

Омский государственный медицинский университет  
Управление Роспотребнадзора в Омской области



**О ВНЕДРЕНИИ В ПРАКТИКУ НОВЫХ  
ПРИЕМОМ АНАЛИЗА И ПРОГНОЗА  
САНИТАРНО-  
ЭПИДЕМИОЛОГИЧЕСКОГО  
БЛАГОПОЛУЧИЯ НАСЕЛЕНИЯ**

**Ж.В. Гудинова, Е.И. Толькова, Г.Н. Жернакова,  
Е.Л. Овчинникова, И.В. Гегечкори**

г. Пермь, 13-15 мая 2015 года

# ОСНОВНОЕ НАПРАВЛЕНИЕ НАУЧНОЙ РАБОТЫ кафедры общей гигиены с курсом ГДП Омского ГМУ – СГМ



- «Совершенствование методологии причинно-следственных связей в гигиене и социально-гигиеническом мониторинге» (под руководством зав. каф. Ж.В. Гудиновой)
- Исследования ведутся с начала 2000-х годов
- Защищено 3 докторских и 4 кандидатских диссертации
- Изучались: инвалидность детей, методология анализа медико-демографических и социально-экономических показателей, климат и погода, профессиональная заболеваемость, сифилис и гонорея у детей
- Разработано и внедрено на федеральном уровне 2 методических документа
- Имеется опыт преподавания основ СГМ студентам и курсантам ФУВ

# ПРОБЛЕМЫ

Анализ баз данных с целью извлечения из них новой и полезной информации – суть и методологическая основа СГМ

**ПРОБЛЕМА 1** – методическая «бедность» традиционной аналитической практики, сочетание космической сложности задач СГМ, имеющих объемных баз данных ФИФ и РИФ в отсутствие адекватного методического аппарата.

**ПРОБЛЕМА 2** – трудности интерпретации и понимания результатов анализа, в т.ч. лицами, принимающими решения.

# АНАЛИТИЧЕСКИЙ АППАРАТ СГМ

1. Методы статистического анализа данных
2. Методы Data Mining
3. Методы прогнозирования
4. Оценка риска как специфический гигиенический метод оценки ситуации и ее прогноза

# 1. МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

описательная статистика

метод перцентилей

методы определения статистической нормы

методы сравнения групп

особые приемы межрегиональных сравнений  
(картографирование, перцентиль-профиль)

анализ динамических рядов

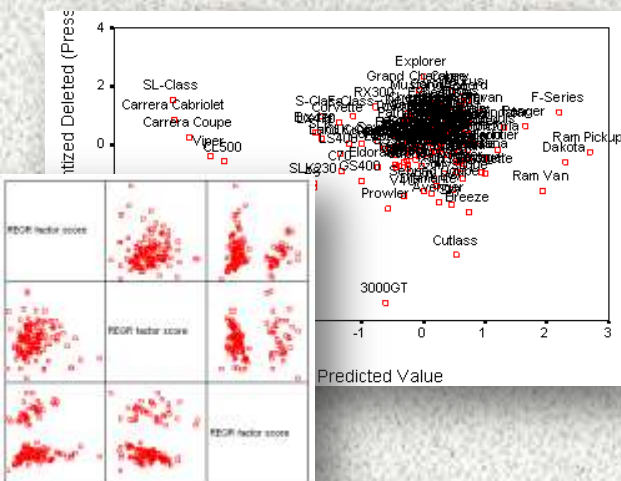
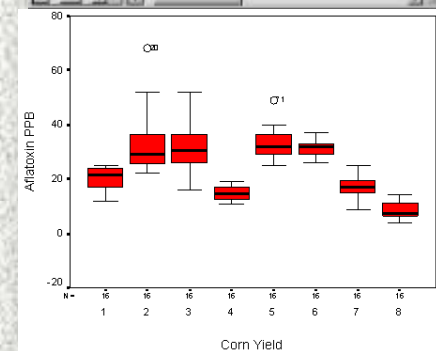
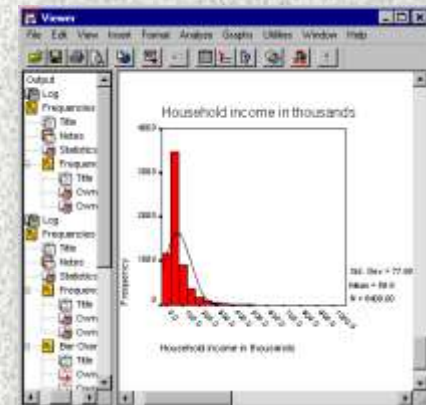
корреляционно-регрессионный анализ

факторный, компонентный анализ

кластерный анализ

множественная регрессия

канонический анализ и др.





# О ПРИМЕНЕНИИ УКАЗАННЫХ МЕТОДОВ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ В ПРАКТИКЕ СГМ

ФЕДЕРАЛЬНАЯ СЛУЖБА ПО НАДЗОРУ В СФЕРЕ ЗАЩИТЫ ПРАВ  
ПОТРЕБИТЕЛЕЙ И БЛАГОПОЛУЧИЯ ЧЕЛОВЕКА

## 2004 год

**2.1.9. СОСТОЯНИЕ ЗДОРОВЬЯ НАСЕЛЕНИЯ В СВЯЗИ С СОСТОЯНИЕМ  
ОКРУЖАЮЩЕЙ ПРИРОДНОЙ СРЕДЫ И УСЛОВИЯМИ ПРОЖИВАНИЯ  
НАСЕЛЕНИЯ**

**СОЦИАЛЬНО - ГИГИЕНИЧЕСКИЙ МОНИТОРИНГ.**

**ИНВАЛИДНОСТЬ ДЕТЕЙ.**

**СБОР, ОБРАБОТКА И АНАЛИЗ ПОКАЗАТЕЛЕЙ.**

Методические рекомендации  
ФЦ/3718 от 01.12.2004



ФЕДЕРАЛЬНАЯ СЛУЖБА ПО НАДЗОРУ В СФЕРЕ ЗАЩИТЫ  
ПРАВ ПОТРЕБИТЕЛЕЙ И БЛАГОПОЛУЧИЯ ЧЕЛОВЕКА

[Документы](#) / [Письма](#)

Приказ №341 от 20.09.2010

## Об утверждении методических рекомендаций по социально- гигиеническому мониторингу

В соответствии с Планом основных организационных мероприятий Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека на 2010 год, в целях совершенствования государственной системы социально-гигиенического мониторинга, п  
р и к а з ы в а ю:

1. Утвердить методические рекомендации «Социально-гигиенический мониторинг. Анализ демографической и социально-экономических показателей на региональном уровне» (приложение).
2. Руководителям Управлений Роспотребнадзора по субъектам Российской Федерации, главным врачам ФГУЗ «Центр гигиены и эпидемиологии в субъектах Российской Федерации» использовать методические рекомендации при проведении социально-гигиенического мониторинга.
3. Контроль за исполнением приказа возложить на заместителя руководителя Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека Н.В. Шестопалова.

Руководитель

Г. Г. Онищенко

Файлы для загрузки:

[Приложение](#) (1425 КБ)

## приказ Г.Г. Онищенко № 341 от 20 сентября 2010 года

# ПРИМЕРНАЯ ПОШАГОВАЯ ТЕХНОЛОГИЯ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ В СГМ

5. Сбор данных. Формирование базы данных.

6. Описательная статистика.

7. Сравнение данных. Методы сравнения групп.

8. Сравнение данных. Картографический анализ.

9. Сравнение данных. Перцентиль-профиль.

11. Анализ динамических рядов. Прогнозирование.

12. Методы классификации. Кластерный анализ.

13. Причинно-следственные связи. Корреляционно-регрессионный анализ.

14. Метод сокращения числа переменных. Факторный анализ.

