


МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Ярославский государственный университет им. П.Г. Демидова

Кафедра информационных и сетевых технологий

УТВЕРЖДАЮ

Декан факультета ИВТ

 Д.Ю. Чалый
« 31 » августа 2020 г.

Рабочая программа дисциплины
«Статистические методы анализа данных»

Направление подготовки
02.04.02 Фундаментальная информатика и информационные технологии

Направленность (профиль)
«Компьютерные науки»

Форма обучения
очная

Программа рассмотрена
на заседании кафедры
от 31 августа 20 20 г.,
протокол № 10

Зав. кафедрой

 Д.Ю. Чалый

Программа одобрена НМК
факультета ИВТ
от 31 августа 20 20 г.,
протокол № 1

Председатель НМК

 Г.В. Шабаршина

Ярославль
2020

1. Цели освоения дисциплины

Целями дисциплины «Статистические методы анализа данных» являются:

- представление о применимости статистического аппарата для решения разного рода прикладных задач в различных сферах деятельности;
- обучение методам анализа и обработки больших объемов информации;
- выработка навыков расчета статистических показателей на реальных примерах, анализ полученных результатов

2. Место дисциплины в структуре ОП магистратуры

Дисциплина «Статистические методы анализа данных» относится к части, формируемой участниками образовательных отношений, ОП магистратуры и является дисциплиной по выбору.

Для изучения данной дисциплины студентам необходимо знать основы дисциплин математического блока, в частности - «Линейную алгебру», базовые понятия «Теории вероятностей и математической статистики», а также владеть методами общей теории статистики, изучаемых в курсе «Статистика» и/или «Прикладная статистика». Кроме того, для успешного выполнения некоторых индивидуальных заданий желательно (но не обязательно!) знать основы «Эконометрики».

Знания и навыки, полученные при изучении дисциплины, студенты смогут использовать при выполнении расчетов в научных исследованиях и в практической части выпускных квалификационных работ, связанных с моделированием и анализом массовых явлений различной природы.

3. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОП магистратуры

Процесс изучения дисциплины направлен на формирование следующих элементов компетенций в соответствии с ФГОС ВО, ОП ВО и приобретения следующих знаний, умений, навыков и (или) опыта деятельности:

Код компетенции	Формулировка компетенции	Перечень планируемых результатов обучения
Профессиональные компетенции		
ПК-3	Способность понимать и применять в научно-исследовательской и прикладной деятельности современный математический аппарат, основные законы естествознания, современные языки программирования и программное обеспечение; операционные системы и сетевые технологии	<u>Знать</u> основные подходы к анализу и моделированию массовых процессов и явлений.. <u>Уметь</u> анализировать качество моделей, их надежность и предсказательную способность, и выбирать наиболее эффективные из них. <u>Владеть навыками</u> расчета параметров математико-статистических моделей и оценки их качественных характеристик.

4. Объем, форма реализации, структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зач. ед., 108 акад. час.

Дисциплина реализуется с применением дистанционных образовательных технологий (ДОТ), предоставляемых образовательной площадкой MOOK ЯрГУ им. П.Г. Демидова (DemidOnline).

Все основные элементы курса преподаются в дистанционной форме в рамках онлайн курса «Статистические методы анализа данных», размещённого по ссылке:

https://demidonline.uniyar.ac.ru/courses/course-v1:DemidOnline+SM01x+Oct_2020/course/

№ п/п	Темы (разделы) дисциплины, их содержание	Семестр	Виды учебных занятий, включая самостоятельную работу студентов, и их трудоемкость (в академических часах)						Формы текущего контроля успеваемости Форма промежуточной аттестации (по семестрам)
			Контактная работа						
			лекции	практические	лабораторные	консультации	аттестационные испытания	самостоятельная работа	
1.	Статистические показатели, характеризующие совокупность и распределение	1	3		3			4	Тест по теме №1 онлайн-курса
2.	Критерии согласия и проверка статистических гипотез	1	3		3			4	Тест по теме №2 онлайн-курса
3.	Корреляционно-регрессионный анализ: закономерности, тенденции и прогнозы. Проблемы регрессионного анализа	1	4		4			6	Тест по теме №3 онлайн-курса
4.	Многомерные методы: факторный, кластерный и дискриминантный анализ	1	4		4			6	Тест по теме №4 онлайн-курса
5.	Примеры анализа и моделирования реальных социально-экономических процессов	1	3		3	6		14	Индивидуальное расчетное задание
						2	0,5	33,5	Экзамен Итоговый тест онлайн-курса
	Всего		17		17	6	0,5	67.5	

