

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ**

Методические указания к лабораторным работам
и организации самостоятельной работы
для студентов направления
«Государственное и муниципальное управление»
(уровень магистратуры)

2018

Лепихина З. П. Информационные технологии статистического анализа данных: Методические указания к лабораторным работам и организации самостоятельной работы для студентов направления «Государственное и муниципальное управление» (уровень магистратуры) / З.П.Лепихина. – Томск, 2018. – 34 с.

© Томский государственный университет систем управления и радиоэлектроники,
2018

© Лепихина З.П., 2018

Оглавление

1 ВВЕДЕНИЕ.....	4
2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРОВЕДЕНИЮ ЛАБОРАТОРНЫХ РАБОТ.....	5
2.1 ЛАБОРАТОРНАЯ РАБОТА «Первичный анализ данных».....	5
2.2 ЛАБОРАТОРНАЯ РАБОТА «Первичный анализ данных на компьютере с использованием встроенных функций».....	6
2.3 ЛАБОРАТОРНАЯ РАБОТА «Изучение взаимосвязи признаков методом аналитической группировки».....	6
2.4 ЛАБОРАТОРНАЯ РАБОТА «Построение и анализ типологии объектов».....	9
2.5 ЛАБОРАТОРНАЯ РАБОТА «Построение и анализ структурных группировок.....	11
2.6 ЛАБОРАТОРНАЯ РАБОТА «Построение информативной системы признаков».....	12
2.7 ЛАБОРАТОРНАЯ РАБОТА «Анализ таблиц сопряженностей».....	13
2.8 ЛАБОРАТОРНАЯ РАБОТА «Исследование взаимосвязей и зависимостей числовых данных».....	14
2.9 ЛАБОРАТОРНАЯ РАБОТА «Прогнозирование динамики развития социально-экономических процессов».....	15
3. ОРГАНИЗАЦИЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ.....	17
3.1 ОБЩИЕ ПОЛОЖЕНИЯ.....	17
3.2. РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА.....	17
3.3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ИЗУЧЕНИЮ ТЕМ ДИСЦИПЛИНЫ.....	18
ПРИЛОЖЕНИЯ.....	27

1 Введение

Целями дисциплины являются формирование математической культуры магистрантов, знаний и навыков применения статистических методов и алгоритмов в социально-экономическом анализе для принятия эффективных управленческих решений на основе использования современных информационных технологий и программных средств.

Задачи изучения дисциплины:

– сформировать у студентов знания, умения и навыки в использовании методов получения статистической информации и умение критически оценивать информацию;

– сформировать способность правильно выбирать статистические методы анализа результатов исследования в соответствии с его целями, задачами, гипотезами и имеющимися данными;

развить необходимые навыки использования информационных технологий решения конкретных исследовательских и управленческих задач **Целями дисциплины** являются формирование математической культуры магистрантов, знаний и навыков применения статистических методов и алгоритмов в социально-экономическом анализе для принятия эффективных управленческих решений на основе использования современных информационных технологий и программных средств.

В данных Методических указаниях содержится:

- краткое изложение теоретического материала по теме, варианты заданий и порядок выполнения лабораторных работ;
- рекомендации по организации самостоятельной работы.

Лабораторные работы выполняются с использованием табличного процессора MS Excel LibreOffice Calc (OpenOffice Calc, LibreOffice Calc). Форма контроля выполнения лабораторной работы: демонстрация преподавателю расчетов и результатов анализа, собеседование, ответы на вопросы, выполнение дополнительных заданий.

При самостоятельной работе и подготовке к лабораторным занятиям студенту следует повторить теоретический материал по конспекту лекций и источникам, приведенным в разделе «Рекомендуемая литература», а также пользоваться информацией, представленной в статистических сборниках, в научной литературе и Интернете

2 Методические указания к проведению лабораторных работ

2.1 ЛАБОРАТОРНАЯ РАБОТА «Первичный анализ данных»

Цель работы:

Представление статистических данных, построение вариационных рядов, вычисление средних величин и показателей вариации.

Исходные данные. Исследован размер заработной платы работников предприятия. Данные представлены в таблице (Приложение 1)

Порядок выполнения лабораторной работы

- 1) В соответствии с вариантом выбрать данные из таблицы исходных данных.
- 2) Упорядочить исходные данные (провести сортировку по возрастанию)
- 3) На основе *исходных* данных определить:
 - а) среднее значение показателя, моду и медиану
 - б) размах вариации, среднее линейное отклонение, дисперсию, стандартное отклонение, коэффициент вариации
- 4) На основе исходных данных построить *дискретный* вариационный ряд и определить:
 - а) среднее значение показателя, моду и медиану
 - б) размах вариации, среднее линейное отклонение, дисперсию, стандартное отклонение, коэффициент вариации
 - в) первый и третий квартили
 - г) построить диаграммы распределения работников по заработной плате.
- 5). На основе исходных данных построить *интервальный* вариационный ряд с равными интервалами. Число интервалов задано в каждом варианте. Определить:
 - а) среднее значение показателя, моду и медиану
 - б) размах вариации, среднее линейное отклонение, дисперсию, стандартное отклонение, коэффициент вариации
 - в) первый и третий квартили
 - г) построить диаграммы распределения работников по заработной плате.
- 6) Провести сравнительный анализ полученных результатов.
- 7) Оформить отчет.

2.1 ЛАБОРАТОРНАЯ РАБОТА «Первичный анализ данных на компьютере с использованием встроенных функций»

Цель работы:

Первичный анализ данных на компьютере в среде Microsoft Excel. Вычисление статистических характеристик показателей с использованием встроенных функций.