Фрагмент оглавления МР «СГМ. Анализ медико-демографических и социально-экономических показателей ...» (приказ №345 от 2010 г., ГАРАНТ)

# О МЕТОДАХ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ



# Фрагмент базы данных «Показатели здоровья населения и среды обитания в районах Омской области»

STATISTICA - [Данные: база\_грюнер (22x by 33c)]

Times New Roman 11 В I U Add to Workbook Добавление в отчет

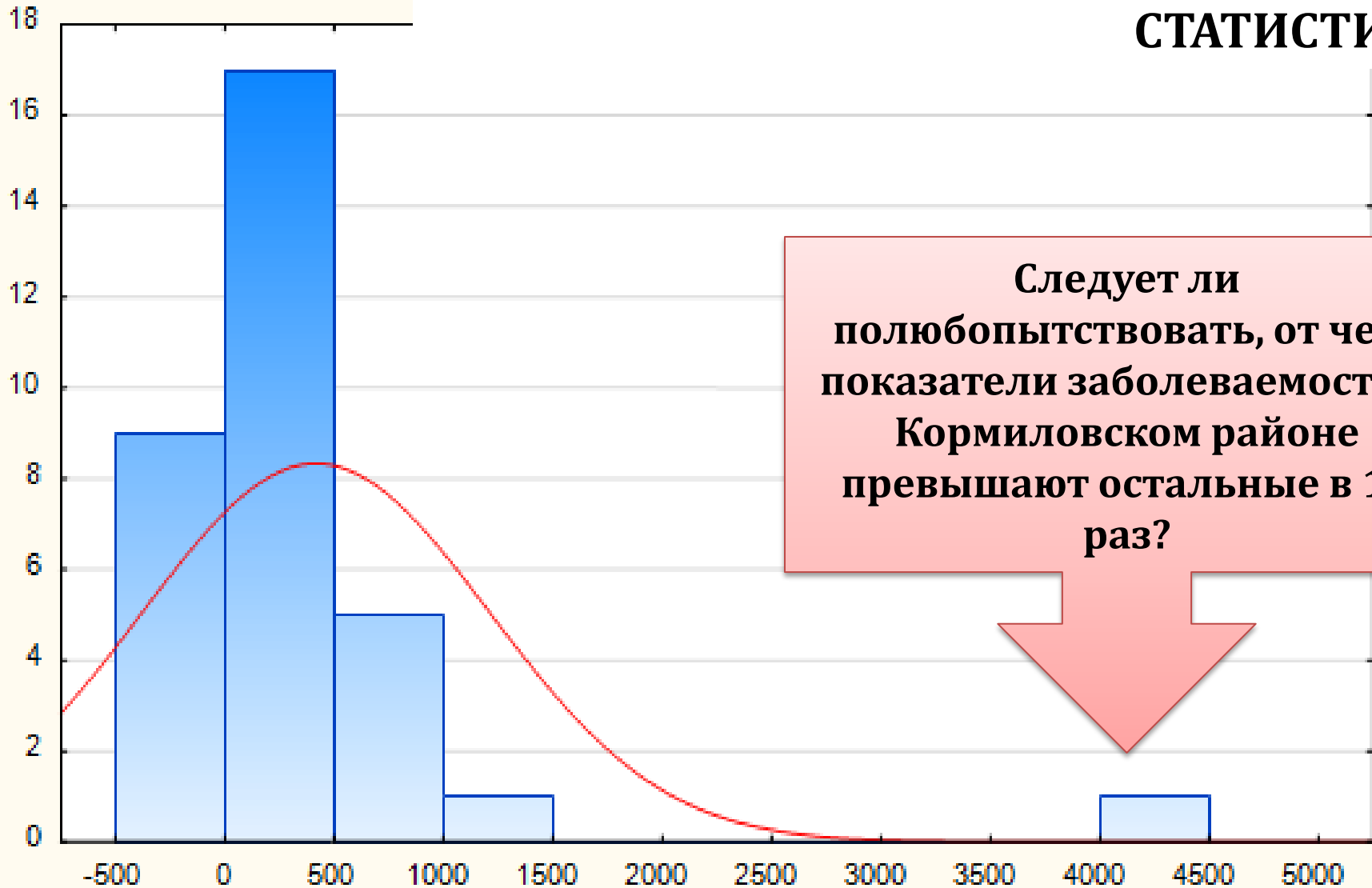
Файл Редактирование Просмотр Вставка Формат Статистика Графы Инструменты Данные Окно Справка

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	геогра	Об заб	Об заб	Об ИД	Млад с	Рожда	Смерт	Плотнс	Ср МП	Врачи	Койки	Амб пс	Зарпла	Произе	Дс
Большеуковский	1	109884	84151	257,7	17,9	10	14	1	112	20	113	6,1	2212	133	
Знаменский	1	63720	48217	264	18,4	12	14	4	119	22	116	5,4	2289	142	
Седельниковский	1	95901	73217	242	17,8	11	14	3	109	21	99	4,9	2100	79	
Тарский	1	104962	67268	337	13,5	10	17	3	138	30	132	6,5	2540	144	
Тевризский	1	85598	65619	225	15,4	12	14	2	88	20	106	4,9	2358	137	
Усть-Ишимский	1	86814	64726	228	18,1	13	16	2	92	16	140	4,6	2134	165	
Большереченский	2	116693	91808	298	10,6	10	14	8	88	22	82	6,2	2168	118	
Горьковский	2	81395	52143	242	12,6	11	13	9	86	18	118	8,1	2051	68	
Колосовский	2	46234	27654	245	18,4	11	14	4	96	20	113	6,2	1747	76	
<b>Крутинский</b>	2	119130	65368	162	16,8	11	15	4	102	17	108	5,9	2126	93	
Муромцевский	2	126184	62269	237	10,0	9	16	4	98	22	100	8,1	1976	93	
Называевский	2	90427	71731	202	13,1	12	14	6	79	18	97	6,0	2392	152	
Нижнеомский	2	65244	84413	172	19,1	12	13	6	90	18	94	4,5	1880	56	
Саргатский	2	102183	111670	281	13,5	11	14	6	93	25	104	6,7	1996	90	
Тюкалинский	2	93205	87855	228	16,3	11	14	5	77	16	107	5,8	2249	119	
Азовский	3	127514	110374	167	12,0	14	10	15	72	23	70	6,4	2665	16	
Исилькульский	3	99567	82080	177	17,8	14	13	18	86	22	102	6,2	2389	124	
Калачинский	3	105609	84158	222	13,9	10	14	17	84	24	90	7,1	2852	115	
Кормиловский	3	102439	113155	236	13,8	11	13	14	97	27	95	7,7	2555	86	
Любинский	3	134415	97963	324	14,2	12	14	13	75	20	85	6,4	2228	81	
Марьяновский	3	98375	68949	182	15,6	13	14	17	84	21	96	6,0	2246	116	
Москаленский	3	103869	98772	185	16,1	15	12	14	80	22	83	6,1	1897	86	
Омский	3	131755	99265	200	14,3	11	13	26	56	21	51	5,6	3334	65	