Содержание разделов дисциплины

(с указанием формы проведения):

1. Статистические показатели, характеризующие совокупность и распределение
 - 1.1. Основные понятия и формулы прикладной статистики: показатели центра распределения, вариации и формы распределения
 - 1.2. Задачи выборочного наблюдения, ошибки выборки
 - 1.3. Показатели взаимосвязи и зависимости
2. Критерии согласия и проверка статистических гипотез
 - 2.1. Функция и плотность распределения, мат.ожидание и дисперсия основных видов распределений
 - 2.2. Критерии согласия для проверки различных статистических гипотез
3. Корреляционно-регрессионный анализ: закономерности, тенденции и прогнозы. Проблемы регрессионного анализа
 - 3.1. Корреляционный анализ: частные, парные и множественные коэффициенты корреляции и проверка их значимости
 - 3.2. Регрессионный анализ: парная и множественная регрессия, интерпретация моделей; тенденции и прогнозы. Оценка качества модели, ее надежности и предсказательной способности. Использование фиктивных переменных
 - 3.3. Спецификация модели и проблемы регрессионного анализа: мультиколлинеарность, гетероскедастичность, автокорреляция
4. Многомерные методы: факторный, кластерный и дискриминантный анализ
 - 4.1. Кластерный анализ: сущность, расстояния и методы кластеризации
 - 4.2. Дискриминантный анализ: цели и различные подходы; вычисление дискриминантной функции
 - 4.3. Факторный анализ: задачи и возможности; метод главных компонент
5. Примеры анализа и моделирования реальных социально-экономических процессов:
 - анализ результатов ЕГЭ,
 - моделирование стоимости жилья,
 - анализ факторов развития региона и др.

Все темы можно изучить дистанционно на площадке DemidOnline в рамках курса «Статистические методы анализа данных» или очно с элементами дистанционного обучения в рамках собственной образовательной среды LMS Moodle ЯрГУ.

Тестирование по темам и итоговое тестирование проводится дистанционно на площадке DemidOnline в рамках курса «Статистические методы анализа данных».

5. Образовательные технологии, используемые при осуществлении образовательного процесса по дисциплине

Формы преподавания дисциплины «Статистических методов анализа данных» достаточно традиционны.

Это *лекции* (первая – вводная и «инструктивная», остальные – академические), как наиболее эффективный по времени метод передачи большого объема материала большой группе обучаемых. Как правило, студенты конспектируют излагаемый на доске и/ или с помощью проектора материал. Составление конспекта лекций и дальнейшая работа с ним при подготовке к занятиям выступает как значительная часть процесса обучения. Курс выстроен таким образом, что конспекты охватывают практически весь учебный материал по «Статистическим методам анализа данных» (за исключением тех моментов, где предполагается, что для выполнения выбранного магистрантами индивидуального задания нужно найти какие-то «особые» методы расчета и анализа показателей, и студенты должны сделать это самостоятельно).

Для удобства восприятия и повышения заинтересованности студентов лекционный материал курса «Статистические методы анализа данных» реализован Автором в виде презентаций PowerPoint, однако данный материал студентам заранее не выдается (чтобы

иметь возможность скорректировать презентации с учетом особенностей чтения лекций на данном конкретном потоке, и «из педагогических соображений» ☺); презентации выкладываются (точнее - становятся доступными) в Электронном университете MOODLE ЯрГУ по мере изучения (т.е. после прочтения соответствующей лекции).

Лабораторные занятия с лекциями обычно дополняют друг друга. Проводятся в академических группах под руководством преподавателя. Целями практических занятий являются разъяснение студентам теоретического материала, изложенного на лекции, через решение упражнений и задач, а также получение ими навыков вычислительной работы. Здесь преподавание строится на разумном для каждой темы сочетании коллективной работы группы с самостоятельной индивидуальной работой студентов.