Исходные данные. Основные социально-экономические показатели субъектов СФО РФ представлены в таблице (Приложение 2)

орядок выполнения лабораторной работы

- 1) По номеру варианта выбрать из таблицы Приложения 2 столбец, содержащий значения показателя.
- 2) Представить графически (столбиковая диаграмма) значения показателя у субъектов.
- 3) Пользуясь статистическими процедурами Excel, определить:
 - Максимальное и минимальное значение признака (МАКС, МИН)
 - Среднее значение (СРЗНАЧ), медиану (МЕДИАНА), моду (МОДА)
 - Дисперсию (ДИСПР) и среднее квадратическое отклонение (СТАНДОТКЛОН)
- 4) Вычислить коэффициент вариации. Сделать выводы относительно однородности совокупности.
- 5) Определить удельный вес каждого субъекта в общем объеме признака в СФО
- 6) Представить графически (круговая диаграмма) структуру совокупности.
- 6) Провести сравнительный анализ полученных результатов.
- 7) Оформить отчет

2.3 ЛАБОРАТОРНАЯ РАБОТА «Изучение взаимосвязи признаков методом аналитической группировки»

Цель работы:

Построение и оценка качества группировки. Изучение взаимосвязи признаков методом аналитической группировки.

Определить вариант лабораторной работы и выбрать данные из таблицы ПРИЛОЖЕНИЯ 3

Вариант	Фактор (номер показателя)	Результат-ВРП	Вариант	Фактор (номер показателя)	Результат-ВРП
1.	2	1	6.	7	1
2.	3	1	7.	8	1
3.	4	1	8.	9	1
4.	5	1	9.	10	1
5.	6	1	10.	11	1

Исходные данные

Сформировать исходную таблицу, содержащую названия регионов и указанные в варианте показатели социально-экономического развития регионов СФО (ПРИЛОЖЕНИЕ 3).

Регион	Показатель (фактор)	ВРП (результат)
	x	y
...

Порядок выполнения лабораторной работы

- 1) Провести сортировку по значению фактора.
- 2) Провести по всей совокупности для каждого признака расчет среднего значения, дисперсии, стандартного отклонения, коэффициента вариации
- 3) Построить *точечную* диаграмму зависимости результата от фактора.
- 4) Провести группировку регионов по значению фактора, выделив 3 группы: «Малые», «Средние», «Крупные». Границы группировочного показателя задать самостоятельно и уметь обосновать их..

Для каждой группы определить и занести в табл.1:

- частоту группы,
- групповые средние значения показателей x и y
- групповые дисперсии показателя x и y ,
- групповые коэффициенты вариации показателей x и y .

Таблица 1. Статистические характеристики группировки

Группа	Интервалы признака-фактора	Числота группы	Признак – фактор x			Признак – результат y		
			f_j	Среднее	Дисперсия	Коэффициент вариации	Среднее	Дисперсия
Малые								
Средние								
Крупные								

- 5) Дать оценку качества построенной группировки по признаку-фактору. При расчете коэффициента детерминации R^2 рассчитать межгрупповую дисперсию.
- 6) Провести анализ наличия связи, направления связи между x и y
- 7) По величинам b_{yx} ..определить линейность (нелинейность) связи между x и y
- 8) Рассчитать по формуле межгрупповую дисперсию по показателю – фактору
- 9) Дать оценку силы связи на основе расчета коэффициента детерминации R^2 . При расчете коэффициента детерминации R^2 рассчитать:
 - среднюю групповых дисперсий группировки по признаку-результату;
 - используя правило сложения дисперсий, вычислить межгрупповую дисперсию для группировки по признаку-результату
- 10) вычислить эмпирическое корреляционное отношение по формуле (7).
- 11) Провести анализ полученных результатов.
- 12) Оформить отчет.

2.4 ЛАБОРАТОРНАЯ РАБОТА «Построение и анализ типологии объектов»

Цель работы:

Исследование типологии социально-экономических явлений. Методы кластер-анализа. Использование базовых алгоритмов кластер-анализа в программной системе «Statistica».

Определить вариант лабораторной работы и выбрать данные из таблицы ПРИЛОЖЕНИЯ 3

Вариант	Фактор (номер показателя)	Результат-ВРП	Вариант	Фактор (номер показателя)	Результат-ВРП
6.	2	1	6.	7	1
7.	3	1	7.	8	1
8.	4	1	8.	9	1
9.	5	1	9.	10	1
10.	6	1	10.	11	1

Исходные данные

Сформировать в EXCEL исходную таблицу, содержащую названия регионов и указанные в варианте показатели социально-экономического развития регионов СФО (ПРИЛОЖЕНИЕ 3).

Регион	Показатель (фактор)	ВРП (результат)
	x	y
...

Порядок выполнения лабораторной работы (часть 1)

1) Изучить основные положения работы с программной системой STATISTICA:

- Запуск системы
- Создание файла в системе
- Добавление, удаление, копирование переменных и объектов
- Первичный анализ данных в системе

- Переключение модулей,
- 2). Запустить модуль **STA_BAS.EXE**.
 - 3) Скопировать файл исходных данных из EXCEL в систему STATISTICA, предварительно увеличив число строк (случаев – *case*) до нужного количества.
 - 4) Определить средствами STATISTICA описательные статистики показателей (среднее, дисперсия и др.), коэффициент корреляции
 - 5) Построить диаграммы распределения регионов по значению признаков..

Порядок выполнения лабораторной работы (часть 2)

- 1) Переключиться на модуль кластерного анализа **Cluster Analysis** .
 - 1.1.Провести кластерный анализ, используя *иерархический алгоритм* (Joining (tree clustering) единственной связи (**Single Linkage**)). Определить вариант разбиения на 3 класса и результаты разбиения записать в табл.1
 - 1.2.Провести кластерный анализ, используя *иерархический алгоритм* (Joining (tree clustering) Варда (**Ward**)). Определить вариант разбиения на 3 класса и результаты разбиения записать в табл.1.
 - 1.3.Провести кластерный анализ на 3 класса, используя метод К-средних (**k-means clustering**) и результаты разбиения записать в табл.1.

Таблица 1.Варианты разбиения регионов на классы

	Названия регионов, входящих в классы		
	Класс № 1	Класс №2	Класс №3
Single Linkage			
Ward			
k-means clustering			

- 2) Сравнить результаты, полученные тремя алгоритмами кластер-анализа

3.По результатам метода К-средних:

- 3.1.внести в исходный файл (EXCEL) дополнительный столбец с номером кластера
- 3.2.Построить точечную диаграмму в пространстве двух признаков с изображением принадлежности регионов кластерам. Например,

3.3. рассчитать средние значения и коэффициенты вариации классификационных признаков по совокупности в целом и по кластерам.