Показатели регионального  
информационного фонда СГМ

Число районов

# О НЕОБХОДИМОСТИ ИСПОЛЬЗОВАНИЯ ГИСТОГРАММЫ В ОПИСАТЕЛЬНОЙ СТАТИСТИКЕ

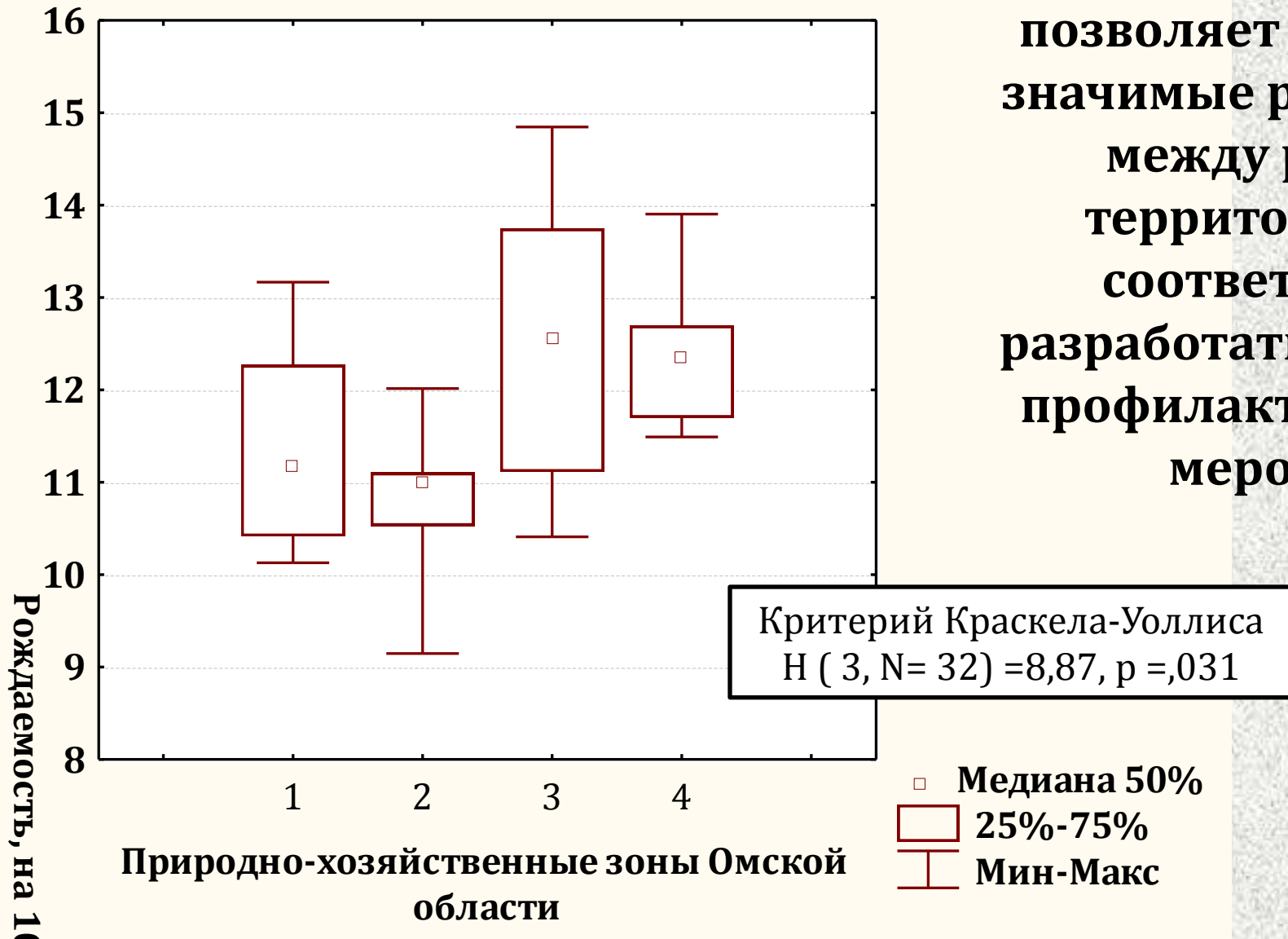


Автор: Ж. В. Гудинова, 2013

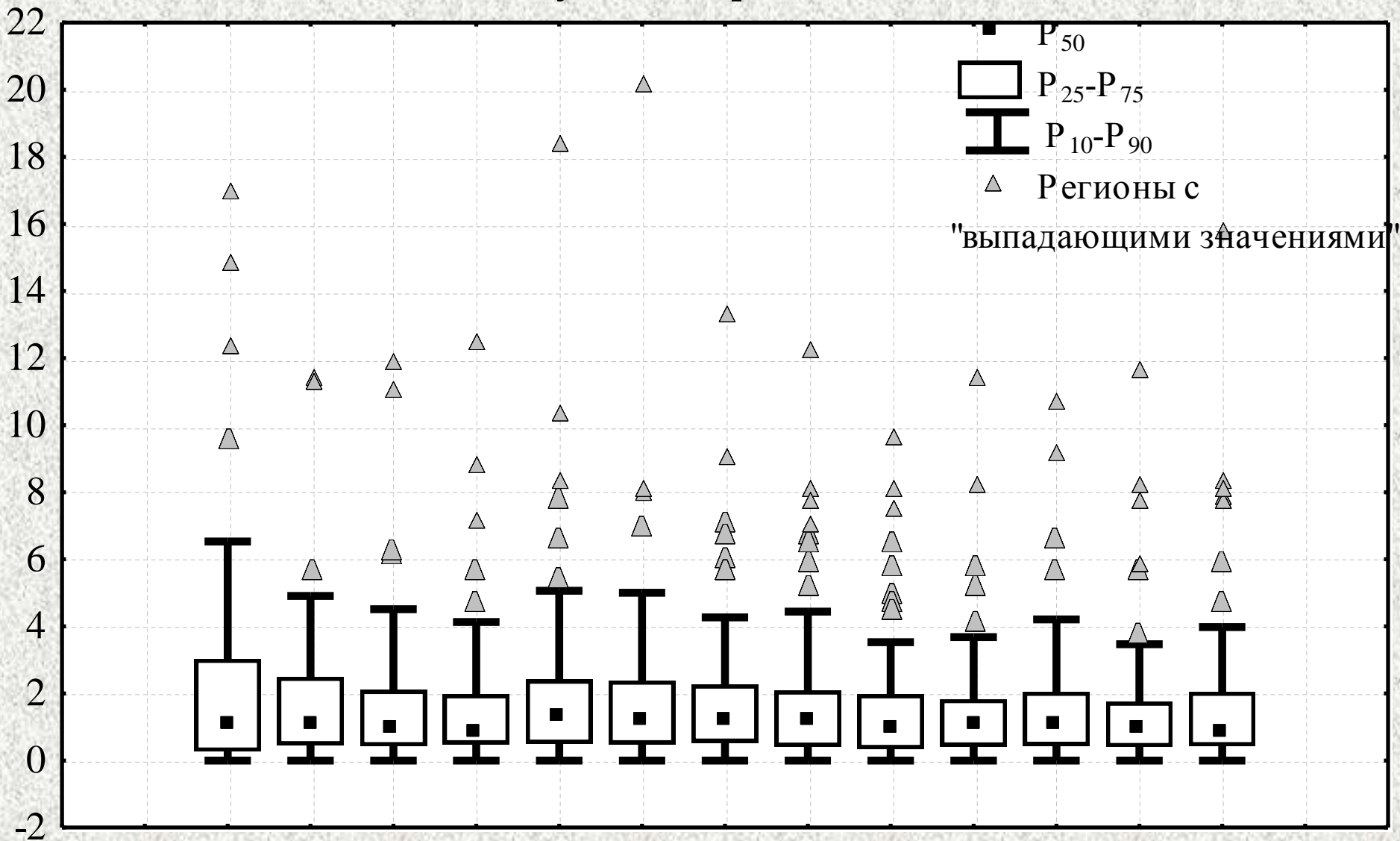
Заболеваемость, на 10 000 подростков

# Сравнение групп – важная часть анализа данных в СГМ

позволяет выявить значимые различия между разными территориями и, соответственно, разработать разные профилактические мероприятия



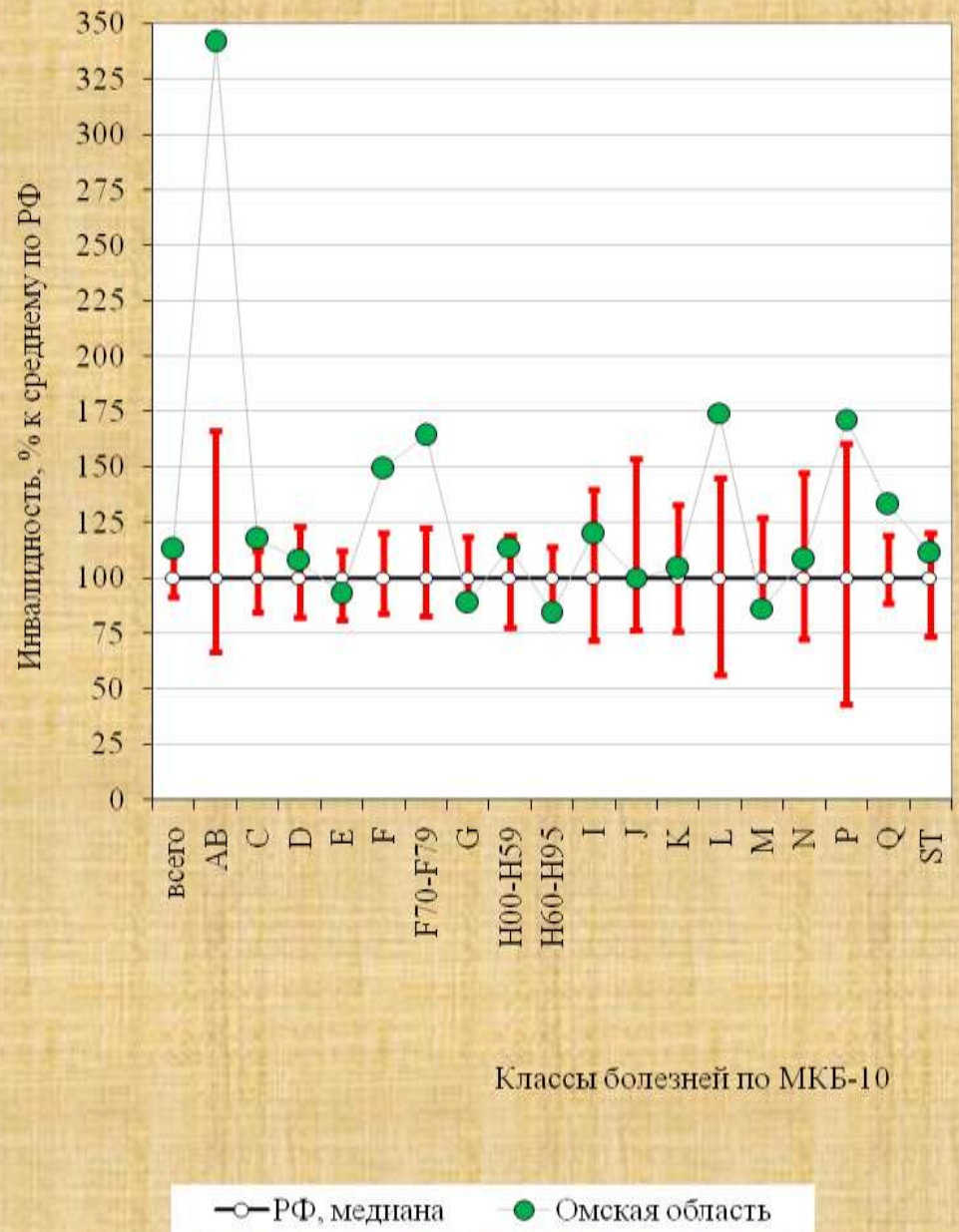
# МЕТОД ПЕРСЕНТИЛЕЙ на примере профессиональной заболеваемости в совокупности регионов России в 1997-2009