Лабораторные занятия проводятся в компьютерном классе или в аудитории с электронной доской (и/или с компьютером и проектором) с целью показать студентам, как производить соответствующие расчеты средствами MS Excel.

Групповые консультации проводятся перед контрольными мероприятиями (контрольные работы, зачет, экзамен) для большой группы студентов с целью систематизации знаний и устранению имеющихся сложностей с пониманием материала общего характера.

Индивидуальные консультации проводятся регулярно для желающих с целью устранения имеющихся у студентов проблем с материалом частного характера.

Самостоятельная работа студентов реализуется:

1. Непосредственно в процессе аудиторных занятий при выполнении текущих заданий и контрольных работ.
2. В контакте с преподавателем вне рамок расписания – на консультациях по учебным вопросам, при ликвидации задолженностей, при выполнении индивидуальных заданий и т.д.
3. В библиотеке, дома, посредством поиска в сети Интернет и т.д. при выполнении студентом контрольных работ и индивидуальных заданий, выдаваемых на дом.

6. Перечень электронных ресурсов и информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения и информационных справочных систем (при необходимости)

В процессе осуществления образовательного процесса используются:

- для формирования материалов лекционных и практических занятий, заданий для промежуточной и текущей аттестации – программы пакета Microsoft Office (MS Word, MS Excel, MS PowerPoint);
- материалы онлайн-курса «Статистические методы анализа данных», размещённого на образовательной онлайн-площадке ЯрГУ DemidOnline по ссылке: https://demidonline.uniyar.ac.ru/courses/course-v1:DemidOnline+SM01x+Oct_2020/course/
- для размещения материалов лекций, заданий и тестов – «Электронный университет MOODLE ЯрГУ»: <https://moodle.uniyar.ac.ru/>
- для поиска другой учебной литературы – электронная библиотека учебных материалов ЯрГУ: http://www.lib.uniyar.ac.ru/opac/bk_cat_find.php
- для расчетов (в некоторых индивидуальных заданиях, где это требуется) также можно использовать бесплатно распространяемый специальный статистический пакет MicroTSP и/или другие доступные стат.пакеты.

7. Перечень основной и дополнительной учебной литературы, ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

а) **основная:**

1. Васильева, Э. К., Статистика [Электронный ресурс] : учебник / Э. К. Васильева, В. С. Лялин, М., Юнити-Дана, 2015, 399с
2. Аскеров, П. Ф., Общая и прикладная статистика : учебник для вузов / П. Ф. Аскеров, Р. Н. Пахунова, А. В. Пахунов, М., ИНФРА-М, 2014, 271с

б) **дополнительная:**

1. Спиридонова, Е. М., Использование электронных таблиц EXCEL в эконометрических расчетах : метод. указания, Ярославль, ЯрГУ, 2001, 23с
2. Айвазян С. А. Прикладная статистика и основы эконометрики: учебник для вузов. / С. А. Айвазян - М.: ЮНИТИ, 1998. - 1022 с.

в) **ресурсы сети «Интернет»**

1. Электронный учебник по статистике StatSoft (Москва, StatSoft, Inc. (2012)): <http://statsoft.ru/home/textbook/default.htm>
2. База примеров реальных задач: <http://statsoft.ru/solutions/ExamplesBase/tasks/>

8. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Университетская библиотека (филиал, находящийся в учебном корпусе) обеспечивает студентов имеющимися в наличии учебниками и методическими указаниями в соответствии с принятыми нормативами. Кроме того, студенты получают электронный вариант учебных материалов (презентации лекций, пособия и данные для расчетов) непосредственно у преподавателя или скачивают их из «Электронного университета MOODLE ЯрГУ».

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине включает в свой состав специальные помещения:

- учебные аудитории для проведения занятий лекционного типа;
- учебные аудитории для проведения лабораторных работ;
- учебные аудитории для проведения групповых и индивидуальных консультаций,
- учебные аудитории для проведения текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы;
- помещения для хранения и профилактического обслуживания технических средств обучения.

