



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»  
(ДГТУ)

Факультет «ИиВТ»

Кафедра «Информационные системы в строительстве»

## **Математическое моделирование и методы математической статистики**

*Методические рекомендации к самостоятельной работе и контрольной работе  
магистрантов очной и заочной формы обучения  
направления подготовки*

*25.04.01 «Техническая эксплуатация летательных аппаратов и двигателей»  
программа «Техническая эксплуатация авиационной техники»*

Ростов-на-Дону  
2021

Методические рекомендации составила ктн Батурина Н.Ю.

Методические рекомендации по самостоятельной работе магистрантов по дисциплине «Математическое моделирование и методы математической статистики» разработаны в соответствии с федеральным государственным образовательным стандартом высшего образования - магистратура по направлению подготовки 25.04.01 Техническая эксплуатация летательных аппаратов и двигателей, утвержденного приказом Министерства образования и науки Российской Федерации от 10.01.2018 №19 и составлены на основании учебного плана по программе подготовки 25.04.01 Техническая эксплуатация авиационной техники, утвержденного Ученым советом ДГТУ 26.02.2021 (протокол №8).

Методические рекомендации одобрены на заседании кафедры «ИСвС»  
Зав. кафедрой: Ляпин А.А.

## ОГЛАВЛЕНИЕ

1. ОБЩИЕ ПОЛОЖЕНИЯ .....	4
2. ВИДЫ И ФОРМЫ ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ.....	5
3. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ И ВЫПОЛНЕНИЮ ВНЕАУДИТОРНОЙ САМОСТОЯТЕЛЬНОЙ РАБОТЫ.....	6
4. ВОПРОСЫ ЭКЗАМЕНА ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ».....	13
5. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ .....	15
6. ТЕКСТЫ ЛЕКЦИЙ. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ВЫПОЛНЕНИЯ ЛАБОРАТОРНЫХ РАБОТ И КОНТРОЛЬНОЙ РАБОТЫ	15

# 1. ОБЩИЕ ПОЛОЖЕНИЯ

Дисциплина «Математическое моделирование и методы математической статистики» направлена на подготовку магистрантов к самостоятельной деятельности в области организации и проведения научных исследований и имеет цель дать необходимые для практической деятельности знания по математическому моделированию и методам математической статистики.

В результате освоения дисциплины «Математическое моделирование и методы математической статистики» у обучающегося должны сформироваться, определенные учебным планом компетенции, и магистрант должен:

Знать: классификацию задач математического моделирования; методологию математического моделирования, статистического анализа и обработки данных.

Уметь: разрабатывать и классифицировать математические модели; применять регрессионный, дисперсионный, компонентный анализ, математическую теорию планирования экспериментов, пользуясь современными программными средствами, при решении задач профессиональной деятельности.

Владеть: навыками использования методов математического моделирования, статистической обработки данных и программных средств при решении прикладных задач профессиональной деятельности.

Для достижения поставленных целей изучения дисциплины учебным планом предусмотрены лекции, практические занятия, лабораторные занятия, самостоятельная работа и оценка результатов обучения по дисциплине.

Для оценивания результатов обучения по дисциплине «Математическое моделирование и методы математической статистики» предусмотрены следующие виды контроля:

- текущий контроль (осуществление контроля всех видов аудиторной и внеаудиторной деятельности обучающегося с целью получения первичной информации о ходе усвоения отдельных элементов содержания дисциплины);
- промежуточная аттестация (оценивается уровень и качество подготовки по дисциплине в целом).

Текущий контроль успеваемости предназначен для регулярного и систематического оценивания хода освоения студентами дисциплины и выполнения всех видов учебной работы, предусмотренных учебным планом образовательной программы. Для обучающихся по заочной форме обучения текущий контроль не предусмотрен.

Текущий контроль успеваемости по дисциплине «Математическое моделирование и методы математической статистики» предусматривает две контрольные точки в течение учебного периода (семестра), каждая из которых обобщает результаты освоения отдельного тематического блока дисциплины. Контрольная точка (КТ) является сроком или датой завершения тематического блока и оценки достижения его образовательных целей. КТ не обязательно сопровождается контрольным мероприятием и может выставляться преподавателем по заранее установленным критериям на основании суммарных итогов текущей работы, степени и качества выполнения заданий. Итоги текущего контроля успеваемости по каждой контрольной точке вносятся преподавателем в электронный журнал.

Результаты текущего контроля подводятся по шкале балльно-рейтинговой системы. Регламент балльно-рейтинговой системы определен Положением о системе «Контроль успеваемости и рейтинг обучающихся». Оценивание результатов учебной работы обучающихся осуществляется в баллах по всем видам такой работы с применением 100-балльной шкалы. По суммарному результату текущего контроля может быть выставлена итоговая оценка по дисциплине если обучающийся набрал по текущему контролю необходимые и достаточные баллы для выставления оценки автоматом.

Промежуточная аттестация обучающихся в форме экзамена – установление уровня

достижения компетенций с помощью оценивания промежуточных и окончательных результатов обучения по дисциплине предусмотренных учебным планом. Порядок проведения промежуточной аттестации регламентируется «Положением о промежуточной аттестации обучающихся».

Методика и порядок оценивание результатов обучения по дисциплине представлены в документе «Оценочные материалы (оценочные средства) для проведения текущего контроля и промежуточной аттестации по дисциплине «Математическое моделирование и методы математической статистики».

## **2. ВИДЫ И ФОРМЫ ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ**

Самостоятельная работа является одной из важнейших составляющих образовательного процесса и представляет собой процесс активного, целенаправленного приобретения обучающимися новых знаний, умений без непосредственного участия преподавателя, характеризующийся предметной направленностью, эффективным контролем и оценкой результатов их деятельности. Самостоятельная работа является обязательной для каждого студента и определяется учебным планом.

*Цели самостоятельной работы:*

- углубление и расширение теоретических знаний;
- систематизация и закрепление полученных теоретических знаний и практических умений студентов;
- формирование умений использовать техническую, нормативную, правовую, справочную и специальную литературу;
- развитие познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развитие исследовательских навыков.

Учебный процесс включает два вида самостоятельной работы: аудиторную и внеаудиторную. Основные формы и виды самостоятельной внеаудиторной работы студентов включают:

- самостоятельная работа с источниками информации;
- составление опорного конспекта;
- подготовка доклада, реферата, презентации, отчета и т.п.;
- выполнение упражнений и заданий;
- подготовка к практическим занятиям;
- подготовку ко всем видам контроля, в том числе к экзамену;
- работу в студенческих научных обществах, кружках, семинарах и т.п.;
- участие в работе факультативов, семинаров, конференциях и т.п.;
- участие в научной и научно-методической работе;
- другие виды деятельности, организуемой и осуществляемой университетом и органами студенческого самоуправления.

Результатом самостоятельной работы является устный или письменный отчет студента в форме: сообщения, доклада, реферата, творческой работы и т.д.

Объем времени, отведенный на самостоятельную работу (в академических часах), определяется учебным планом и конкретизируется в рабочей программе учебной дисциплины с ориентировочным распределением по темам и указанием рекомендуемых форм выполнения самостоятельной работы и форм ее контроля.

Самостоятельная работа, не предусмотренная образовательной программой и учебным планом, осуществляется студентами инициативно с целью реализации собственных учебных и научных интересов.

### **3. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ И ВЫПОЛНЕНИЮ ВНЕАУДИТОРНОЙ САМОСТОЯТЕЛЬНОЙ РАБОТЫ**

#### **3.1. Лекционные занятия**

Лекционные занятия проводятся в аудитории с применением мультимедийного оборудования. Лекции имеют цель - систематизацию основы научных знаний по дисциплине и концентрации внимания обучающихся на наиболее актуальных проблемах, вопросах и положениях. В ходе лекций преподаватель излагает и разъясняет основные, наиболее сложные вопросы темы, а также теоретически связанные с ней проблемы. В ходе лекции, обучающиеся не ведут конспектирование учебного материала, процесс которого требует значительного времени от общего времени, отводимого на лекцию. В случае необходимости (например, не понял суть, требуется уточнение и т.п.) обучающиеся делают пометки в рабочей тетради. Данный подход, во-первых, облегчает работу студента на лекции, во-вторых, позволяет преподавателю более полно изложить и разъяснить учебный материал, и, в-третьих, появляется время для ответов на вопросы студентов, возникших в ходе лекции и выборочного опроса студентов в начале лекции с целью оценки усвоения предыдущего учебного материала.

В заключении лекции преподаватель дает рекомендации на подготовку к практическим, лабораторным занятиям и обязательные к выполнению указания на самостоятельную работу. Обучающимся в электронном виде выдаётся презентация занятия, текст лекции и указываются рекомендуемые источники информации, что в совокупности необходимо для отражения в конспекте учебного материала. Отдельные темы предлагаются для самостоятельного изучения также с обязательным отражением в конспекте.

Самостоятельная работа с лекционным материалом, в первую очередь заключается в составлении конспекта, и её следует проводить в тот же день. Рекомендуется запись осуществлять на одной стороне листа, а другую оставлять для дальнейшей проработки учебного материала. Начинать самостоятельную работу следует с просмотра презентации лекции, что позволит восстановить в памяти учебный материал занятия. Далее рекомендуется работа с текстом лекции и источниками информации, разрешая и уточняя в ходе чтения, возникшие ранее затруднения и вопросы, с одновременным оформлением конспекта.

Конспект позволяет привести в систему знания и ускоряет повторение материала, экономит время при повторном, после определенного перерыва, обращении к ранее изученной теме. Конспектирование – один из самых сложных этапов самостоятельной работы. Каких-либо единых, пригодных для каждого студента методов и приемов конспектирования не существует. Однако это не исключает соблюдения некоторых, наиболее оправдавших себя общих правил, с которыми преподаватель и обязан познакомить студентов.

Главное в конспекте – не его объем, а содержание. Умение излагать материал сжато и кратко приходит с опытом и знаниями, но их накоплению помогает соблюдение простых правил: не торопиться записывать при первом же чтении, необходимо вносить в конспект лишь то, что стало ясным.

Конспект не должен быть «слепым», безликим, состоящим из сплошного текста. Особо важные места, яркие примеры выделяются цветным подчеркиванием, взятием в рамочку, оттенением, пометками на полях специальными знаками, чтобы как можно быстрее найти нужное положение. В процессе составления конспекта для наглядности обязательно применять различные выделения, подзаголовки, создавая блочную структуру конспекта.

Это делает конспект легко воспринимаемым, удобным для работы. Работая над конспектом, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, работа с лекционным материалом позволит глубоко овладеть знаниями.

Изучение дисциплины следует начинать с учебника, поскольку учебник - это книга, в которой изложены основы научных знаний по определенному предмету в соответствии с целями и задачами обучения, установленными программой и требованиями дидактики. Предварительное чтение направлено на выявление в тексте незнакомых терминов и поиск их значения в справочной литературе. Сквозное чтение предполагает прочтение материала от начала до конца. Сквозное чтение литературы из приведенного списка дает возможность обучающемуся сформировать тезаурус основных понятий из изучаемой области и свободно владеть ими. Выборочное чтение имеет целью поиск и отбор материала. В рамках данной дисциплины выборочное чтение, как способ освоения содержания, должно использоваться при подготовке к практическим занятиям по соответствующим темам. Аналитическое чтение - это критический разбор текста с последующим его конспектированием. Целью изучающего чтения является глубокое и всестороннее понимание учебной информации.

Выбор вида ведения конспекта зависит от характера изучаемого материала и целей работы с ним. Если содержание материала несложное, легко усваиваемое, можно ограничиться составлением плана. План - это схема прочитанного материала, краткий перечень вопросов, отражающих структуру и последовательность материала. Подробно составленный план может заменить конспект. Если материал содержит новую и трудно усваиваемую информацию, целесообразно его максимально законспектировать. Конспект - это систематизированное, логичное изложение источника.

Различают четыре типа конспектов:

- План-конспект - это развернутый детализированный план, в котором достаточно подробные записи приводятся по тем пунктам плана, которые нуждаются в пояснении.
- Текстуальный конспект - это воспроизведение наиболее важных положений и фактов источника.
- Свободный конспект - это четко и кратко (изложенные) основные положения в результате глубокого осмысливания материала. В нем могут присутствовать цитаты, тезисы; часть материала может быть представлена планом.
- Тематический конспект - составляется на основе изучения ряда источников и дает более исчерпывающий ответ по каждому вопросу.

Выбор вида ведения конспекта обучающийся осуществляет самостоятельно, исходя из своих личностных качеств, но с учётом указаний преподавателя.

### **3.2. Практические занятия**

Целью практических занятий является получение базовых навыков по применению теоретических знаний. Это необходимо при решении всевозможных задач на различных этапах практической деятельности.

*Задачи практических занятий:*

- закрепление и углубление знаний;
- создание практических навыков и умений в практической деятельности и повседневной жизни для: поиска, анализа и использования информации;
- анализа конкретных ситуаций и условий их реализации; изложения и аргументации собственных суждений по практической ситуации;
- развитие стремления и способности к самостоятельному исследованию изучаемых реальностей, их критической оценки.

Подготовка к практическим занятиям включает повторение теоретического (лекционного) материала и выполнение указаний преподавателя на подготовку к практическим

занятиям. Значительную помощь в подготовке к практическим занятиям оказывают тематические сайты.

Практические занятия по дисциплине проводятся в различных формах. В частности, в форме дискуссий, круглых столов, анализа конкретных ситуаций с последующим выполнением индивидуальных заданий. Тематика и содержание индивидуальных заданий формируются индивидуально в соответствии с темами магистерских диссертаций и рекомендациями научных руководителей.

### **3.3. Лабораторные занятия**

Лабораторное занятие – одна из основных форм организации учебного процесса, направленная на творческое усвоение теоретических основ учебной дисциплины и получение практических навыков исследования путем постановки, проведения, обработки и представления результатов эксперимента на основе практического использования различных средств (наблюдения, измерения, контроля, вычислительной техники), приобретения навыков опыта творческой деятельности.

Лабораторная работа - конкретное учебное задание по изучаемой дисциплине, выполняемое на лабораторном занятии.

Цель лабораторного занятия - практическое освоение обучающимися содержания и методологии изучаемой дисциплины при использовании специальных средств.

Основными задачами лабораторных занятий являются:

- приобретение опыта решения учебно-исследовательских и реальных практических задач на основе изученного теоретического материала;
- приобретение опыта проведения эксперимента, в том числе вычислительного;
- овладение новыми методиками экспериментирования в соответствующей отрасли науки, техники и технологии;
- приобретение умений и навыков эксплуатации технических средств оборудования;
- формирование умений обработки результатов проведенных исследований;
- анализ и обсуждение полученных результатов и формулирование выводов;
- выработка способности логического осмысления самостоятельно полученных знаний;
- обеспечение рационального сочетания коллективной и индивидуальной форм обучения.

Преподаватель, который проводит лабораторные занятия, имеет право определять содержание лабораторных работ, выбирать методы и средства проведения лабораторных исследований, наиболее полно отвечающие их особенностям и обеспечивающие высокое качество учебного процесса.

Лабораторное занятие состоит из следующих элементов: вводная часть, основная и заключительная.

Вводная часть обеспечивает подготовку обучающихся к выполнению заданий работы. В ее состав входят:

- формулировка темы, цели и задач занятия, обоснование его значимости в профессиональной подготовке обучающихся;
- изложение теоретических основ работы;
- характеристика состава и особенностей заданий работы и объяснение методов (способов, приемов) их выполнения;
- характеристика требований к результату работы;

- инструктаж по технике безопасности при эксплуатации технических средств;
- проверка готовности обучающихся выполнять задания работы;
- указания по самоконтролю результатов выполнения заданий обучающимися.

Основная часть включает процесс выполнения лабораторной работы, оформление отчета и его защиту. Она может сопровождаться дополнительными разъяснениями по ходу работы, устранением трудностей при ее выполнении, текущим контролем и оценкой результатов отдельных обучающимися, ответами на вопросы обучающихся.

Заключительная часть содержит:

- подведение общих итогов занятия;
- оценку результатов работы отдельных обучающихся;
- ответы на вопросы обучающихся;
- выдачу рекомендаций по устранению пробелов в системе знаний и умений обучающихся, по улучшению результатов работы;
- сбор отчетов обучающихся для проверки, изложение сведений, касающихся подготовки к выполнению следующей работы.

В ряде случаев может быть целесообразным вынести отдельные практические (лабораторные) занятия для самостоятельного внеаудиторного выполнения. Особенно эффективно использовать такие формы работы при формировании профессиональных компетенций, связанных с получением, переработкой и систематизацией информации. Преимущество этой формы заключается в возможности подготовки индивидуальных заданий и последующего обсуждения и оценивания результатов их выполнения на аудиторных занятиях.

### **3.4. Подготовка рефератов**

Реферат представляет собой сочинение на определенную тему, включающее обзор соответствующих научных и нормативных источников или изложение сути монографии, статьи, то есть, первичного документа (или его части) с основными фактическими сведениями и выводами. Реферат является самостоятельной творческой работой, выполненной и оформленной аналогично требованиям, изложенным в Положении «Правила оформления и требования к содержанию курсовых проектов (работ) и выпускных квалификационных работ». Рекомендуемый объем реферата – от 10 до 15 печатных страниц.

При работе над рефератом изначально необходимо составить план, который раскрывал бы заявленную проблему. Единой схемы составления плана не существует, главное в том, чтобы его пункты соответствовали логической схеме. Первый элемент плана – введение, далее следует основная часть, заключение, список используемых источников.

Во введении необходимо обосновать актуальность проблемы, наметить пути ее анализа и решения, дать краткий аналитический обзор литературы и обосновать структуру изложения материала.

Основная часть должна представлять последовательное изложение вопросов плана, каждому из вопросов должен предшествовать заголовок. Рекомендуется весь материал равномерно распределить по вопросам, а изложение каждого вопроса завершить кратким выводом. Содержание каждого из разделов (параграфов) должно раскрывать его название.

В заключении необходимо сделать концептуальный вывод по всей работе, который раскрывал бы главную проблему.

Библиографический список завершает работу. В него необходимо включить все источники информации, с которыми ознакомился автор, а не только те, которые были цитированы в тексте работы. Они должны быть сгруппированы следующим образом:

- нормативные акты (по юридической силе в нисходящем порядке, внутри группы равных по юридической силе – по дате издания);
- специальная литература (в алфавитном порядке);

- неопубликованные материалы;
- практические материалы.

Текст полностью написанной и оформленной работы подлежит тщательной проверке. Ошибки и опiski, как в тексте, так и в цитатах и в научно-справочном аппарате не допустимы.

Результаты работы студент должен публично представить. Предварительно подготовив тезисы доклада, студент в течение 5-7 минут должен кратко изложить основные положения своей работы. Чтобы выступление было удачным, оно должно хорошо восприниматься на слух, быть интересным для слушателей. При выступлении приветствуется активное использование мультимедийного сопровождения доклада (презентация, видеоролики, аудиозаписи).

### **3.5. Подготовка эссе**

Эссе - самостоятельная творческая письменная работа, представляющая собой развернутое и аргументированное изложение точки зрения обучающегося по предложенной теме.

Структурно эссе состоит из четырех элементов:

1. Введение. Во введении излагается в общем виде та позиция, которую предполагается отстаивать в основной части эссе.
2. Основная часть. В основной части представлены подробные ответы на вопрос или излагается позиция, подтверждаемая теоретическими аргументами и эмпирическими данными.
3. Заключение. В заключении резюмируются главные идеи основной части, подводящие к предполагаемому ответу на вопрос или заявленной точке зрения, делаются выводы.
4. Список использованных источников. Указывается литература, которая используется при подготовке эссе.

Мнение автора эссе по проблеме излагаются в форме кратких тезисов. Мнение должно быть подкреплено доказательствами - поэтому за тезисом следуют аргументы. Аргументы - это факты, явления общественной жизни, события, жизненные ситуации и жизненный опыт, научные доказательства, ссылки на мнения ученых и др. Лучше приводить два-три аргумента в пользу каждого тезиса: один аргумент кажется неубедительным. Таким образом, эссе приобретает кольцевую структуру (количество тезисов и аргументов зависит от темы, избранного плана, логики развития мысли).

При подготовке основной части эссе рекомендуется использовать формулу П-О-П-С:

- П - положение (утверждение);
- О - объяснение;
- П - пример, иллюстрация;
- С - суждение (итоговое).

### **3.6. Подготовка научного доклада**

Научный доклад представляет собой исследование по конкретной проблеме, изложенное перед аудиторией слушателей. Работа по подготовке научного доклада включает не только знакомство с литературой по избранной тематике, но и самостоятельное изучение определенных вопросов.

Научный доклад может быть подготовлен для выступления на занятии, конференции научного общества, или в рамках круглого стола. В любом случае успешное выступление во многом зависит от правильной организации самого процесса подготовки научного доклада. Подготовка научного доклада включает несколько этапов работы:

1. выбор темы доклада;
2. подготовка материалов;

3. работа над текстом доклада;
4. подготовка к выступлению.

Подготовка к научному докладу начинается с выбора темы. Не следует выбирать слишком широкую тему научного доклада. Это связано с ограниченностью докладчика во времени. Доклад должен быть рассчитан на 7-15 минут. Работа по подбору материалов для доклада связана с изучением соответствующей тематике литературы.

В процессе работы над текстом доклада обучающиеся глубже постигают вопросы изучаемого предмета, поскольку:

- анализируются различные точки зрения, факты и события;
- ведется научно обоснованная полемика;
- обобщается материал;
- лаконично излагаются мысли;
- правильно оформляется работа с составлением плана библиографии и систематизацией информации.

*Текст научного доклада должен включать три основные части:*

- введение - краткое знакомство слушателей с обсуждаемой проблемой;
- основную часть - логическое продолжение вопросов, обозначенных автором во введении. В этой части доклада раскрывается тема выступления, приводятся необходимые доказательства (аргументы);
- заключение - обобщение основной мысли и идеи выступления. В заключении можно кратко повторить основные выводы и утверждения, прозвучавшие в основной части доклада.

Научный доклад представляет собой устное произведение и чтение вслух подготовленного текста недопустимо. К особенностям хорошего выступления можно отнести следующее:

- до и после важных мыслей следует делать паузу;
- для большего акцента рекомендуется менять тон голоса и тембр речи (это сделает речь более выразительной);
- необходимо сохранять контакт с аудиторией во время выступления.

Таким образом, Вступление представляет собой краткое знакомство слушателей с обсуждаемой в докладе проблемой. Во вступлении должна быть сформулирована главная мысль, которая затем займет центральное место.

Основная часть доклада является логическим продолжением вопросов, обозначенных автором во введении. Именно в этой части доклада предстоит раскрыть тему выступления, привести необходимые доказательства (аргументы).

Для того чтобы правильно построить основную часть своего доклада, необходимо составить ее подробный план. Важность составления такого плана связана с основной задачей автора. Он должен в течение времени, отведенного на основную часть, суметь представить и изложить авторскую точку зрения по обозначенной в теме доклада проблеме. Наличие подробного плана позволяет выполнить эту задачу, дает возможность автору в сжатой форме донести свои идеи до аудитории и уложиться в установленный регламент.

Заключение имеет целью обобщить основные мысли и идеи выступления. В заключении можно кратко повторить основные выводы и утверждения, прозвучавшие в основной части доклада. На заключение можно возложить также функцию обобщения всего представленного докладчиком материала.

После доклада студент отвечает на вопросы. На основе обсуждения написанного и доложенного реферата ему выставляется соответствующая оценка.

### **3.7. Подготовка презентации**

Презентация – документ или комплект документов, предназначенный для представления чего-либо (организации, проекта, продукта и т.п.). Цель презентации – донести до целевой аудитории полноценную информацию об объекте презентации в удобной и наглядной форме. Мультимедийные презентации используются для того, чтобы выступающий смог на большом экране наглядно продемонстрировать материалы к своему сообщению.

*Электронные презентации выполняются в программе MS PowerPoint в виде слайдов в следующем порядке:*

- титульный лист с заголовком темы и автором презентации;
- план презентации (5-6 пунктов - это максимум);
- основная часть (не более 10 слайдов);
- заключение (вывод).

*Общие требования к стилевому оформлению презентации*

Для оформления слайдов презентации рекомендуется использовать простые шаблоны без анимации, соблюдать единый стиль оформления всех слайдов. Не рекомендуется на одном слайде использовать более 3 цветов: один для фона, один для заголовков, один для текста. Шрифт, выбираемый для презентации должен обеспечивать читаемость на экране и соответствовать выбранному шаблону оформления. Не следует использовать разные шрифты в одной презентации.

Презентация не должна быть меньше 10 слайдов. Первый слайд – это титульный лист, на котором обязательно должны быть представлены: полное наименование образовательного учреждения, тема работы, фамилия, имя, отчество студента, фамилия, имя, отчество руководителя. Следующим слайдом должна быть цель работы и решаемые в ней задачи. Алгоритм дальнейшего выстраивания презентации соответствует логической структуре работы и отражает последовательность ее этапов. Для оптимального отбора содержания материала работы в презентации необходимо выделить ключевые понятия, теории, проблемы, которые раскрываются в презентации в виде схем, диаграмм, таблиц. На каждом слайде определяется заголовок по содержанию материала. Все слайды должны быть выдержаны в одном стиле и пронумерованы.

### **3.8. Работа с учебными пособиями, монографиями, периодикой**

Грамотная работа с источниками информации, особенно если речь идет о научной литературе, предполагает соблюдение ряда правил, овладению которыми необходимо настойчиво учиться. При работе с источниками информации недопустим формальный, поверхностный подход. Не механическое заучивание, не простое накопление цитат, выдержек, а сознательное усвоение прочитанного, осмысление его, стремление дойти до сути – вот главное правило.

Другое правило – соблюдение при работе над книгой определенной последовательности. Вначале следует ознакомиться с оглавлением, содержанием предисловия или введения. Это дает общую ориентировку, представление о структуре и вопросах, которые рассматриваются в книге. Следующий этап – чтение. Первый раз целесообразно прочитать книгу с начала до конца, чтобы получить о ней цельное представление. При повторном чтении происходит постепенное глубокое осмысление каждой главы, критического материала и позитивного изложения, выделение основных идей, системы аргументов, наиболее ярких примеров и т. д.

Непременным правилом чтения должно быть выяснение незнакомых слов, терминов, выражений, неизвестных имен, названий. Для этой цели можно завести специальную тетрадь или блокнот. Важная роль в связи с этим принадлежит библиографической подготовке студентов. Она включает в себя умение активно, быстро пользоваться научным аппаратом книги, справочными изданиями, каталогами, умение вести поиск необходимой информации, обрабатывать и систематизировать ее. Научная методика работы с литературой предусматривает также ведение конспекта прочитанного.

### 3.9. Подготовка к промежуточной аттестации

Промежуточная аттестация по дисциплине – экзамен. Направлен на проверку конкретных результатов обучения, выявление овладения обучающимися системой знаний, умений и навыков, компетенций, полученных в процессе изучения дисциплины.

*Основные критерии оценки:*

- уровень освоения студентом теоретического материала;
- умение студента использовать теоретические положения при выполнении практических задач;
- четкость и обоснованность изложения ответа.

Подготовка обучающихся к промежуточной аттестации включает в себя два этапа: самостоятельная работа в течение семестра и непосредственная подготовка в дни, предшествующие промежуточной аттестации.

Основным источником подготовки к экзамену должен являться конспект лекций, где учебный материал должен быть представлен в систематизированном виде, основные положения его детализируются, подкрепляются современными фактами и информацией, которые в силу новизны не вошли в опубликованные печатные источники.

В ходе подготовки к экзамену студентам необходимо обращать внимание не только на уровень запоминания, но и на степень понимания излагаемых проблем. Особое внимание при подготовке к экзамену необходимо уделить терминологии, т.к. успешное овладение любой дисциплиной предполагает усвоение основных понятий, их признаков и особенности.