# ПЕРСЕНТИЛЬ-ПРОФИЛЬ

## Выводы по размерам границ статистической нормы (штрихи красного цвета):

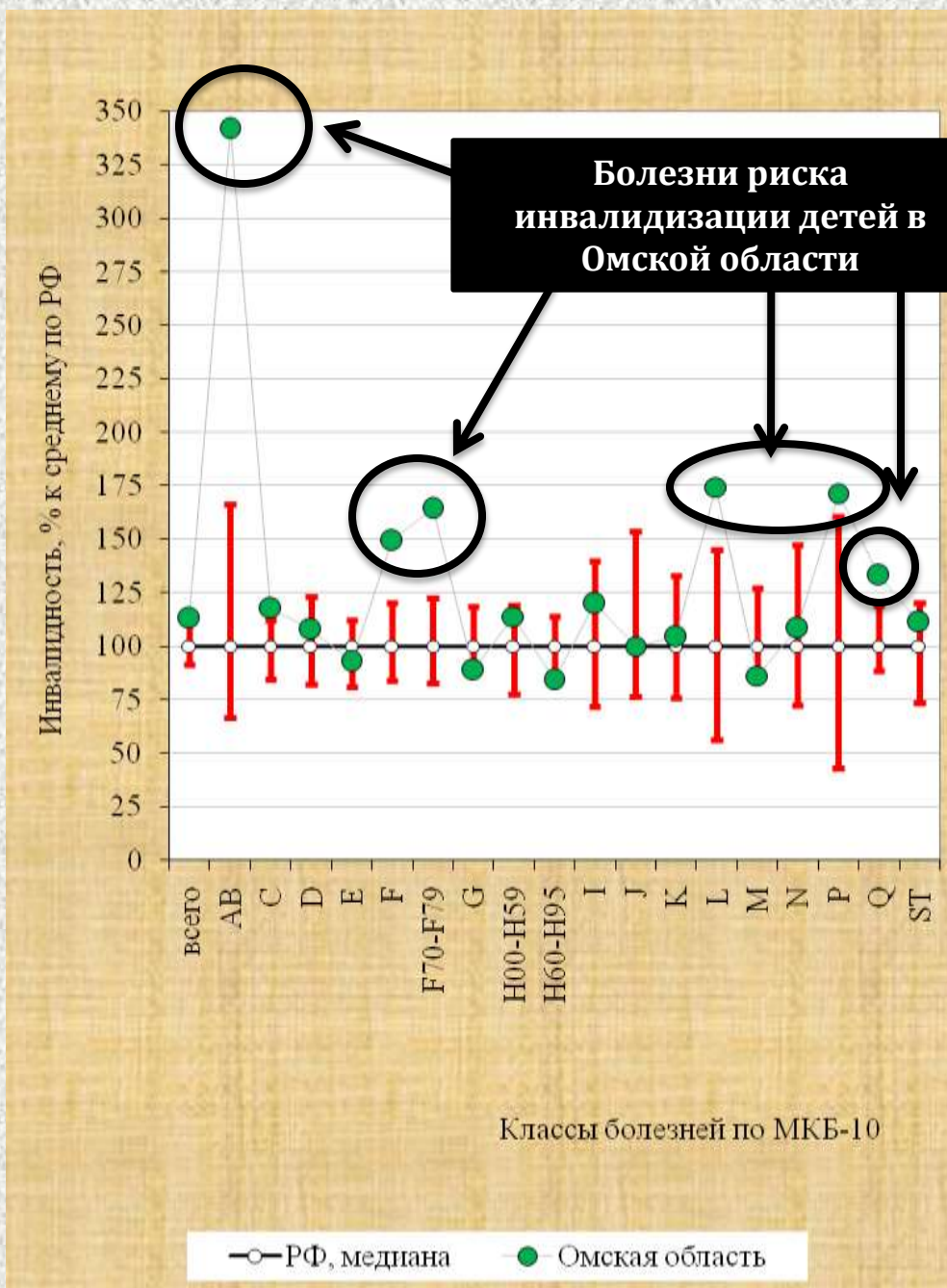
1. В первоочередном исследовании и уточнении нуждаются показания к определению инвалидности по классам инфекционных болезней (АВ), отдельных состояний в перинатальном периоде (Р), болезней кожи (L), и некоторых других.
2. Если показания четкие и определенные, то, очевидно, причину разброса данных следует искать в средовых влияниях.



# Персентиль-профиль

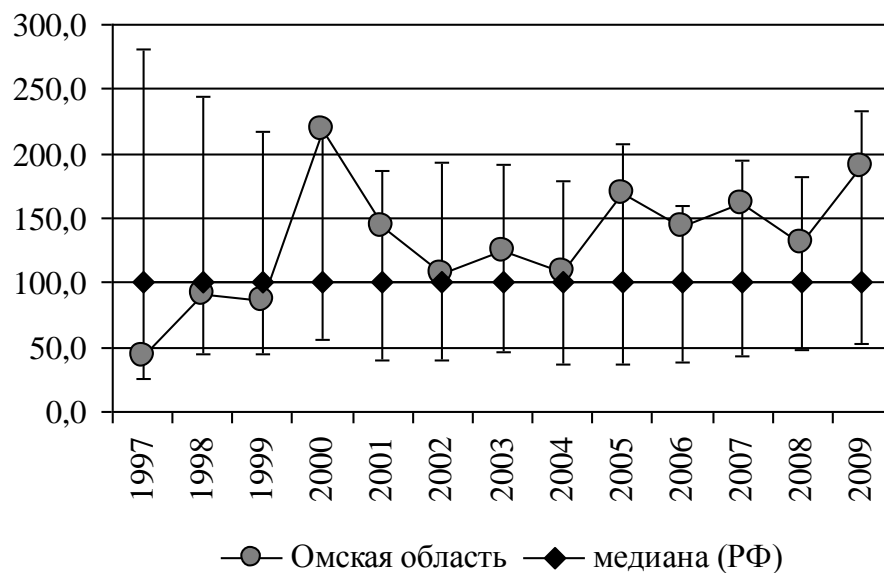
Выводы:

3. Для Омской области патологией риска инвалидизации детей являются инфекционные болезни (АВ), психические расстройства (F) и умственная отсталость (F70-F79), болезни кожи (L), отдельные состояния, возникающие в перинатальном периоде (P) и врожденные пороки развития (Q).
4. Эти болезни нуждаются в первоочередном исследовании и профилактике
5. Эти болезни (показатели) могут быть целевыми индикаторами долгосрочных программ Правительства.