Специальные помещения укомплектованы средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий, хранящиеся на электронных носителях и обеспечивающие тематические иллюстрации, соответствующие рабочим программам дисциплин.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации, а также к материалам онлайн-курсов, размещенных на образовательной онлайн-площадке ЯрГУ DemidOnline.

Число посадочных мест в лекционной аудитории больше либо равно списочному составу потока, а в аудитории для лабораторных занятий – списочному составу группы обучающихся.

Все очные аудиторные занятия по Статистическим методам анализа данных проводятся в ауд. 309 7-го учебного корпуса, в которой установлена интерактивная доска, используемая для демонстрации презентаций лекционного материала, тестов и приемов практической работы по обработке данных (в MS Excel).

Автор(ы) :

доцент кафедры ИСТ, д.э.н., доцент _____



Е.М. Спиридонова

**Приложение №1 к рабочей программе дисциплины
«Статистические методы анализа данных»**

**Фонд оценочных средств
для проведения текущей и промежуточной аттестации студентов по дисциплине**

1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Обязательными формами текущего контроля знаний студентов являются «индивидуальное расчетное задание» и тестовые работы, которые выполняются с применением электронного обучения и дистанционных образовательных технологий, по мере изучения соответствующих тем.

**1.1. Контрольные задания и иные материалы,
используемые в процессе текущего контроля успеваемости**

ТЕМЫ индивидуальных заданий (с примерами)

Надо выбрать одну тему, проанализировать данные и ответить на поставленные вопросы. К темам 1, 2, 3, 4 и 7 прилагаются файлы в формате .xls с исходными данными. Кроме того, в темах 4 и 7 можно использовать более актуальные данные, собранные самостоятельно (из открытых источников информации), что поощряется более высокими баллами. Задания №5 и №6 предполагают исключительно «собственные» данные.

Отчет необходимо оформить в виде «статьи» с результатами анализа (5-7 стр.)



✦ 1. Анализ абитуриентов

А. Рассчитать основные характеристики распределения баллов ЕГЭ абитуриентов ф–та ИВТ на момент окончания подачи документов (2017-2019 гг. – 26 июля):

- по сумме баллов и отдельно по предметам;
- отдельно по направлениям (по приоритету).
- ✓ Есть ли различия в средних по направлениям (по сумме и отдельным предметам)?
- ✓ Есть ли различия в распределениях баллов (по сравнению с «общим»)?

✦ 1. Анализ абитуриентов

Б. Произвести те же расчеты, что и в п.А, для зачисленных на бюджетные места.

- ✓ Есть ли различия в средних по направлениям (по сумме и отдельным предметам)?
- ✓ Есть ли различия в распределениях баллов (по сравнению с «общим»)?
- ✓ Есть ли отличия в средних и распределениях «поступивших» и «подавших документы»?
- ✓ Продемонстрировать отличия на графиках.

1. Анализ абитуриентов

- ✦ для ВКР: Написать программный модуль (макрос *MS Excel*, *Python* или др.), позволяющий ставить соотв. «статусы» абитуриентам в рейтинговом списке (*.xls).