Ответы на теоретические вопросы должны быть даны в соответствии с формулировкой вопроса и содержать не только изученный теоретический материал, но и собственное понимание проблемы. В ответах желательно привести примеры из практики. Если в процессе подготовки к промежуточной аттестации во время самостоятельной работы над изучением теоретического материала или при решении задач у обучающегося возникают вопросы, разрешить которые самостоятельно не удастся, необходимо обратиться к преподавателю для получения у него разъяснений или указаний. В своих вопросах обучающийся должен четко выразить, в чем он испытывает затруднения, характер этого затруднения. За консультацией следует обращаться и в случае, если возникнут сомнения в правильности ответов на вопросы самопроверки.

## **4. ВОПРОСЫ ЭКЗАМЕНА ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ»**

1. Понятие модели, цели моделирования. Системный и классический подходы при моделировании.
2. Стадии разработки моделей. Классификация моделей.
3. Основные этапы построения математической модели. Математическая модель системы, ее составляющие. Системные свойства модели.
4. Типовые математические схемы моделирования систем.
5. Непрерывно-детерминированные модели (D-схемы).
6. Непрерывно-стохастические модели (Q-схемы). Марковские случайные процессы.
7. Граф состояний системы массового обслуживания. Уравнения Колмогорова.
8. Классификация систем массового обслуживания и их математические модели.
9. Статистический анализ. Генеральная совокупность и выборка. Репрезентативность выборки. Дискретный и интервальный вариационные ряды распределения.
10. Точечные оценки характеристик распределения. Несмещенность, эффективность,

- состоятельность оценок.
11. Интервальные оценки. Доверительный интервал и доверительная вероятность.
  12. Коэффициенты ковариации, корреляции и детерминизации, их свойства.
  13. Точечные оценки коэффициентов ковариации и корреляции.
  14. Парная регрессия. Линейная модель парной регрессии. Нахождение параметров по методу наименьших квадратов.
  15. Нелинейная регрессия. Виды уравнений нелинейной регрессии и критерии их выбора.
  16. Дисперсионный анализ. Общая, факторная и остаточная суммы, связь между ними.
  17. Общая, факторная и остаточная дисперсии, приходящиеся на одну степень свободы. F –критерий Фишера.
  18. Оценка существенности уравнения регрессии в целом о F -критерию Фишера и по критерию средней ошибки аппроксимации.
  19. Оценка существенности параметров уравнения регрессии по t -критерию Стьюдента.
  20. Временные ряды. Автокорреляция уровней временного ряда. Построение трендов временного ряда.
  21. Множественная регрессия. Отбор факторов при построении уравнения множественной регрессии.
  22. Линейная модель множественной регрессии. Система уравнений для определения параметров на основе метода наименьших квадратов.
  23. Множественная корреляция и ее матрица.
  24. Нелинейные модели множественной регрессии. Линеаризация уравнений.
  25. Множественная регрессия в стандартизованном масштабе. Смысл стандартизованных коэффициентов регрессии.
  26. Свойства остатков регрессионной модели (закон распределения, несмещенность, эффективность, состоятельность, гомоскедастичность, коррелированность).
  27. F -критерий Фишера для оценки значимости уравнения множественной регрессии. Критерий средней ошибки аппроксимации.
  28. Инструментальные средства для проведения регрессионного, корреляционного, дисперсионного анализа.
  29. Компонентный анализ и факторный анализ, их применения. Алгебраические основы метода.
  30. Матрицы нагрузок, счетов и ошибок при компонентном анализе.
  31. Кластерный анализ, расстояние между объектами кластерами.
  32. Иерархическая кластеризация, ее методы.
  33. Кластеризация методом k-средних.
  34. Инструментальные средства для проведения компонентного и кластерного анализа.
  35. Математические методы планирования экспериментов Математическая модель объекта исследования при планировании эксперимента. Функция отклика.
  36. Кодирование (нормирование) факторов. Типы планирования. Матрица плана эксперимента.
  37. Нахождение коэффициентов полинома функции отклика и оценка его значимости.
  38. Инструментальные средства для реализации математических методов планирования эксперимента.

## 5. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Основная литература				
	Авторы, составители	Заглавие	Издательство, год	Количество
Л1.1	Шинаков, Ю.С.	Изучение элементов и технологии применения подсистемы моделирования динамических процессов SIMULINK (MATLAB R2014b): практикум	Москва: Московский технический университет связи и информатики, 2015	ЭБС
Л1.2	Завьялов, О.Г., Подповетная, Ю.В.	Теория вероятностей и математическая статистика с применением Excel и Maxima: учебное пособие	Москва: Прометей, 2018	ЭБС
Л1.3	Кремер Наум Шелевич	Математическая статистика: Учебник и практикум	Москва: Издательство Юрайт, 2019	ЭБС
Л1.4	Дьяконов, В.П.	MATLAB R2006/2007/2008 + Simulink 5/6/7. Основы применения	Москва: СОЛОН-ПРЕСС, 2017	ЭБС
Л1.5	Балдин Константин Васильевич, Башлыков Виктор Николаевич	Теория вероятностей и математическая статистика: Учебник	Москва: Издательско-торговая корпорация "Дашков и К", 2020	ЭБС
Л1.6	Алибеков, И.Ю.	Теория вероятностей и математическая статистика в среде MATLAB: учебное пособие для вузов	Санкт-Петербург: Лань, 2021	ЭБС
Дополнительная литература				
	Авторы, составители	Заглавие	Издательство, год	Количество
Л2.1	Васильев, А.Н.	Matlab [Электронный ресурс]: самоучитель	Санкт-Петербург: Наука и Техника, 2015	ЭБС
Л2.2	Лагутин Михаил Борисович	Наглядная математическая статистика: Учебное пособие	Москва: Лаборатория знаний, 2019	ЭБС
Л2.3	Коткин, Г.Л., Попов, Л.К.	Компьютерное моделирование физических процессов с использованием MATLAB: учебное пособие	Новосибирск: Новосибирский государственный университет, 2017	ЭБС
Методические разработки				
	Авторы, составители	Заглавие	Издательство, год	Количество
Л3.1		Теория вероятностей и математическая статистика: метод. указания к практическим занятиям	Ростов н/Д.: ИЦ ДГТУ, 2018	ЭБС
Л3.2		Методические указания для практических занятий «Создание приложений с графическим интерфейсом пользователя в системе MATLAB» по дисциплине «Компьютерные технологии в приборостроении»	Ростов н/Д.: ИЦ ДГТУ, 2018	ЭБС

## 6. ТЕКСТЫ ЛЕКЦИЙ. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ВЫПОЛНЕНИЯ ЛАБОРАТОРНЫХ РАБОТ И КОНТРОЛЬНОЙ РАБОТЫ

**Л 1. Введение в дисциплину  
«Математическое моделирование и  
методы математической статистики»**

***Цель*** изучения дисциплины «Математическое моделирование и методы математической статистики» состоит в освоении методологии и технологии моделирования (в первую очередь компьютерного), включая статистические модели

# Структура лекции

- Понятие модели, математические модели, этапы моделирования, классификация моделей, типовые математические схемы моделирования
- Непрерывно-детерминированные модели (D-схемы).  
Моделирование динамических систем. ЛР-1
- Непрерывно-стохастические модели (Q-схемы).  
Моделирование технологического участка. ЛР-2
- Задания 1,2 контрольной работы

# Понятие модели

**Моделирование** – это замещение исследуемого объекта (оригинала) его условным образом или другим объектом(моделью) и изучение свойств оригинала путем исследования свойств модели.

**Требования к модели:**

**Валидация** – адекватность модели реальному объекту.

**Верификация** - соответствие структуры и логики модели замыслу моделирования.

(Валидация подтверждает, что «вы создали правильный продукт», а верификация подтверждает, что «вы создали продукт таким, каким и намеревались его сделать»)

**Математическое моделирование** - это идеальное научное знаковое формальное моделирование, при котором описание объекта осуществляется на языке математики, а исследование модели проводится с использованием математических методов.

# Параметры и переменные математической модели

В общем случае параметры и переменные, описывающие состояние и поведение объекта моделирования разбиваются на четыре типа

- совокупность входных (управляемых) воздействий на объект  $X$ ;
- совокупность воздействий внешней среды (неуправляемых)  $U$ ;
- совокупность внутренних (собственных) параметров объекта  $I$ ;
- совокупность выходных характеристик  $Y$ .

## **Этапы моделирования**

**Процесс разработки и реализации модели включает четыре этапа:**

- постановка задачи и построение концептуальной модели;**
- разработка модели: алгоритм и программная реализация модели;**
- проведение компьютерных экспериментов с моделью;**
- анализ и интерпретация полученных результатов.**

## **1) Построение концептуальной (содержательной) модели включает:**

- **постановку целей и задач моделирования.**
- **структуризацию модели (рекомендуется производить с привлечением методологии и средств системного моделирования, например, универсального языка моделирования UML, технологии IDEF0.**
- **выдвижение гипотез и определение законов, влияющих на объект.**
- **определение параметров и переменных модели.**
- **обоснование выбора показателей и критериев эффективности системы.**
- **составление содержательного описания модели по концептуальному описанию.**

## **2) Разработка модели: алгоритм и программная реализация модели включают:**

- **построение логической схемы алгоритма.**
- **получение математических соотношений, содержащее аналитическую часть в виде явных функций и имитационную часть в виде моделирующего алгоритма (математическая модель).**
- **проверку достоверности алгоритма.**
- **выбор вычислительных и программных средств разработки - среда и язык программирования;**
- **программирование.**

### **3) Проведение машинных экспериментов с моделью и интерпретация:**

- **планирование машинного эксперимента.**
- **проведение модельных экспериментов согласно разработанному плану.**
- **представление результатов моделирования (таблицы, графики, диаграммы, мнемосхемы и т.д.)**

### **4) Анализ и интерпретация результатов моделирования:**

- ***определение* разного рода характеристик случайных величин и процессов, выполнение анализов - дисперсионного, регрессионного, факторного и др.**

**перевод результатов в термины *предметной области*.**

**выдача рекомендаций по функционированию или оптимизации объекта**

# Классификация математических моделей

- по методам решения
- по параметрам модели
- по виду оператора модели
- по сложности объекта моделирования
- по степени абстракции
- по целям моделирования

# Классификация математических моделей по методам решения



Для создания **аналитических моделей** используются математические методы и универсальные математические пакеты:

MATLAB, MathCAD, Maple, Mathematica, MS Excel, Scientific Workplace и др

Основаниями для применения **алгоритмических моделей** являются :

- невозможность решить задачу аналитически из-за ее чрезмерной математической сложности или громоздкости;
- требуется визуализация динамики происходящих в ней процессов;

При **численном подходе** совокупность математических соотношений модели заменяется конечномерным аналогом. Чаще всего это достигается дискретизацией исходных соотношений.

**Имитационные модели** являются разновидностью алгоритмических моделей и реализуют наиболее сложные и громоздкие алгоритмы описания объектов и систем.

# Классификация ММ по параметрам модели



# Классификация ММ по виду оператора модели



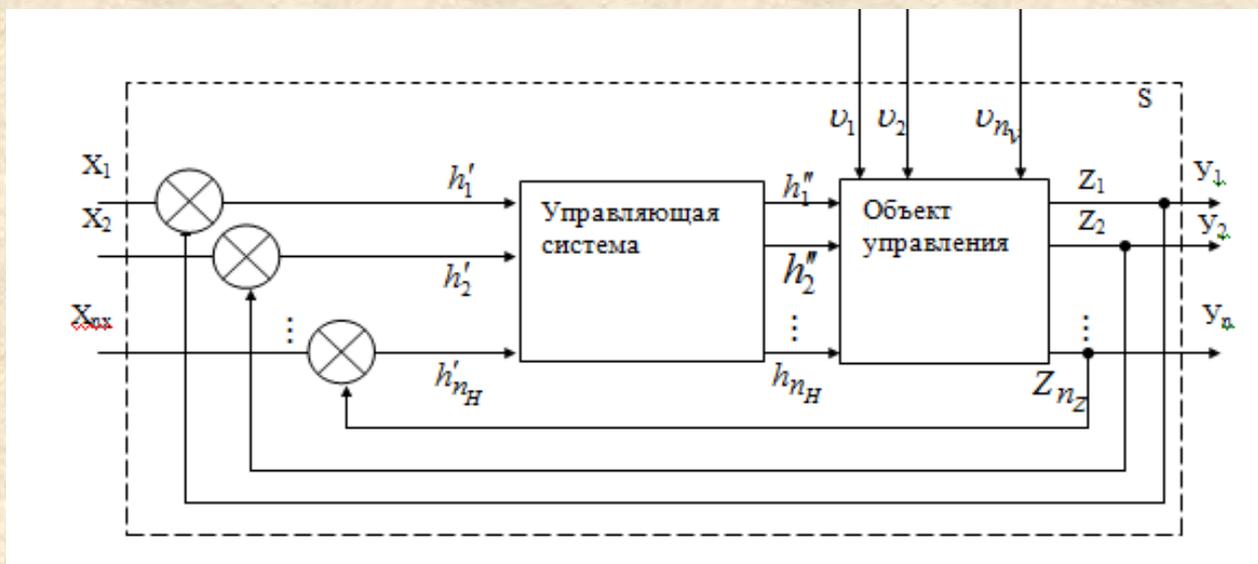
# Типовые математические схемы моделирования систем

- 1) непрерывно-детерминированные или D-схемы (дифференцированные уравнения и системы ДУ);
- 2) дискретно-детерминированные или F-схемы (конечные автоматы);
- 3) дискретно-стохастические или P-схемы (вероятностные автоматы);
- 4) непрерывно-стохастические или Q-схемы (системы СМО)
- 5) Сетевые или N-схемы (сети Петри)
- 6) Обобщенные, универсальные или A-схемы (агрегативные системы).

# Непрерывно-детерминированные модели (D-схемы)

- **Непрерывно-детерминированные модели** используются для анализа и проектирования динамических систем с непрерывным временем, процесс функционирования которых описывается детерминированными зависимостями. Данный класс систем может быть описан с помощью дифференциальных, интегральных, интегрально-дифференциальных и общих функциональных уравнений.
- Обычно в математических моделях, описываемых дифференциальными уравнениями, в качестве независимой переменной, от которой зависят неизвестные искомые функции, служит время  $t$ . Математические схемы такого вида отражают динамику изучаемой системы и носит название *D-схем*, от английского слова dynamic (динамика).
- Использование D-схем позволяет формализовать процесс функционирования непрерывно-

## Система автоматического управления (САУ) как пример непрерывно-детерминированной модели



**Структурная схема САУ**

вектор входных воздействий  $x$ ; вектор возмущающих воздействий  $v$ ; вектор сигналов ошибок  $h'$ ; вектор управляющих воздействий  $h''$ ; вектор состояний системы  $z$ ; вектор выходных переменных  $y$

# Математическая модель САУ – D схема

$$F(y^{(n)}, y^{(n-1)}, y, x^{(m)}, x^{(m-1)}, x) = 0$$
$$y(t_0) = y_0, y'(t_0) = y_{01}, \dots$$

$$\frac{d\vec{y}}{dt} = \vec{f}(\vec{y}, t), \quad \vec{y}(t_0) = \vec{y}_0$$

**D - схемы являются математическим аппаратом теории систем автоматического управления.**

**САУ – это совокупность программно-технических средств, которые обеспечивают достижение объектом управления заданной цели.**

**Насколько точно он достигает заданной цели можно судить по вектору выходных переменных  $y$ . Разница между заданным  $y$  и действительным искаженным  $y$  составляет ошибку управления  $h'$ .**

# Понятие устойчивости системы

$$\frac{d\vec{y}}{dt} = \vec{f}(\vec{y}, t), \quad \vec{y}(t_0) = \vec{y}_0 \quad \Rightarrow \quad \vec{y}(t) - \text{решение}$$

$$\frac{d\vec{y}}{dt} = \vec{f}(\vec{y}, t), \quad \vec{y}(t_0) = \vec{y}_{\varepsilon 0} \quad \Rightarrow \quad \vec{y}_{\varepsilon}(t) - \text{возм. решение}$$

$$|\vec{y}_{\varepsilon 0} - \vec{y}_0| < \sigma \quad \Rightarrow \quad |\vec{y}_{\varepsilon}(t) - \vec{y}(t)| < \varepsilon,$$

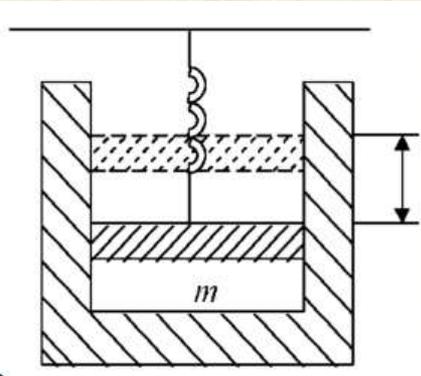
то устойчивое решение

## Линейные динамические системы. Случай свободных и вынужденных колебаний

В случае свободных колебаний, если правая часть ДУ  $x(t)=0$  возможны незатухающие, затухающие колебания и отсутствие колебаний.

При взаимодействии изучаемой системы с внешней средой появляется входное воздействие  $x(t)$ , которое в различных системах может быть внешней силой, источником энергии, и т. д. Такая система описывается ДУ с правой частью. В этом случае в системе может быть резонанс.

$$h_0 \frac{d^2 z}{dt^2} + h_1 \frac{dz}{dt} + h_2 \cdot z = x(t)$$



# **Нелинейные динамические системы**

## **Модель Лотки – Вольтерра (хищник-жертва)**

Модель Лотки – Вольтерра (амер. и итальян. математики нач. 20в), называемая классической моделью взаимодействия популяций, позволяет решить практические задачи, связанные с конкурентными процессами в различных областях человеческой деятельности, в частности,

- модели взаимодействия загрязнения с окружающей средой;
- модель классовой борьбы;
- модель военных действий;
- вирусная модель инфекционного заболевания;
- модель распространения эпидемий, включая модель заражения вирусом компьютеров;
- модель взаимодействия когнитивных и/или эмоциональных мод мозга.

# Система ДУ модели Лотки – Вольтерра

Модель «хищник-жертва», в которой рассматриваются популяции хищников и популяции жертв, взаимодействующих в какой-то среде: жертвы едят растительность, хищники — жертв, описывается системой ДУ

$$\begin{cases} \dot{x} = (\alpha - \beta y) x \\ \dot{y} = (-\gamma + \delta x) y \end{cases}$$

$$x(0) = x_0, y(0) = y_0$$

- $x$  — численность жертв (травоядных);
- $y$  — численность хищников;
- $\alpha$  — вероятность того, что травоядные размножатся;
- $\beta$  — вероятность того, что травоядное будет съедено хищником;
- $\gamma$  — вероятность того, что хищник умрет от голода;
- $\delta$  — вероятность того, что хищнику хватит еды на дальнейшее размножение.

$$\begin{cases} \dot{x} = (\alpha - \beta y) x \\ \dot{y} = (-\gamma + \delta x) y \end{cases}$$

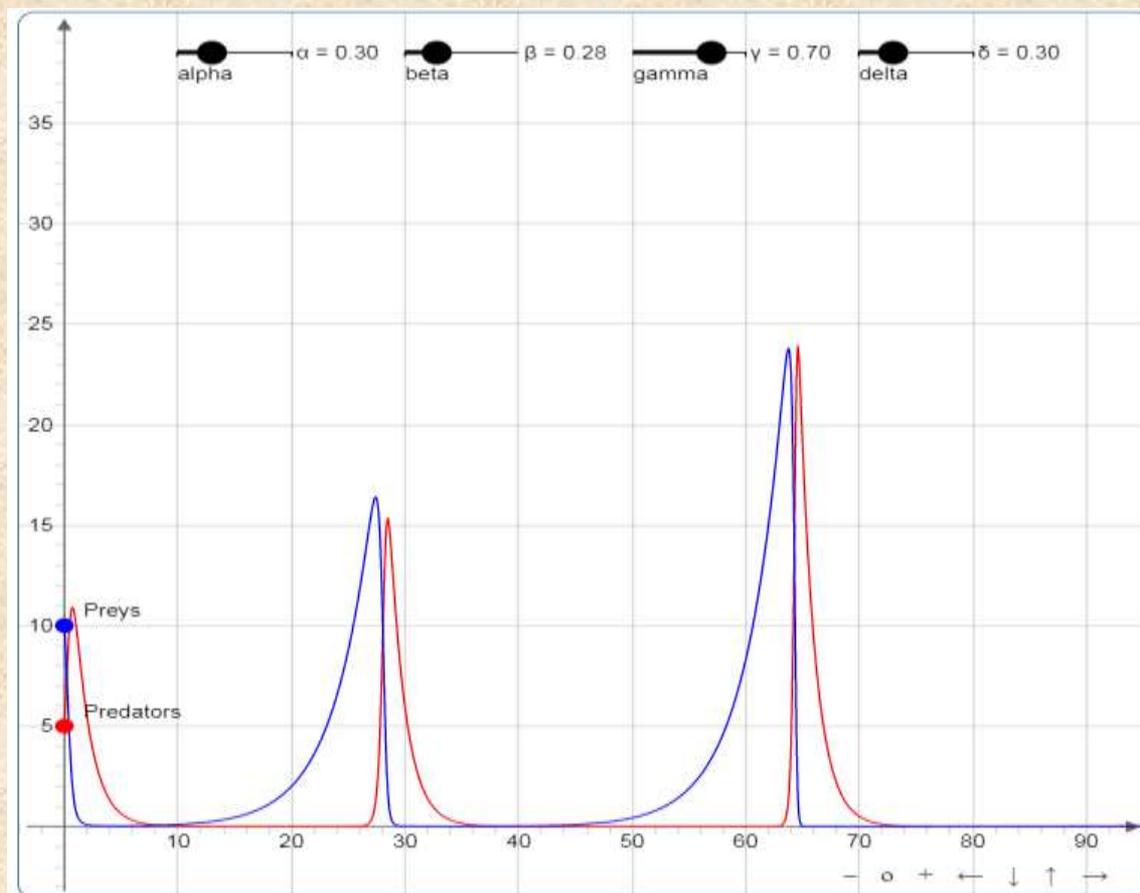
Из системы сразу следует, что если жертв нет ( $x = 0$ ), то хищники будут вымирать экспоненциально с неким начальным коэффициентом ( $\gamma$  согласно уравнению).

▶ Иллюстрация

$$\dot{y} = -\gamma \cdot y \Rightarrow y = C_1 \cdot e^{-\gamma \cdot t}, C_1 \in \mathbb{R}$$

Схожую ситуацию получаем при полном отсутствии хищников ( $y = 0$ ):

## Зависимости численности хищников и жертв от времени



## Особые точки

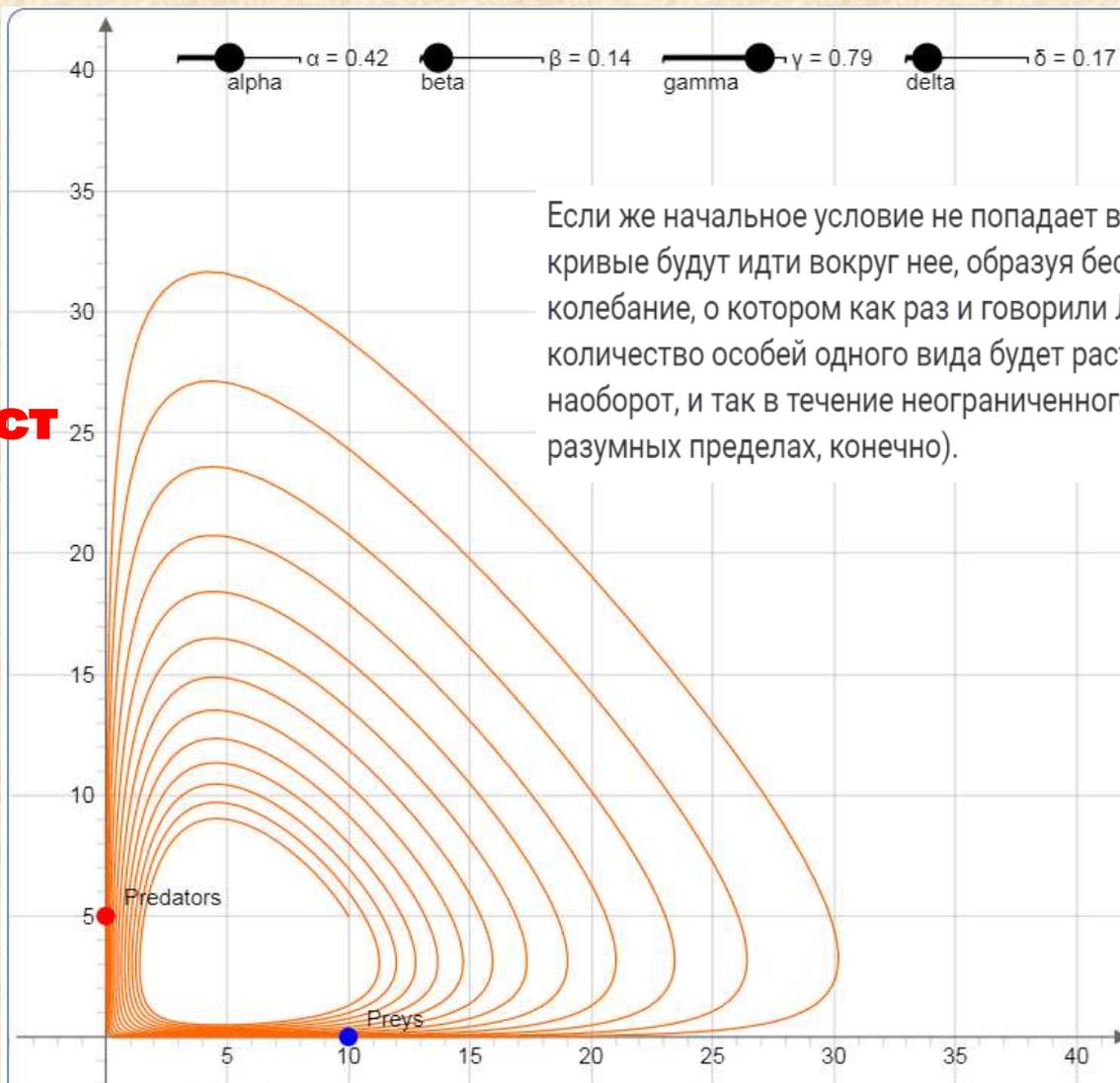
Найдем особые точки, которыми обладает система:

$$\begin{cases} (\alpha - \beta y) x = 0 \\ (-\gamma + \delta x) y = 0 \end{cases} \longrightarrow \begin{cases} \alpha x = \beta xy \\ \gamma y = \delta xy \end{cases} \xrightarrow{x \neq 0, y \neq 0} \begin{cases} y(0) = \frac{\alpha}{\beta} \\ x(0) = \frac{\gamma}{\delta} \end{cases}$$

Понятно, что при  $x(0) = 0, y(0) = 0$  особой точкой будет как раз  $(0, 0)$ , но этот случай не интересен, так как в нулевой момент времени животные обоих видов отсутствуют и, что логично, дальше не появляются.

Гораздо более интересные вещи происходят в ненулевом случае. В

# Фазовый портрет - зависимость $x$ от $y$



Если же начальное условие не попадает в особую точку, фазовые кривые будут идти вокруг нее, образуя бесконечное циклическое колебание, о котором как раз и говорили Лотка и Вольтерра. То есть количество особей одного вида будет расти, другого – падать, затем наоборот, и так в течение неограниченного количества времени (в разумных пределах, конечно).

# ЛР-1. Моделирование конкуренции

Постановка задачи. Исследовать противодействие двух сторон, если известно, что оно описывается моделью Лотки-Вольтерра. Известны

$x_0$  и  $y_0$  - начальные численности сторон;

$\alpha$  - вероятность увеличения численности первой стороны за счет собственных ресурсов;

$\beta$  - вероятность того, что вторая сторона уничтожит единицу; первой стороны;

$\gamma$  - вероятность того единица второй стороны погибнет из- за влияния первой стороны;

$\delta$  - вероятность увеличения численности второй стороны за счет первой стороны.