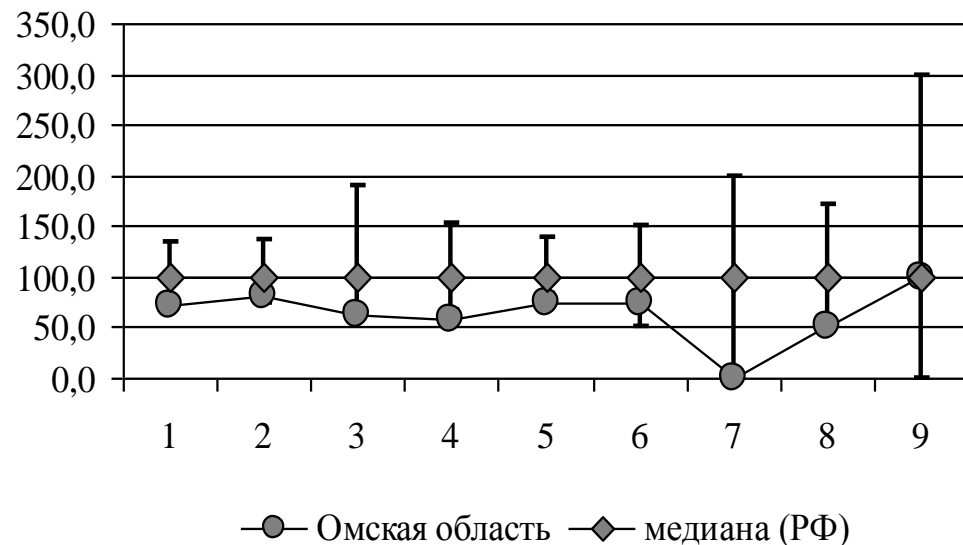


# Применение перцентиль-профиля в СГМ

## Профессиональная заболеваемость



## Условия труда



**Вывод: сравнение позиций Омской области по двум логически связанным между собой признакам вызывает вопросы по поводу высоких оценок условий труда (шум, излучения и проч.) в регионе в ряду других регионов страны**

# ПЕРСЕНТИЛЬ-ПРОФИЛЬ – МЕТОДИЧЕСКАЯ ОСНОВА СОЦИАЛЬНО- ГИГИЕНИЧЕСКИХ ПАСПОРТОВ РАЙОНОВ ОМСКОЙ ОБЛАСТИ

Паспортизация районов области проводилась Управлением Роспотребнадзора по Омской области.

Идея формирования социально-гигиенических паспортов предполагает емкое изложение комплексной систематизированной информации, которое сводится к перечню четко обозначенных проблем здоровья и среды обитания на конкретной территории.

Информация данного паспорта может использоваться муниципалитетами в Планах социально-экономического развития территории.

Автор: Ж. В. Гудинова, 2013



Разработку и ежегодную актуализацию Паспортов возможно реализовать при малых кадровых и материальных ресурсах.



# ВАЖНО

- Информация о СЭБН, как и информационный фонд СГМ представляют собой большой комплекс показателей
- Понятно, что и здоровье и среда характеризуется многими показателями, понятно, что отдельный показатель здоровья формируется под действием многих факторов среды
- Однако в традиционной практике анализа данных практически отсутствуют многомерные методы исследования (кластерный, факторный, компонентный виды анализа, множественная регрессия и проч.)

# Кластерный анализ

## КЛАССИФИКАЦИОННЫЕ ХАРАКТЕРИСТИКИ:

- 1) заболеваемость детей,
- 2) инвалидность детей,
- 3) смертность населения,
- 4) плотность населения,
- 5) обеспеченность врачами,
- 6) обеспеченность амбулаторной помощью,
- 7) производство,
- 8) дороги,
- 9) благоустройство жилья,
- 10) питьевая вода.



# Результаты канонической корреляции двух множеств ( $R_{can} = 0,75$ , $R^2_{can} = 0,57$ , $p = 0,002$ )

## Множество показателей «Здоровье населения»



## Множество показателей «Среда обитания»



# О методах Data Mining

# О необходимости Data mining в гигиене (СГМ)

- если есть объемные базы данных (как ФИФ), то их нужно анализировать
- нужны обоснованные управленческие решения, планирование деятельности и научной работы
- сейчас деятельность по анализу данных носит весьма субъективный, хаотичный, не целенаправленный характер: исследуются отдельные, чаще общепринятые, элементы системы «среда обитания – здоровье населения», не выделяются местные региональные проблемы в общероссийском ряду и, следовательно, не находят их решения.
- так же хаотично планируется и осуществляется научная работа, почему и справедливо критикуется

## DATA MINING (датамайнинг)

- Data Mining – это собирательное название, используемое для обозначения совокупности методов обнаружения в базах данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.
- Data Mining предполагает, что наука об извлечении содержания из массивов данных становится все более изощренной, а задачи, за которые берутся мастера такого поиска, – все более человечными.
- Термин введен Григорием Пятецким-Шапиро в 1989 году.

# Знания, получаемые в ходе DATA MINING

- **ранее не известные**, обязательно **новые** (а не подтверждающие какие-то ранее полученные сведения);
- **нетривиальные** — те, которые нельзя просто так увидеть (при непосредственном визуальном анализе данных или при вычислении простых статистических характеристик), неочевидные, неожиданные;
- **практически полезные** — те, которые представляют ценность для исследователя или потребителя;
- **доступные для интерпретации** — их можно легко представить в наглядной для пользователя форме и легко объяснить в терминах предметной области; кроме того, полученные знания должны быть логически объяснимы.

# Термины- синонимы:



1. добыча данных (Data Mining, датамайнинг)
2. интеллектуальный анализ
3. глубинный анализ
4. выявление знаний в базах данных
5. обнаружение полезных знаний в базах данных  
(а традиционные статистические показатели не всегда полезны – средняя температура по больнице)



# Различия статистического анализа и Data mining

**Традиционные статистические методы** предполагают некоторые априорные представления (гипотезу), а Data Mining предполагает обнаружение ранее неизвестных знаний (позволяет выдвигать гипотезы), причем очень быстро

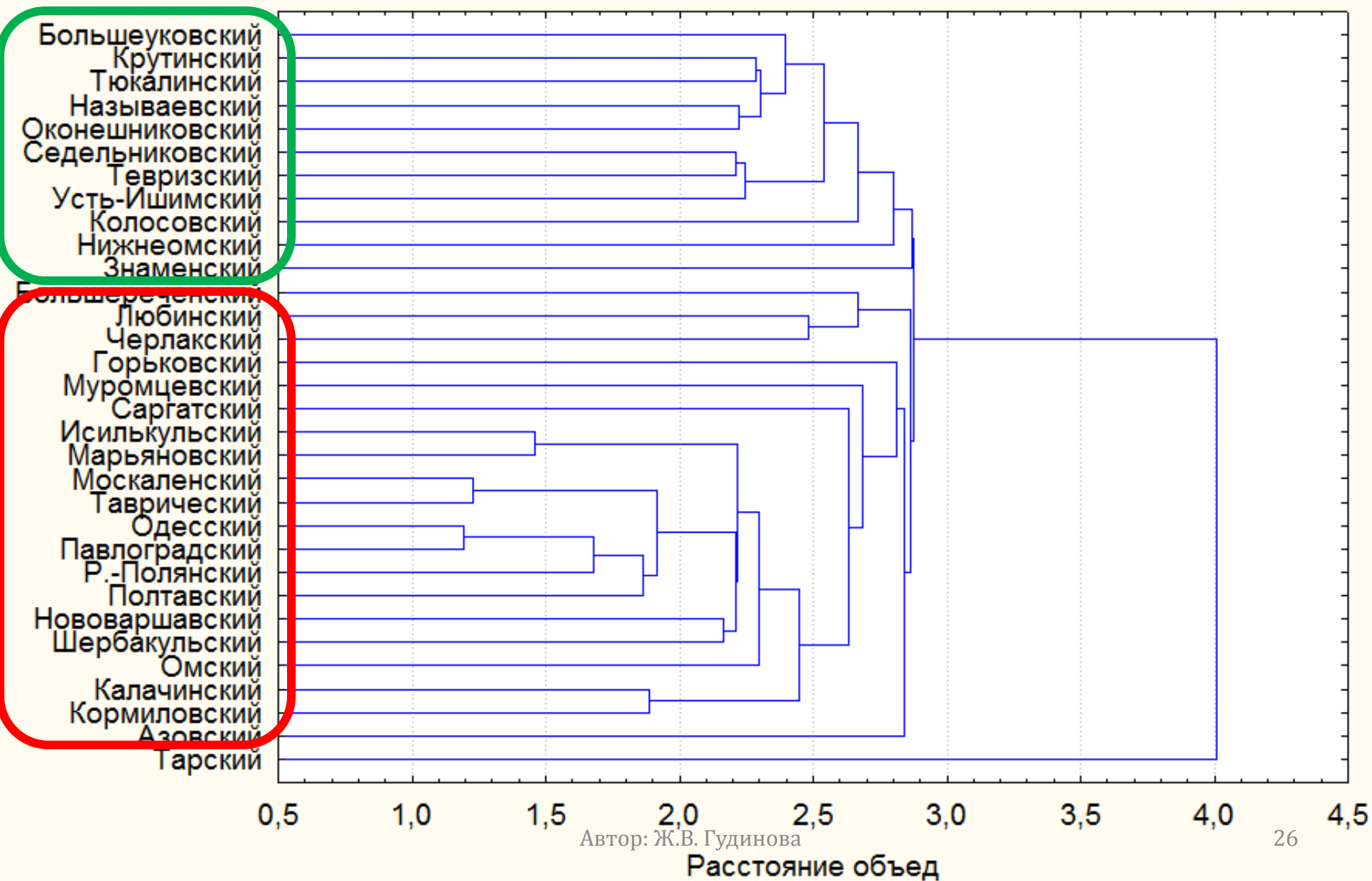
## СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА

