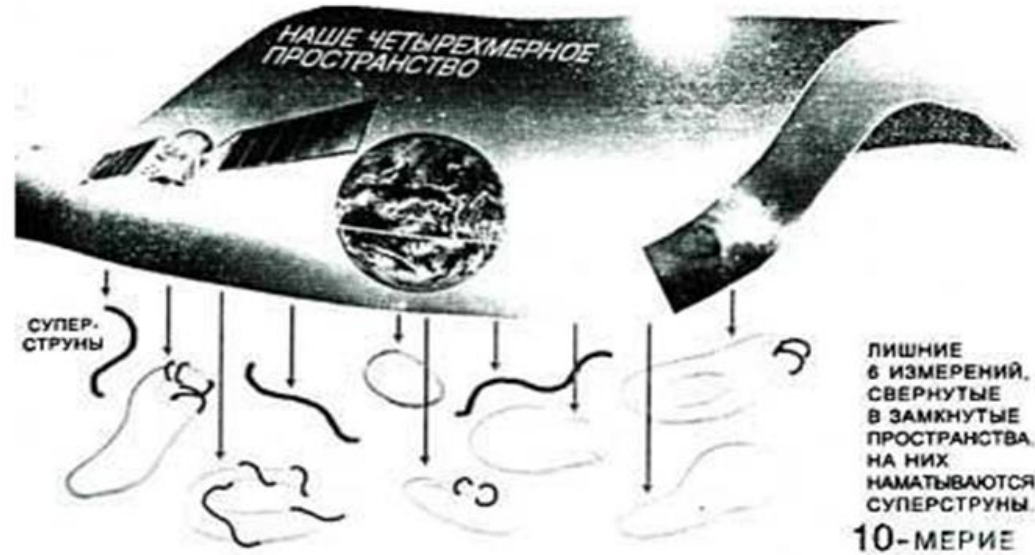
Дж. Нейсбит



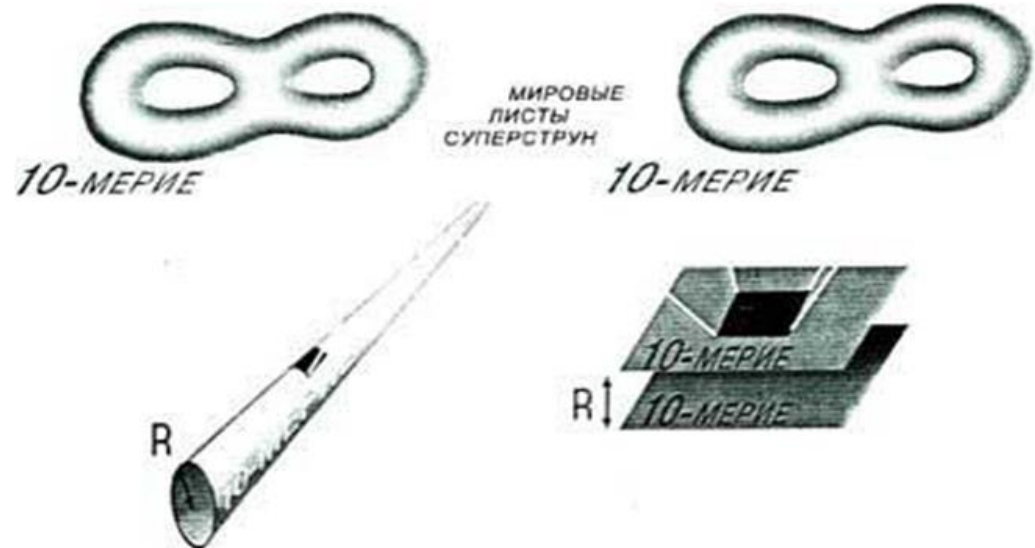
«Физический аналог» особенностей нашего восприятия - теория суперструн

7 ± 2

Социум движется в многомерном пространстве и прошлые мегатренды «сворачиваются» вдоль новых ...