3.4. Дать содержательную интерпретацию результатов (оценить: уровень развития регионов по классификационным признакам, наличие «естественного расслоения», различие средних значений, однородность групп, взаимосвязь признаков).

4) Оформить отчет.

2.5 ЛАБОРАТОРНАЯ РАБОТА «Построение и анализ структурных группировок»

Цель работы:

Исследование структуры социально-экономических процессов, построение структурной группировки, анализ структурных сдвигов.

Исходные данные.

Определить вариант лабораторной работы и выбрать данные из таблицы ПРИЛОЖЕНИЯ 4

<i>Вариант 1</i> : 2005, 2008 годы по сравнению с 2000 годом
<i>Вариант 2</i> : 2006, 2009 годы по сравнению с 2000 годом
<i>Вариант 3</i> : 2007, 2010 годы по сравнению с 2000 годом
<i>Вариант 4</i> : 2008, 2011 годы по сравнению с 2000 годом
<i>Вариант 5</i> : 2006, 2008 годы по сравнению с 2005 годом
<i>Вариант 6</i> : 2006, 2009 годы по сравнению с 2005 годом
<i>Вариант 7</i> : 2007, 2010 годы по сравнению с 2005 годом
<i>Вариант 8</i> : 2008, 2011 годы по сравнению с 2007 годом
<i>Вариант 9</i> : 2009, 2011 годы по сравнению с 2007 годом
<i>Вариант 10</i> : 2010, 2011 годы по сравнению с 2007 годом

Порядок выполнения лабораторной работы

- 1) На основе исходных данных построить структурные группировки для категории «прибывшие в РФ»
- 2) Определить среднее линейное и среднее квадратическое изменение структуры прибывших в РФ.
- 3) Построить диаграммы структуры прибывших в исследуемых и базовом году
- 4) На основе исходных данных построить структурные группировки для категории «выбывшие из РФ»
- 5) Определить среднее линейное и среднее квадратическое изменение структуры выбывших из РФ.

- 6) Построить диаграммы структуры выбывших в исследуемых и базовом году
- 7) Сделать выводы относительно изменения структур.
- 8) Оформить отчет.

2.6 ЛАБОРАТОРНАЯ РАБОТА «Построение информативной системы признаков»

Цель работы:

Построение информативной системы признаков. Снижение размерности признакового пространства. Применение алгоритмов факторного анализа для построения интегрированных показателей.

Исходные данные. Определить вариант лабораторной работы и выбрать данные из таблицы ПРИЛОЖЕНИЯ 3

Вариант	Номера исследуемых показателей			Вариант	Номера исследуемых показателей		
11.	2	1	11	6.	7	1	2
12.	3	1	10	7.	8	1	3
13.	4	1	9	8.	9	1	4
14.	5	1	8	9.	10	1	5
15.	6	1	7	10.	11	1	6

Исходные данные

Сформировать в EXCEL исходную таблицу, содержащую названия регионов и указанные в варианте показатели социально-экономического развития регионов СФО (ПРИЛОЖЕНИЕ 3).

Регион	Показатель	Показатель	Показатель
	x	y	z
...	

Порядок выполнения лабораторной работы

1). Запустить модуль STA_FAC.EXE.

Скопировать файл исходных данных из EXCEL в систему STATISTICA, предварительно увеличив число строк (случаев – *case*) до нужного количества.

2).. Определить средствами STATISTICA описательные статистики показателей (среднее, дисперсия и др.), коэффициент корреляции

- 3). Построить диаграммы распределения регионов по значению признаков.
- 4). Провести факторный анализ:
 - на стартовой панели модуля Factor Analysis (Факторный анализ) выберите все 3 переменные
 - задайте метод выделения факторов (по умолчанию принимается метод Главных компонент),
 - число факторов (максимальное число факторов в случае трех переменных равно 3)
- 5). В окне Factor Analysis Results проанализировать результаты факторного анализа:
 - объясненная дисперсия
 - собственные значения
 - таблица факторных нагрузок
 - общности решения
 - значения факторов для каждого региона.
- 6). Сделать содержательную интерпретацию полученных результатов.
- 7). Оформить отчет

2.7 ЛАБОРАТОРНАЯ РАБОТА «Анализ таблиц сопряженностей»

Цель работы

Методы исследования взаимосвязей нечисловых данных. Анализ таблиц сопряженностей в программных системах EXCEL и STATISTICA,

Исходные данные

. В файле ДляЛАБ_ТАБЛИЦЫ.xls приведены данные 50 респондентов о предпочитаемых напитках. При этом данные закодированы следующим образом: первая переменная ПОЛ (1- мужской, 2-женский), НАПИТОК (1-pepsi, 2-cola).

В соответствии с вариантом отобрать для анализа 20 анкет (строк).

№ варианта	1	2	3	4	5
Строки	2-21	5-24	10-29	15-34	20-39
№ варианта	6	7	8	9	10
Строки	25-44	30-49	32-51	3-22	6-25

Порядок выполнения лабораторной работы.

1. Построить четырехклеточную таблицу вида

ПОЛ	НАПИТОК	
	1-pepsi	2-cola
Мужской (1)	n_{11}	n_{12}
Женский (2)	n_{21}	n_{21}

Для удобства подсчета n_{11} , n_{12} , n_{21} , n_{21} провести сортировку данных по первой переменной, а затем по второй.

Заполнение таблицы 2x2 можно провести «вручную» или воспользоваться функцией СЧЁТЕСЛИ.

2). Провести одномерный анализ данных: рассчитать частоты, относительные частоты по каждому признаку, построить диаграммы различных видов (гистограммы, секторные; линейные и др.

3). провести анализ таблицы 2x2: установить наличие связи, рассчитать коэффициенты ассоциации, контингенции

4). Объяснить полученные результаты

5) Запустить модуль базовых характеристик **STA_BAS.EXE**.

6) Создать файл с исходными данными: скопировать данные из EXCEL

7) **Провести анализ таблицы сопряженностей 2x2:**

Используя процедуру < **Tables and banners** >, построить таблицу сопряженности 2x2, получить Хи-квадрат, сравнить его с табличным, получить коэффициенты Φ , Крамера, контингенции

8) Объяснить полученные результаты

9) Оформить отчет.