Всего бюджетных мест: 20 Проходной балл на текущий момент: 206

№	id	ФИО	Оригинал	Согласие	Рейтинг	Статус абитуриента
1	43277	Суздалева Анна Сергеевна	Да	Да	213	Проходит в рамках особой квоты
2	42583	Данилов Павел Сергеевич	Да	Да	169	Проходит в рамках целевой квоты
3	47087	Смирнова Анастасия Александровна	Нет	Нет	285	Приоритет на другом направлении, нет оригиналов документов
4	46873	Гасанов Ислам Магеррамович	Нет	Нет	275	Приоритет на другом направлении, нет оригиналов документов
5	43542	Мельникова Таисия Михайловна	Нет	Нет	270	Приоритет на другом направлении, нет оригиналов документов
6	42487	Кузьмина Анастасия Николаевна	Да	Нет	268	Проходит по конкурсу в первую волну на другое направление
7	44876	Аробелидзе Жужуна Тамазовна	Да	Нет	266	Проходит по конкурсу в первую волну на другое направление
8	43376	Юров Сергей Валерьевич	Нет	Нет	264	Приоритет на другом направлении, нет оригиналов документов
9	44661	Мельник Альбина Александровна	Да	Да	261	Проходит по конкурсу в первую волну
10	42307	Приходько Елизавета Анатольевна	Да	Да	261	Проходит по конкурсу в первую волну
11	45257	Ромашов Владислав Игоревич	Нет	Нет	260	Приоритет на другом направлении, нет оригиналов документов
12	43423	Мулик Андрей Алексеевич	Да	Нет	257	Проходит по конкурсу в первую волну на другое направление
13	45587	Тарелкина Наталья Андреевна	Да	Нет	257	Проходит по конкурсу в первую волну на другое направление
14	43020	Задоров Кирилл Юрьевич	Да	Нет	256	Проходит по конкурсу в первую волну на другое направление
15	46594	Барышникова Анна Андреевна	Нет	Нет	256	Приоритет на другом направлении, нет оригиналов документов
16	44594	Кузьмина Анастасия Александровна	Да	Да	254	Проходит по конкурсу в первую волну
17	43411	Титов Федор Михайлович	Нет	Нет	254	Приоритет на другом направлении, нет оригиналов документов
18	43473	Борисова Анастасия Дмитриевна	Да	Нет	253	Проходит по конкурсу в первую волну на другое направление
19	44722	Широкова Анна Владимировна	Да	Да	251	Проходит по конкурсу в первую волну
20	44694	Гулев Александр Германович	Нет	Да	247	Нет оригиналов документов

2. Анализ студентов

👉 Данные у Д.Ю.!

Есть данные о результатах ЕГЭ и успеваемости студентов 1–го, 2–го (и 3–го?) курсов ф–та ИВТ.

- ✓ Есть ли зависимость между средней успеваемостью и суммой баллов ЕГЭ?
- ✓ Есть ли зависимость между результатами ЕГЭ по математике, информатике (и русскому языку) и успеваемостью по соотв. блокам дисциплин?
- ✓ Меняется ли эта зависимость со временем?
- ✓ * Вычислить вероятность отчисления студентов, поступивших с относительно низкими баллами ЕГЭ.

3. Анализ клиентов банка

Есть данные об «активности» клиентов одного из отделений банка. Некоторые из клиентов «засыпают».

➤ Необходимо выяснить, от каких факторов (размер остатков на счете, их динамика, жалобы и др.) это зависит.

С целью повышения качества работы был реализован Проект «Мобильный консультант».

- ✓ Изменилась ли выявленная ранее зависимость и в какую сторону?
- ✓ Подтвердить позитивные изменения (если они есть) с помощью соотв. стат.тестов.

4. Анализ недвижимости

Есть данные о ценах на недвижимость:

- А). *очень старые!* (1998 г.) по квартирам Москвы, \$
- Б). по нежилым помещениям Челябинска (2015 г.), Р
- В). по квартирам Ярославля (Авито, 2017 г.), Р

➤ Построить регрессионные модели, взяв в качестве зависимой переменной (Y):

- стоимость квартиры/объекта;
- стоимость квадратного метра.

✓ Изменился ли набор факторных переменных, их значимость, направление/форма зависимости?