ЛИБО



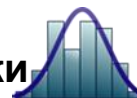


Мегатренды развития общества
(к устойчивому существованию в рамках
которых должно стремиться общество)
определены тысячелетия назад –
это десять заповедей.

Ю.И. Бершицкий

- 1) Постепенно теряется различие между реальным и виртуальным.
- 2) Приоритет в жизни получают «физические» люди, предпочитающие быстрые решения.
- 3) Информационные технологии, смещают общество в сферу услуг.
- 4) Информационные технологии – предмет быта.
- 5) Люди платят и получают деньги за виртуальные действия.
- 6) Современные аналитики среднего звена занимаются увеличением посещаемости рекламного сайта компании.
- 7) Особое внимание уделяется концепции *Big Data* в рамках изучения социальных сетей, промышленности, бизнеса.
- 8) Мы живем в обществе потребления, ориентированном на «символы» и «таблоиды».
- 9) *Переход к информационному обществу требует избавиться от устаревшей системы образования, которая готовила кадры для индустриального общества.*
- 10) Приобретает актуальность реальный сектор экономики.

«Мегатренды» 1-9 – наша среда, отравленная IT (Дж.Нейсбит)



«Мегатренды»

формируют **социально-экономическое пространство**, в котором каждый индивидуум в результате лексикографического упорядочивания выбирает оси координат в качестве контекста для оценки того, или иного события.



Чёрный лебедь, птица из рода лебедей (Cygnus) семейства утиных (Anatidae). До XVIII века была неизвестна науке.

Реализуется вероятностная модель языка:

$p(x/\mu)$, где x – событие, а μ – рассматриваемый контекст, $p(\mu)$ – плотность вероятности смысла слова или «тенденция к осуществлению смысла слова», $p(x/\mu)$ – функция правдоподобия события x в контексте смысла слова μ , соответствующего, выбранной оси координат (В.В. Налимов).

Целью социально ориентированного общества является «антихрупкое» (Н. Талеб) **управление** (учитывающее риск и негативные последствия как для себя, так и для других) **формированием социально-экономического пространства и соответствующей системы ценностей для осознания и оценки явлений окружающего мира.**



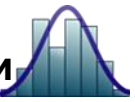
В качестве эксперимента по составлению социального портрета интернет-сообщества посредством выявления его мегатрендов магистрантами ФПИ Юрием Михалевичем и Еленой Шкуропат была разработана программа, реализующая алгоритмы *Web Mining* и *Text Mining*, производящая сбор данных сообщества социальной сети *vk.com* и последующих их анализ.



Обработка данных в разработанной программе включает в себя три этапа:

- загрузка сообщений и комментариев из указанного сообщества социальной сети вконтакте (также загружаются дата публикации и, в случае, если эта информация опубликована, идентификатор автора);
- формирование структурированной базы данных сообщений и комментариев (используется СУБД *MongoDB*);
- частотный анализ слов; формирование коллекций в базе данных, содержащих информацию о частоте встречи конкретных слов.

Таким образом, разработанная программа реализует практически все этапы анализа. Завершающим этапом является ручной разбор данных о частоте встречаемости групп и словосочетаний. На основании этого анализа можно выделить основные тренды, присущие данному сообществу.