- Описательная статистика, корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов

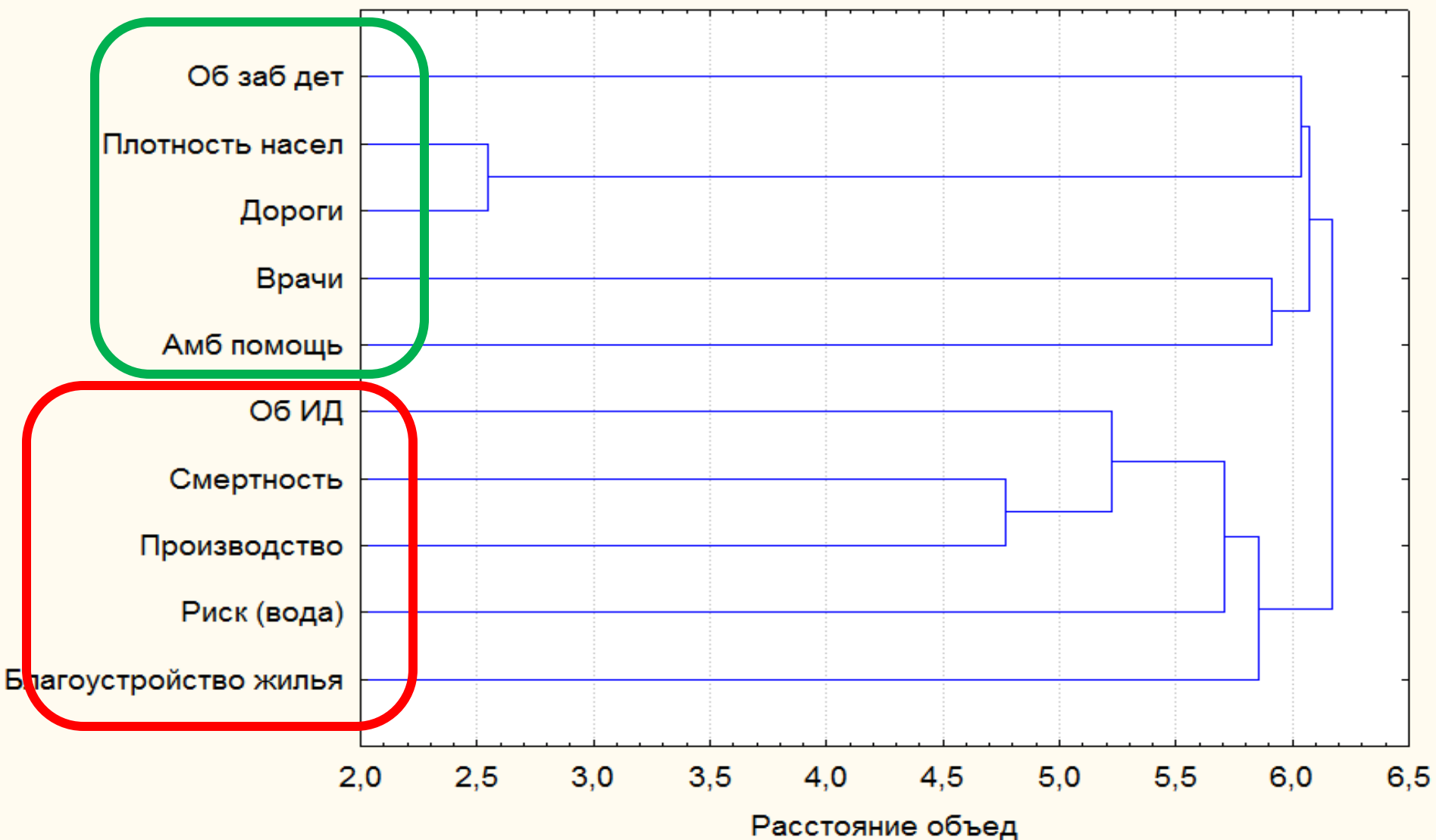
## МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

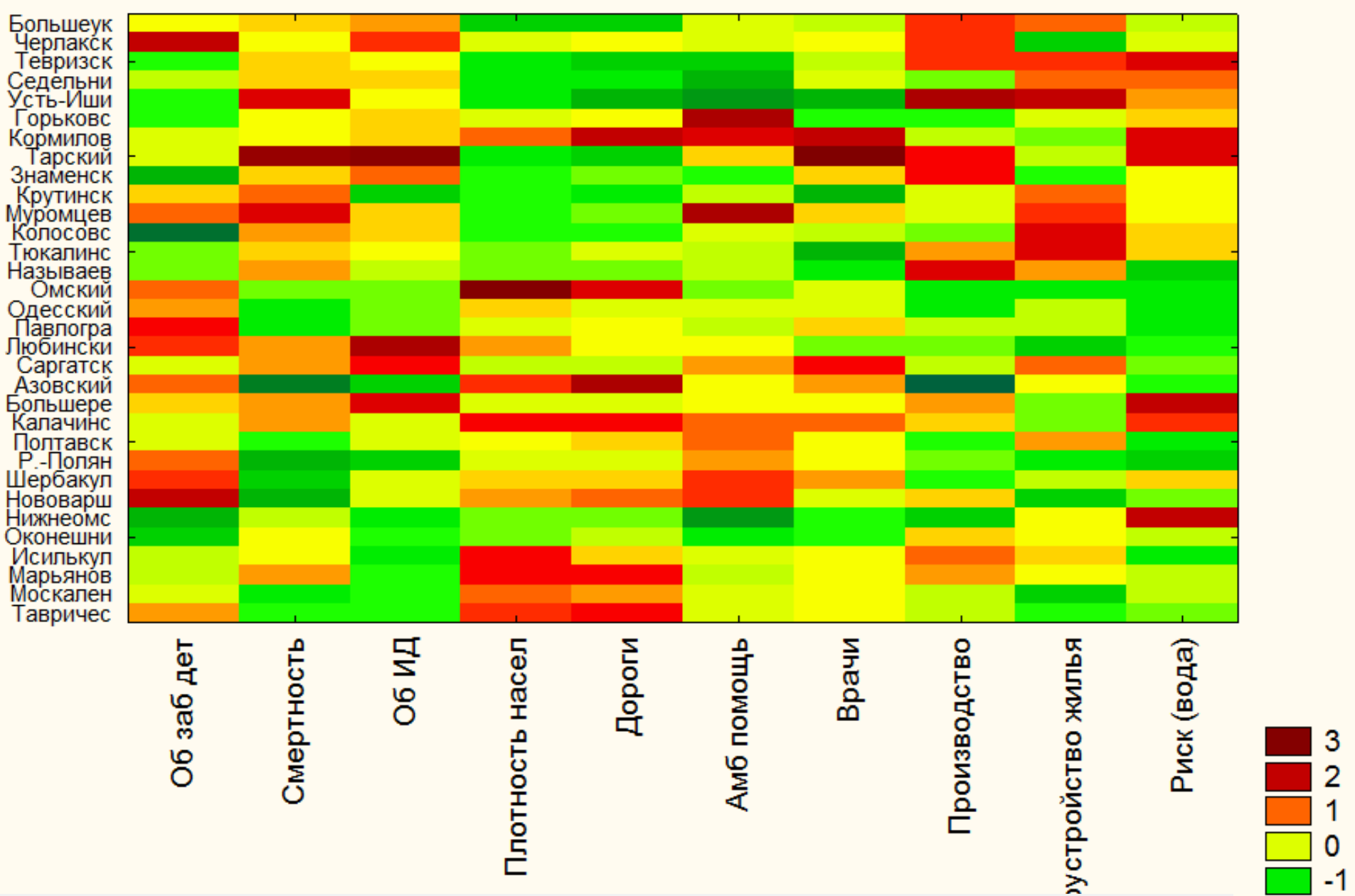
- Методы ассоциации, кластеризации, классификации, моделирования и прогнозирования, основанные на применении деревьев решений, искусственных нейронных сетей, генетических алгоритмов, эволюционного программирования, ассоциативной памяти, нечеткой логики