2.8 ЛАБОРАТОРНАЯ РАБОТА «Исследование взаимосвязей и зависимостей числовых данных»

Цель работы:

Исследование взаимосвязей и зависимостей числовых данных. Применение алгоритмов регрессионного анализа в программной системе STATISTICA.

Исходные данные. Определить вариант лабораторной работы и выбрать данные из таблицы ПРИЛОЖЕНИЯ 3

Вариант	Фактор (номер показателя)	Результат-ВРП	Вариант	Фактор (номер показателя)	Результат- ВРП
1.	2	1	6.	7	1
2.	3	1	7.	8	1
3.	4	1	8.	9	1
4.	5	1	9.	10	1
5.	6	1	10.	11	1

Порядок выполнения лабораторной работы

1) Запустить модуль **STA_REG.EXE**.
 2) Сформировать исходную таблицу, содержащую названия регионов и указанные в варианте показатели социально-экономического развития регионов СФО
 3) Для проведения регрессионного анализа (линейная регрессия) использовать модуль Multiple Regression. В стартовом диалоговом окне модуля при помощи кнопки **Variables** указываются зависимая (dependent) и независимые(ая) (independent) переменные. В поле **Input file** указывается тип файла с данными:

4) По окончании выполнения процедуры в окне Multiple Regressions Results (результаты регрессионного анализа) отобразятся результаты расчетов:

- Multiple R- коэффициент множественной корреляции;
- RI (R^2)- коэффициент детерминации;
- F- F-критерий используется для проверки значимости регрессии.
- df - число степеней свободы для F-критерия;
- p- вероятность нулевой гипотезы для F-критерия;
- Intercept– оценка свободного члена уравнения;
- Beta- -коэффициенты уравнения.

2.9 ЛАБОРАТОРНАЯ РАБОТА «Прогнозирование динамики развития социально-экономических процессов»

Цель работы:

Анализ и прогнозирование динамики развития социально-экономических процессов. Построение моделей временного ряда в MS Excel

Исходные данные. В соответствии с номером варианта необходимо выбрать данные из таблицы ПРИЛОЖЕНИЯ 5 исходных данных о численности персонала инновационных предприятий.

Вариант 1	Республика Алтай	Вариант 7	Красноярский край
Вариант 2	Республика Бурятия	Вариант 8	Иркутская область
Вариант 3	Республика Тыва	Вариант 9	Кемеровская область
Вариант 4	Республика Хакасия	Вариант 10	Новосибирская область
Вариант 5	Алтайский край	Вариант 11	Омская область
Вариант 6	Забайкальский край	Вариант 12	Томская область

Порядок выполнения лабораторной работы

- 1) построить линейную модель для выбранных данных,
- 2) Для построенной в пункте (1) провести оценку модели по критерию Фишера, рассчитать коэффициент детерминации.
- 3) Построить график динамики исходных и выровненных значений.
- 4) Рассчитать точечный прогноз на следующий период
- 5) 5). Пользуясь средствами MS Excel, провести исследования различного вида моделей тренда. Рассмотреть линейную модель, полиномы 2-й и т.д. степеней, помещая на график уравнение тренда и значения коэффициента детерминации. Сделать прогноз на следующие периоды для разных трендов.

3. ОРГАНИЗАЦИЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

3.1 ОБЩИЕ ПОЛОЖЕНИЯ

Цель самостоятельной работы по дисциплине – повышение эффективности изучения теоретической части дисциплины и полноценной работы на лабораторных занятиях, а также получение навыков самостоятельного проведения поиска, сбора и анализа информации по отдельным направлениям статистического анализа данных.

В соответствии с рабочей программой дисциплины «Информационные технологии статистического анализа данных» на проведение самостоятельной работы студенту отводится 108 часов (ПРИЛОЖЕНИЕ 6), включая 36 на самостоятельную подготовку к сдаче экзамена.

Более половины времени (55 часов) студенту необходимо использовать на подготовку к аудиторным занятиям и контрольным работам.. При проработке лекционного материала по каждой теме магистрант должен внимательно ознакомиться с конспектом лекций, а затем для углубленного изучения материала следует обратиться к литературным источникам (учебникам, учебным пособиям, монографиям, статьям, статистическим сборникам), а также материалам, размещенным в сети Интернет. Для закрепления материала темы необходимо ответить на предлагаемые в пособиях вопросы и прорешать задачи по теме.

3.2. РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Основная литература

- 1) Прасолов А.В. Математические методы экономической динамики [Электронный ресурс] : учебное пособие. — Электрон. дан. — СПб. : Лань, 2015. — 350 с. **Гриф УМО** — Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=67480
- 2) Управление данными [Электронный ресурс] : учебник / В.В. Цехановский, В.Д. Чертовской. — Электрон. дан. — СПб. : Лань, 2015. — 432 с. **Гриф УМО** — Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=65152

Дополнительная литература

- 3) Симчера В.М Методы многомерного анализа статистических данных. М.: Финансы и статистика,2008. – 400с. ISBN:978-5-279-

03184-4. Гриф УМО. -[Электронный ресурс]. — URL:

http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=1005

4) Божко В.П. Информационные технологии в статистике. М.: Финансы и статистика, 2011. – 152 с. ISBN 978-5-279-03514-4. Гриф УМО. [Электронный ресурс] :

http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=5686

5) Информационные системы и технологии в экономике и управлении : Учебное пособие для вузов / Санкт-Петербургский государственный университет экономики и финансов ; ред. : В. В. Трофимов. - 2-е изд., перераб. и доп. - М. : Высшее образование, 2007. - 480 с. : ил. - (Университеты России). - В библиотеке ТУСУРа: 20 экз.

6) Васильев А.Н. Числовые расчеты в Excel [Электронный ресурс] : учебное пособие. — Электрон. дан. — СПб. : Лань, 2014. — 598 с. — Режим доступа:

http://e.lanbook.com/books/element.php?pl1_id=45683

7) Бураков, П.В. Корпоративные информационные системы [Электронный ресурс] : учебное пособие. — Электрон. дан. — СПб. : НИУ ИТМО (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики), 2014. — 100 с. — Режим доступа:

http://e.lanbook.com/books/element.php?pl1_id=70882

8) Боровиков В.П. Прогнозирование в системе STATISTICA в среде Windows: Основы теории и интенсивная практика на компьютере : Учебное пособие для вузов / Владимир Павлович Боровиков, Григорий Иванович Ивченко. - М. : Финансы и статистика, 2000. - 384 с. В библиотеке ТУСУР: 8 экз.