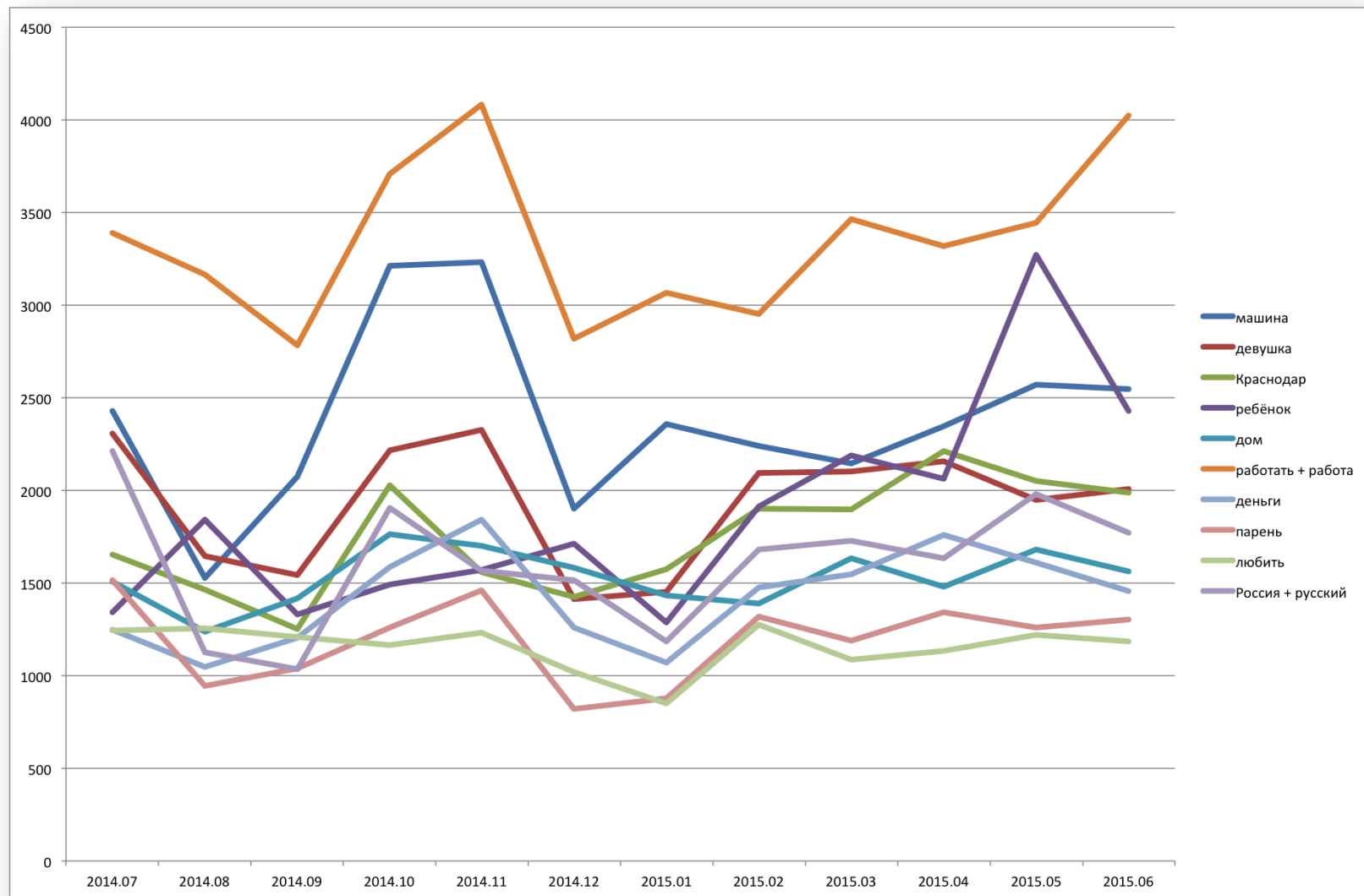
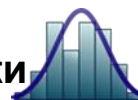
Для корректного осуществления частотного анализа были произведены следующие манипуляции:

- исключены из текста знаки препинания, лишние пробелы, табуляции, переводы строк, смайлики;
- сформированы словари «мусорных» (не несущих полезной нагрузки) слов (местоимения, союзы), эти слова необходимо исключить из статистической выборки; однако, имеет смысл включить их в словосочетания;
- произведён морфологический разбор анализируемых слов; при осуществлении анализа были рассмотрены «нормализованные» версии слов.