# Варианты ЛР-1

№ Вар	a	b	g	d	x0	y0
1	0,488914	0,100314	0,753498	0,21263	14	7
2	0,472371	0,141938	0,796177	0,221151	12	7
3	0,413766	0,188328	0,770184	0,245064	20	9
4	0,450032	0,174374	0,779615	0,255413	10	7
5	0,488864	0,18416	0,71404	0,265664	20	9
6	0,438669	0,158387	0,741475	0,248161	18	7
7	0,431325	0,186943	0,743488	0,259278	14	5
8	0,450157	0,105735	0,712854	0,276487	22	9
9	0,470246	0,124836	0,722991	0,203917	10	5
10	0,492532	0,18834	0,771464	0,24176	14	5
11	0,408731	0,172655	0,716888	0,275764	16	5
12	0,488795	0,179863	0,766086	0,26455	16	6
13	0,469365	0,127686	0,754029	0,292608	20	7
14	0,47	0,180791	0,782603	0,271177	18	9
15	0,459559	0,170367	0,758359	0,293844	10	8
16	0,485876	0,104341	0,717343	0,255161	12	7
17	0,496317	0,121382	0,744528	0,297104	10	5
18	0,490736	0,155674	0,769477	0,217412	16	9
19	0,476873	0,184403	0,710144	0,292283	16	7
20	0,429086	0,110683	0,719147	0,299938	16	8

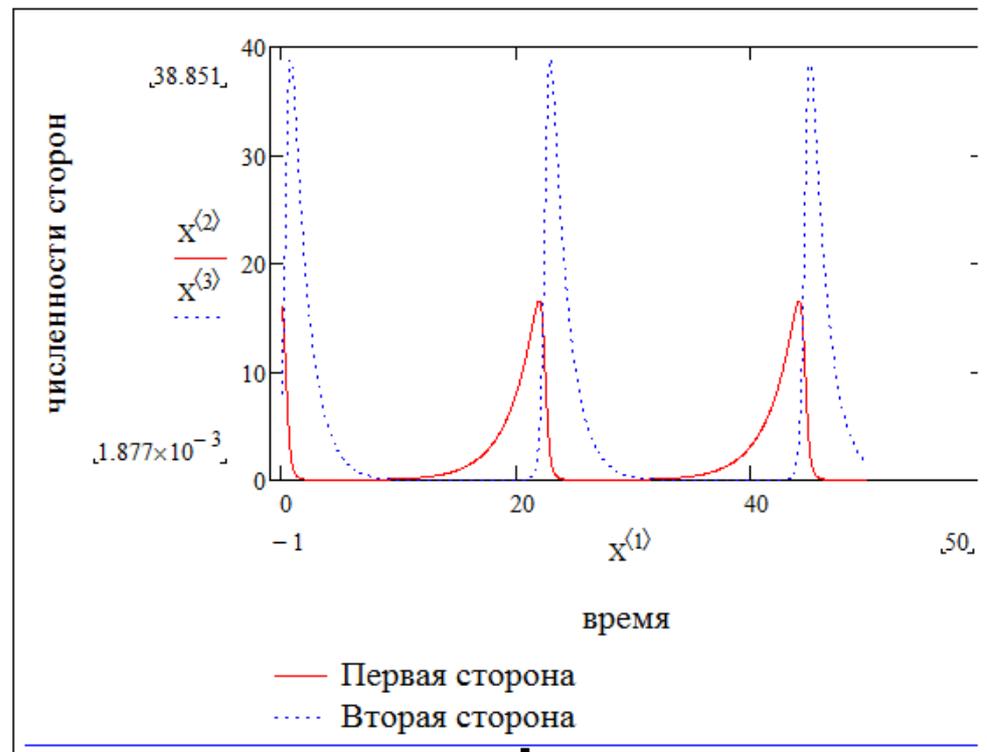
# Выполнение в Mathcad

ORIGIN := 1

$\alpha := 0.43$      $\beta := 0.11$      $\gamma := 0.72$      $\delta := 0.3$

$$x_0 := \begin{pmatrix} 16 \\ 8 \end{pmatrix} \quad f(t, x) := \begin{bmatrix} (\alpha - \beta \cdot x_2) x_1 \\ (-\gamma + \delta \cdot x_1) x_2 \end{bmatrix}$$

$X := \text{rkfixed}(x_0, 0, 50, 1000, f)$



# Фазовый портрет

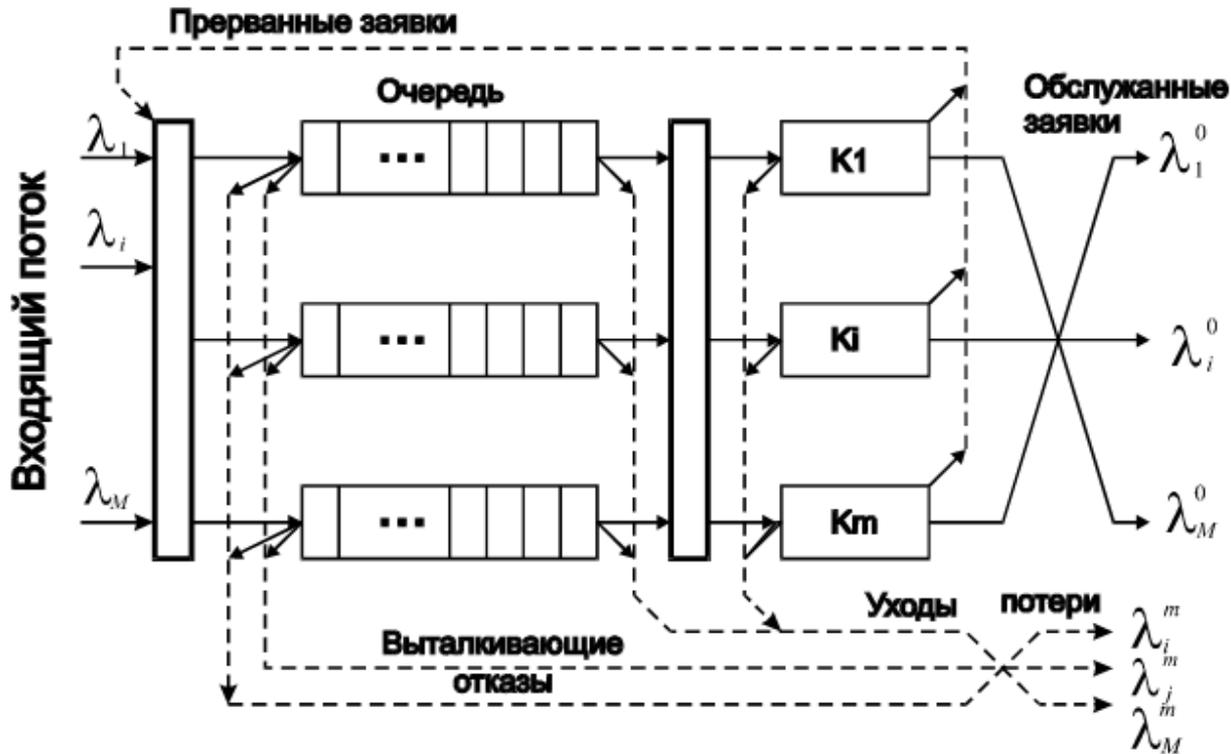
Зависимость  
численности  
одной  
стороны от  
численности  
другой  
стороны



# Непрерывно-стохастические модели (Q-схемы). Моделирование систем массового обслуживания

- **При непрерывно-стохастическом подходе** в качестве типовых математических схем применяется системы массового обслуживания (англ. queueing system), которые будем называть Q-схемами.
- **СМО (системы массового обслуживания)** - это *модели систем*, в которые в случайные моменты времени извне или изнутри поступают заявки (требования), которые должны быть обслужены системой. Длительность обслуживания чаще всего случайна.
- **В качестве СМО** могут быть представлены различные по своей физической природе производственные, технические и другие системы, например, системы поставок продукции, обслуживания клиентов, обработки информации ЭВМ и т.д.

# Обобщенная структура СМО



Основными элементами СМО являются:

1. входной поток заявок;
2. очередь;
3. каналы обслуживания;
4. выходной поток заявок (обслуженные заявки).

## Классификация СМО



# Показатели эффективности СМО

**Предмет теории массового обслуживания – построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, правила работы, характер потока заявок) с интересующими нас характеристиками – показателями эффективности СМО.**

**Эти показатели описывают способность СМО справляться с потоком заявок.**

**Показатели можно разделить на группы:**

- показатели, характеризующие систему в целом,**
- вероятностные характеристики,**
- экономические показатели.**

- **показатели, характеризующие систему в целом:**
  - среднее число заявок, обслуживаемых СМО в единицу времени
  - среднее число заявок в единицу времени, получивших отказ
  - среднее число занятых каналов
  - среднее число заявок в очереди
  - среднее время ожидания обслуживания
- **вероятностные характеристики:**
  - вероятность того, что заявка будет обслужена ( $P_{\text{обс}}$ )
  - вероятность того, что получит отказ в обслуживании ( $P_{\text{отк}}$ )
  - вероятность того, что все каналы свободны ( $p_0$ )
  - вероятность того, что определенное число их занято ( $p_k$ )
  - вероятность наличия очереди
- **экономические показатели:**
  - стоимость потерь, связанных с уходом не обслуженной заявки
  - экономический эффект, полученный в результате обслуживания заявки

# Марковские случайные процессы

- Задачи массового обслуживания, сформулированные математически, обычно сводятся к изучению случайных процессов.
- В системе  $S$  протекает **случайный процесс**, если она с течением времени меняет свое состояние (переходит из одного состояния в другое), причем, заранее неизвестным случайным образом.
- Случайный процесс, протекающий в системе, называется **Марковским**, если для любого момента времени  $t_0$  вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент  $t_0$  и не зависят от того, когда и как система пришла в это состояние.
- Математический анализ работы СМО очень облегчается, если процесс этой работы Марковский. Иначе математическое описание процесса очень усложняется. На практике Марковские процессы в чистом виде обычно не встречаются. Имеются процессы, для которых влиянием «предыстории» можно пренебречь. При изучении таких процессов можно применять Марковские модели
- Приведенный далее математический аппарат описывает Марковские процессы.

# Марковские процессы с дискретными состояниями и непрерывным временем

- Процесс называется *процессом с дискретным состоянием*, если его возможные состояния  $S_1, S_2, \dots$  можно заранее определить, и переход системы из состояния в состояние происходит «скачком», практически мгновенно.
- Процесс называется *процессом с непрерывным временем*, если моменты возможных переходов из состояния в состояние не фиксированы заранее, а неопределенны, случайны и могут произойти в любой момент.

# Поток событий

Каждая СМО предназначена для обслуживания потока требований

- **Поток событий** – последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени. Примеры: поток отказов и поток восстановлений, поток вызовов на телефонной станции, поток сообщений, поток данных и т.д.
- **Интенсивность потока событий** – это среднее число событий, приходящееся на единицу времени.
- Поток событий называется **стационарным**, если его вероятностные характеристики не зависят от времени.
- Поток событий называется **поток без последствий**, если для любых двух непересекающихся участков времени число событий, попадающих на один из них, не зависит от того, сколько событий попало на другой.
- Поток событий называется **ординарным**, если события в нем появляются поодиночке, а не группами по несколько сразу.
- Поток событий называется **простейшим (или стационарным пуассоновским)**, если он обладает сразу тремя свойствами:
  - 1) стационарен, 2) ординарен, 3) не имеет последствий.

# Показательный закон распределения

- Для простейшего потока с интенсивностью  $\lambda$  интервал между соседними событиями имеет так называемое *показательное (экспоненциальное) распределение*

$$f(t) = \lambda e^{-\lambda t}$$

- Для случайной величины, имеющей показательное распределение, математическое ожидание есть величина, обратная интенсивности, а среднее квадратичное отклонение равно математическому ожиданию

$$m_T = \sigma_T = 1/\lambda$$

# Уравнения Колмогорова для вероятностей состояний. Финальные вероятности состояний

- Рассматривая Марковские процессы с дискретными состояниями и непрерывным временем, подразумевается, что все переходы системы  $S$  из состояния в состояние происходят под действием простейших потоков событий (потоков вызовов, потоков отказов, потоков восстановлений и т.д.).
- На графе состояний системы у каждой дуги проставляют интенсивности того потока событий, который переводит систему по данной дуге.
- Для нахождения всех вероятностей состояний как функций времени составляются и решаются уравнения Колмогорова (Эрланга) – особого вида уравнения, в которых неизвестными функциями являются вероятности состояний  $p_i(t)$  или финальные (предельные) вероятности.
- Финальные вероятности состояний  $\tilde{p}_i$  – это вероятности состояний при  $t \rightarrow \infty$ .

# Уравнения Колмогорова (Эрланга) для n-канальной СМО с отказами

## Граф состояний

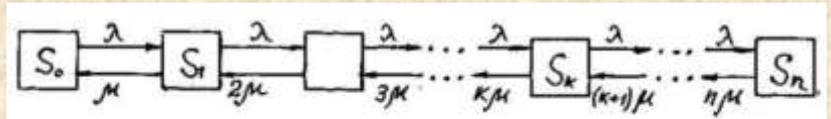
$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t),$$

.....

$$\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu) p_k(t) + (k+1)\mu p_{k+1}(t) \quad (0 < k < n),$$

.....

$$\frac{dp_n(t)}{dt} = \lambda p_{n-1}(t) - n\mu p_n(t).$$



Состояния n-канальной системы:

- $S_0$  - система полностью свободна;
- $S_1$  - занят один канал, остальные каналы свободны;
- $S_2$  - занято два канала, остальные каналы свободны;
- .....
- $S_i$  - занято  $i$  каналов, остальные каналы свободны;
- .....
- $S_n$  - заняты все  $n$  каналов.

**Правило составления уравнений.** В каждом уравнении системы в правой его части стоит **со знаком минус** вероятность данного состояния  $p_i$ , умноженная на суммарную интенсивность всех потоков, **ведущих из данного состояния**, а **со знаком плюс** стоит сумма произведений интенсивностей всех потоков, **входящих в  $i$ -е состояние**, на вероятности тех состояний, из которых эти потоки исходят.

## Уравнения Эрланга в установившемся режиме

$$\begin{aligned} -\lambda p_0 + \mu p_1 &= 0, \\ \lambda p_0 - (\lambda + \mu) p_1 + 2\mu p_2 &= 0, \\ \lambda p_1 - (\lambda + 2\mu) p_2 + 3\mu p_3 &= 0, \\ \dots \\ \lambda p_{k-1} - (\lambda + k\mu) p_k + (k+1)\mu p_{k+1} &= 0 \\ \dots \\ \lambda p_{n-2} - [\lambda + (n-1)\mu] p_{n-1} + n\mu p_n &= 0, \\ \lambda p_{n-1} - n\mu p_n &= 0. \end{aligned}$$

$$\sum_{k=0}^n p_k = 1.$$

## Показатели эффективности n – канальной СМО с отказами

**Абсолютная пропускная способность:**

$$A = \lambda \left[ 1 - \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!} \right], \text{ шт/ед.времени,}$$

где  $n$  – количество каналов СМО;

*Среднее количество заявок, находящихся в системе,  $N_{\text{СМО}}$  в данном случае равно числу занятых каналов, т. к. каждый канал может одновременно обслуживать только одну заявку, а очереди нет. Коэффициент занятости каналов равен отношению числа занятых каналов к общему числу каналов  $K_z = n_z/n$ .*

*Среднее время нахождения заявок в системе  $T_{\text{СМО}}$  может быть найдено через  $N_{\text{СМО}}$  по формуле Литтла:*

$$T_{\text{СМО}} = \frac{N_{\text{СМО}}}{\lambda} = \frac{n_z}{\lambda}. \quad (3.8)$$

$$p_n = \frac{\lambda^n}{n! \mu^n} p_0.$$

**Относительная пропускная способность:**

$$Q = 1 - \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!}.$$

Напомним, что это средняя доля заявок, обслуживаемых системой. При этом

$$A = \lambda Q;$$

$$Q = 1 - P_{\text{отк}}.$$

**Вероятность отказа:**

$$P_{\text{отк}} = \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!}.$$

Напомним, что это вероятность того, что заявка покинет СМО необслуженной. Очевидно, что  $P_{\text{отк}} = 1 - Q$ .

**Среднее число занятых каналов** (среднее число заявок, обслуживаемых одновременно):

$$\bar{k} = \frac{\lambda}{\mu} \left[ 1 - \left( \frac{\lambda}{\mu} \right)^n \frac{p_0}{n!} \right].$$

## **ЛР-2. Технологическая система (участок)**

**Постановка задачи.** Система  $S$  состоит из двух станков, каждый из которых в случайный момент времени может выйти из строя (отказаться), после чего мгновенно начинается ремонт узла, тоже продолжающийся заранее неизвестное, случайное время. Возможны следующие состояния системы:

- $S_0$  - оба станка исправны;
- $S_1$  - первый станок ремонтируется, второй исправен;
- $S_2$  - второй станок ремонтируется, первый исправен;
- $S_3$  - оба станка ремонтируются.

Переходы системы  $S$  из состояния в состояние происходят практически мгновенно, в случайные моменты выхода из строя того или иного станка или окончания ремонта.  $\lambda_{ij}$  –интенсивности переходов заданы.

Разработать модель технологического участка в Mathcad для установившегося режима и для неустойчившегося режима.

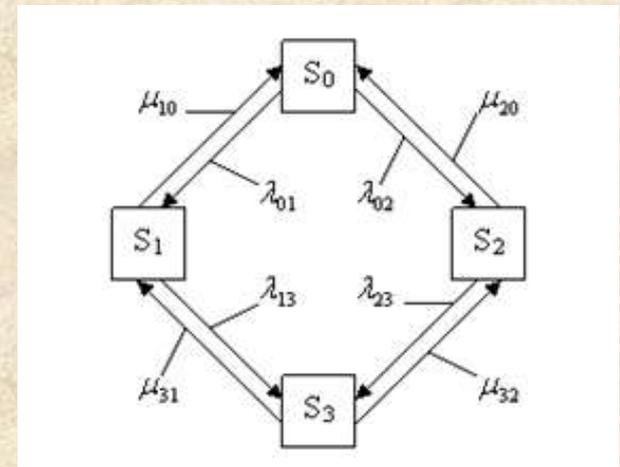
Оценить среднюю эффективность работы системы и загрузку ремонтных органов.

# Уравнения Эрланга в установившемся режиме. Определение финальных

Граф  
СОСТОЯНИЙ

$$\begin{cases} (\lambda_{01} + \lambda_{02})p_0 = \mu_{10}p_1 + \mu_{20}p_2 \\ (\mu_{10} + \lambda_{13})p_1 = \lambda_{01}p_0 + \mu_{31}p_3 \\ (\mu_{20} + \lambda_{23})p_2 = \lambda_{02}p_0 + \mu_{32}p_3 \\ (\mu_{31} + \mu_{32})p_3 = \lambda_{13}p_1 + \lambda_{23}p_2 \end{cases}$$

$$p_0 + p_1 + p_2 + p_3 = 1$$



# Оценка эффективности работы системы по вероятностям

Исходные данные:

$$\lambda_{01} = \lambda_{23} = 1; \lambda_{13} = \lambda_{02} = 2; \mu_{10} = \mu_{32} = 2; \mu_{20} = \mu_{31} = 3.$$

$$p_0 = 6/15 = 0,4; p_1 = 3/15 = 0,2; p_2 = 4/15 \cong 0,27; p_3 = 2/15 \cong 0,13.$$

Знание финальных вероятностей может помочь оценить среднюю эффективность работы системы и загрузку ремонтных органов: в стационарном режиме система  $S$  в среднем 40 % времени будет проводить в состоянии  $S_0$  (оба станка исправны), 20 % - в состоянии  $S_1$  (первый станок ремонтируется, второй работает), 27 % - в состоянии  $S_2$  (второй станок ремонтируется, первый работает), 13% - в состоянии  $S_3$  (оба станка ремонтируются).

Пусть система  $S$  в состоянии  $S_0$  (полностью исправна) приносит в единицу времени доход 8 условных единиц, в состоянии  $S_1$  – доход 3 условные единицы, в состоянии  $S_2$  – доход 5 условных единиц, в состоянии  $S_3$  – не приносит дохода. Тогда в предельном, стационарном режиме средний доход в единицу времени будет равен  $W = 0,4 * 8 + 0,2 * 3 + 0,27 * 5 = 5,15$  условных единиц.

# Варианты ЛР-2

№ Вар	$\lambda_{01}$	$\lambda_{23}$	$\lambda_{13}$	$\lambda_{02}$	$\mu_{10}$	$\mu_{32}$	$\mu_{20}$	$\mu_{31}$
1	1,3	1,1	2,2	2,2	2,0	2,1	3,0	3,2
2	1,2	1,2	2,1	2,1	2,0	2,3	3,3	3,0
3	1,1	1,1	2,0	2,4	2,1	2,0	3,3	3,1
4	1,1	1,4	2,2	2,0	2,3	2,0	3,2	3,1
5	1,0	1,1	2,2	2,0	2,1	2,2	3,0	3,1
6	1,0	1,4	2,3	2,3	2,3	2,0	3,2	3,2
7	1,1	1,1	2,1	2,2	2,1	2,1	3,2	3,2
8	1,3	1,1	2,3	2,3	2,0	2,3	3,0	3,0
9	1,2	1,3	2,1	2,2	2,2	2,1	3,2	3,1
10	1,3	1,2	2,3	2,2	2,2	2,1	3,1	3,2
11	1,0	1,1	2,0	2,1	2,2	2,1	3,1	3,3
12	1,0	1,2	2,1	2,1	2,1	2,1	3,1	3,0
13	1,2	1,2	2,2	2,0	2,2	2,2	3,2	3,1
14	1,2	1,2	2,2	2,0	2,1	2,3	3,2	3,3
15	1,1	1,1	2,1	2,3	2,2	2,2	3,1	3,2
16	1,0	1,1	2,3	2,2	2,3	2,0	3,0	3,3
17	1,2	1,2	2,3	2,3	2,0	2,1	3,3	3,3
18	1,1	1,1	2,1	2,2	2,2	2,1	3,3	3,0
19	1,1	1,1	2,2	2,1	2,0	2,2	3,2	3,1
20	1,1	1,0	2,0	2,2	2,2	2,2	3,3	3,1

## Технологическая система

№ Вар	$\lambda_{01}$	$\lambda_{23}$	$\lambda_{13}$	$\lambda_{02}$	$\mu_{10}$	$\mu_{32}$	$\mu_{20}$	$\mu_{31}$
20	1,1	1,0	2,0	2,2	2,2	2,2	3,3	3,1

$$\lambda_{01} := 1.1 \quad \lambda_{23} := 1.0 \quad \lambda_{13} := 2 \quad \lambda_{02} := 2.2$$

$$\mu_{10} := 2.2 \quad \mu_{32} := 2.2 \quad \mu_{20} := 3.3 \quad \mu_{31} := 3.1 \quad \text{Исходные данные}$$

ORIGIN := 0

*Определение правых частей системы и начальных условий*

$$p := \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{-- начальные вероятности}$$

$$D(t,p) := \begin{bmatrix} -(\lambda_{01} + \lambda_{02}) \cdot p_0 + \mu_{10} \cdot p_1 + \mu_{20} \cdot p_2 \\ -(\mu_{10} + \lambda_{13}) \cdot p_1 + \lambda_{01} \cdot p_0 + \mu_{31} \cdot p_3 \\ -(\mu_{20} + \lambda_{23}) \cdot p_2 + \lambda_{02} \cdot p_0 + \mu_{32} \cdot p_3 \\ -(\mu_{31} + \mu_{32}) \cdot p_3 + \lambda_{13} \cdot p_1 + \lambda_{23} \cdot p_2 \end{bmatrix}$$

$t1 := 0 \quad t2 := 3 \quad k := 1000$       Решение системы дифференциальных уравнений

$Z := \text{rkfixed}(p, t1, t2, k, D)$

**Комментарий** к процедуре `rkfixed(p, t1, t2, k, D)`:

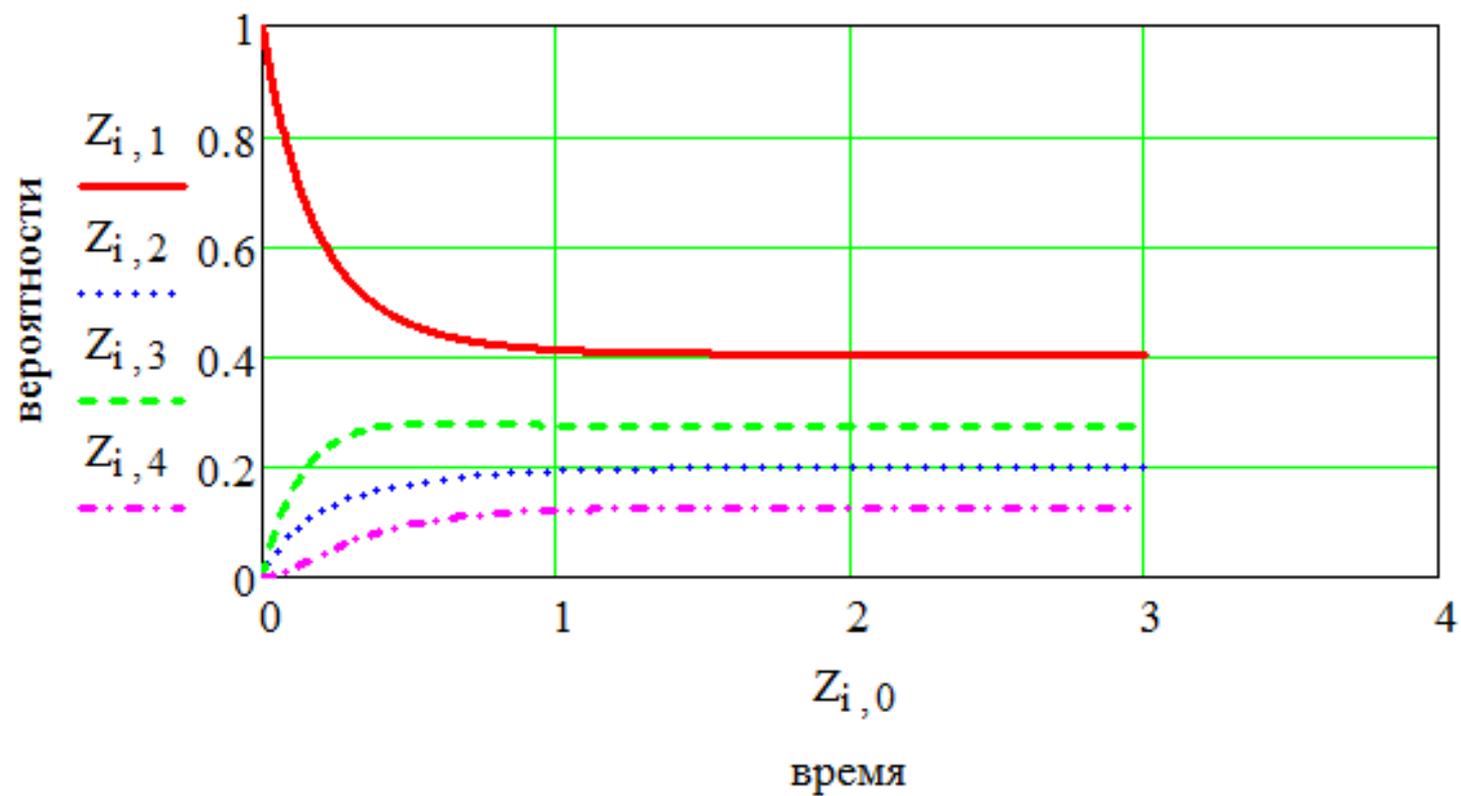
$p$  – вектор вероятностей размерности  $n$ , где  $n$  – порядок системы;  
 $t1, t2$  – граничные точки интервала, в котором ищется решение;  
 $k$  – число точек, в которых ищется решение,  
 $(k+1) \cdot (1+n)$  – размерность матрицы результата, возвращаемой `rkfixed`.