9) Лепихина З.П. Основы социального прогнозирования: Учебное пособие. – Томск: Томский межвузовский центр дистанционного образования, 2006.–112с. (Имеется в ФДО, библиотеке ТУСУРа –2 экз., на каф.АОИ).

3.3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ИЗУЧЕНИЮ ТЕМ ДИСЦИПЛИНЫ

Тема «Основные положения анализа данных и специфика социально-экономических данных»

Содержание темы

Информационное обеспечение органов государственной власти. Основные задачи государственной статистики. Современное состояние и перспективы развития информационных технологий в российской статистике. Природа социально-экономической информации, виды и

способы ее представления. Измерительные шкалы. Статистические показатели и статистические величины. Исследование статистических закономерностей: этапы, методы, показатели. Выборочный метод в изучении социально-экономических явлений и процессов. Data mining как современное направление прикладного статистического анализа данных. Основные понятия и типы задач, решаемых Data Mining..

Литература: [1, 2, 4, 5].

Методические указания по изучению темы

Существует множество определений термина «анализ данных». В настоящее время большинство авторов подчеркивают, что Data Mining - мультидисциплинарная область, возникшая и развивающаяся на базе таких наук как прикладная статистика, распознавание образов, искусственный интеллект, теория баз данных и др. При изучении методологических основ следует обратить внимание на связь прогностики с законами диалектики, на необходимость системного анализа исследуемого явления, процесса, объекта. Важно понять необходимость статистического анализа данных при принятии управленческих задач, сходство и различие понятий «данные», «информация», «прогноз», «план».

Теория статистических показателей занимает одно из центральных мест в общей теории статистики, так как отчетность организаций, планирование, исследовательская и аналитическая работа, моделирование и прогнозирование базируется на использовании различных систем статистических показателей. Все используемые в статистической практике показатели по форме выражения классифицируются на абсолютные, относительные и средние.

При изучении математических методов, применяемых при анализе данных, следует уяснить связь формальную постановку задачи с содержательной. Важным является правильный выбор метода, выбор вида модели для исследуемой предметной области.

Следует обратить внимание, что важными вопросами являются вопросы формирования выборочной совокупности, определение способа отбора единиц совокупности, вычисление ошибок выборки и построение доверительных интервалов выборочных характеристик, расчет необходимого объема выборки.

Вопросы и упражнения для самопроверки:

1. Охарактеризуйте понятия «информация» и «данные».
2. Сформулируйте задачи Data Mining.
3. Охарактеризуйте стадии статистического исследования?

4. Проведите классификацию статистических признаков
5. Приведите примеры признаков, измеренных в номинальной шкале
6. . Укажите преимущества выборочного наблюдения перед сплошным наблюдением
7. Запишите соотношение средней и предельной ошибок выборки

Тема «Выявление типологии социально-экономических явлений»

Содержание темы

Роль типологии в социально-экономических исследованиях. Группировка и сводка материалов статистического наблюдения. Виды статистических группировок и задачи, решаемые методом группировок. Классификация и характеристика основных задач распознавания образов. Методы кластеризации и дискриминации. Постановка задачи кластеризации. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации. Базовые алгоритмы кластеризации.

Литература: [1, 3, 9].

Методические указания по изучению темы

Группировкой называется разбиение единиц изучаемой совокупности на однородные группы по определенным существенным признакам. При изучении темы важно уяснить, что группировки не являются самоцелью, а позволяют выявить социально-экономические типы явления, изучить структуру явлений и изменение структуры, выявить связи и зависимости между признаками.

Необходимо отметить, что метод группировок тесно связан с представлением данных в виде групповых и комбинационных таблиц, а также с графическим представлением структуры совокупности, ее частей и соотношений между ними.

Различия в целевом назначении группировки выражаются в классификации группировок: типологические, структурные, аналитические. В процессе изучения каждого вида важно увидеть различия между видами группировок, общее и связи между видами.

Оценить качество проведенной группировки можно различными способами. В частности, на основе правила сложения дисперсий: полная дисперсия признака равна сумме средней из групповых дисперсий и межгрупповой дисперсии. Если исходная совокупность может быть разбита на группы и группировка проведена правильно, то группы однородны (дисперсия внутри групп достаточно мала), а полная дисперсия образуется в основном за счет различий в значениях признака,

относящихся к разным группам (межгрупповая дисперсия относительно велика). В этом случае значение коэффициента детерминации близко к единице.

Многомерная группировка (размерность признакового пространства равна p) относится к задачам распознавания образов. Задача распознавания образов состоит в следующем: объекты, обладающие типичными для всех чертами, образуют характерный класс, называемый образом. Требуется построить некоторый оптимальный алгоритм, который бы позволил по этим типичным чертам определить принадлежность данного объекта к тому или иному классу, то есть распознать образ. В данном курсе рассматриваются алгоритмы кластер – анализа: отыскание набора кластеров (классов, групп) $\{S_i\}$ в заданном признаковом пространстве $\{x_i\}$ с известной решающей функцией $\delta_{i=\alpha}$. Для усвоения материала студенту необходимо познакомиться с основными определениями, идеями алгоритмов и прорешать предложенные задачи.

Вопросы и упражнения для самопроверки:

1. Сформулируйте принцип построения группировок.
2. Приведите примеры простой и сложной группировок.
3. Какой вывод можно сделать, если квадратичный показатель структурных сдвигов равен нулю?
4. Предположим, проведена аналитическая группировка на 3 группы. Поясните, почему при линейной зависимости должно выполняться равенство

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y_3 - y_2}{x_3 - x_2}.$$

5. Если средняя внутригрупповых дисперсий равна полной дисперсии признака в совокупности, чему равна межгрупповая дисперсия?
6. В таблице приведена информация о городах, описанных двумя признаками:

x – количество студентов: тыс.чел.

y – число театров

	Пичтаун	Эпльвилль	Черриль	Грейп	Лемонск
x	6	7	0	2	1
y	3	3	0	1	0

Используя алгоритмы ближайшего соседа, дальнего соседа, медианной связи, проведите разбиение объектов на 2 класса. Постройте дендрограмму.