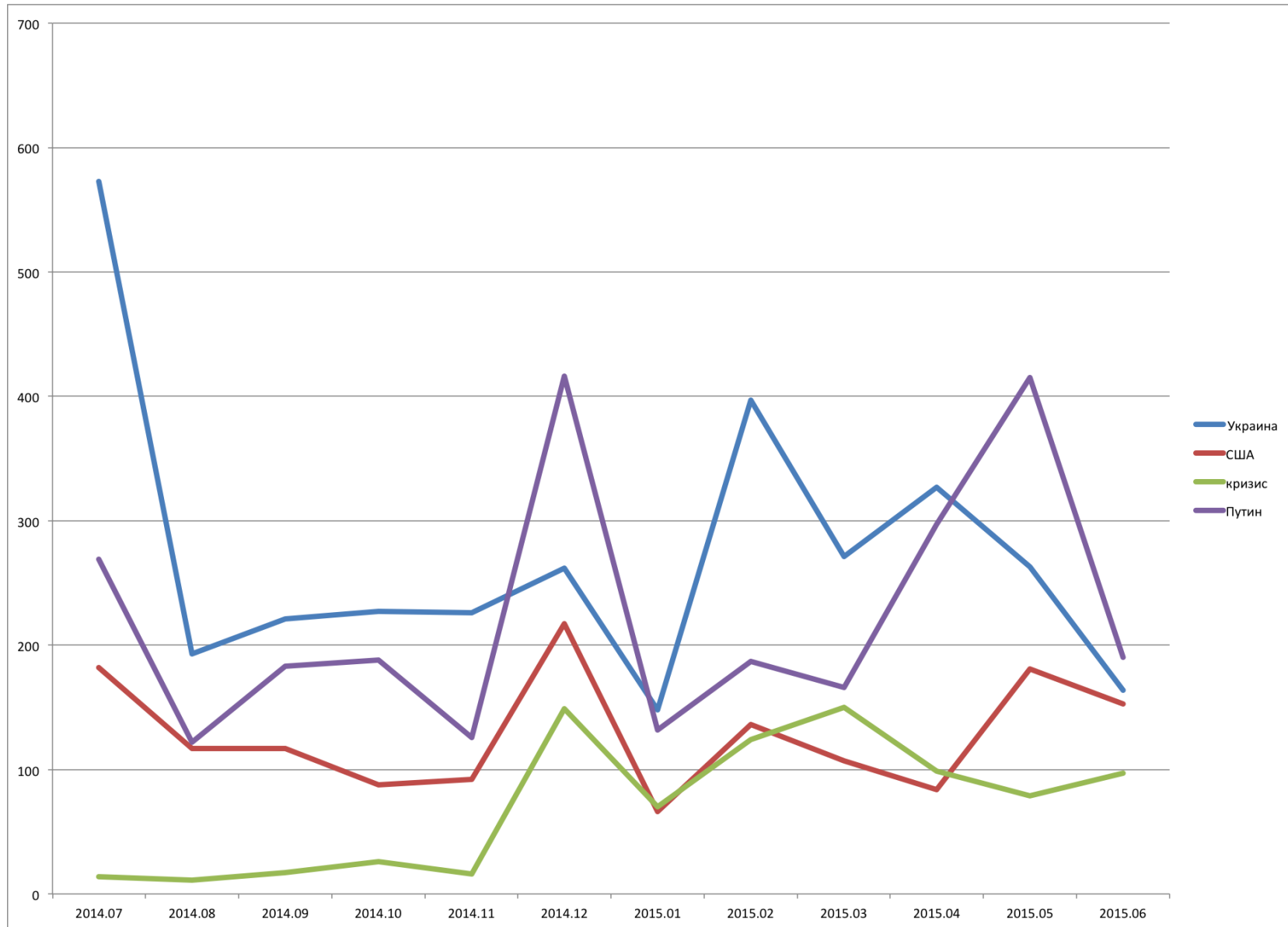
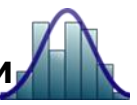


По результатам анализа выявлены следующие «мегатренды» (в скобках указано количество упоминаний в за ~3,5 года существования сообщества):

- работать + работа (72292);
- машина (55994);
- девушка (50610);
- Россия + русский (50257);
- Краснодар (42838);
- ребёнок (40553);
- дом (36635);
- деньги (33345);
- парень (30305);
- любить (28924).