На Графике    $t=Z_{i0}$     $P_0=Z_{i1}$     $P_1=Z_{i2}$     $P_2=Z_{i3}$     $P_3=Z_{i4}$

# Решение в Mathcad

$i := 0..k$

График зависимости вероятностей от времени



- $p_0$
- .....  $p_1$
- - -  $p_2$
- .....  $p_3$

# Решение в установившемся режиме

+

Решение системы алгебраических уравнений

Given

$p_0 := 1$      $p_1 := 0$      $p_2 := 1$      $p_3 := 0$     начальные условия

$$-(\lambda_{01} + \lambda_{02}) \cdot p_0 + \mu_{10} \cdot p_1 + \mu_{20} \cdot p_2 = 0$$

$$-(\mu_{10} + \lambda_{13}) \cdot p_1 + \lambda_{01} \cdot p_0 + \mu_{31} \cdot p_3 = 0$$

$$-(\mu_{20} + \lambda_{23}) \cdot p_2 + \lambda_{02} \cdot p_0 + \mu_{32} \cdot p_3 = 0$$

$$-(\mu_{31} + \mu_{32}) \cdot p_3 + \lambda_{13} \cdot p_1 + \lambda_{23} \cdot p_2 = 0$$

$$\text{Find}(p_0, p_1, p_2, p_3) = \begin{pmatrix} 0.404 \\ 0.199 \\ 0.271 \\ 0.126 \end{pmatrix}$$

Финальные вероятности при  $t$   
стремящемся к бесконечности

# **Задания 1,2 контрольной работы**

- **Задание 1 соответствует ЛР-1**
- **Задание 2 соответствует ЛР-2**

# **Лекция 2**

**Тема:**

**«Регрессия корреляция,  
автоматизация обработки  
многомерных данных»**

# Структура лекции

2

- Основные теоретические положения для множественной регрессии и корреляции
- Постановка и решение задачи множественной регрессии и корреляции в Excel . Нахождение стандартизованного уравнения и определение степени влияния факторов на результативный признак (ЛР-3)
- Автоматизация обработки данных, используя VBA (ЛР-4)
- Задания 3,4 контрольной работы

# Введение

- ▶ Методы анализа и представления данных основываются на многомерных статистических методах, включающих корреляционный, регрессионный, дисперсионный, компонентный и кластерный анализ.
- ▶ При проведении экспериментальных исследований для сбора статистической информации бывает важно получение достоверной информации при возможно меньшем числе экспериментов. Это достигается благодаря применению математических методов планирования экспериментов.
- ▶ Пакеты прикладных программ, такие как, надстройки Excel, MatLab Statistics ToolBox, Statistica, StateGraphics, SAS и др., снимают трудности в применении многомерных статистических методов, так как позволяют автоматизировать процессы сбора, анализа и обработки многомерных данных.

# СТАТИСТИЧЕСКАЯ И КОРРЕЛЯЦИОННАЯ ЗАВИСИМОСТИ

4

- ▶ В математике мы привыкли к тому, что речь идет о функциональной зависимости, когда каждому значению одной переменной соответствует вполне определенное значение другой.
- ▶ Очень часто между переменными величинами существуют зависимости, когда каждому значению одной переменной соответствует не какое-то определенное, а множество возможных значений другой переменной (или определенное условное распределение другой переменной). Такая зависимость получила название статистической, вероятностной.
- ▶ Возникновение такой связи обусловлено тем, что зависимая переменная подвержена влиянию неконтролируемых или неучтенных факторов, а также случайных ошибок.
- ▶ Когда каждому значению одной переменной  $x$  соответствует определенное условное математическое ожидание (среднее значение) другой  $M_x(Y)$  то такая зависимость называется корреляционной, а уравнение  $M_x(Y) = f(x)$  называется уравнением регрессии

# Выборочная кривая регрессии

5

- В статистической практике уравнение регрессии обычно не известно, а имеется выборка пар значений объема  $n$ :  $(x_i, y_i)$ ,  $i=1,2,\dots,n$ .
- В этом случае речь может идти о приближенном выражении функции регрессии или аппроксимации по выборке  $(x_i, y_i)$ . Такой оценкой является выборочная линия регрессии, ее уравнение:

$$\hat{y} = f(x, a, b_1, \dots, b_q)$$

- парная регрессия

$$\hat{y} = f(x_1, x_2, x_3, \dots, x_p, a, b_1, \dots, b_q)$$

- множественная регрессия

$$a, b_1, \dots, b_p$$

- параметры, подлежащие определению

$$|\hat{y}_i - y_i| = \varepsilon_i$$

- отклонения (остатки) во всех  $n$  точках

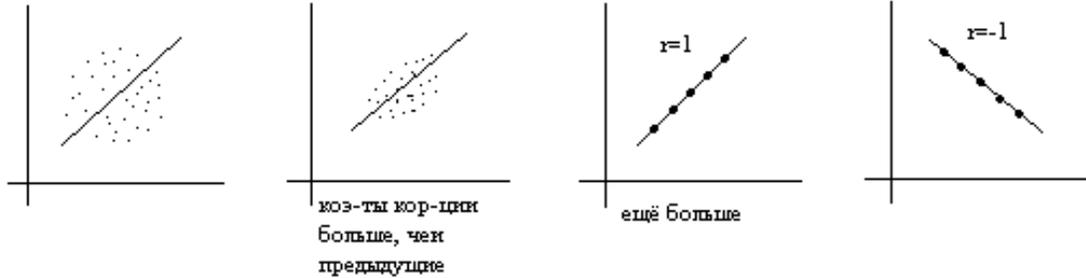
## Этапы нахождения регрессионной модели

- **1 шаг:** спецификация модели, т.е. выбор вида уравнения регрессии
- **2 шаг:** выбрать наилучшие значения параметров
- **3 шаг:** оценка качества модели по критериям

# Коэффициент корреляции

7

Коэффициент корреляции является показателем тесноты связи между  $x$  и  $y$



Формула вычисления

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2}}$$

Свойства коэффициента корреляции

Значения коэффициента корреляции  $-1 \leq r \leq 1$

Корреляционная связь представляет линейную функциональную зависимость при  $r = \pm 1$

Линейная связь отсутствует при  $r = 0$

Близость к нулю не означает отсутствия связи между признаками, она может оказаться достаточно тесной, но нелинейной

# Постановка задачи анализа многомерных данных

8

- ▶ Результативный признак  $y$  (затраты на ТО летательного аппарата) зависит от трех комплексных факторов  $x_1, x_2, x_3$ . Фактор  $x_1$  определяет тип ЛА; фактор  $x_2$  определяет виды эксплуатационных и подготовительных работ; фактор  $x_3$  определяет задействованные материально-технические ресурсы и человеческий фактор.
- ▶ На основе статистических измерений получена таблица зависимости значений результативного признака  $y$  от значений факторов  $x_1, x_2, x_3$  (варианты)

**Требуется** разработать математическую модель, позволяющую оценивать затраты на ТО в зависимости от различных факторов.

**1 часть:** корреляционный, дисперсионный, регрессионный анализ и автоматизация представления данных (ЛР-3,4)

**2 часть:** компонентный, кластерный анализ (ЛР-5,6)

# Варианты ЛР-3-6

		ВАРИАНТ 1		
№	y	x1	x2	x3
1	65,84	2,83	10,50	1,30
2	185,72	5,13	17,39	1,60
3	71,50	2,62	9,87	1,90
4	120,75	3,53	12,60	2,20
5	211,79	4,81	16,44	2,50
6	133,74	3,43	12,30	2,80
7	112,65	2,92	10,76	3,10
8	246,60	4,70	16,09	3,40
9	164,77	3,49	12,46	3,70
10	243,06	4,34	15,02	4,00
11	110,33	2,43	9,30	4,30
12	130,14	2,65	9,94	4,60
13	379,96	5,25	17,74	4,90
14	381,80	5,12	17,37	5,20
15	158,10	2,76	10,28	5,50
16	181,11	2,95	10,86	5,80
17	255,96	3,65	12,95	6,10
18	347,54	4,36	15,08	6,40
19	199,14	2,91	10,73	6,70
20	198,98	2,83	10,50	7,00

		ВАРИАНТ 2		
№	y	x1	x2	x3
1	190,29	4,09	3,20	14,26
2	136,04	3,19	3,40	11,56
3	192,73	3,92	3,60	13,75
4	95,17	2,33	3,80	9,00
5	178,73	3,55	4,00	12,65
6	166,66	3,31	4,20	11,92
7	237,78	4,10	4,40	14,29
8	272,18	4,38	4,60	15,13
9	146,78	2,82	4,80	10,47
10	138,53	2,65	5,00	9,94
11	221,51	3,60	5,20	12,80
12	379,37	5,01	5,40	17,04
13	140,98	2,51	5,60	9,53
14	148,33	2,55	5,80	9,66
15	293,54	4,04	6,00	14,11
16	194,61	2,99	6,20	10,98
17	420,39	4,93	6,40	16,78
18	359,47	4,39	6,60	15,16
19	311,25	3,93	6,80	13,79
20	468,43	5,04	7,00	17,13

		ВАРИАНТ 3		
№	y	x1	x2	x3
1	434,63	5,07	17,21	6,30
2	210,61	3,05	11,16	6,60
3	270,18	3,54	12,63	6,90
4	386,35	4,39	15,18	7,20
5	401,29	4,41	15,22	7,50
6	460,57	4,72	16,16	7,80
7	285,77	3,36	12,09	8,10
8	388,02	4,06	14,17	8,40
9	306,36	3,39	12,16	8,70
10	222,98	2,64	9,91	9,00
11	564,40	4,90	16,69	9,30
12	324,68	3,33	11,98	9,60
13	479,67	4,26	14,77	9,90
14	497,23	4,29	14,86	10,20
15	365,91	3,42	12,27	10,50
16	291,09	2,86	10,57	10,80
17	359,78	3,27	11,81	11,10
18	394,46	3,43	12,30	11,40
19	553,46	4,25	14,76	11,70
20	587,31	4,36	15,08	12,00

## **Этапы проведения корреляционного и регрессионного анализа данных**

- **Исследование взаимного влияния факторов**
- **Выбор вида уравнения регрессии**
- **Нахождение параметров уравнения регрессии (МНК)**
- **Определение значимости уравнения регрессии**

# Интеркорреляция факторов

11

Применительно к множественной регрессии, необходимо до получения уравнения регрессии произвести отбор факторов. Факторы, включаемые в модель должны быть:

- Количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.
- Факторы не должны быть интеркоррелированы (коллинеарны) и тем более находиться в точной функциональной связи. Это может привести к тому, что система уравнений МНК может оказаться плохо обусловленной и повлечь за собой неустойчивость и ненадежность оценок коэффициентов регрессии.

Мультиколлинеарность факторов, когда более чем 2 фактора связаны между собой линейной зависимостью. Для оценки мультиколлинеарности может использоваться определитель корреляционной матрицы  $\text{Det}$ . Если бы факторы не коррелировали между собой, то  $\text{Det}(R)=1$  (идеальный случай).

Предпочтительно, чтобы  $\text{Det}(R)$  был близок к единице.

# Формулы расчета определителя корреляционной матрицы

12

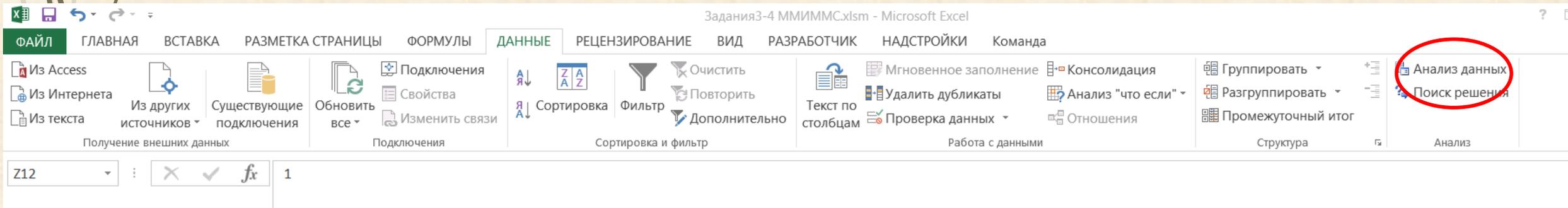
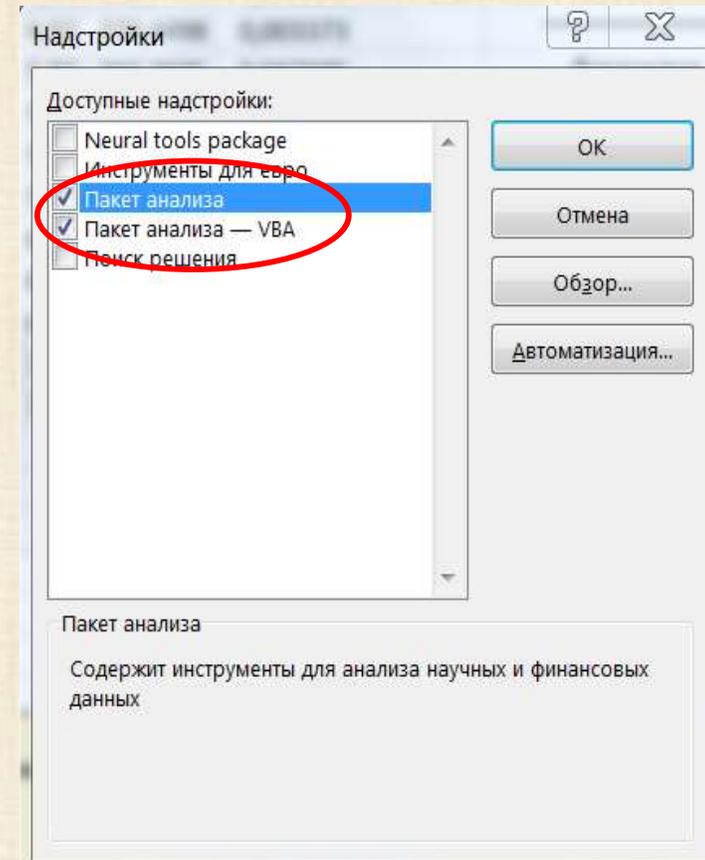
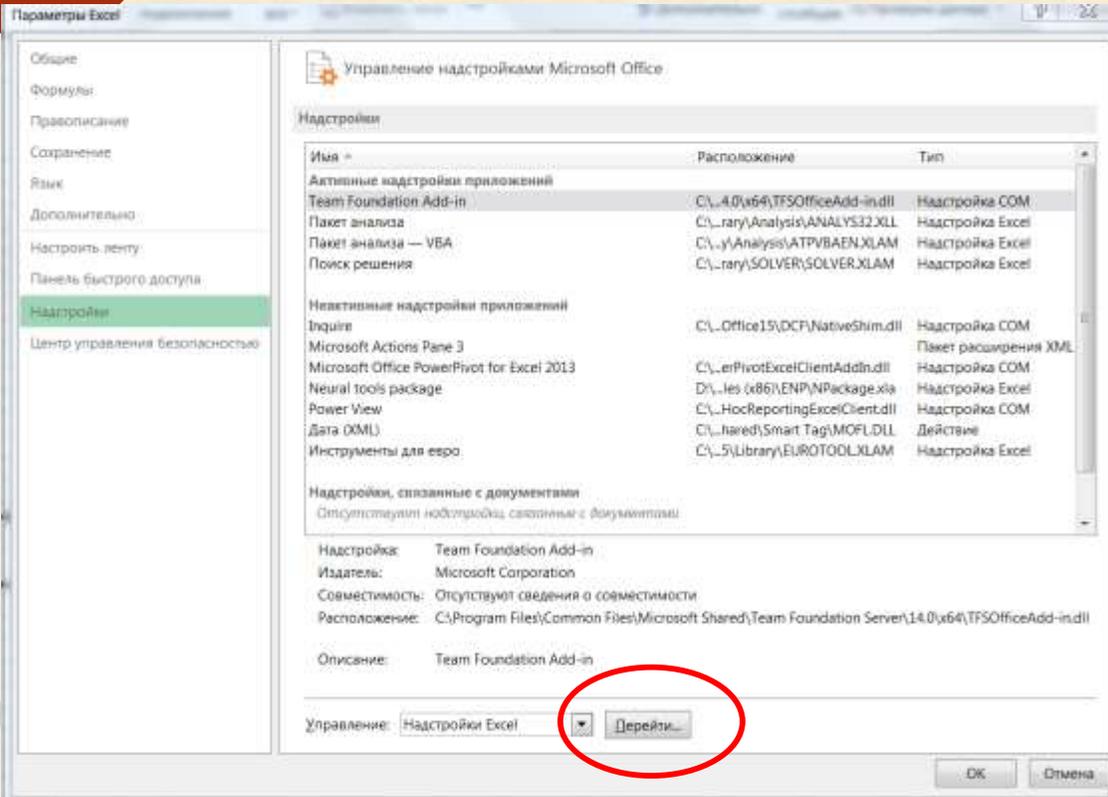
$$r_{x_i x_j} = \frac{\overline{x_i x_j} - \overline{x_i} \cdot \overline{x_j}}{\sigma_{x_i} \sigma_{x_j}}$$

$$\text{Det}|R| = \begin{vmatrix} r_{x_1 x_1} & r_{x_2 x_1} & r_{x_3 x_1} & \dots \\ r_{x_1 x_2} & r_{x_2 x_2} & r_{x_3 x_2} & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix}$$

## **ЛР-3. Корреляционный, дисперсионный, регрессионный анализ**

# Подключение надстроек **Пакет анализа** и **пакет анализа VBA**

14



# Определение матрицы корреляции в Excel (ЛР-3)

118 : X ✓ fx {=МОПРЕД(J12:L14)}

	A	B	C	D	E	F	G	H	I	J	K	L	M
4													
5		№	y	x1	x2	x3				x1	x2	x3	
6		1	57,48	2,58	9,75	1,78			x1	1			
7		2	110,18	3,72	13,15	2,18			x2	1	1		
8		3	203,89	5,16	17,47	2,30			x3	0,179845	0,179845	1	
9		4	85,25	2,80	10,39	2,66							
10		5	134,51	3,61	12,83	2,57							
11		6	99,87	2,81	10,44	3,12				x1	x2	x3	
12		7	195,88	4,22	14,66	3,41			x1	1	1	0,179845	
13		8	134,39	3,16	11,48	3,95			x2	1	1	0,179845	
14		9	223,84	4,26	14,79	4,62			x3	0,179845	0,179845	1	
15		10	289,68	4,85	16,55	4,73							
16		11	275,38	4,55	15,65	4,92							
17		12	254,14	4,19	14,56	5,55							
18		13	161,75	2,99	10,97	5,58			Выход: факторы x1, x2 коллинеарны				
19		14	329,68	4,67	16,01	5,57			Det	-4,09232E-21			
20		15	292,14	4,20	14,61	5,64			Определитель близок к нулю				
21		16	384,80	4,89	16,66	6,73							

Анализ данных

Инструменты анализа

- Двухфакторный дисперсионный анализ без повторений
- Корреляция**
- Ковариация
- Описательная статистика
- Экспоненциальное сглаживание
- Двухвыборочный F-тест для дисперсии
- Анализ Фурье
- Гистограмма
- Скользящее среднее
- Генерация случайных чисел

OK Отмена Справка

Корреляция

Входные данные

Входной интервал: \$D\$5:\$F\$25

Группирование:  по столбцам  по строкам

Метки в первой строке

Параметры вывода

Выходной интервал: \$I\$5

Новый рабочий лист:

Новая рабочая книга

OK Отмена Справка

# Регрессия. Виды многомерного уравнения регрессии

16

$$\hat{y} = a + b_1 x_1 + \dots + b_p x_p$$

линейная

$$\hat{y} = a \cdot x_1^{b_1} \cdot x_2^{b_2} \dots \cdot x_p^{b_p}$$

степенная

$$\hat{y} = e^{a+b_1 x_1 + \dots + b_p x_p}$$

показательная

$$\hat{y} = (a + b_1 x_1 + \dots + b_p x_p)^{-1}$$

обратная

## Линеаризация степенной зависимости

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \dots \cdot x_p^{b_p}$$

$$\ln y = \ln a + b_1 \ln x_1 + b_2 \ln x_2 \dots + b_p \ln x_p$$

$$Y = \ln y, \quad A = \ln a, \quad X_i = \ln x_i,$$

$$y = \exp(\ln a + b_1 \ln x_1 + b_2 \ln x_2 \dots + b_p \ln x_p)$$

## Линеаризованное уравнение

$$Y = A + b_1 X_1 + \dots + b_p X_p$$



# Регрессия в Excel

## Линейная зависимость

### Анализ данных

#### Инструменты анализа

Ковариация  
Описательная статистика  
Экспоненциальное сглаживание  
Двухвыборочный F-тест для дисперсии  
Анализ Фурье  
Гистограмма  
Скользящее среднее  
Генерация случайных чисел  
Ранг и перцентиль  
**Регрессия**

OK

Отмена

Справка

### Регрессия

#### Входные данные

Входной интервал Y:

\$C\$5:\$C\$25

Входной интервал X:

\$D\$5:\$F\$25

 Метки Уровень надежности:

95 %

 Константа - ноль

#### Параметры вывода

 Выходной интервал:

\$B\$31

 Новый рабочий лист: Новая рабочая книга

#### Остатки

 Остатки Стандартизованные остатки График остатков График подбора

#### Нормальная вероятность

 График нормальной вероятности

OK

Отмена

Справка

G6 = \$C\$47+\$C\$48\*D6+\$C\$49\*E6+\$C\$50\*F6

№	y	x1	x2	x3	y предск	COA
1	57,48	2,58	9,75	1,78	29,0182	0,495128
2	110,18	3,72	11,15	2,18	130,357	0,183086
3	203,89	5,16	12,56	2,30	247,51	0,213915
4	85,25	2,80	13,97	2,66	70,3593	0,174645
5	134,51	3,61	15,37	2,57	131,009	0,026026
6	99,87	2,81	16,78	3,12	83,958	0,159318
7	195,88	4,22	18,18	3,41	203,652	0,039697
8	134,39	3,16	19,59	3,95	135,176	0,005868
9	223,84	4,26	21,00	4,62	242,414	0,082996
10	289,68	4,85	22,40	4,73	291,476	0,006211
11	275,38	4,55	20,81	4,92	274,45	0,003373
12	254,14	4,19	19,22	5,55	266,261	0,047695
13	161,75	2,99	17,62	5,58	173,174	0,070619
14	329,68	4,67	16,03	5,57	307,191	0,068201
15	292,14	4,20	14,43	5,64	273,418	0,064073
16	384,80	4,89	12,84	6,73	362,246	0,058614
17	257,46	3,67	11,25	6,29	252,777	0,018173
18	358,17	4,45	9,65	6,61	325,642	0,090823
19	330,80	4,12	8,06	7,68	334,067	0,009883
20	164,45	2,46	6,47	7,89	209,557	0,274307

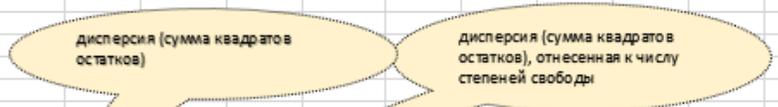
компоненты дисперсии	сумма квадратов	число степеней свободы	дисперсия на одну степень свободы
Общая	$\sum (y - \bar{y})^2$	n - 1	$S_{\text{общ}}^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$
Факторная	$\sum (\hat{y}_i - \bar{y})^2$	m	$S_{\text{факт}}^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{m}$
Остаточная	$\sum (y - \hat{y}_i)^2$	n - m - 1	$S_{\text{ост}}^2 = \frac{\sum (y - \hat{y}_i)^2}{n - m - 1}$

**COA=10,4%**

Решение с помощью пакета анализа в Excel

Вывод ИТОГОВ

Регрессионная статистика	
Множественный R	0,97517
R-квадрат	0,95096
Нормированный R-квадрат	0,94177
Стандартная ошибка	23,431
Наблюдения	20



Дисперсионный анализ

	df	SS	MS	F	значимость F
Регрессия	3	170342,817	56780,9	103,42366	1,1E-10
Остаток	16	8784,20895	549,013		
Итого	19	179127,026			

коэффициент статистика Р-значение нижние 95% верхние 95% средние 95,0%

	коэффициент	статистика	Р-значение	нижние 95%	верхние 95%	средние 95,0%	проверка уравнения для средних
Y-пересечение	-224,623	30,7652722	-7,30118	1,778E-06	-289,842	-159,4033	-224,623
x1	79,3928	7,02866186	11,2956	4,917E-09	64,4927	94,29291	64,4927
x2	-0,64857	1,27876454	-0,50718	0,6189458	-3,35943	2,062293	-3,35943
x3	30,799	5,06959543	10,0336	2,622E-08	24,2917	37,30624	24,2917

Вывод остатка

Наблюдение	у	у предск	Остатки
1	57,48	29,0182	28,4618

$$y = -224,6 + 79,4x_1 - 0,6x_2 + 30,8x_3$$

# Толкование результатов суммы квадратов отклонений, дисперсии, R

$$y_i - \bar{y} = y_i - \bar{y} + \hat{y}_i - \hat{y}_i = \left( y_i - \hat{y}_i \right) + \left( \hat{y}_i - \bar{y} \right)$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n \left( \hat{y}_i - \bar{y} \right)^2 + \sum_{i=1}^n \left( y_i - \hat{y}_i \right)^2$$

*общ*  
TSS
*факторн*  
RSS
*остат*  
ESS

суммы квадратов отклонений  
(остатков)

$$\frac{TSS}{n-1} = D_{общ}, \quad \frac{RSS}{k_{1факт}} = D_{факт}, \quad \frac{ESS}{n-2} = D_{остат}$$

дисперсии D, отнесенные  
к числу степеней свободы:  
общая, факторная, остаточная

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}$$

квадрат коэффициента детерминизации

# Критерии оценки значимости уравнения в целом

21

Проверить значимость уравнения регрессии - значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным, достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной. Основные критерии для оценки значимости уравнения в целом:

$$F = \frac{D_{\text{факт}}}{D_{\text{остат}}} > F_{\text{табл}}(\alpha, k_{1\text{факт}}, k_{2\text{остат}})$$

критерий Фишера

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS} > 0,9$$

коэффициент детерминизации

$$COA = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

<8% средняя ошибка аппроксимации

# Регрессия в Excel

## Степенная зависимость

При расчете `упредк = урегр` учитывать, что коэффициенты найдены для линеаризованной зависимости

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
5		№	y	x1	x2	x3	у предск	COA			Ln(y)	ln(x1)	ln(x2)	ln(x3)
6		1	57,48	2,58	9,75	1,78	58,9528	0,025692			4,05137	0,94863	2,27691	0,57907
7		2	110,18	3,72	11,15	2,18	117,478	0,066201			4,70215	1,31307	2,41169	0,77816
8		3	203,89	5,16	12,56	2,30	199,666	0,020737			5,3176	1,64029	2,53044	0,83346
9		4	85,25	2,80	13,97	2,66	86,7901	0,018098			4,44556	1,02827	2,63658	0,97972
10		5	134,51	3,61	15,37	2,57	123,853	0,079228			4,90164	1,28366	2,73252	0,94204
11		6	99,87	2,81	16,78	3,12	97,0308	0,028419			4,60386	1,03429	2,82006	1,13814
12		7	195,88	4,22	18,18	3,41	190,814	0,025847			5,27748	1,43954	2,90055	1,22771
13		8	134,39	3,16	19,59	3,95	135,849	0,010871			4,90073	1,15034	2,97504	1,37392
14		9	223,84	4,26	21,00	4,62	238,764	0,066692			5,41091	1,44967	3,04437	1,53049
15		10	289,68	4,85	22,40	4,73	294,294	0,01594			5,66877	1,57892	3,1092	1,55336
16		11	275,38	4,55	20,81	4,92	275,881	0,001825			5,61815	1,51497	3,0354	1,59238
17		12	254,14	4,19	19,22	5,55	266,685	0,049366			5,53788	1,43227	2,95572	1,7132
18		13	161,75	2,99	17,62	5,58	161,235	0,003194			5,08606	1,09544	2,86914	1,71886
19		14	329,68	4,67	16,03	5,57	319,22	0,031714			5,79811	1,54116	2,77435	1,71678
20		15	292,14	4,20	14,43	5,64	276,338	0,054078			5,67722	1,43609	2,66962	1,72936
21		16	384,80	4,89	12,84	6,73	397,582	0,033216			5,95272	1,58677	2,55262	1,90585
22		17	257,46	3,67	11,25	6,29	246,514	0,042503			5,55085	1,2991	2,4201	1,83853
23		18	358,17	4,45	9,65	6,61	346,076	0,03377			5,88101	1,49208	2,2673	1,88827
24		19	330,80	4,12	8,06	7,68	347,736	0,051203			5,80151	1,41654	2,08686	2,03886
25		20	164,45	2,46	6,47	7,89	163,809	0,003884			5,1026	0,90013	1,86653	2,06588
26														
27								0,033124						