Тема «Информационно-программные средства статистического анализа данных»

Содержание темы

Технология анализа социально-экономической информации средствами статистических пакетов. Обзор рынка инструментов статистического анализа данных.

Средства массового применения (Microsoft Excel), специализированные инструментальные средства (Statistica, SPSS и др.), средства создания информационных хранилищ и обеспечения различного вида анализа. Анализ и обработка экспериментальных данных с помощью пакетов «Statistica», «SPSS».

Литература: [2, 4, 6, 8].

Методические указания по изучению темы

Аспекты проблемы анализа и необходимые для их разрешения функции нашли выражение в соответствующих программных продуктах. Соответственно средства автоматизации анализа представлены в различных видах. Имеются комплексные информационно-аналитические системы, выполняющие в той или иной степени функции в соответствии с рассмотренными аспектами.

Как правило, все инструментальные средства, предназначенные для автоматизации аналитических работ, приспособлены для обработки многомерных массивов информации; имеют также возможность импорта/экспорта данных в другие операционные среды, развитые средства визуального двумерного (2D) и трехмерного (3D) представления информации. Развитой системой в отношении возможностей, предоставляемых модулями и автономными подсистемами интеллектуального анализа, является инструментарий ИАС фирмы США «SAS Institute Inc».

Необходимо отметить следующее обстоятельство. Целевые программные продукты весьма дороги и пока малодоступны для массового российского потребителя. Выходом из этого положения является использование редко применяемых на практике возможностей массовых программных инструментальных средств Excel, Mathcad, Stadia, Statistica и др. Пакет «Statistica» — с его помощью возможно решение

всех задач статистического анализа, а также выявления закономерностей и прогнозирования.

Вопросы и упражнения для самопроверки:

1. Сформулируйте понятие «информационно-аналитические системы».
2. Проведите обзор рынка программных систем статистического анализа данных.
3. Проведите сравнение состава статистических процедур в пакетах Excel и Statistica.
4. Рассмотрите реализацию регрессионного анализа в пакетах Excel и Statistica.
5. Какой пакет вы использовали при решении задачи прогнозирования динамического ряда? Ответ обоснуйте.

Тема «Методы снижения размерности в задачах анализа данных»

Содержание темы

Проблемы и критерии информативности признаков. Логико-эвристические методы формирования признакового пространства. Статистические методы формирования признакового пространства. Методы снижения размерности. Метод главных компонент. Факторный анализ, основные проблемы его применения. Экспертно-статистический метод построения единого сводного показателя..

Литература: [2, 4, 6, 8].

Методические указания по изучению темы

В многомерном статистическом анализе каждый объект описывается вектором, размерность которого произвольна. Однако уже в трехмерном пространстве человеку трудно анализировать скопления точек. Поэтому вполне естественным является желание перейти от многомерной выборки к данным небольшой размерности, чтобы «на них можно было посмотреть». Кроме того, студенту важно понять, что могут быть малоинформативные переменные, которые лишь мешают статистическому анализу. Поэтому желательно избавиться от таких факторов.

При анализе многомерных данных обычно рассматривают не одну, а множество задач, в частности, по-разному выбирая независимые и зависимые переменные. Поэтому задачу снижения размерности можно рассмотреть в следующей формулировке. Дана многомерная

выборка. Требуется перейти от нее к совокупности векторов меньшей размерности, максимально сохранив структуру исходных данных, по возможности не теряя информации, содержащихся в данных. Задача конкретизируется в рамках каждого конкретного метода снижения размерности.

Метод главных компонент является одним из наиболее часто используемых методов снижения размерности.

Для визуального анализа данных часто используют проекции исходных векторов на плоскость первых двух главных компонент.

Различные алгоритмы факторного анализа объединены тем, что во всех них происходит переход к новому базису в исходном p -мерном пространстве. Важным является понятие «нагрузка фактора», применяемое для описания роли исходного фактора (переменной) в формировании определенного вектора из нового базиса.

Новая идея по сравнению с методом главных компонент состоит в том, что на основе нагрузок происходит разбиение факторов на группы. В одну группу объединяются факторы, имеющие сходное влияние на элементы нового базиса. Затем из каждой группы рекомендуется оставить одного представителя. Иногда вместо выбора представителя расчетным путем формируется новый фактор, являющийся центральным для рассматриваемой группы. Снижение размерности происходит при переходе к системе факторов, являющихся представителями групп. Остальные факторы отбрасываются.

Вопросы и упражнения для самопроверки:

1. Чем обоснована необходимость решения задачи снижения размерности признакового пространства?
2. Определите сущность задачи снижения размерности.
3. Сформулируйте различия компонентного и факторного анализов.
4. Перечислите проблемы факторного анализа.
5. Какими методами определяется число главных факторов?
6. Сформулируйте идею варимакс-вращения.
7. Рассмотрите реализацию факторного анализа в программной системе Statistica.

Тема «Исследование взаимосвязей социально-экономических явлений и прогнозирование»

Содержание темы

Математические задачи и методы в проблематике экспертного оценивания, шкалирования, латентного анализа, контент-анализа. Ме-

тоды исследования взаимосвязей различных типов данных. Анализ таблиц сопряженностей. Анализ ранговых признаков. Методы исследования взаимосвязей и зависимостей числовых данных. Методы прогнозирования социально-экономических показателей.

Литература: [2, 4, 6, 9].

Методические указания по изучению темы

Применимость методов анализа социально-экономических явлений ограничена уровнем измерения переменных или шкалой, в которой измеряются переменные (шкала наименований, интервальная шкала, порядковая, шкала отношений).

Изучая методы анализа нечисловой информации, студенту необходимо понять, что поиск статистической связи между номинальными признаками – не самоцель. Например, выявленная в ходе опроса отрицательная связь между номинальными признаками «Занятия спортом» и «Курение» позволяет сделать прогноз «При развитии массового спорта снизится число курящих».