# РЕЗУЛЬТАТЫ КЛАСТЕРНОГО АНАЛИЗА: классификация районов Омской области по показателям СЭБН (по строкам)



# РЕЗУЛЬТАТЫ КЛАСТЕРНОГО АНАЛИЗА: классификация районов Омской области по показателям СЭБН (по столбцам)





# РЕЗУЛЬТАТЫ КЛАСТЕРНОГО АНАЛИЗА (диаграмма двухвходового объединения)

# О МЕТОДАХ ПРОГНОЗИРОВАНИЯ

# АКТУАЛЬНОСТЬ ВОПРОСА

- Очевидно высока – в виду определения и задач СГМ (анализ и прогнозирование)
- Очевидна сложность прогнозирования – высший пилотаж
- Поэтому специалистам необходимо овладеть рядом понятий в сфере прогностики
- **ПРОГНОЗНЫЙ ФОН** – это совокупность внешних по отношению к объекту прогнозирования условий, существенных для решения задачи прогноза (Рабочая книга по прогнозированию / Под ред. И.В. Бестужев-Лада. – М., Мысль, 1982)

# ВИДЫ ОБЪЕКТОВ ПРОГНОЗИРОВАНИЯ:



социальные

научно-технические

экономические

# **ВИДЫ ПРОГНОЗОВ:**



**по размеру**

**по содержанию**

**по периоду упреждения**

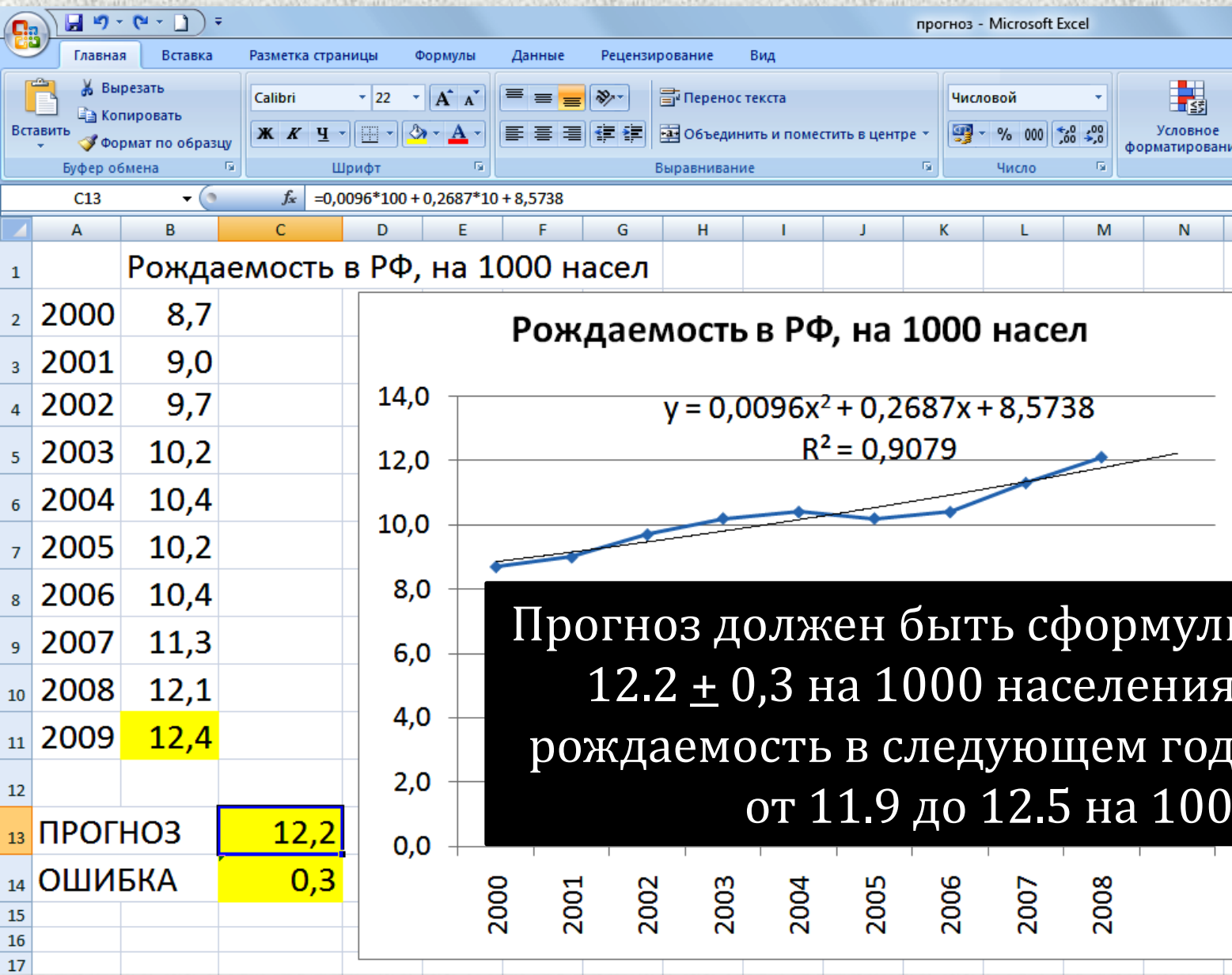


# Виды прогнозов: по размеру

**ТОЧЕЧНЫЙ** – прогноз, результат которого представлен в виде единственного значения характеристики объекта прогнозирования без указания доверительного интервала

**ИНТЕРВАЛЬНЫЙ** – прогноз, результат которого представлен в виде доверительного интервала характеристики объекта прогнозирования для заданной вероятности осуществления прогноза.