**COA=3,3%**

Решение с помощью пакета анализа в Excel

Вывод: COA=3,3% < 8% -лучше линейной

Вывод ИТОГОВ

Регрессионная статистика

Множественный R	0,99702
R-квадрат	0,99404
Нормированный R-квадрат	0,99292
Стандартная ошибка	0,04436
Наблюдения	20

дисперсия (сумма квадратов остатков)

дисперсия (сумма квадратов остатков), отнесенная к числу степеней свободы

Дисперсионный анализ

	df	SS	MS	F	значимость F
Регрессия	3	5,252445224	1,75082	889,67688	5,3E-18
Остаток	16	0,031486759	0,00197		
Итого	19	5,283931983			

a= 10,5921

Линеаризованное уравнение

$$Y = A + b_1 X_1 + \dots + b_p X_p$$

$$y = \exp(\ln a + b_1 \ln x_1 + b_2 \ln x_2 + \dots + b_p \ln x_p)$$

$$a = \exp(A)$$

$$y = 10,59 \cdot x_1^{1,52} \cdot x_2^{-0,06} \dots \cdot x_3^{0,72}$$

	коэффициент	стандартная ошибка	статистика t	P-значение	верхние 95%	нижние 95%	верхние 95%	нижние 95%	проверка уравнения для средних
Y-пересечение	2,36011	0,093955889	25,1193	2,778E-14	2,16093	2,559287	2,16093	2,55929	2,36011
ln(x1)	1,52229	0,049669576	30,6484	1,225E-15	1,417	1,627589	1,417	1,62759	
ln(x2)	-0,0633	0,032352353	-1,9565	0,0680924	-0,13188	0,005286	-0,13188	0,00529	
ln(x3)	0,71954	0,024089564	29,8692	1,837E-15	0,66847	0,770603	0,66847	0,7706	

Вывод ОСТАТКА

Наблюдение	сказанное	Остатки
------------	-----------	---------

Буфер обмена Шрифт Выравнивание

G6 :  $=EXP(\$C\$47+\$C\$48*D6+\$C\$49*E6+\$C\$50*F6)$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
5		№	y	x1	x2	x3	y предск	COA			Ln(y)			
6		1	57,48	2,58	9,75	1,78	64,4874	0,121982			4,05137			
7		2	110,18	3,72	11,15	2,18	112,325	0,019437			4,70215			
8		3	203,89	5,16	12,56	2,30	211,597	0,037781			5,3176			
9		4	85,25	2,80	13,97	2,66	85,6652	0,004902			4,44556			
10		5	134,51	3,61	15,37	2,57	119,592	0,110908			4,90164			
11		6	99,87	2,81	16,78	3,12	96,0136	0,038604			4,60386			
12		7	195,88	4,22	18,18	3,41	183,807	0,061616			5,27748			
13		8	134,39	3,16	19,59	3,95	131,864	0,018779			4,90073			
14		9	223,84	4,26	21,00	4,62	237,983	0,0632			5,41091			
15		10	289,68	4,85	22,40	4,73	313,738	0,083062			5,66877			
16		11	275,38	4,55	20,81	4,92	282,183	0,024708			5,61815			
17		12	254,14	4,19	19,22	5,55	267,791	0,053718			5,53788			
18		13	161,75	2,99	17,62	5,58	161,291	0,002842			5,08606			
19		14	329,68	4,67	16,03	5,57	318,945	0,032549			5,79811			
20		15	292,14	4,20	14,43	5,64	262,261	0,102263			5,67722			
21		16	384,80	4,89	12,84	6,73	416,857	0,083307			5,95272			
22		17	257,46	3,67	11,25	6,29	228,638	0,111935			5,55085			
23		18	358,17	4,45	9,65	6,61	330,024	0,078586			5,88101			
24		19	330,80	4,12	8,06	7,68	344,194	0,040495			5,80151			
25		20	164,45	2,46	6,47	7,89	176,373	0,072512			5,1026			
26														
27								0,058159						
28														
29														
30		Решение с помощью пакета анализа в Excel			Вывод: COA=6%<8% -лучше линейной									
31		Вывод ИТОГОВ												
32														
33		Регрессионная статистика												
34		Множественный R	0,991											
35		R-квадрат	0,98209											
36		Нормированный R	0,97873											
37		Стандартная ошибка	0,07692											
38		Наблюдения	20											
39		Дисперсионный анализ												
40														
41			df	SS	MS	F	значимость F							
42		Регрессия	3	5,189275031	1,72976	292,38353	3,5E-14							
43		Остаток	16	0,094656952	0,00592									
44		Итого	19	5,283931983										
45														
46		Коэффициенты стандартная ошибка статистика P-значение нижние 95% верхние 95% нижние 95% верхние 95%												
47		Y-пересечение	2,68626	0,100991702	26,5988	1,135E-14	2,47217	2,900355	2,47217	2,90035				
48		x1	0,41584	0,023072655	18,0231	4,731E-12	0,36693	0,464752	0,36693	0,46475				
49		x2	0,00912	0,00419774	2,17318	0,0451291	0,00022	0,018021	0,00022	0,01802				
50		x3	0,17795	0,010076416	17,6599	6,457E-12	0,15659	0,19931	0,15659	0,19931				
51														
52														
53														
54		Вывод ОСТАТКА												
55														
56		Наблюдение	сказанное	Остатки										
57		1	4,16647	-0,115097171										

**COA=5,8%**

дисперсия (сумма квадратов остатков)

дисперсия (сумма квадратов остатков),

МножРегрессияПоказ

# Регрессия в Excel

## Показательная зависимость

$$y = e^{a+b_1x_1+\dots+b_px_p}$$

$$\ln y = a + b_1x_1 + \dots + b_px_p$$

$$Y = \ln y$$

Линеаризованное уравнение

$$Y = a + b_1x_1 + \dots + b_px_p$$

# Свойства остатков

24

В задачу регрессионного анализа входит не только построение самой модели, но и исследование случайных отклонений  $\varepsilon_i$  (остатков). Они должны быть несмещенными, состоятельными, эффективными, или равноизменчивыми, некоррелированными.

Несмещенность означает, что математическое ожидание остатков равно нулю.

Состоятельность оценки характеризуется увеличением ее точности с увеличением выборки.

Эффективность означает оценку, характеризующуюся наименьшей дисперсией остатков.

Условие гомоскедастичности или равноизменчивости означает, что для каждого фактора дисперсия остатков одинакова (проверяется с помощью частных уравнений регрессии)

# Стандартизованное уравнение

25

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m}$$

где  $t_y, t_{x_1}, \dots, t_{x_m}$  – стандартизированные переменные:  $t_y = \frac{y - \bar{y}}{\sigma_y}$ ,

$t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ , для которых среднее значение равно нулю:  $\bar{t}_y = \bar{t}_{x_i} = 0$ , а

среднее квадратическое отклонение равно единице:  $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$ ;  $\beta_i$  – стандартизированные коэффициенты регрессии.

Коэффициенты  $\beta_i$  при стандартизованных (**центрированных**) переменных  $t_{x_i}$  позволяют оценить силу влияния факторов  $x_i$  на результативный признак  $y$ . Чем больше  $\beta_i$ , тем больше влияние. Коэффициенты корреляции  $r_{yx_i}$  в этом отношении менее информативны, так как не учитывают коллинеарность факторов

# Связь между коэффициентами корреляции и коэффициентами стандартизованного уравнения

26

$$\left\{ \begin{array}{l} r_{yx_1} = \beta_1 + \beta_2 r_{x_1x_2} + \beta_3 r_{x_1x_3} + \dots + \beta_m r_{x_1x_m}, \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2 + \beta_3 r_{x_1x_3} + \dots + \beta_m r_{x_1x_m}, \\ \dots \\ r_{yx_m} = \beta_1 r_{x_1x_m} + \beta_2 r_{x_2x_m} + \beta_3 r_{x_3x_m} + \dots + \beta_m, \end{array} \right.$$

# Расчет значений стандартизованных переменных в EXCEL (линейная зависимость)

27

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	
1																									
2									Стандартизованные переменные																
3	№	y	x1	x2	x3				ty	tx1	tx2	tx3													
4	1	57,4762667	2,58216	9,75	1,78439			1	-1,68758	-1,59644	-1,13915	-1,59195													
5	2	110,183783	3,71759	11,15	2,17745			2	-1,13064	-0,18621	-0,82563	-1,37646													
6	3	203,893635	5,15665	12,56	2,30127			3	-0,14045	1,60114	-0,51212	-1,30858													
7	4	85,2473181	2,79622	13,97	2,66372			4	-1,39414	-1,33057	-0,1986	-1,10987													
8	5	134,509953	3,60984	15,37	2,56522			5	-0,8736	-0,32004	0,11491	-1,16387													
9	6	99,8689479	2,8131	16,78	3,12096			6	-1,23964	-1,3096	0,42843	-0,8592													
10	7	195,876359	4,21876	18,18	3,4134			7	-0,22517	0,43626	0,74195	-0,69887													
11	8	134,387779	3,15928	19,59	3,9508			8	-0,87489	-0,87964	1,05546	-0,40425													
12	9	223,836095	4,2617	21,00	4,62046			9	0,07027	0,4896	1,36898	-0,03712													
13	10	289,676689	4,8497	22,40	4,72734			10	0,76598	1,2199	1,68249	0,02147													
14	11	275,378632	4,54926	20,81	4,91542			11	0,6149	0,84675	1,32719	0,12459													
15	12	254,139257	4,1882	19,22	5,54669			12	0,39047	0,39831	0,97189	0,47067													
16	13	161,751246	2,9905	17,62	5,57816			13	-0,58575	-1,08927	0,61658	0,48792													
17	14	329,675398	4,66999	16,03	5,5666			14	1,18863	0,99669	0,26128	0,48158													
18	15	292,135658	4,20424	14,43	5,63707			15	0,79197	0,41822	-0,09402	0,52022													
19	16	384,800268	4,88796	12,84	6,72515			16	1,77111	1,26742	-0,44932	1,11674													
20	17	257,456277	3,66599	11,25	6,28727			17	0,42552	-0,25029	-0,80463	0,87668													
21	18	358,17181	4,44633	9,65	6,60795			18	1,48974	0,71891	-1,15993	1,05249													
22	19	330,798029	4,12284	8,06	7,68185			19	1,2005	0,31712	-1,51523	1,64124													
23	20	164,448185	2,45993	6,47	7,89224			20	-0,55725	-1,74825	-1,87053	1,75658													
24																									
25	ср	217,185579	3,86751	14,8562	4,68817			ср	3E-16	-5,8E-16	0	0													
26	сигм	94,6380014	0,80514	4,48552	1,82404			сигм	1	1	1	1													
27																									
28																									
29																									
30																									

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon,$$

где  $t_y, t_{x_1}, \dots, t_{x_m}$  - стандартизованные переменные:  $t_y = \frac{y - \bar{y}}{\sigma_y}$ ,

$t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ , для которых среднее значение равно нулю:  $\bar{t}_y = \bar{t}_{x_i} = 0$ , а

среднее квадратическое отклонение равно единице:  $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$ ;  $\beta_i$  -

стандартизованные коэффициенты регрессии.

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$$

	A	B	C	D	E	F	G	H	I	J	K
31											
32		ВЫВОД ИТОГОВ									
33											
34		<i>Регрессионная статистика</i>									
35		Множест	0,9751723								
36		R-квадрат	0,95096101								
37		Нормиров	0,9417662								
38		Стандарт	0,24758583								
39		Наблюд	20								
40											
41		<i>Дисперсионный анализ</i>									
42			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значимость F</i>				
43		Регрессия	3	19,0192	6,33974	103,424	1,1E-10				
44		Остаток	16	0,98078	0,0613						
45		Итого	19	20							
46											
47		<i>Коэффициент стандартная статистика Значение нижние 95% верхние 95% нижние 95% (рхние 95,0%</i>									
48		Y-пересеч	6,6925E-16	0,05536	1,2E-14	1	-0,11736	0,11736	-0,11736	0,11736	
49		tx1	0,67543879	0,0598	11,2956	4,9E-09	0,54868	0,8022	0,54868	0,8022	
50		tx2	-0,0307399	0,06061	-0,50718	0,61895	-0,15923	0,09775	-0,15923	0,09775	
51		tx3	0,59361463	0,05916	10,0336	2,6E-08	0,46819	0,71903	0,46819	0,71903	
52											
53											
54											
55		ВЫВОД ОСТАТКА									
56											
57		<i>Наблюдение</i>		<i>дсказанное</i>	<i>Остатки</i>						
58		1	-1,988286	0,3007							
59		2	-0,9174825	-0,21316							
60		3	0,32042146	-0,46087							
61		4	-1,5514516	0,15732							
62		5	-0,9105896	0,03699							
63		6	-1,4077595	0,16812							
64		7	-0,143003	-0,08216							
65		8	-0,8665564	-0,00833							

**Вычисление коэффициентов стандартизованного уравнения регрессии (линейная зависимость) Оценка влияния факторов на результативный признак**



# Вычисление коэффициентов стандартизованного уравнения регрессии (степенная зависимость)

	A	B	C	D	E	F	G	H	I	J	K
29											
30		Решение с помощью пакета анализа в Excel									
31		ВЫВОД ИТОГОВ									
32											
33		<b>Регрессионная статистика</b>									
34		Множественный R	0,99702								
35		R-квадрат	0,99404								
36		Нормированный R	0,99292								
37		Стандартная ошибка	0,08631								
38		Наблюдения	20								
39											
40		<b>Дисперсионный анализ</b>									
41			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значимость F</i>				
42		Регрессия	3	19,8808207	6,62694	889,67688	5,3E-18				
43		Остаток	16	0,11917927	0,00745						
44		Итого	19	20							
45											
46			<i>Коэффициент</i>	<i>стандартная ошибка</i>	<i>статистика t</i>	<i>P-Значение</i>	<i>нижние 95%</i>	<i>верхние 95%</i>	<i>нижние 95,0%</i>	<i>верхние 95,0%</i>	
47		Y-пересечение	-1,3E-15	0,01929858	-6,7E-14	1	-0,04091	0,040911	-0,04091	0,04091	
48		tlx1	0,65698	0,02143594	30,6484	1,225E-15	0,61154	0,70242	0,61154	0,70242	
49		tlx2	-0,04109	0,02100054	-1,9565	0,0680924	-0,08561	0,003432	-0,08561	0,00343	
50		tlx3	0,61225	0,02049779	29,8692	1,837E-15	0,5688	0,655705	0,5688	0,65571	
51											
52											
53											
54		ВЫВОД ОСТАТКА									
55											
56		<i>Наблюдение</i>	<i>исказанное</i>	<i>Остатки</i>							
57		1	-2,31044	-0,0493528							
58		2	-0,96898	-0,124712							
59		3	0,06291	0,04076841							
60		4	-1,55801	-0,0348945							
61		5	-0,86617	0,16058997							
62		6	-1,34101	0,05609001							
63		7	-0,02531	0,05094713							
64		8	-0,68632	-0,0210362							

## **ЛР- 4. Автоматизация представления данных**

# Подключение надстройки **Пакет анализа VBA** и настройка ленты

The image shows the 'Надстройки' (Add-ins) dialog box and the 'Настройка ленты' (Ribbon Configuration) task pane in Microsoft Excel.

**Надстройки (Add-ins):**

- Доступные надстройки (Available add-ins):
  - Neural tools package
  - Инструменты для евро
  - Пакет анализа**
  - Пакет анализа — VBA
  - Поиск решения
- Buttons: ОК, Отмена, Обзор..., Автоматизация...
- Пакет анализа (Packaging Tools):

Содержит инструменты для анализа научных и финансовых данных

**Настройка ленты (Ribbon Configuration):**

- Выбор команд (Select commands):
  - Часто используемые команды (Popular commands)
  - Список команд: Быстрая печать, Вернуть, Все типы диаграмм..., Вставить, Вставить столбцы на лист, Вставить строки на лист, Вставить функцию..., Вставить ячейки..., Вырезать, Выровнять по центру, Границы, Диспетчер имен, Задать, Закрепить области, Копировать, Макросы, Настраиваемая сортировка..., Обновить все, Объединить и поместить в..., Орфография...
- Настроить ленту (Configure ribbon):
  - Основные вкладки (Main tabs):
    - Главная
    - Вставка
    - Разметка страницы
    - Формулы
    - Данные
    - Рецензирование
    - Вид
    - Разработчик** (Developer)
      - Код
      - Надстройки
      - Элементы управления
      - XML
      - Изменение
      - Надстройки**
      - Команда
      - Удаление фона
- Buttons: Добавить >>, << Удалить

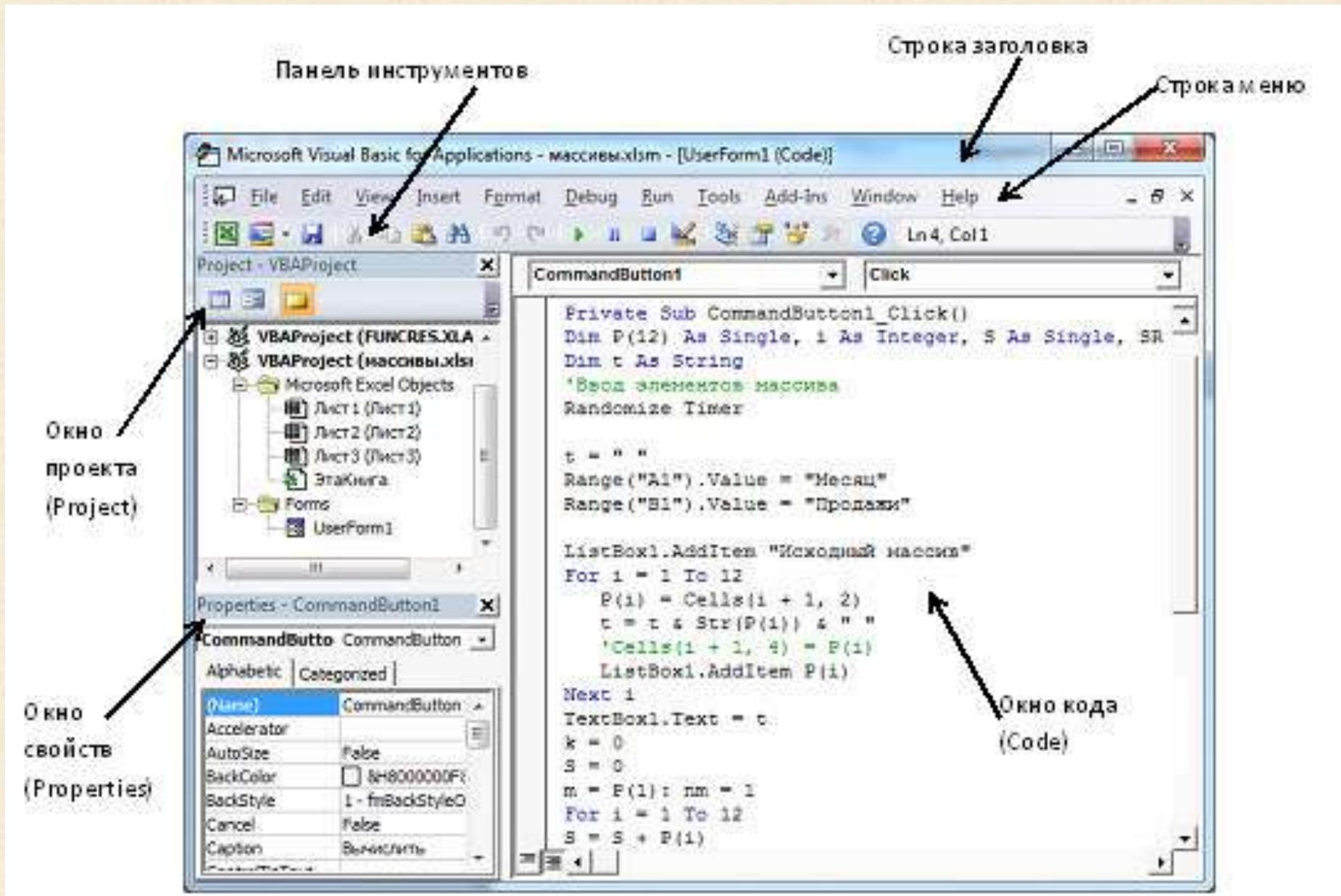
# Окно редактора VBA

33

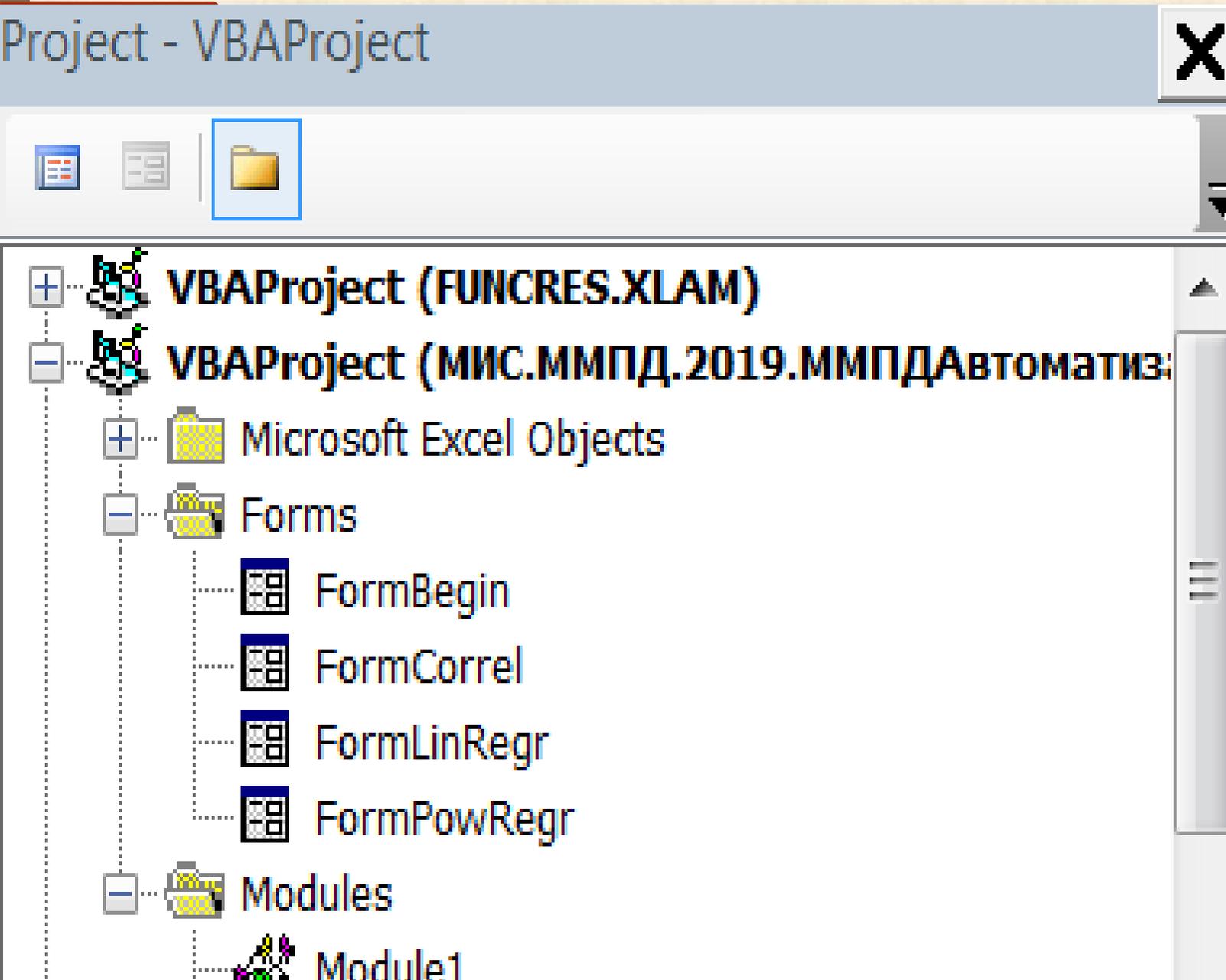
## Запуск редактора Visual Basic Editor (VBE)

- на вкладке **Разработчик** (настроить ленту, если нет вкладки) в группе **Код** кнопка **Visual Basic**.