Методы анализа ранговых корреляций используются для измерения связи между признаками, измеренными в порядковой шкале, а также в экспертных обследованиях для оценки согласованности мнений экспертов и построения интегральной (совокупной) оценки признака. Следует отметить, что понятия «статистическая связь» и «согласованность мнений» имеют в некотором смысле противоположный смысл. Если установлена сильная статистическая связь между признаками и есть возможность перейти от шкалы наименований к ранговой шкале, то вычисленные по тем же данным ранговые коэффициенты будут иметь малую величину.

При изучении данной темы большое внимание уделяется анализу временных рядов и построению модели тренда. Метод экстраполяции тренда - один из пассивных методов прогнозирования, - и называется «наивным» прогнозом, так как предполагает строгую инерционность развития, которая представляется в виде проектирования прошлых тенденций в будущее, а главное — независимость показателей развития от тех или иных факторов. Студенту следует на примере линейной модели уяснить методику построения модели, а для более сложных случаев необходимо использовать статистические пакеты прикладных программ.

Вопросы и упражнения для самопроверки:

1. Укажите различия трендовых и эконометрических моделей.
2. Приведите примеры различных форм связей.
3. Что такое «наивный прогноз»?

4. Определите понятие «ранг».
5. Проведен опрос населения по поводу планирующегося строительства магазина. Из 150 опрошенных мужчин 30 человек высказались против строительства, а из 100 женщин отрицательно отнеслись к строительству 50 человек. Есть ли связь между полом респондентов и их отношением к проекту?
6. Три эксперта оценили инвестиционную привлекательность 5 регионов следующим образом:

	Регион 1	Регион 2	Регион 3	Регион 4	Регион 5
Иванов	1	2	5	4	3
Перов	1	3	4	5	2
Сонин	1	2	4	5	3

Оцените степень согласованности мнений экспертов.

7. Покажите, что для постоянных маргинальных частот разность между наблюдаемой и «независимой» частотами в любой клетке таблицы 2×2 будут равны $\pm D$.

ПРИЛОЖЕНИЕ 1

Таблица Заработная плата работников, тыс.руб.

№ Варианта

<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	
23	23	18	20	30	
20	20	22	18	20	
25	25	32	20	20	
20	20	18	20	18	
20	20	20	23	18	
18	18	16	20	20	
20	20	20	18	18	
18	18	18	22	20	
20	20	24	18	20	
20	20	20	22	23	
23	23	23	32	20	
20	20	20	18	16	
18	18	18	20	22	
22	22	22	23	23	
18	18	16	20	30	
22	22	22	30	20	
32	32	23	20	18	
18	18	20	20	20	
20	20	24	18	20	
16	17	20	18	23	
20	20	20	20	20	
18	18	18	18	18	
20	20	20	20	22	
20	20	18	20	18	
23	23	20	23	22	
20	20	20	20	32	
18	18	23	18	18	
22	22	20	22	20	
16	17	18	23	23	
22	22	22	30	20	
<i>Число равных интервалов</i>	<i>K=4</i>	<i>K=3</i>	<i>K=4</i>	<i>K=3</i>	<i>K=4</i>

Продолжение табл.

<i>№ Варианта</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
23	23	18	20	25	
20	20	22	18	20	
25	25	28	20	20	
20	20	18	20	18	
20	20	20	23	18	
18	18	16	20	20	
20	20	20	18	18	
18	18	18	22	20	
20	20	24	18	20	
20	20	20	22	23	
23	23	23	27	20	
20	20	20	18	18	
18	18	18	20	22	
22	16	22	23	23	
18	18	16	20	23	
22	22	22	27	20	
29	32	23	20	17	
18	18	20	20	20	
20	20	24	18	20	
17	17	20	18	23	
20	20	20	20	20	
18	18	18	18	18	
20	20	20	20	22	
20	20	18	20	18	
23	23	20	23	22	
20	20	20	20	25	
18	18	23	18	18	
22	22	20	22	20	
25	17	18	23	23	
22	22	22	25	20	
<i>Число равных интервалов</i>	<i>K=3</i>	<i>K=4</i>	<i>K=4</i>	<i>K=3</i>	<i>K=4</i>

ПРИЛОЖЕНИЕ 2
ОСНОВНЫЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

	Площадь территории ¹⁾ , тыс. км ²	Численность населения на 1 января 2012 г., тыс. человек	Среднегодовая численность занятых в экономике, тыс. человек	Среднедушевые денежные доходы (в месяц), руб.	Потребительские расходы в среднем на душу населения (в месяц), руб.
	1	2	3	4	5
Сибирский ФО	5145,0	19261,0	9018,6	16567,9	11824,8
Республика Алтай	92,9	208,4	91,7	13836,9	7179,0
Республика Бурятия	351,3	971,4	417,4	15715,5	11340,0
Республика Тыва	168,6	309,4	106,0	10962,8	4944,6
Республика Хакасия	61,6	532,2	239,2	14222,8	9680,5
Алтайский край	168,0	2407,2	1075,6	12499,9	9765,7
Забайкальский край	431,9	1099,4	489,4	15968,8	10572,7
Красноярский край	2366,8	2838,4	1437,5	20145,5	14105,7
Иркутская область	774,8	2424,4	1121,7	16017,2	10580,2
Кемеровская область	95,7	2750,8	1302,0	16666,0	11237,2
Новосибирская область	177,8	2686,9	1305,1	18244,1	14898,1
Омская область	141,1	1974,8	945,5	17247,9	12663,1
Томская область	314,4	1057,7	487,5	16516,0	11199,4

Продолжение Таблицы

	Средне- месячная номиналь- ная начис- ленная заработ- ная плата ра- бот- ников, руб.	Валовой региональ- ный продукт в 2010 г. , млн. руб.	Основные фонды в экономи- ке (по полной учетной стоимости; на конец года) ²⁾ , млн. руб.	Ввод в дейст- вие общей площади жилых домов, тыс. м ²	Оборот розничной торговли, млн. руб.
	6	7	8	9	10
Сибирский ФО	20889,9	4093589,0	10256100	7213,8	2064138
Республика Алтай	15632,4	21635,8	61628	76,6	14312
Республика Бурятия	19924,0	136374,0	430210	304,4	100938
Республика Тыва	19163,1	30601,0	47409	52,4	13742
Республика Хакасия	20689,5	93709,0	292915	156,2	46034
Алтайский край	13822,6	299715,3	757632	663,2	218077
Забайкальский край	21099,6	162100,2	650405	276,9	106366
Красноярский край	25658,6	1050158,5	1815754	1047,1	361607
Иркутская об- ласть	22647,7	539245,6	1975486	755,2	225846
Кемеровская область	20478,8	622513,0	1406912	1082,6	287279
Новосибирская область	20308,5	482026,5	1229181	1505,2	368292
Омская область	19087,8	371218,1	725451	836,7	228595
Томская об- ласть	24001,0	284292,0	863117	457,6	93050