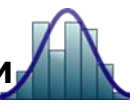
В динамике трендов сообщества «Типичный Краснодар» не прослеживается аномалий (за исключением любопытного пика тренда «ребёнок» в мае 2015 года), из чего можно предположить, что изменение трендов сообщества находится в соответствии с изменением активности сообщества в течении рассматриваемого периода.



Динамика трендов «Украина», «США», «кризис», «Путин» более любопытна и отвечает происходившим политическим событиям

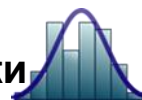


2016-2017 гг.



.....
Наш рассказ о событиях, развернувшихся в основаниях математики в XX в. (**анализа данных в начале 21 века** ...), мы хотели бы закончить следующей притчей. На берегах Рейна в течение многих веков возвышался прекрасный замок. Пауки, обитавшие в подвалах замка, затянули паутиной все его проходы. Однажды сильный порыв ветра разрушил тончайшие нити паутины, и пауки принялись поспешно восстанавливать образовавшиеся бреши: они считали, что замок держится на их паутине!

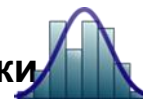
(М. Клайн, Математика.
Утрата определенности.)



Жизнь человека в информационном обществе, сопряженная с большими потоками информации, быстрыми изменениями и пр. **требует** от человека **выделять главные факторы («мегатренды»)** посредством использования *IT, наблюдений «в поле» и использовать их в своей жизни, не теряя при этом человечности и понимания информационных технологий* сквозь призму человеческого бытия, времени, религии и искусства **как дополнения к жизни, обогащенной близостью к природе, физическим трудом, общением с друзьями.**

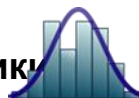
Дж. Нейсбит.

Высокая технология, глубокая гуманность: Технологии и наши поиски смысла



ДОКЛАД ОКОНЧЕН,

СПАСИБО ЗА ВНИМАНИЕ!



**Перечень системных задач лексикографически упорядочиваемых в каждой из наук,
изучающих *целеориентированные системы***

№	Задачи	Результат решения задачи
1	Наблюдаемости	Возможность определения состояния системы по наблюдениям
2	Идентификации	Выявление существенных свойств и характеристик объекта и их взаимосвязи
3	Реализации	Построение компактной модели, согласующейся с реальными данными
4	Управляемости	Возможность влияния на выход системы по входным воздействиям
5	Прогнозирования	Предсказания дальнейшего развития и (или) состояния системы на основе математических моделей и (или) интуиции
6	Оптимизации	Поиск значений переменных, обеспечивающих максимум (минимум) целевой функции
7	Связности	Изучение и проектирование структуры системы с использованием теории графов и топологии
8	Сложности	Возможность описать (построить удовлетворительную модель) за приемлемое время (один из основных подходов)
9	Композиции и декомпозиции	Обеспечение адекватного перехода между уровнями описания системы
10	Самоорганизации	Самосоздание новых устойчивых структур в системах с управляемыми элементами
11	Устойчивости	Возможность сохранения структуры и поведения независимо от возмущающих воздействий
12	Чувствительности	Способность ограниченной реакции системы на возмущения
13	Разнообразия	Качественно-количественная характеристика системы и связей в ней, обеспечивающая самоорганизацию, некоторое множество выходов и входов, устойчивость и прогнозирование (кризисное, посткризисное)
14	Адаптируемости	Способность восприятия системой внешних воздействий без необратимых последствий в поведении путём изменений параметров и, возможно, структур системы
15	Катастроф (кризиса)	Как одна из форм движения системы: изменение-развитие-эволюция
16	Принятия решения	Выбор одной из возможных альтернатив для достижения цели