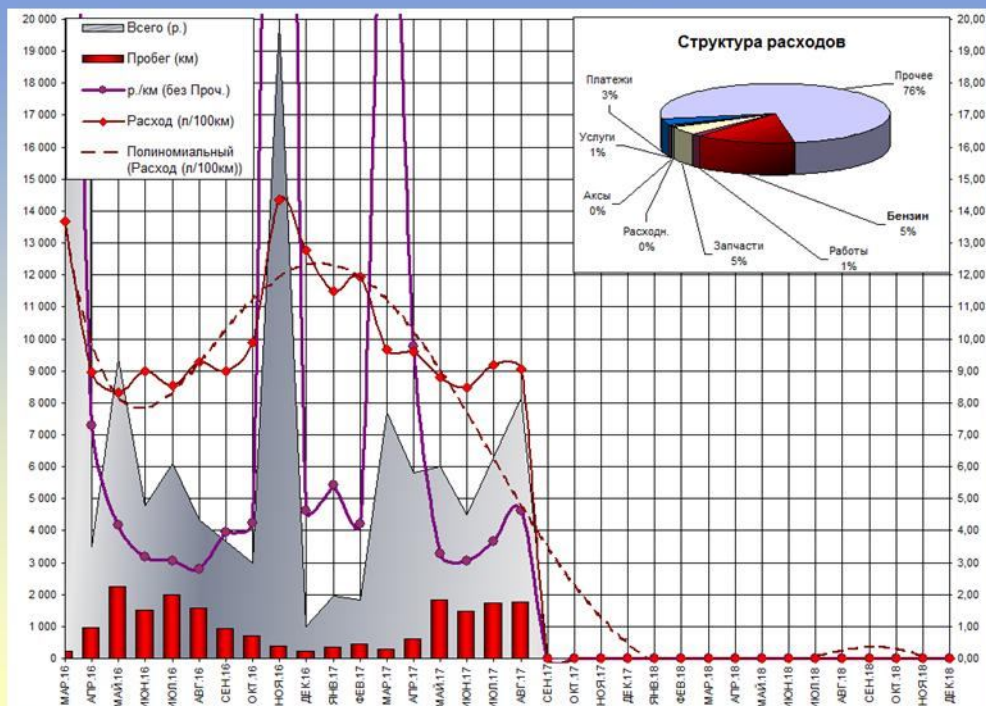
+ для ВКР (и статьи): Собрать «свои» данные!

5. Автокалькулятор

Задача: вести «учет» расходов на автомобиль (бензин, запчасти, работы, штрафы, платежи и др.)

- Рассчитывать текущие и «общие» показатели:
 - расход бензина на 100 км,
 - средний пробег в день/месяц/между заправками,
 - стоимость 1 км, и т.д. и т.п. 😊
- Рассчитывать показатели по месяцам (и годам), строить соответствующие графики.
- Построить уравнения динамики для расхода бензина на 100 км (линейное, с учетом сезонных колебаний, др.?), и выбрать «лучшее» из них. От каких факторов это зависит?

5. Автокалькулятор



5. Автокалькулятор

Дата	Сеть АЗС, марка бензина	Цена 1 л, руб	Сумма, руб	Объем заправки, л	Показания одометра, км	Пробег между заправками	Средний расход на 100 км	"Цена" 1 км, руб	Период между заправ., дн.	Средний пробег, км/день
2016			33 895	989	10 516	10 718	9,2	3,2	286	37
2017			29 051	784	19 220	8 487	9,2	3,4	246	35
2018										
ИТОГО:			62 946	1 772	19 220	19 205	9,2	3,3	532	36
06.01.2017	ГПН, 95	36,10	500	13,85	10 733	122	11,4	4,10	6	20
12.01.2017	ГПН, 95	36,10	1 000	27,70	10 855	240	11,5	4,17	32	8
13.02.2017	ГПН, 95	33,50	838	25,00	11 095	212	11,8	3,95	15	14
28.02.2017	ГПН, 95	36,65	1 000	27,28	11 307	226	12,1	4,42	19	12

ПРОБЕГ	СР.РАСХОД
МАР 16	210
АПР 16	962
МАЙ 16	2 247
ИЮН 16	1 516
ИЮЛ 16	1 995
АВГ 16	1 562
СЕН 16	923
ОКТ 16	700
НОЯ 16	386
ДЕК 16	217
ЯНВ 17	362
ФЕВ 17	438
МАР 17	282
АПР 17	595
МАЙ 17	1 840
ИЮН 17	1 480
ИЮЛ 17	1 725
АВГ 17	1 765
СЕН 17	
ОКТ 17	
НОЯ 17	
ДЕК 17	
ЯНВ 18	
ФЕВ 18	
МАР 18	
АПР 18	
МАЙ 18	
ИЮН 18	
ИЮЛ 18	