- Сочетание **ALT + F11**



# Создание начальной формы



- Зайти в редактор VBA ( ALT+F11 или через Разработчик).
- Insert, UserForm
- Предварительно добавить в Книгу Excel доп. лист, на котором находятся исходные данные

# Настройка свойств формы

35

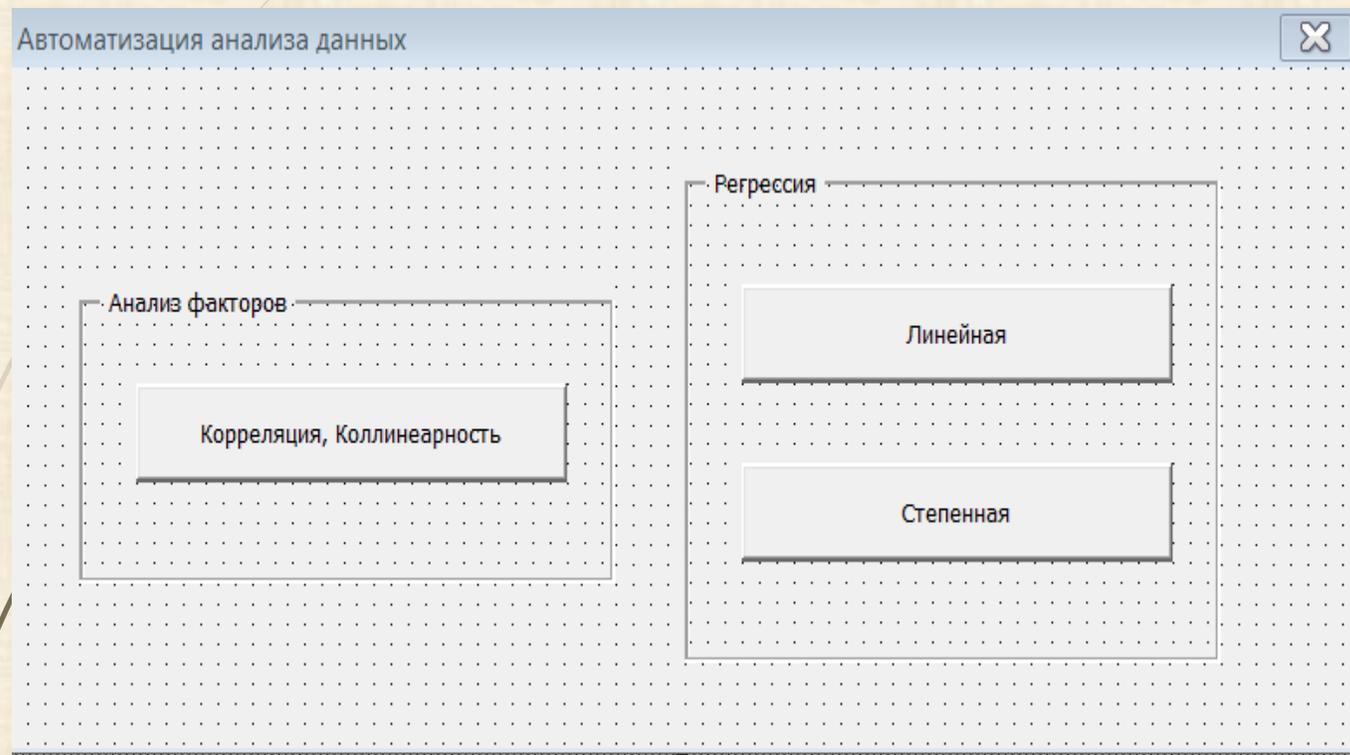
Properties - FormBegin

FormBegin UserForm

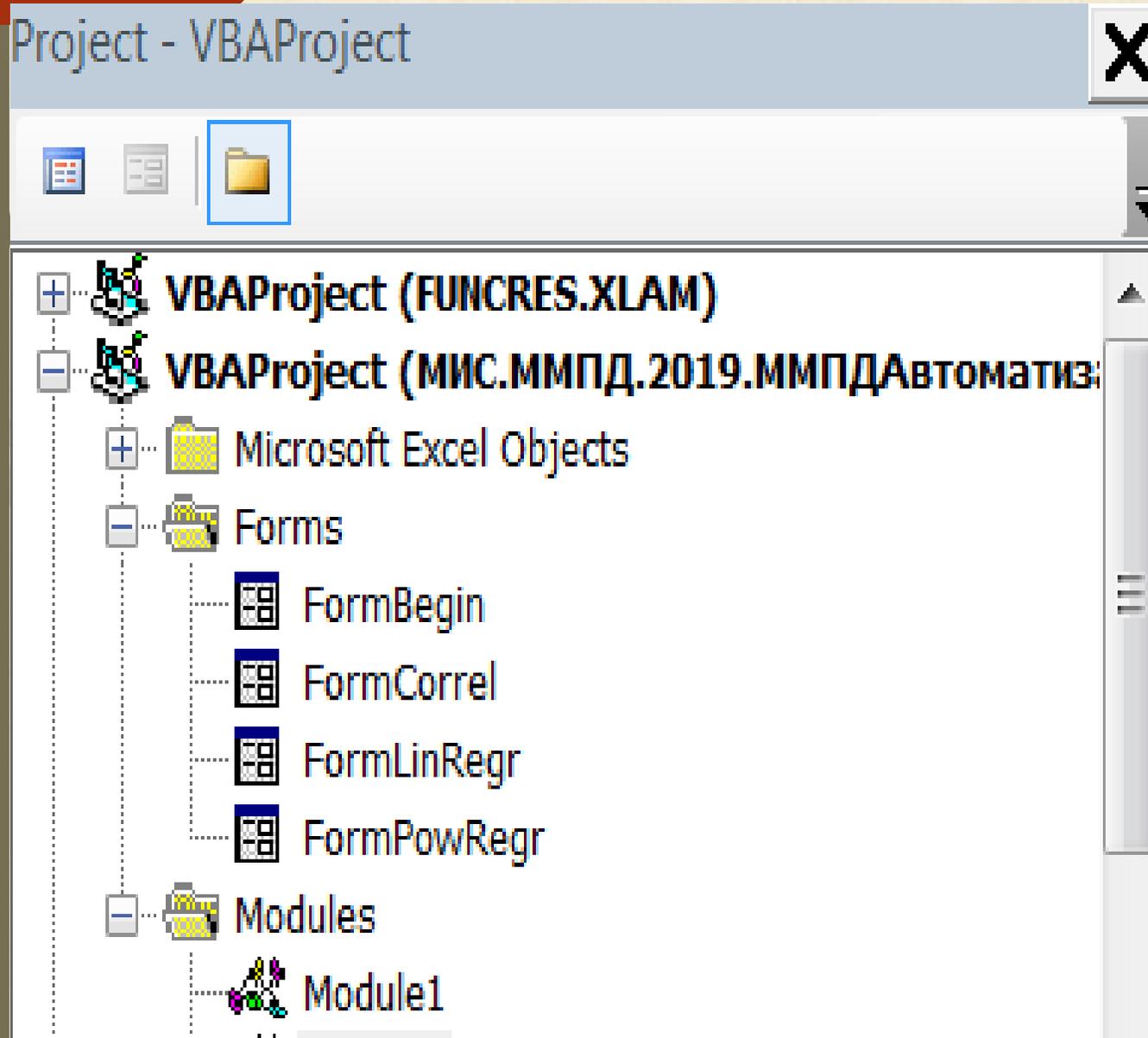
Alphabetic Categorized

(Name)	FormBegin
BackColor	<input type="checkbox"/> &H8000000F&
BorderColor	<input checked="" type="checkbox"/> &H80000012&
BorderStyle	0 - fmBorderStyleNone
Caption	Автоматизация представления данных
Cycle	0 - fmCycleAllForms
DrawBuffer	32000
Enabled	True
Font	Tahoma
ForeColor	<input checked="" type="checkbox"/> &H80000012&
Height	321,6
HelpContextID	0
KeepScrollBarsVisible	3 - fmScrollBarsBoth

# Создание командных кнопок, инструменты ToolBox



Для командных кнопок необходимо задать подходящие Name и Caption. Например, для кнопки с заголовком Caption= «Линейная» имя Name= CommandLinRegr



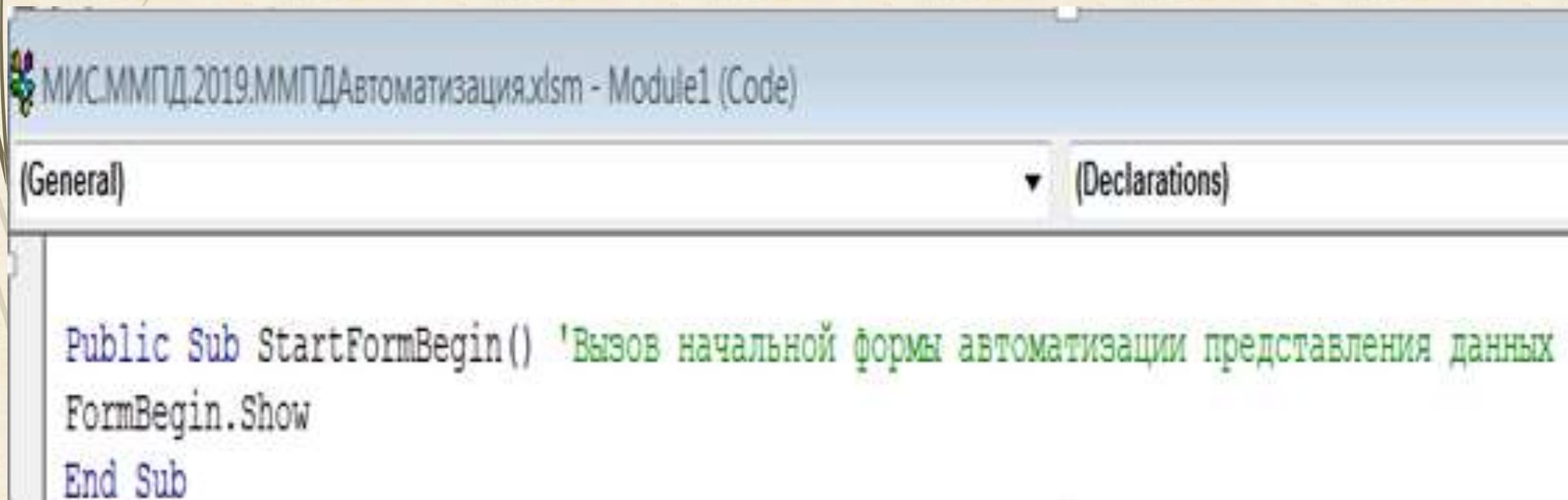
➤ Insert, Module

➤ нет необходимости создавать отдельные стандартные модули для каждой процедуры

➤ При вставке модуля редактор автоматически создает папку для стандартных модулей  
**Modules**

# Процедура вызова стартовой формы в Module1

- ➔ Insert, Procedure, дать имя процедуре (макросу), например, StartFormBegin
- ➔ Метод FormBegin.Show



The screenshot shows the VBA code editor for a module named "МИС.ММПД.2019.ММПДАвтоматизация.xlsm - Module1 (Code)". The editor is set to the "Declarations" view. The code defines a public sub procedure named "StartFormBegin" which calls the "Show" method of the "FormBegin" object. The procedure is enclosed in a comment: "'Вызов начальной формы автоматизации представления данных'".

```
Public Sub StartFormBegin() 'Вызов начальной формы автоматизации представления данных
FormBegin.Show
End Sub
```

- Процедура `StartFormBegin`, записанная в стандартном модуле `Module1`, – макрос, отображаемый в Excel в окне Макросы
- Макрос запускают с помощью панели инструментов, командной кнопки, сочетания клавиш и т.д..
- Чтобы настроить запуск макроса по сочетанию клавиш, нужно перейти из редактора VBA в Excel, вызвать окно параметров макроса `StartFormBegin` (разработчик, макросы, параметры), в котором задать сочетание клавиш, например, `ctrl+shift+A (RU)`, по которому будет запускаться макрос.

# Создать форму линейной регрессии

Линейная регрессия

Перенести данные с листа исходных данных на расчетный лист

Рассчитать коэффициенты уравнения регрессии и среднюю ошибку аппроксимации COA

Вывести COA

COA %

Если COA <8%, то уравнение регрессии в целом значимо

- Создать форму выполнения линейной Insert, UserForm
- Задать заголовок и имя Name=FormLinRegr для формы
- Для каждой кнопки задать заголовок и имя, которое будет использоваться в процедурах обработки событий, и заголовок

# Процедуры модуля формы

41

## FormBegin

- Записать в модуле стартовой формы процедуры вызова вспомогательных форм

(General)

```
' Процедуры модуля формы FormBegin
```

```
Private Sub CommandLinRegr_Click() 'вызов формы линейной регрессии  
FormLinRegr.Show  
End Sub
```

```
Private Sub CommandPowRegr_Click() 'вызов формы степенной регрессии  
FormPowRegr.Show  
End Sub
```

# Значения кнопок

42

Линейная регрессия

Перенести данные с листа исходных данных на расчетный лист

Рассчитать коэффициенты уравнения регрессии и среднюю ошибку аппроксимации COA

Вывести COA

COA %

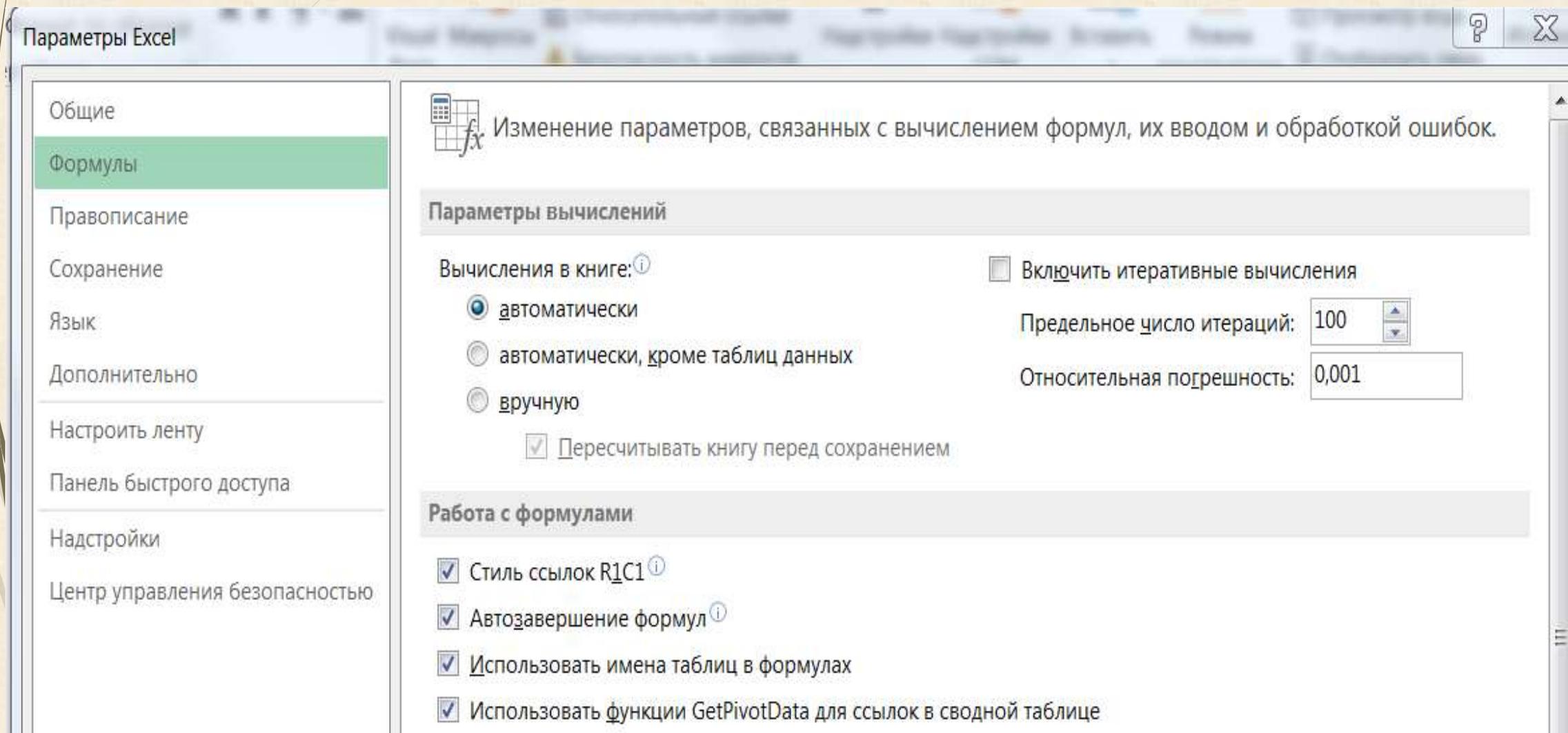
Если COA <8%, то уравнение регрессии в целом значимо

Кнопка «Перенести данные .....» нужна для переноса данных произвольного варианта, записанных на листе «ИсхДанМнож», на расчетный лист линейной регрессии «МножРегрессияЛин»

- Кнопка «Рассчитать коэффициенты ...и COA» должна запускать окно получения уравнения регрессии с помощью вкладки Анализ данных на панели инструментов Excel.
- Кнопка «Вывести COA» должна обеспечить отображение в текстовом поле значения COA в процентах

# Настройка стиля ссылок R1C1

(строки и столбцы листа обозначаются числами)



# Процедура переноса исходных

44

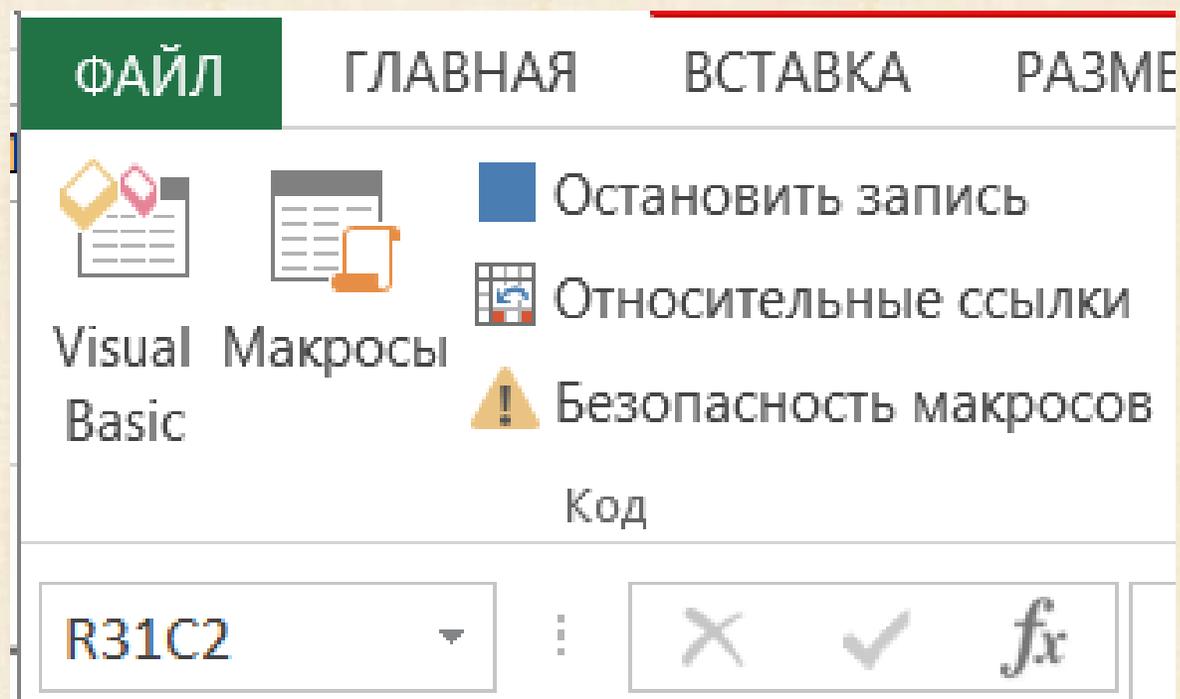
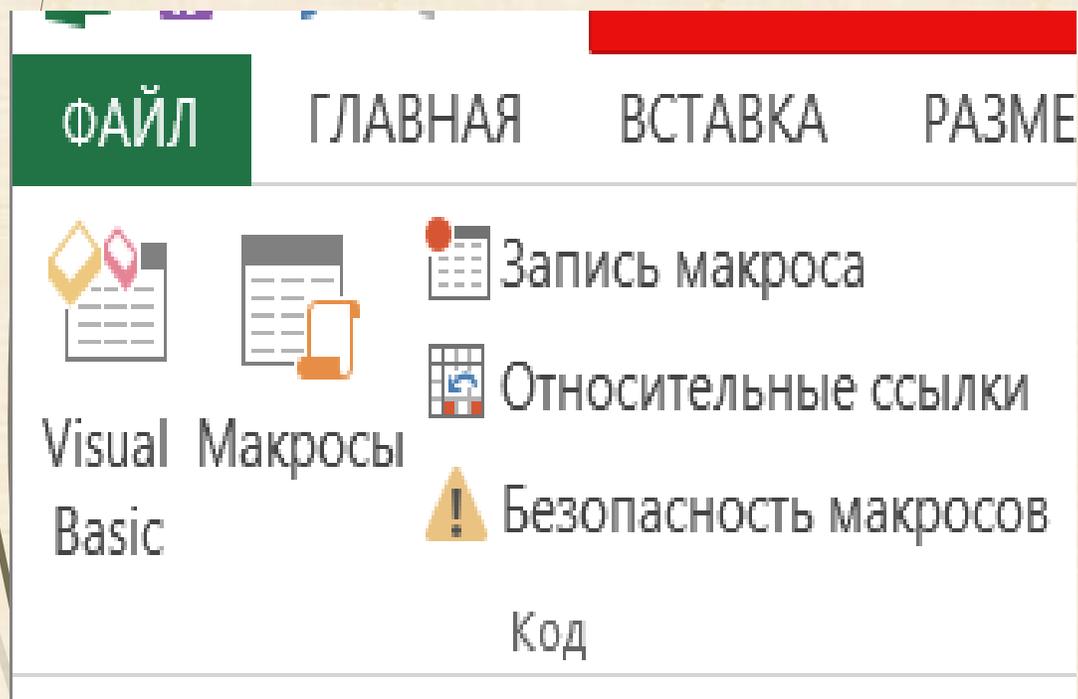
' Процедуры модуля формы FormLinRegr

```
Private Sub CommandSetData_Click() 'неколлинеарные данные перенос исходных данных с листа "ИсхДанМнож" на лист "МножРе:  
For i = 47 To 66  
For j = 3 To 6  
Worksheets("МножРегрессияЛин").Cells(i - 41, j) = Worksheets("ИсхДанМнож").Cells(i, j)  
Next j  
Next i  
End Sub
```

	1	2	3	4	5	6		1	2	3	4	5	6
45							5	№	y	x1	x2	x3	
46		№	y	x1	x2	x3	6	1	57,48	2,58	9,75	1,78	
47		1	57,48	2,58	9,75	1,78	7	2	110,18	3,72	11,15	2,18	
48		2	110,18	3,72	11,15	2,18	8	3	203,89	5,16	12,56	2,30	
49		3	203,89	5,16	12,56	2,30	9	4	85,25	2,80	13,97	2,66	
50		4	85,25	2,80	13,97	2,66	10	5	134,51	3,61	15,37	2,57	
51		5	134,51	3,61	15,37	2,57	11	6	99,87	2,81	16,78	3,12	
52		6	99,87	2,81	16,78	3,12	12	7	195,88	4,22	18,18	3,41	
53		7	195,88	4,22	18,18	3,41	13	8	134,39	3,16	19,59	3,95	
54		8	134,39	3,16	19,59	3,95	14	9	223,84	4,26	21,00	4,62	
55		9	223,84	4,26	21,00	4,62	15	10	289,68	4,85	22,40	4,73	
56		10	289,68	4,85	22,40	4,73	16	11	275,38	4,55	20,81	4,92	
57		11	275,38	4,55	20,81	4,92	17	12	254,14	4,19	19,22	5,55	
58		12	254,14	4,19	19,22	5,55	18	13	161,75	2,99	17,62	5,58	
59		13	161,75	2,99	17,62	5,58	19	14	329,68	4,67	16,03	5,57	
60		14	329,68	4,67	16,03	5,57	20	15	292,14	4,20	14,43	5,64	
61		15	292,14	4,20	14,43	5,64	21	16	384,80	4,89	12,84	6,73	
62		16	384,80	4,89	12,84	6,73	22	17	257,46	3,67	11,25	6,29	
63		17	257,46	3,67	11,25	6,29	23	18	358,17	4,45	9,65	6,61	
64		18	358,17	4,45	9,65	6,61	24	19	330,80	4,12	8,06	7,68	
65		19	330,80	4,12	8,06	7,68	25	20	164,45	2,46	6,47	7,89	
66		20	164,45	2,46	6,47	7,89							

# Автоматическая запись макроса

45



# Процедура Регрессии, полученная с помощью автоматической записи макроса

46

## Процедуры модуля формы FormLinRegr

```
Private Sub CommandCalcRegr_Click() ' нахождение уравнения линейной регрессии и расчет COA,  
' для получения кода используются автоматическая запись макроса  
Application.Run "ATPVBAEN.XLAM!Regress", ActiveSheet.Range("$C$5:$C$25"),  
    ActiveSheet.Range("$D$5:$F$25"), False, True, , ActiveSheet.Range("$B$31")  
    , True, False, False, False, , False  
End Sub
```

автоматически записывается в Modules,  
затем перенести в процедуру формы  
FormLinRegr

# Процедура вывода COA в текстовое поле

Процедуры модуля формы FormLinRegr

```
Private Sub CommandGetError_Click() ' отображение COA в заданном формате в текстовом поле
    TextError.Value = Format(Worksheets("МножРегрессияЛин").[R27C8] * 100, "###.###")
End Sub
```

# Результаты автоматических расчетов

6 7 8 9

Автоматизация анализа данных

Анализ факторов

Корреляция, Коллинеарность

Вычисление коэффициентов парной корреляции

Внести данные с листа исходных данных

Рассчитать коэффициенты парной корреляции

Вывести коэффициенты корреляции  $K_{ij}$

$K_{x1x2}$  1,00

$K_{x1x3}$  0,18

$K_{x2x3}$  0,18

Определитель корреляционной матрицы Det -4,09E-21

Если Det близок к 1, то факторы не коллинеарны, в противном случае - коллинеарны

Линейная регрессия

Перенести данные с листа исходных данных на расчетный лист

Рассчитать коэффициенты уравнения регрессии и среднюю ошибку аппроксимации COA

Вывести COA

COA % 10,46%

Если COA <8%, то уравнение регрессии в целом значимо

## Задания 3,4 контрольной работы

- Задание 3 соответствует ЛР-3
- Задание 4 соответствует ЛР-4

(исходные данные по вариантам представлены в на листе  
Варианты ЛР3-6)

# **Лекция 3**

## **Компонентный анализ и кластерный анализ**

# Структура лекции

- **Компонентный анализ: Алгебраические и статистические основы компонентного анализа, ЛР-5**
- **Кластерный анализ (иерархический и метод k-средних) ЛР-6**
- **Задания 5,6 контрольной работы**

- Компонентный анализ является основным методом факторного анализа, позволяет определить структурную зависимость между факторами
- Идея метода заключается в замене сильно коррелированных переменных (факторов) новыми переменными (главными компонентами), между которыми корреляция отсутствует
- В результате его использования получается сжатое описание меньшего объема, несущее почти всю информацию, содержащуюся в исходных данных
- Главные компоненты получаются из исходных переменных как линейные комбинации исходных переменных

# Алгебраические и статистические основы Компонентного анализа

4

- Отыскание собственных значений корреляционной матрицы векторов центрированных и нормированных по столбцам исходных данных
- Выбор среди собственных значений существенных ( $>1$ )
- Отыскание соответствующих собственных векторов (главных компонент) и **матрицы нагрузок**, столбцами которой являются нормированные собственные векторы
- Отыскание **матрицы счетов (весов) или результатов**, строки которой - ортогональные проекции строк исходных данных на главные компоненты

# Алгебраические и статистические основы Компонентного анализа

5

- Собственные значения корреляционной матрицы исходных данных равны **дисперсиям столбцов матрицы счетов** и показывают вклад соответствующей компоненты в суммарную дисперсию
- Данные, спроецированные на главные компоненты, характеризуются **наибольшими дисперсиями** (рассеяниями), поэтому позволяют сохранить максимум информации по исходным данным, при этом уменьшив размерность задачи

# Матрица нагрузок P и матрица счетов T

- $X_0$  - центрированная и нормированная **матрица данных (m×n)**, элементы которой

$$x_{0ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}}$$

$$X_0 = \begin{pmatrix} x_{011} & \dots & x_{0n1} \\ \dots & \dots & \dots \\ x_{0m1} & \dots & x_{0mn} \end{pmatrix}$$

- $R(X_0)$  – **корреляционная матрица по  $X_0$  (n×n)**
  - $P$  – **матрица нагрузок (n×k)**, столбцы которой – главные компоненты, собственные векторы для  $R(X_0)$ , соответствующие существенным ( $>1$ ) собственным значениям
- $k$  - количество главных компонент

$$P = \begin{pmatrix} p_{11} & \dots & p_{1k} \\ \dots & \dots & \dots \\ p_{n1} & \dots & p_{nk} \end{pmatrix}$$

# Матрица счетов $T$ и матрица ошибок $E$

7

- $T = X_0 \cdot P$  – матрица счетов ( $m \times k$ ), сжатая матрица данных, элементы которой в строках – проекции строк  $X_0$  на главные компоненты

$$T = \begin{pmatrix} t_{11} & \dots & t_{1k} \\ \dots & \dots & \dots \\ t_{m1} & \dots & t_{mk} \end{pmatrix}$$

- $P^T$  – транспонированная для  $P$  матрица нагрузок, размерность ( $k \times n$ )
- $E = X_0 - T \cdot P^T$  – матрица ошибок ( $m \times n$ ), должна быть близка к 0-матрице

# Создание m-файла в MatLab для выполнения компонентного анализа

## Алгоритм решения

- New, script, имя КомпАнализ
- Внести исходные данные
- Найти  $X_0$  и  $R(X_0)$  (ф-и **mean, std, corrcoef**)
- Найти собственные значения и векторы (ф-я **eig**)
- Найти матрицу **нагрузок** P
- Найти матрицу **счетов** T
- Найти матрицу **ошибок** E

## Компонентный анализ (ЛР-5)

Создадим m-файл для выполнения компонентного анализа в соответствии со своим вариантом (варианты ЛР 3-6 приведены в Лекции 2)