# ПРИМЕР РАСЧЕТА ИНТЕРВАЛЬНОГО ПРОГНОЗА

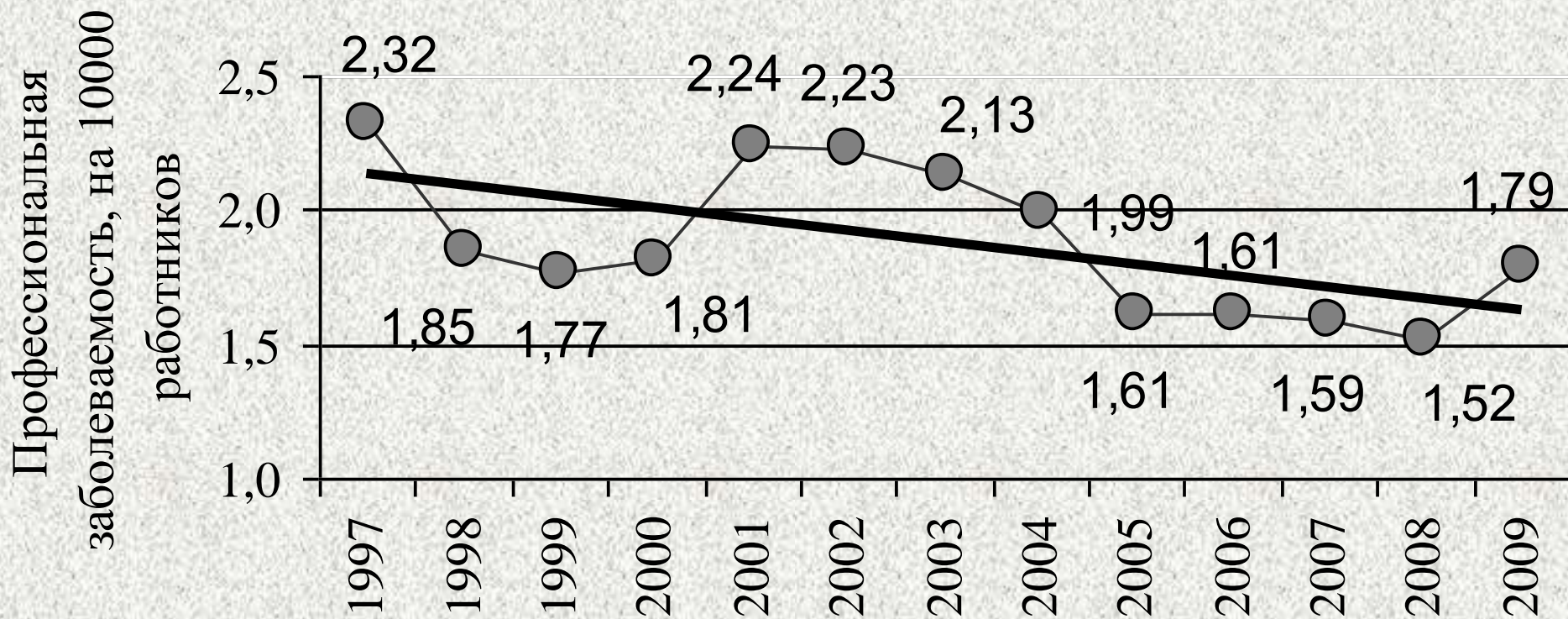


# Виды прогнозов: по содержанию

**ПОИСКОВЫЙ** – прогноз, содержанием которого является определение возможных состояний объекта прогнозирования в будущем.

**НОРМАТИВНЫЙ** – прогноз, содержанием которого является определение путей и сроков достижения возможных состояний объекта прогнозирования в будущем, принимаемых в качестве цели

# ПРИМЕР ПОИСКОВОГО ПРОГНОЗА профессиональной заболеваемости в РФ



(Жернакова Г.Н., Гудинова Ж.В.)

# ПРИМЕР НОРМАТИВНОГО ПРОГНОЗА профессиональной заболеваемости в зависимости от условий труда в регионах РФ

$$y = 0,86 + 0,22 * X_1 - 0,49 * X_2 + 0,07 * X_2^2 + 0,92 * X_3^2, \text{ где}$$

$X_1$  – удельный вес работающих под воздействием повышенной запыленности воздуха рабочей зоны

$X_2$  – удельный вес работающих, занятых тяжелым физическим трудом

$X_3$  – удельный вес работающих на оборудовании, не отвечающем требованиям безопасности

$$(R=0,75, R^2_{\text{скорр.}}=0,54, p=0,000)$$

(Жернакова Г.Н., Гудинова Ж.В.)

# Виды прогнозов: по периоду упреждения

- **ОПЕРАТИВНЫЙ** – прогноз до 1 месяца:.
- **КРАТКОСРОЧНЫЙ** – прогноз от 1 месяца до 1 года
- **СРЕДНЕСРОЧНЫЙ** – прогноз от 1 года до 5 лет
- **ДОЛГОСРОЧНЫЙ**– прогноз от 5 лет до 15 лет
- **ДАЛЬНЕСРОЧНЫЙ** – прогноз свыше 15 лет

# ЭКСПЕРТНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ



Экспертные методы прогнозирования базируются на экспертной информации (мнении экспертов).

Выделяют методы индивидуальной и групповой экспертной оценки, методы интервью по заданной схеме «вопрос – ответ», метод экспертных комиссий, дельфийский метод.

Эти методы практически не используются в нашей сфере, хотя явно имеют определенный потенциал

# О наших попытках решения проблем педагогического характера



# Итак, ...

В СГМ требуются современные, эффективные и наглядные методы анализа и способы представления его результатов

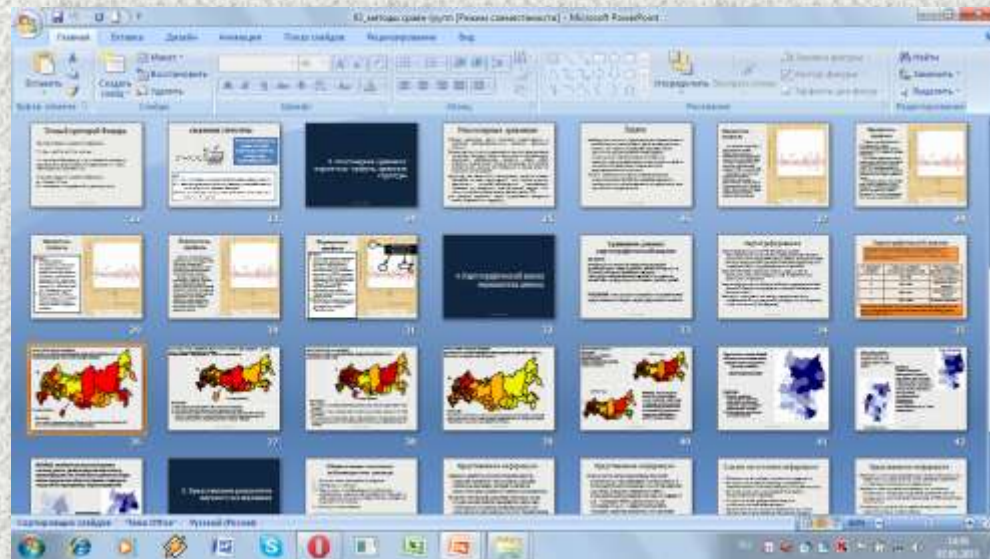
Применение таких методов в практике СГМ должно внедряться в технологическом, пошаговом, общепринятом и утвержденном виде, возможно в виде неких аналитических шаблонов

Требуются подготовленные специалисты и программные средства

# ПРОБЛЕМА 3 – преподавания СГМ

- Отдельной дисциплины по СГМ на додипломном этапе нет
- СГМ как фрагмент общей гигиены в соответствии с ФГОС-3 будет преподаваться с 2017 года
- Учебников по СГМ нет (Общая гигиена, социально-гигиенический мониторинг: руководство к практическим занятиям. Раздел "Общая гигиена": Учебное пособие/П.И. Мельниченко и др. М., 2014.)
- В настоящее время СГМ студентам преподается на отдельных лекциях, занятиях и элективных курсах на разных кафедрах МПФ (гигиенических и эпидемиологических)
- Формирование цельного понимания СГМ затруднительно
- Так же на разных кафедрах преподаются студентам основы анализа данных

# Наши разработки в преподавании основ статистического анализа



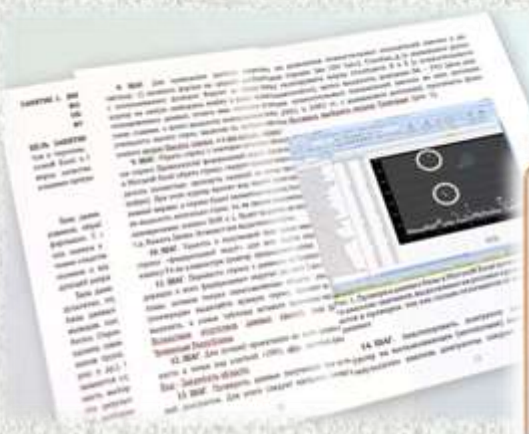
- обучающий курс тематического усовершенствования «Дружелюбная статистика: анализ и прогнозирование: пошаговые инструкции» (с 2006 года)
- с 2009 года мы обучаем всех аспирантов ОмГМУ
- подготовлено 2 преподавателя курса, защищено 16 диссертаций после обучения на курсе

# Издано учебное пособие

содержит пошаговые инструкции решения типичных для медиков задач в распространенных форматах Microsoft Excel и STATISTICA, данных в «дружелюбной» форме, учитывающей характер мышления и уровень математических знаний медиков.



В учебном пособии рассмотрены вопросы статистического анализа данных в формате Microsoft Excel и STATISTICA, даны пошаговые инструкции по решению типичных для медиков задач в распространенных форматах Microsoft Excel и STATISTICA, даны данные в «дружелюбной» форме, учитывающей характер мышления и уровень математических знаний медиков.



# ВИДЕОЛЕКЦИИ по статистике



В настоящее время заканчивается монтаж серии видеолекций курса с последующим размещением на информационно-образовательном портале ОмГМА



1. База данных



2. Описательная статистика



3. Методы сравнения



4. Корреляционно-регрессионный анализ



5. Кластерный анализ

# СГМ мог бы дать больше

- Промедление в применении методов анализа баз данных, на наш взгляд, существенно тормозит развитие гигиенической науки и практики (социально-гигиенического мониторинга).
- Накопленные базы данных предполагают их всесторонний анализ с последующим корректным обоснованием оптимизации СЭБН

**СПАСИБО**