5. Автокалькулятор

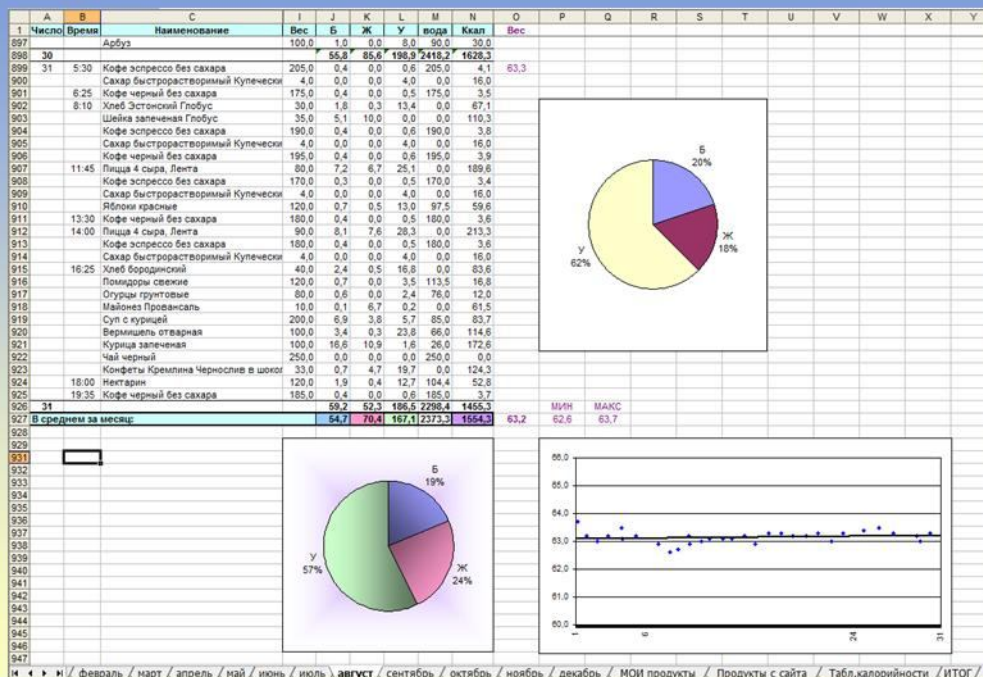
+ для ВКР: Сделать онлайн-калькулятор и разместить его на соотв.ресурсе, чтобы им могли пользоваться и другие автолюбители.

6. Калькулятор калорий

Задача: вести «учет» потребленных продуктов питания, в разрезе килокалорий, белков–жиров–углеводов, воды (а при необходимости – и разл.микроэлементов) с целью контроля веса (для спортсменов и/или последователей ЗОЖ). 😊

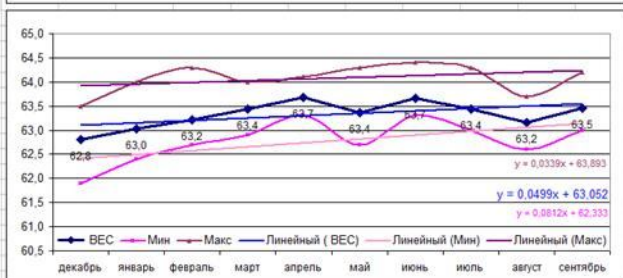
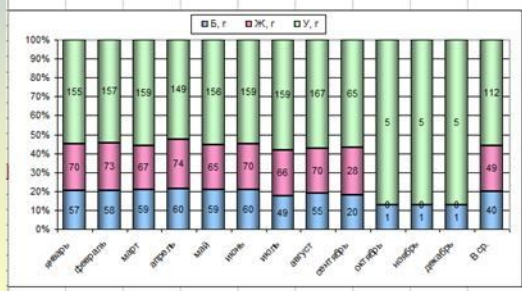
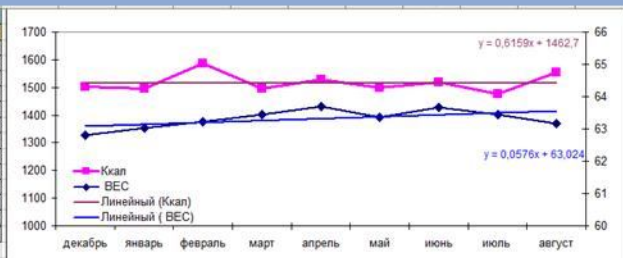
- Рассчитывать показатели ежедневно и по месяцам, строить соответствующие графики (в динамике).
- Построить уравнения зависимости (за достаточно длит.период) веса от Ккал, БЖУ, в т.ч. – с лагами.
- ✓ Оценить качество уравнений, значимость факторов. Что больше влияет на вес? (у всех – м.б. по-разному!)
- ✓ * На основании регрессионного анализа дать рекомендации по изменению рациона питания (для достижения какой-то цели).