**Код:**

**Ввод  
комментариев  
(disp, %),  
исходных  
данных X  
Нахождение  
X0 и R(X0)  
ф-и  
mean,  
std,  
corrcoef**

## Случай коллинеарных факторов

```
1 - disp('исходная матрица данных')%вывод комментариев в командном окне
2 - X=[2.582159403  9.746478209 1.784386847
3     3.717586719 13.15276016 2.177451969
4     5.156650524 17.46995157 2.301271129
5     2.796219378 10.38865813 2.663718058
6     3.60983943  12.82951829 2.565220808
7     2.813099187 10.43929756 3.120964854
8     4.218760074 14.65628022 3.413400664
9     3.159280469 11.47784141 3.950804217
10    4.261704878 14.78511463 4.620459139
11    4.849698185 16.54909455 4.727342169
12    4.549262511 15.64778753 4.915422902
13    4.188204708 14.56461412 5.546694077
14    2.990496859 10.97149058 5.578163639
15    4.669988848 16.00996654 5.566599794
16    4.204235719 14.61270716 5.637071047
17    4.887957364 16.66387209 6.725146587
18    3.665990378 12.99797113 6.287265111
19    4.446334701 15.3390041  6.607954967
20    4.122841144 14.36852343 7.681852146
21    2.459931334 9.379794003 7.89223892
22    ]
23 - [m n] = size(X) %нахождение размерности матрицы
24 - XMean = mean(X) %нахождение среднего значения по столбцам по умолчанию
25 - XStd = std(X) %нахождение выборочной дисперсии по столбцам, делится на n-1,
26     % а в генеральной на n
27 - X0 = (X - repmat(XMean,m,1))./ repmat(XStd,m,1)% центрирование и нормировка
28     %матрицы, применение функции повторения матрицы, n строк, 1 столбец
29 - X0Mean = mean(X0) %проверка центрирования
30 - X0Std = std(X0)
31 - RX0=corrcoef(X0) %вычисление корреляционной матрицы
```

```
32 - disp(' Матрицы собств векторов и собственных знач по возрастанию собств значений')
33 - [U D] = eig(RX0) %нахождение матрицы координат собственных векторов (матрицы
34 %нагрузок) U и преобразованной корреляционной матрицы V в базисе из
35 %собственных векторов (собственные значения по возрастанию)
36 - disp(' Матрица нагрузок по убыванию собственных значений')
37 - P=[U(:,3) U(:,2)] %выбор существенных компонент с наибольшими собственными
38 %значениями (дисперсиями)
39 - disp(' Матрица счетов')
40 - T=X0*P %нахождение матрицы весов (новый вектор данных)
41 - PT=P.' % транспонирование матрицы нагрузок
42 - disp(' Матрица ошибок')
43 - E=X0-T*P.' % проверка - получение исходной матрицы
44 - Dfli=flipud(diag(D))%зеркальное отображение относительно середины
45 - disp(' вклад компонент одной, двух, трех, начиная с существенных')
46 - cumsum(flipud(diag(D))) / sum(diag(D)) %определение вклада каждой
47 %компоненты, использование функции flipud, которая изменяет порядок
48 %следования элементов на противоположный
49 - disp('корреляционная матрица сжатых данных')
50 - RT=corrcoef(T) %вычисление корреляционной матрицы преобразованных данных
51 - covT=cov(T)
```

**ф-я eig, матрицы нагрузок P, счетов T, ошибок E**

Command Window

исходная матрица данных

X =

2.5822	9.7465	1.7844
3.7176	13.1528	2.1775
5.1567	17.4700	2.3013
2.7962	10.3887	2.6637
3.6098	12.8295	2.5652
2.8131	10.4393	3.1210
4.2188	14.6563	3.4134
3.1593	11.4778	3.9508
4.2617	14.7851	4.6205
4.8497	16.5491	4.7273
4.5493	15.6478	4.9154
4.1882	14.5646	5.5467
2.9905	10.9715	5.5782
4.6700	16.0100	5.5666
4.2042	14.6127	5.6371
4.8880	16.6639	6.7251
3.6660	12.9980	6.2873
4.4463	15.3390	6.6080
4.1228	14.3685	7.6819
2.4599	9.3798	7.8922

m =

20

Command Window

n =

3

XMean =

3.8675 13.6025 4.6882

XStd =

0.8261 2.4782 1.8714

X0 =

-1.5560	-1.5560	-1.5516
-0.1815	-0.1815	-1.3416
1.5606	1.5606	-1.2754
-1.2969	-1.2969	-1.0818
-0.3119	-0.3119	-1.1344
-1.2764	-1.2764	-0.8374
0.4252	0.4252	-0.6812
-0.8574	-0.8574	-0.3940
0.4772	0.4772	-0.0362
1.1890	1.1890	0.0209
0.8253	0.8253	0.1214
0.3882	0.3882	0.4588
-1.0617	-1.0617	0.4756
0.9715	0.9715	0.4694

fx

**RUN and  
ADVANCE****Результаты  
выполнени  
я**

X0Mean =

1.0e-15 \*

-0.5995   -0.6994   0.1332

X0Std =

1.0000   1.0000   1.0000

RX0 =

1.0000   1.0000   0.1798

1.0000   1.0000   0.1798

0.1798   0.1798   1.0000

Матрицы собств векторов и собственных знач по возрастанию собств значений

U =

0.7071   0.1648   0.6876

-0.7071   0.1648   0.6876

-0.0000   -0.9724   0.2331

D =

0.0000   0   0

0   0.9390   0

*fx*

0   0   2.0610

# Матрицы собственных векторов и значений по возрастанию

Матрица нагрузок по убыванию собственных значений

P =

0.6876	0.1648
0.6876	0.1648
0.2331	-0.9724

Матрица счетов

T =

-2.5016	0.9959
-0.5624	1.2448
1.8489	1.7548
-2.0357	0.6244
-0.6934	1.0003
-1.9507	0.3936
0.4260	0.8026
-1.2709	0.1005
0.6478	0.1925
1.6401	0.3716
1.1633	0.1540
0.6408	-0.3181
-1.3492	-0.8125
1.4454	-0.1362
0.6788	-0.3587
1.9526	-0.6512
-0.1363	-0.9114
1.2028	-0.7666
0.7980	-1.4537
-1.9443	-2.2267

**Матрицы  
нагрузок P  
и счетов T**

Матрица ошибок

E =

1.0e-09 \*

0.1708	-0.1707	0
-0.7876	0.7876	0
-0.2197	0.2197	0
0.8970	-0.8970	0
-0.1563	0.1563	0.0000
0.2785	-0.2785	0
0.0466	-0.0466	-0.0000
-0.6487	0.6487	0.0000
0.4157	-0.4157	-0.0000
0.4363	-0.4363	-0.0000
0.1211	-0.1211	-0.0000
0.4217	-0.4217	0
-0.6260	0.6260	0.0000
0.2745	-0.2745	-0.0000
-0.9971	0.9971	0
-0.2159	0.2159	0
0.5679	-0.5679	0
0.1225	-0.1225	0
0.0004	-0.0004	0
-0.1016	0.1016	0.0000

**Матрица  
ошибок E**

```
Dfli =
```

```
2.0610  
0.9390  
0.0000
```

```
вклад компонент одной, двух, трех, начиная с существенных
```

```
ans =
```

```
0.6870  
1.0000  
1.0000
```

```
корреляционная матрица сжатых данных
```

```
RT =
```

```
1.0000 -0.0000  
-0.0000 1.0000
```

```
covT =
```

```
2.0610 -0.0000  
-0.0000 0.9390
```

```
fx >>
```

**Расположение  
собств векторов по  
убыванию**

**Определение  
вкладов компонент**

**Определение  
корреляционной и  
ковариационной  
матриц сжатых  
данных**

# Случай неколлинеарных факторов

17

Command Window

RX0 =

1.0000	0.2792	0.1798
0.2792	1.0000	-0.2411
0.1798	-0.2411	1.0000

Матрицы собств векторов и собственных знач по возрастанию собств значений

U =

-0.5739	0.5996	-0.5578
0.6178	-0.1301	-0.7755
0.5375	0.7897	0.2958

D =

0.5309	0	0
0	1.1763	0
0	0	1.2928

Матрица ошибок

E =

0.3599	-0.3874	-0.3371
0.6394	-0.6884	-0.5989
1.0844	-1.1675	-1.0157
-0.0248	0.0267	0.0232
0.2075	-0.2234	-0.1943
-0.3101	0.3339	0.2905
0.0938	-0.1009	-0.0878
-0.5256	0.5658	0.4923
-0.3048	0.3281	0.2855
-0.1963	0.2114	0.1839
-0.2243	0.2415	0.2101
-0.3495	0.3763	0.3274
-0.7094	0.7638	0.6645
0.0848	-0.0913	-0.0795
0.0103	-0.0111	-0.0097
0.2263	-0.2437	-0.2120
-0.0659	0.0709	0.0617
0.3152	-0.3393	-0.2952
0.1320	-0.1421	-0.1236
-0.4429	0.4768	0.4148

Dfli =

1.2928
1.1763
0.5309

fx

**Вывод:**  
**Матрица**  
**ошибок не**  
**нулевая**

```
Dfli =
```

```
1.2928  
1.1763  
0.5309
```

```
вклад компонент одной, двух, трех, начиная с существенных
```

```
ans =
```

```
0.4309  
0.8230  
1.0000
```

```
корреляционная матрица сжатых данных
```

```
RT =
```

```
1.0000 -0.0000  
-0.0000 1.0000
```

```
covT =
```

```
1.2928 -0.0000  
-0.0000 1.1763
```

fx

## Выводы:

**Только три  
компоненты  
обеспечивают  
суммарную  
дисперсию**

**Размерность  
пространства  
данных не  
уменьшить**

- ▶ Кластерный анализ предназначен для разбиения исходных данных на группы, таким образом, чтобы элементы, входящие в одну группу были максимально «схожи», а элементы из разных групп были максимально «отличными» друг от друга.
- ▶ Кластерный анализ – группа методов, используемых для классификации объектов или событий в относительно гомогенные (однородные) группы, которые называют кластерами (clusters).

# Различие между факторным (компонентным) и кластерным анализом

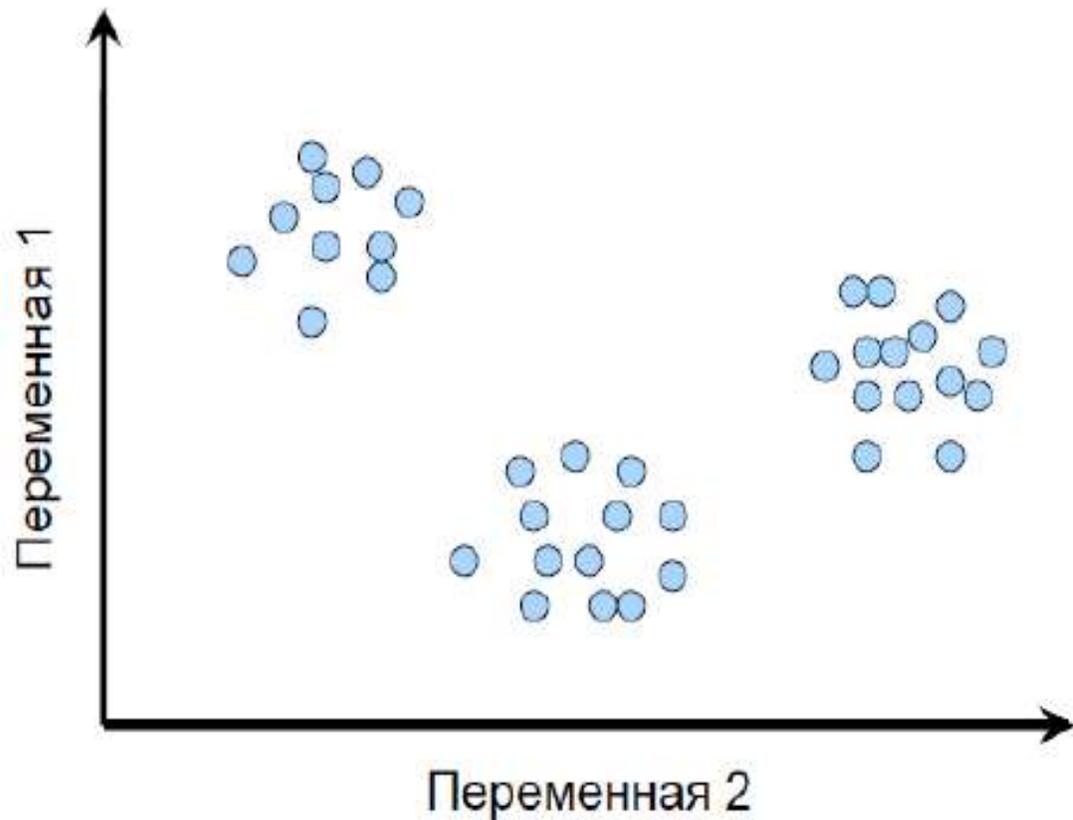
- В факторном анализе группируются столбцы, т.е. цель – анализ структуры множества факторов и выявление обобщенных факторов.
- В кластерном анализе группируются строки, т.е. цель – анализ структуры множества объектов.

## Геометрическая интерпретация

- Каждый объект (респондент) – точка в пространстве признаков.
- Задача кластерного анализа – выделение «сгущений» точек, разбиение совокупности на однородные подмножества объектов (сегментация).

# Как очертить границу кластеров?

## Кластерный анализ в теории



## Кластерный анализ на практике



## КЛАСТЕРНЫЙ АНАЛИЗ

```
graph TD; A[КЛАСТЕРНЫЙ АНАЛИЗ] --> B[Иерархический]; A --> C[Анализ k-средних];
```

### Иерархический

При маленьком количестве наблюдений

### Анализ k-средних

При большом количестве наблюдений

# Этапы кластерного анализа

Формулировка проблемы



Выбор способа измерения расстояния  
(определение меры сходства объектов)



Выбор метода кластеризации



Принятие решения о количестве кластеров



Интерпретация и профилирование кластеров



Оценка достоверности кластеризации

# ИЕРАРХИЧЕСКИЙ КЛАСТЕРНЫЙ АНАЛИЗ

25

- Первоначально каждый объект образует свой отдельный кластер.
- На каждом последующем шаге два соседних (близких по выбранной мере) кластера объединяются в один.
- Этот процесс продолжается до тех пор, пока не останется только один кластер.

# Мера их близости между объектами

26

## или кластерами- расстояние

- **Расстояние**– это всегда неотрицательное вещественное число, для которого выполняется свойство коммутативности и неравенство треугольника.
- **При кластерном анализе** вначале рассматриваются расстояния между объектами, затем между группами объектов (кластерами).

Название	Формула	Обозначения
1. Евклидово расстояние	$d_2(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ki} - x_{kj})^2}$	<p><math>i, j</math> – номера объектов;  <math>p</math> – количество переменных, характеризующих объект;  <math>k</math> – номер признака;  <math>x_{ki}</math> – значение <math>k</math>-го признака у <math>i</math>-го объекта;  <math>S</math> – ковариационная</p>
2. $L_1$ -норма	$d_1(x_i, x_j) = \left[ \sum_{k=1}^p  x_{ki} - x_{kj}  \right]$	
3. Сюрремум норма	$d_\infty(x_i, x_j) = \sup_{k=1,2,\dots,p} \left\{  x_{ki} - x_{kj}  \right\}$	
4. $L_p$ -норма	$d_p(x_i, x_j) = \left[ \sum_{k=1}^p  x_{ki} - x_{kj}  \right]^{1/p}$	

# Методы кластеризации

28

Методы кластеризации различают по способу вычисления расстояний между кластерами. Существуют следующие основные методы кластеризации:

- Межгрупповая связь (Between-groups linkage)
- Ближайший сосед (Nearest neighbor)
- Самый дальний сосед (Furthest neighbor)
- Центроидная кластеризация (Centroid clustering)
- Медианная кластеризация (Median clustering)
- Метод Варда (Уорда) (Ward's method)

## Подробнее о методах кластеризации

**МЕЖГРУППОВАЯ СВЯЗЬ.** Расстояние между кластерами равно среднему значению всех расстояний между всеми возможными парами точек из обоих кластеров. Данный метод устанавливается по умолчанию.

**МЕТОД БЛИЖАЙШЕГО СОСЕДА.** Расстояние между кластерами равно расстоянию между двумя ближайшими соседями классов.

**МЕТОД ДАЛЬНЕГО СОСЕДА.** Межкластерное расстояние равно расстоянию между самыми дальними объектами.

# Подробнее о методах кластеризации

**ЦЕНТРОИДНЫЙ МЕТОД.** Расстояние между двумя кластерами определяется как евклидово расстояние между центрами (средними) этих кластеров

**МЕДИАННЫЙ МЕТОД.** Межкластерное расстояние равно расстоянию между точками с медианными значениями признаков в классах.

**МЕТОД УОРДА.** В качестве целевой функции применяется сумма квадратов расстояний между каждой точкой и средней по кластеру, содержащему этот объект. На каждом шаге объединяются такие два кластера, которые приводят к минимальному увеличению целевой функции.

# Виды стандартизации данных, применяемой при кластеризации

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{\sigma_{X_j}}$$

Z-стандартизация

Из значений вычитается среднее и затем они делятся на стандартное отклонение.

Разброс от -1 до 1

Линейным преобразованием переменных добиваются разброса значений от -1 до 1.

Разброс от 0 до 1

Линейным преобразованием переменных добиваются разброса значений от 0 до 1.

Максимум 1

Значения переменных делятся на их максимум.

Среднее 1

Значения переменных делятся на их среднее.

Стандартное отклонение 1

Значения переменных делятся на стандартное отклонение.

# Как принимаются решения о числе кластеров

32

- ▶ Когда мера расстояния между двумя кластерами увеличивается скачкообразно, процесс объединения в новые кластеры необходимо остановить. Иначе будут объединены кластеры, находящиеся на большом расстоянии друг от друга.
- ▶ Оптимальным считается число кластеров равное разности количества объектов и количества шагов, после которого расстояние увеличивается скачкообразно.
- ▶ Размеры кластеров должны быть значимыми.

# Интерпретация и профилирование кластеров

- Интерпретация и профилирование кластеров включает проверку кластерных центроидов.
- Центроиды—средние значения объектов по каждой из переменных. Центроиды позволяют описывать кластеры.

# КЛАСТЕРНЫЙ АНАЛИЗ МЕТОДОМ k-СРЕДНИХ

**(неиерархический, итерационный)**

**Суть:** Сначала определяются центры кластеров, а затем группируют все объекты в пределах заданных от центров пороговых значений.

## Недостатки:

- Чувствительность к выбросам
- Необходимо заранее задавать количество кластеров, а не как в иерархическом анализе, получать это в качестве результата

## Достоинства:

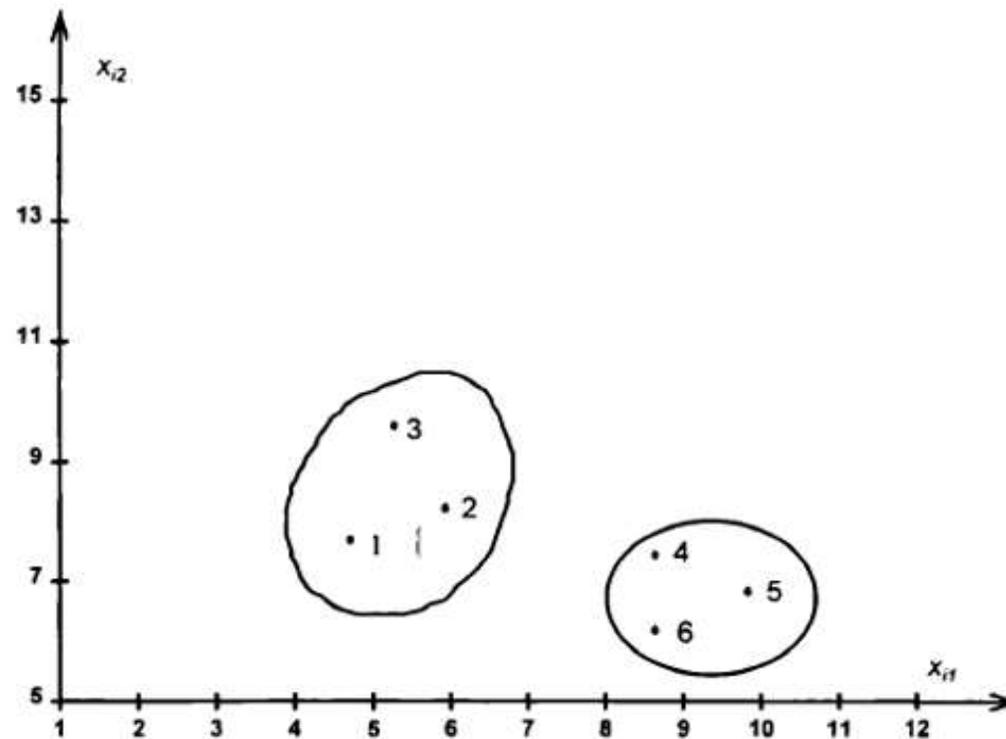
- Простота использования: В качестве метрики используется Евклидово расстояние
- Возможность наглядной интерпретации кластеров с использованием графика «Средних значений в кластерах»

# Пример аналитического решения (иерархический кластерный анализ)

**Пример.** Провести классификацию шести объектов по двум признакам (данные приведены в таблице и на рисунке).

№ объекта $i$	1	2	3	4	5	6
$x_{i1}$	5	6	5	10	11	10
$x_{i2}$	10	12	13	9	9	7

**Решение.** Если изобразить объекты как точки на координатной плоскости, то по евклидову расстоянию, очевидно, что выделяется два кластера. Рассмотрим, как прийти к этому результату аналитически.



Изначально считаем, что каждый объект образует один кластер. Составим матрицу евклидовых расстояний (сходств и различий)

Так как расстояния от всех объектов до 4-го меньше, чем до 5-го, то вычеркиваем пятый столбец и пятую строку в предыдущей матрице

$$R_1 = \{ \rho(x_i, x_j) \} = \begin{pmatrix} 0 & 2,24 & 3 & 5,10 & 6,08 & 5,83 \\ 2,24 & 0 & 1,41 & 5 & 5,83 & 6,40 \\ 3 & 1,41 & 0 & 6,40 & 7,21 & 7,81 \\ 5,10 & 5 & 6,40 & 0 & 1 & 2 \\ 6,08 & 5,83 & 7,21 & 1 & 0 & 2,24 \\ 5,83 & 6,40 & 7,81 & 2 & 2,24 & 0 \end{pmatrix}$$

Из матрицы расстояний следует, что четвертый и пятый объекты наиболее близки  $\rho_{4,5} = 1,00$  и поэтому объединяются в один кластер.

После объединения объектов имеем пять кластеров:

Номер кластера	1	2	3	4	5
Состав кластера	(1)	(2)	(3)	(4,5)	(6)

# Расчет в Excel

	A	B	C	D	E	F	G	H
1								
2	i\j		1	2	3	4	5	6
3		x1i\ x1j	5	6	5	10	11	10
4	1	5	0	1	0	5	6	5
5	2	6	-1	0	-1	4	5	4
6	3	5	0	1	0	5	6	5
7	4	10	-5	-4	-5	0	1	0
8	5	11	-6	-5	-6	-1	0	-1
9	6	10	-5	-4	-5	0	1	0
10								
11	i\j		1	2	3	4	5	6
12		x2i\ x2j	10	12	13	9	9	7
13	1	10	0	2	3	-1	-1	-3
14	2	12	-2	0	1	-3	-3	-5
15	3	13	-3	-1	0	-4	-4	-6
16	4	9	1	3	4	0	0	-2
17	5	9	1	3	4	0	0	-2
18	6	7	3	5	6	2	2	0
19								
20	i\j		1	2	3	4	5	6
21		Rij(1)						
22	1		0.00	2.24	3.00	5.10	6.08	5.83
23	2		2.24	0.00	1.41	5.00	5.83	6.40
24	3		3.00	1.41	0.00	6.40	7.21	7.81
25	4		5.10	5.00	6.40	0.00	1.00	2.00
26	5		6.08	5.83	7.21	1.00	0.00	2.24
27	6		5.83	6.40	7.81	2.00	2.24	0.00

· Евклидово расстояние

$$d_2(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ki} - x_{kj})^2}$$

Таким образом, расстояние  $\rho_{1,(4,5)}$  равно расстоянию от объекта 1 до ближайшего к нему объекта, входящего в кластер  $S_{(4,5)}$  т.е.  $\rho_{1,(4,5)} = \rho_{1,4} = 5,10$ . Тогда матрица расстояний

$$R_2 = \begin{pmatrix} 0 & 2,24 & 3 & 5,10 & 5,83 \\ 2,24 & 0 & 1,41 & 5 & 6,40 \\ 3 & 1,41 & 0 & 6,40 & 7,81 \\ 5,10 & 5 & 6,40 & 0 & 2 \\ 5,83 & 6,40 & 7,81 & 2 & 0 \end{pmatrix}$$

Объединим второй и третий объекты, имеющие наименьшее расстояние  $\rho_{2,3} = 1,41$ . После объединения объектов имеем четыре кластера:

$$S_{(1)}, \quad S_{(2,3)}, \quad S_{(4,5)}, \quad S_{(6)}.$$

## Расчет в Excel

28								
29	$i \setminus j$		1	2	3	4,5	6	
30		$R_{ij}(2)$						
31	1		0.00	2.24	3.00	5.10	5.83	
32	2		2.24	0.00	1.41	5.00	6.40	
33	3		3.00	1.41	0.00	6.40	7.81	
34	4,5		5.10	5.00	6.40	0.00	2.00	
35	6		5.83	6.40	7.81	2.00	0.00	

Вновь найдем матрицу расстояний. Для того чтобы рассчитать расстояние до кластера  $S_{(2,3)}$  воспользуемся матрицей расстояний  $R_2$ . Например, расстояние между кластерами  $S_{(4,5)}$  и  $S_{(2,3)}$  равно

Проведя аналогичные расчеты, получим

$$R_3 = \begin{pmatrix} 0 & 2,24 & 5,10 & 5,83 \\ 2,24 & 0 & 5 & 6,40 \\ 5,10 & 5 & 0 & 2 \\ 5,83 & 6,40 & 2 & 0 \end{pmatrix}$$

Объединим кластеры  $S_{(4,5)}$  и  $S_{(6)}$ , расстояние между которыми, согласно матрице  $R_3$ , наименьшее  $\rho_{(4,5),6} = 2$ . В результате получим три кластера

$$S_{(1)}, \quad S_{(2,3)} \quad \text{и} \quad S_{(4,5,6)}.$$

Матрица расстояний будет иметь вид:

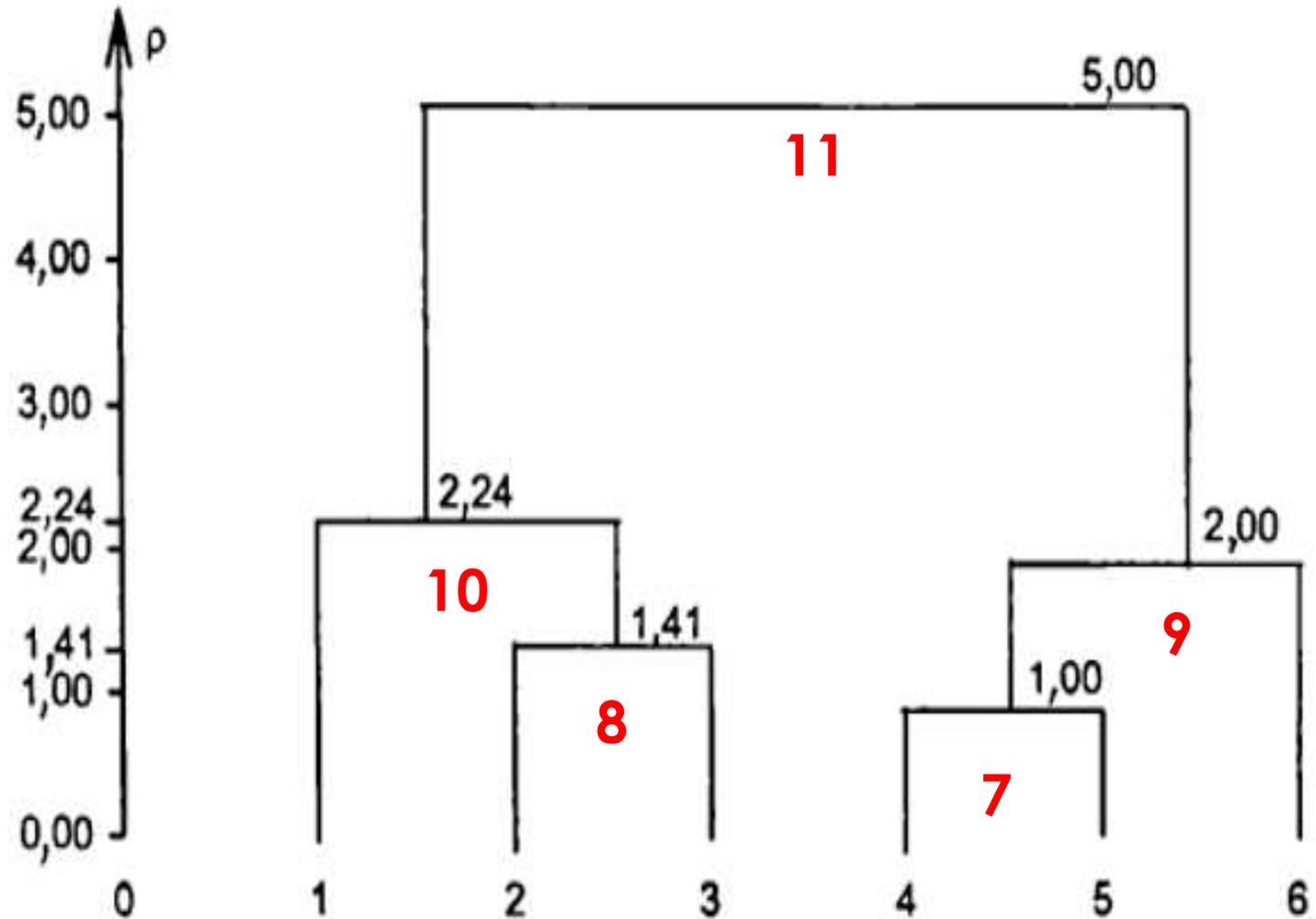
$$R_4 = \begin{pmatrix} 0 & 2,24 & 5,10 \\ 2,24 & 0 & 5 \\ 5,10 & 5 & 0 \end{pmatrix}$$

Объединим теперь кластеры  $S_1$  и  $S_{(2,3)}$ , расстояние между которыми  $P_{1,(2,3)} = 2,24$ . В результате получим два кластера:  $S_{(1,2,3)}$  и  $S_{(4,5,6)}$ , расстояние между которыми, найденное по принципу “ближайшего соседа”,  $P_{(1,2,3),(4,5,6)} = 5$ .