ПРИЛОЖЕНИЕ 3

Таблица - Основные показатели развития регионов Сибирского федерального округа

	Валовой региональный продукт в 2010 г., млн. руб.	Площадь территории ¹ , тыс. км ²	Численность населения на 1 января 2012 г., тыс. человек	Среднегодовая численность занятых в экономике, тыс. человек	Среднедушевые денежные доходы (в месяц), руб.	Потребительские расходы в среднем на душу населения (в месяц), руб.	Среднемесячная номинальная начисленная заработная плата работников, руб.	Основные фонды в экономике (по полной учетной стоимости; на конец года) ² , млн. руб.
	1	2	3	4	5	6	7	8
Республика Алтай	21635,8	92,9	208,4	91,7	13836,9	7179,0	15632,4	61628
Республика Бурятия	13637,4	351,3	971,4	417,4	15715,5	11340,0	19924,0	430210
Республика Тыва	30601,0	168,6	309,4	106,0	10962,8	4944,6	19163,1	47409
Республика Хакасия	93709,0	61,6	532,2	239,2	14222,8	9680,5	20689,5	292915
Алтайский край	29971,5,3	168,0	2407,2	1075,6	12499,9	9765,7	13822,6	757632
Забайкальский край	16210,2	431,9	1099,4	489,4	15968,8	10572,7	21099,6	650405
Красноярский край	10501,58,5	2366,8	2838,4	1437,5	20145,5	14105,7	25658,6	181575,4
Иркутская область	53924,5,6	774,8	2424,4	1121,7	16017,2	10580,2	22647,7	197548,6
Кемеровская область	62251,3,0	95,7	2750,8	1302,0	16666,0	11237,2	20478,8	140691,2
Новосибирская область	48202,6,5	177,8	2686,9	1305,1	18244,1	14898,1	20308,5	122918,1
Омская область	37121,8,1	141,1	1974,8	945,5	17247,9	12663,1	19087,8	725451
Томская область	28429,2,0	314,4	1057,7	487,5	16516,0	11199,4	24001,0	863117

Продолжение Таблицы

Объем отгруженных товаров собственного производства по ВЭД, млн. руб.			Продукция сельского хозяйства - всего, млн. руб.	в том числе		Ввод в действие общей площади жилых домов, тыс. м ²	Оборот розничной торговли, млн. руб.
добыча полезных ископаемых	обрабатывающие производства	Производство и распределение электроэнергии, газа и воды		растениеводства	животноводства		
9	10	11	12	13	14	15	16
862	1296	1654	8020	1205	6815	76,6	14312
12808	51115	19826	13044	4605	8439	304,4	100938
3376	514	2868	4648	827	3821	52,4	13742
26536	56595	21462	9371	3216	6155	156,2	46034
6041	189279	31991	93784	46670	47114	663,2	218077
40377	14365	18311	15154	3433	11721	276,9	106366
266636	628113	95432	68598	32708	35890	1047,1	361607
129795	299406	81275	43610	18748	24862	755,2	225846
507993	385413	85949	38044	19621	18423	1082,6	287279
19674	249816	61592	60425	26860	33565	1505,2	368292
4411	529355	36013	66911	33535	33376	836,7	228595
137513	100598	25617	19420	7225	12195	457,6	93050

ПРИЛОЖЕНИЕ 4

Таблица – Международная миграция по годам (человек)

	2000	2005	2006	2007	2008	2009	2010	2011
Прибыло в РФ -								
всего	359330	177230	186380	286956	281614	279907	191656	356535
в том числе:								
из стран СНГ	326561	163101	170851	263277	261170	261495	171940	310549
из стран ЕС	6259	5386	5246	6617	6231	5667	6394	11413
из других стран	26510	8743	10283	17062	14213	12745	13322	34573

	2000	2005	2006	2007	2008	2009	2010	2011
Выбыло из РФ -								
всего	145720	69798	54061	47013	39508	32458	33578	36774
в том числе :								
в страны СНГ	80510	35418	34669	30726	25542	20326	21206	22568
в страны ЕС	45064	25236	12182	10521	8795	7509	7185	7749
в другие страны	20146	9144	7210	5766	5171	4623	5187	6457

ПРИЛОЖЕНИЕ 5
ЧИСЛЕННОСТЬ ПЕРСОНАЛА, ЗАНЯТОГО НАУЧНЫМИ ИССЛЕДОВАНИЯМИ И РАЗРАБОТКАМИ

вариант		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	Республика Алтай	90	97	105	122	124	111	132	168	173	156	158	172
2	Республика Бурятия	1209	1200	1249	1299	1278	1231	1003	985	954	969	952	1144
3	Республика Тыва	285	323	343	334	330	327	332	416	414	425	416	415
4	Республика Хакасия	128	125	120	108	107	282	205	198	193	166	149	148
5	Алтайский край	3427	2941	2919	2980	2870	2732	2775	2731	2267	2054	1955	2182
6	Забайкальский край	680	775	840	823	779	509	608	520	512	335	322	313
7	Красноярский край	7196	7637	7486	7303	7087	7102	6846	6685	6287	6299	6475	6748
8	Иркутская область	5295	5408	5387	5184	4983	4829	4557	4910	4897	4919	4912	5075
9	Кемеровская область	2090	1916	1820	1594	1557	1476	1511	1496	1327	1336	1258	1231
10	Новосибирская область	25168	25494	25537	25888	25077	24791	23438	22561	21597	21622	21615	21569
11	Омская область	8872	9442	9478	9825	10127	9367	8983	7246	6961	6622	6125	5002
12	Томская область	8037	7913	7768	7916	8175	8229	8257	8511	8374	8560	8687	8795