# Матрица иерархического дерева бинарных кластеров

42

7	4	5	1
8	2	3	1,41
9	7	6	2
10	1	8	2,24
11	9	10	5



## Кластерный анализ (ЛР-6)

Создадим m-файл для выполнения кластерного анализа в соответствии со своим вариантом (варианты ЛР 3-6 приведены в Лекции 2)

# Создание m-файла в MatLab для выполнения кластерного анализа

## Алгоритм решения

- New, script, имя КластАнализ
- Внести исходные данные
- Нахождение расстояний между объектами (ф-я **pdist**)
- Иерархическая кластеризация: Формирование кластеров и построение дендрограммы (ф-и **linkage, dendrogram**)
- Кластеризация по межгрупповым средним (ф-я **kmeans**)
- Построение трехмерного графика точек кластеров (**plot3**)

```

КластернАнализ.m x +
1 - disp('исходная матрица данных')%вывод комментариев в командном окне
2 - X=[2.582159403 9.746478209 1.784386847
3 3.717586719 11.15276016 2.177451969
4 5.156650524 12.559042111 2.301271129
5 2.796219378 13.965324062 2.663718058
6 3.60983943 15.371606013 2.565220808
7 2.813099187 16.777887964 3.120964854
8 4.218760074 18.184169915 3.413400664
9 3.159280469 19.590451866 3.950804217
10 4.261704878 20.996733817 4.620459139
11 4.849698185 22.403015768 4.727342169
12 4.549262511 20.809297719 4.915422902
13 4.188204708 19.21557967 5.546694077
14 2.990496859 17.621861621 5.578163639
15 4.669988848 16.028143572 5.566599794
16 4.204235719 14.434425523 5.637071047
17 4.887957364 12.840707474 6.725146587
18 3.665990378 11.246989425 6.287265111
19 4.446334701 9.653271376 6.607954967
20 4.122841144 8.059553327 7.681852146
21 2.459931334 6.465835278 7.89223892] % неколлинеарн данные
22 - disp('нахождение размерности матрицы данных')% вывод в окно заголовка
23 - [m n] = size(X) % выводятся в командное окно
24 - Y=pdist(X); % нахождение расстояний m*(m-1)/2 различных расстояний
25 % ; означает в командное окно не выводить
26 - disp('матрица иерархического дерева бинарных кластеров размерности (m-1)*3')
27 - Z=linkage(Y)% Z выдается в командное окно
28 - disp(' m+i - индекс кластера, где i - номер строки кластера в матрице Z')
29 - disp('первый и второй столбцы - индексы кластеров, которые объединяются')
30 - disp('третий столбец - расстояние между объединяющимися кластерами')
31 - dendrogram(Z) % построение дендрограммы в отдельном окне

```

**Код:****Ввод  
комментариев  
(disp, %)****ИСХОДНЫХ  
данных X****ф-и  
pdist  
linkage,  
dendrogram**

матрица иерархического дерева бинарных кластеров размерности  $(m-1) * 3$

Z =

9.0000	11.0000	0.4526
10.0000	21.0000	1.5280
4.0000	5.0000	1.6277
14.0000	15.0000	1.6619
6.0000	23.0000	1.7092
12.0000	22.0000	1.7518
17.0000	18.0000	1.8033
7.0000	8.0000	1.8409
1.0000	2.0000	1.8497
26.0000	28.0000	1.9082
19.0000	27.0000	1.9488
13.0000	30.0000	1.9938
25.0000	32.0000	2.0097
3.0000	29.0000	2.0159
16.0000	24.0000	2.0473
31.0000	35.0000	2.0555
20.0000	36.0000	2.3129
33.0000	37.0000	2.3153
34.0000	38.0000	2.7714

$m+i$  - индекс кластера, где  $i$  - номер строки  
 первый и второй столбцы - индексы кластеров  
 третий столбец - расстояние между объединяющимися

разделение иерархического дерева на заданное количество кластеров  
 кластеризация по межгрупповым средним

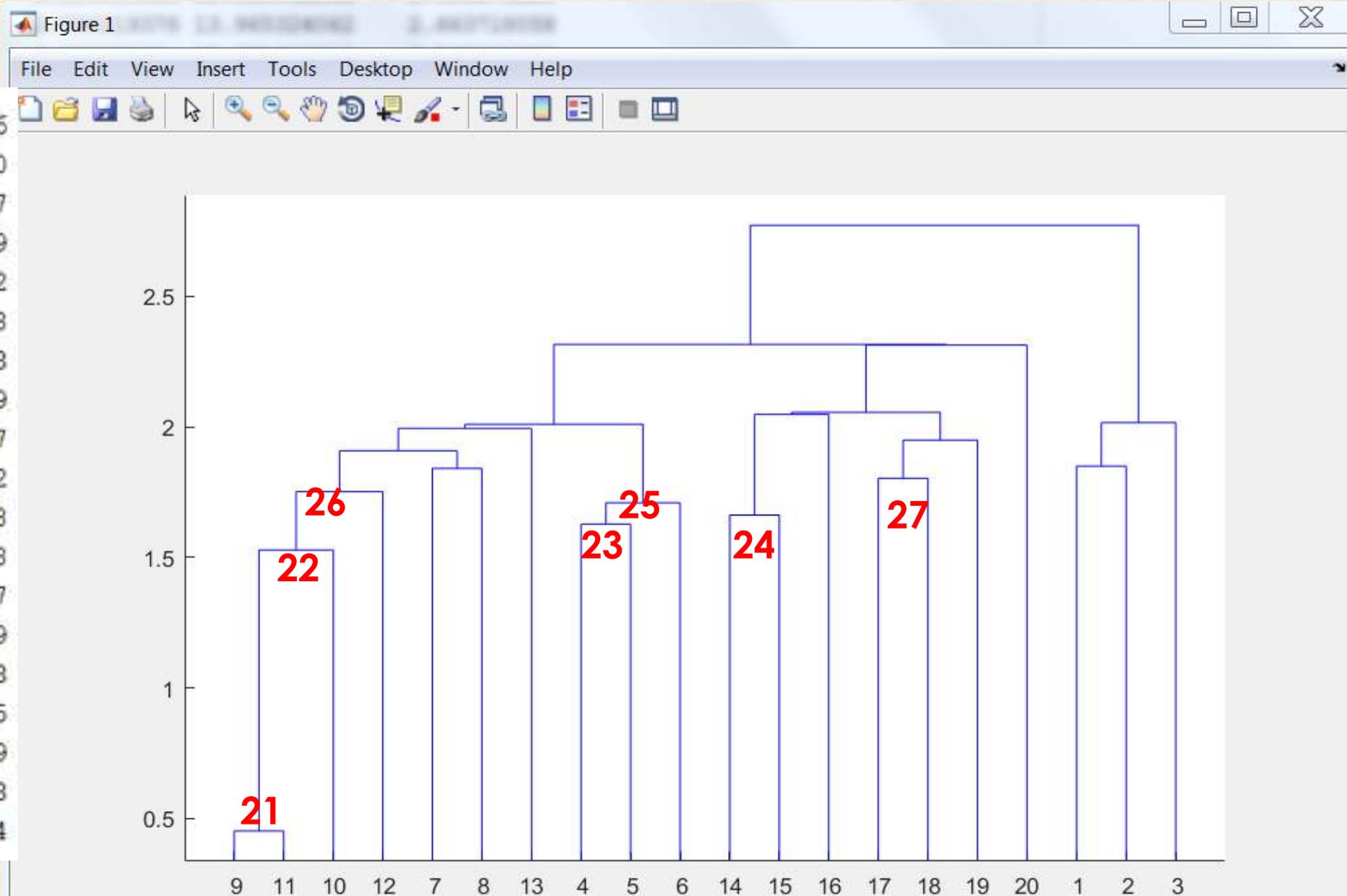
## RUN and ADVANCE

- ▶  $m+i$  - индекс кластера, где  $i$  - номер строки кластера в матрице  $Z$  (разм  $(m-1)*3$  ), где  $m$ -число объектов
- ▶ первый и второй столбцы  $Z$ -индексы кластеров, которые объединяются
- ▶ третий столбец - расстояние между объединяющимися кластерами

# Дендрограмма

47

9.0000	11.0000	0.4526
10.0000	21.0000	1.5280
4.0000	5.0000	1.6277
14.0000	15.0000	1.6619
6.0000	23.0000	1.7092
12.0000	22.0000	1.7518
17.0000	18.0000	1.8033
7.0000	8.0000	1.8409
1.0000	2.0000	1.8497
26.0000	28.0000	1.9082
19.0000	27.0000	1.9488
13.0000	30.0000	1.9938
25.0000	32.0000	2.0097
3.0000	29.0000	2.0159
16.0000	24.0000	2.0473
31.0000	35.0000	2.0555
20.0000	36.0000	2.3129
33.0000	37.0000	2.3153
34.0000	38.0000	2.7714



## Код для вывода матрицы иерархического дерева бинарных кластеров с номерами кластеров

```
31 - NUM=m+1:2*m-1; %задание массива-строки порядковых номеров кластеров
32 - NUMT=NUM.'; % .' - транспонирование, ; означает блокировку вывода в окно
33 - disp('матрица номеров кластеров и иерархического дерева бинарных кластеров')
34 - disp('размерности (m-1) *4')
35 - NUMT_Z=[NUMT Z]
36 -
37 -
38 -
39 -
40 -
41 -
42 -
43 -
44 -
45 -
```

# Матрица иерархического дерева бинарных кластеров с номерами кластеров

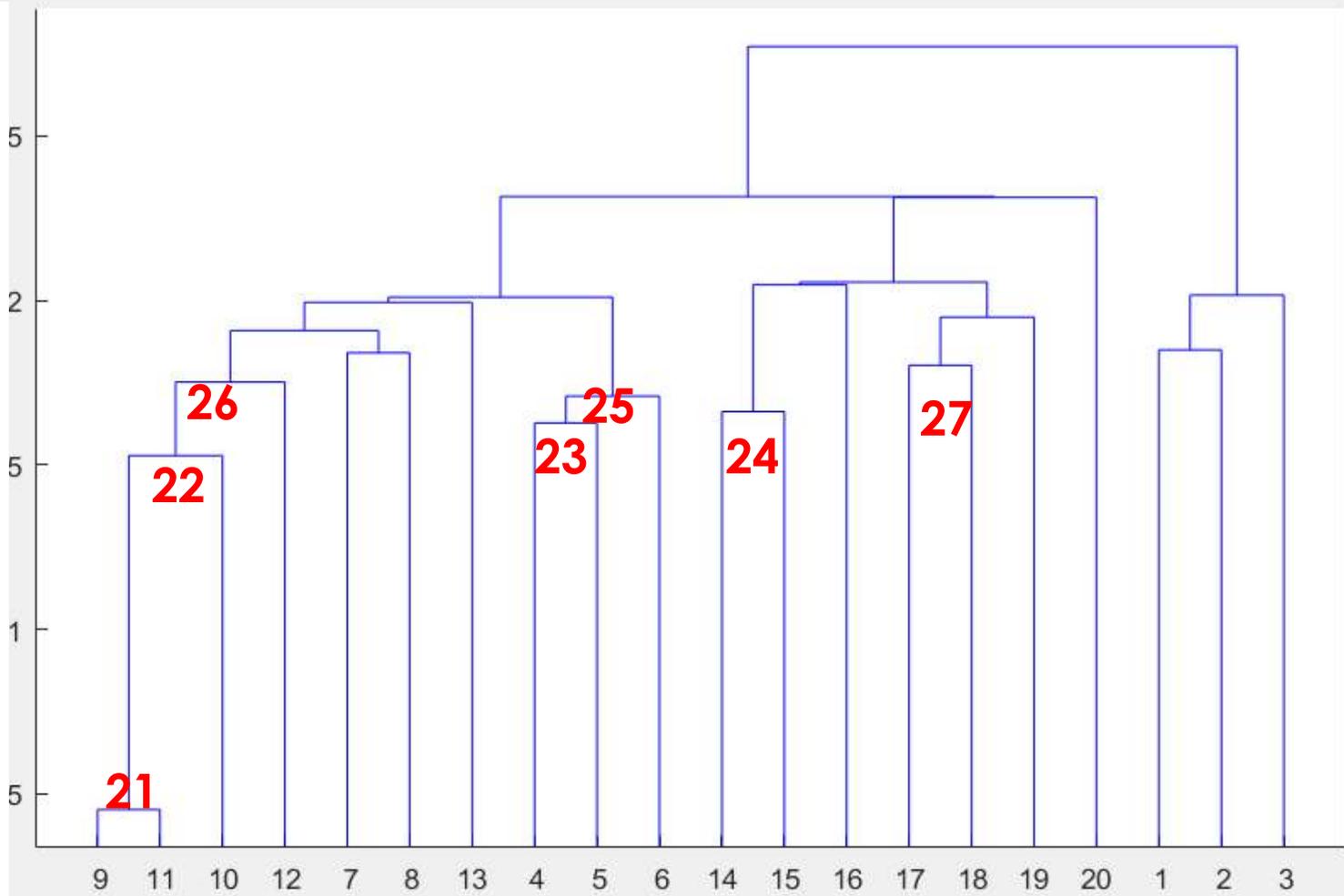
49



Command Window

NUMT\_Z =

21.0000	9.0000	11.0000	0.4526
22.0000	10.0000	21.0000	1.5280
23.0000	4.0000	5.0000	1.6277
24.0000	14.0000	15.0000	1.6619
25.0000	6.0000	23.0000	1.7092
26.0000	12.0000	22.0000	1.7518
27.0000	17.0000	18.0000	1.8033
28.0000	7.0000	8.0000	1.8409
29.0000	1.0000	2.0000	1.8497
30.0000	26.0000	28.0000	1.9082
31.0000	19.0000	27.0000	1.9488
32.0000	13.0000	30.0000	1.9938
33.0000	25.0000	32.0000	2.0097
34.0000	3.0000	29.0000	2.0159
35.0000	16.0000	24.0000	2.0473
36.0000	31.0000	35.0000	2.0555
37.0000	20.0000	36.0000	2.3129
38.0000	33.0000	37.0000	2.3153
39.0000	34.0000	38.0000	2.7714



## Кластеризация по межгрупповым средним (kmeans), Параметры **IDX, C, SUMDIST**

```
31 -  
32 -  
33 -  
34 -  
35 -  
36 -  
37 - NUM=1:m; %задание массива-строки порядковых номеров  
38 - NUMT=NUM.'; % .' - транспонирование, ; означает блокировку вывода в окно  
39 - disp(' ')  
40 - disp('кластеризация по межгрупповым средним, функция kmeans')  
41 - [IDX,C,SUMDIST]=kmeans(X,5);  
42 - clust=[NUMT IDX]  
43 - disp('C - центры кластеров и SUMDIST - суммарные расстояния до центров')  
44 - C  
45 - SUMDIST
```

кластеризация по межгрупповым средним

clust =

```

1 1
2 1
3 1
4 1
5 1
6 1
7 2
8 2
9 2
10 2
11 2
12 2
13 2
14 4
15 4
16 4
17 5
18 5
19 5
20 3

```

C - центры кластеров и SUMDIST - суммарные рас

C =

3.4459	13.2622	2.4355
4.0311	19.8316	4.6789
2.4599	6.4658	7.8922
4.5874	14.4344	5.9763
4.0784	9.6533	6.8590

SUMDIST =

40.2518
23.6099
0
6.1675
6.4543

fx >>

► **CLUST** – массив номеров объектов и номеров кластеров (массив **IDX**), к которым объекты принадлежат

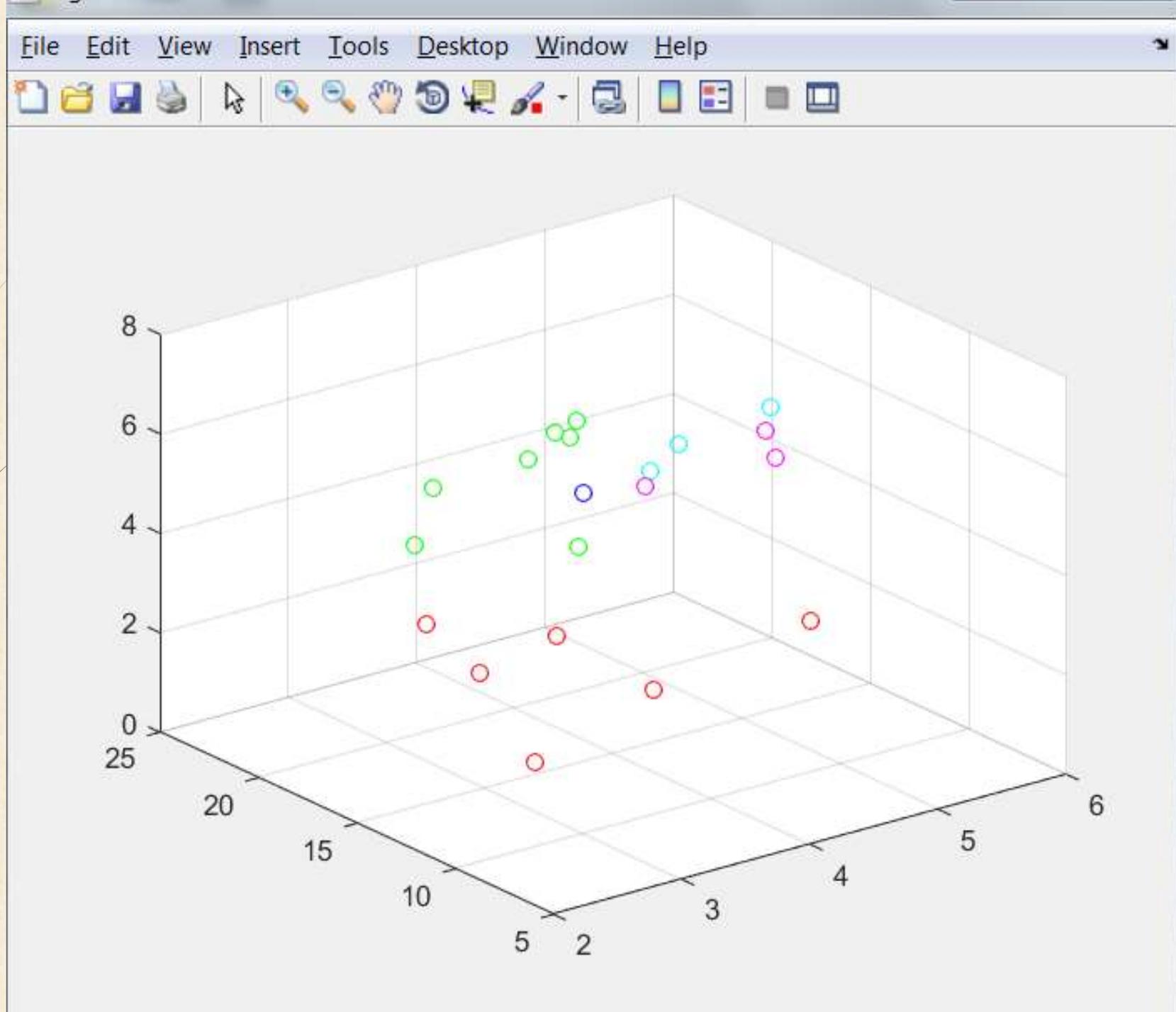
► **C** – координаты центроидов

► **SUMDIST** – суммарные расстояния от центроидов до объектов в кластерах

# Построение множества кластеров

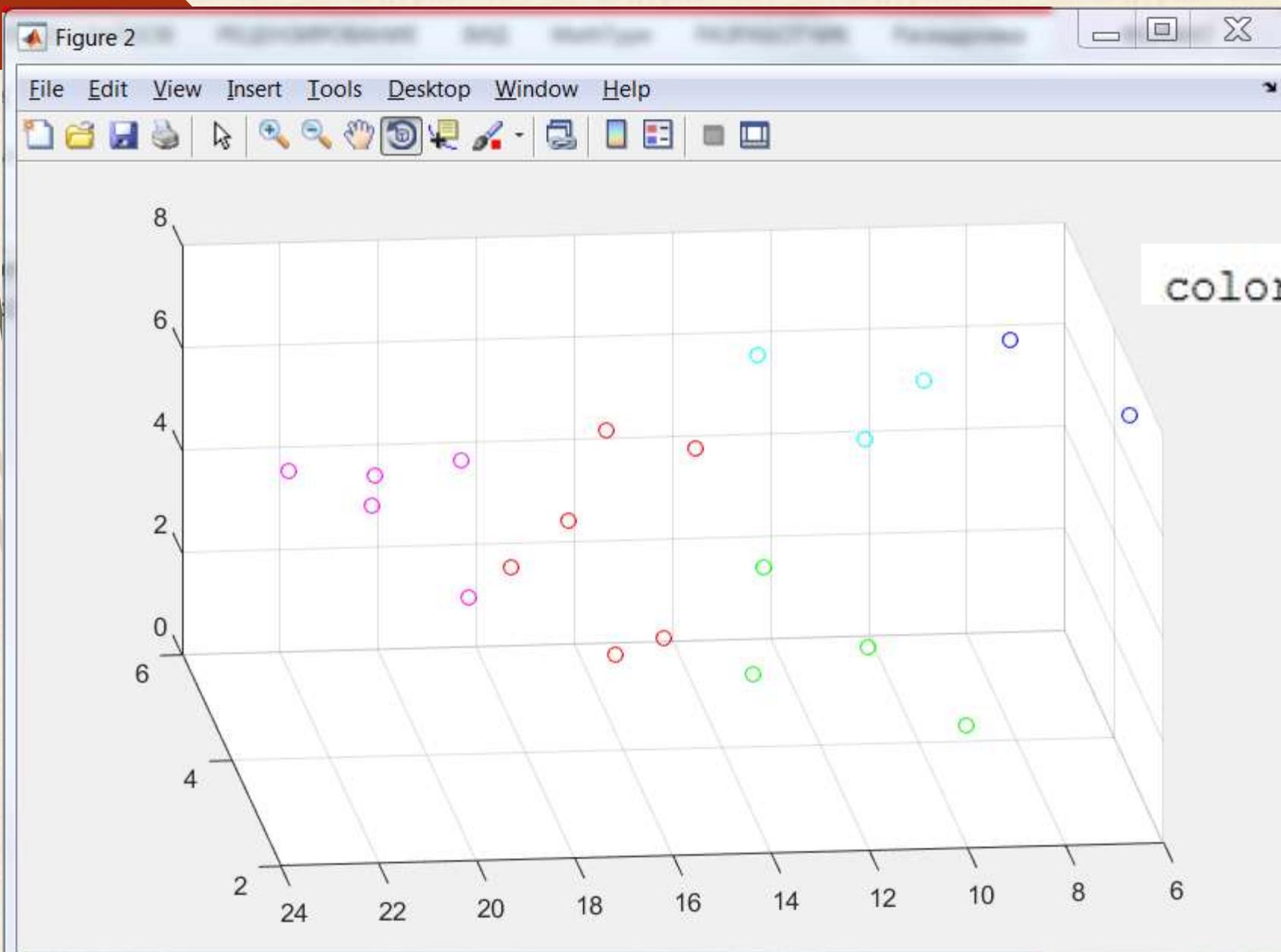
## Функции **figure** , **plot3**, **for**, **switch**, **hold on**, **grid on**

```
46 - X1=X(:,1); X2=X(:,2); X3=X(:,3); % массивы столбцов данных
47 - color=['r' 'g' 'b' 'c' 'm']; % массив разных цветов для разных кластеров
48 - figure % создает новое графическое окно, старое графическое окно остается
49 - % построение трехмерного графика кластеров в отдельном окне
50 - for i=1:1:m % цикл по объектам
51 -     switch IDX(i) % переключатель
52 -         case 1
53 -             plot3(X1(i),X2(i),X3(i), 'Color',color(1), 'Marker', 'o')
54 -         case 2
55 -             plot3(X1(i),X2(i),X3(i), 'Color',color(2), 'Marker', 'o')
56 -         case 3
57 -             plot3(X1(i),X2(i),X3(i), 'Color',color(3), 'Marker', 'o')
58 -         case 4
59 -             plot3(X1(i),X2(i),X3(i), 'Color',color(4), 'Marker', 'o')
60 -         case 5
61 -             plot3(X1(i),X2(i),X3(i), 'Color',color(5), 'Marker', 'o')
62 -     end
63 - hold on % сохранение текущего графика и добавление нового графика в
64 - % текущем окне
65 - end
66 - grid on % построение координатной сетки
```



# Результаты кластеризации

Используется  
вращение  
для увеличения  
наглядности



```
color=['r' 'q' 'b' 'c' 'm'];
```

SUMDIST =

22.8353

14.3963

2.6747

5.9484

9.3370

## Задания 5,6 контрольной работы

55

- ▶ Задание 5 соответствует ЛР-5 (компонентный анализ)
  - ▶ Задание 6 соответствует ЛР-6 (кластерный анализ)
- (исходные данные по вариантам представлены в на листе Excel Варианты ЛР3-6)

## Методические указания к контрольной работе

- ▶ Контрольная работа включает выполнение заданий, сформулированных в ЛР1-6.
- ▶ ЛР5-6 представлены в этой презентации, остальные в презентациях, соответствующих Лек1-2 . Исходные данные выбирать в соответствии со своим вариантом. *Правило выбора номера варианта контрольной работы:* номером варианта служит число  $n$ , равное сумме двух последних цифр зачетной книжки студента, если это не ноль. Если же две последние цифры номера зачетной книжки студента нули, то следует полагать  $n=19$ .