6. Калькулятор калорий



6. Калькулятор калорий

Месяц	Б, г	Ж, г	У, г	вода	Ккал	ВЕС	Мин	Макс
декабрь	52,0	70,1	147,9	2173,1	1503,5	62,8	61,9	63,5
январь	57,3	69,6	154,8	2240,9	1496,7	63,0	62,4	64,0
февраль	58,0	72,6	156,9	2187,7	1585,0	63,2	62,7	64,3
март	58,8	67,0	158,8	2450,7	1495,3	63,4	62,9	64,0
апрель	60,1	73,6	148,8	2349,2	1529,3	63,7	63,3	64,1
май	58,8	65,4	155,8	2367,0	1500,7	63,4	62,7	64,3
июнь	59,7	70,5	159,0	2290,5	1516,9	63,7	63,3	64,4
июль	48,6	66,2	158,9	2346,0	1476,9	63,4	63,0	64,3
август	54,7	70,4	167,1	2373,3	1554,3	63,2	62,6	63,7
сентябрь	20,4	28,3	64,8	1128,3	638,2	63,5	63,0	64,2
октябрь	0,7	0,0	4,9	359,6	22,6	64,3	64,3	64,3
ноябрь	0,7	0,0	5,1	368,2	23,4	62,9	62,9	62,9
декабрь	0,7	0,0	5,0	361,5	22,7	62,9	62,9	62,9
В ср	39,9	48,6	111,7	1735,2	1071,8	63,4	63,0	64,0



6. Калькулятор калорий

+ для ВКР: Сделать онлайн-калькулятор и разместить его на соотв.ресурсе, чтобы им могли пользоваться и другие заинтересованные лица. 😊

Например:

<http://dietadiary.com/calories-calculator/online>

- * Разработать программный модуль расчета калорийности (и БЖУ) готовых блюд в зависимости от способа приготовления (?)
- * Добавить аналог «экспертной системы» – предлагать меню на неделю / на каждый день в соответствии с целями (снижение веса/ набор мышечной массы) и с учетом индивидуальных предпочтений (по блюдам) и особенностей (на основе анализа зависимостей).

7. Анализ школ по ЕГЭ

Цель: «классифицировать» (и/или упорядочить) школы (г. Ярославля) по результатам ЕГЭ. 😊

Школы г. Ярославля	Кластер (2012 г.)	Математика												Русский язык											
		2008	Ранг-08	2009	Ранг-09	2010	Ранг-10	2011	Ранг-11	2012	Ранг-12	2013	Ранг-13	2008	Ранг-08	2009	Ранг-09	2010	Ранг-10	2011	Ранг-11	2012	Ранг-12	2013	Ранг-13
МОУ СОШ № 6	5	49,38	17	35,95	59	43,95	39	41,95	31	50,67	19	49,75	30	54,53	18	52,31	78	57,73	49	63,95	23	70,25	15	67,75	63
МОУ СОШ № 10	4	37,43	55	37,76	65	37,61	66	40,44	52	41,02	51	45,42	49	60,19	41	56,67	57	55,90	63	56,61	52	59,25	58	63,35	43
МОУ СОШ № 11	4	40,51	49	38,73	63	41,30	63	50,76	16	43,30	39	47,38	45	62,17	31	57,73	61	59,25	41	72,13	5	64,28	33	67,19	30
МОУ СОШ № 17	5	30,88	79	35,82	70	37,62	64	41,27	48	47,59	28	44,00	55	52,38	70	57,11	55	54,14	66	49,09	64	51,45	46	57,45	64
МОУ СОШ № 20	4	43,84	37	44,21	42	45,08	31	48,71	20	42,38	43	51,25	29	64,71	17	52,46	20	55,00	18	64,84	19	63,24	41	67,00	32
МОУ СОШ № 27	4	43,84	36	48,82	25	42,45	48	48,95	24	43,82	37	49,71	37	58,38	53	60,08	36	62,55	20	61,67	31	60,45	49	63,48	42
МОУ СОШ № 29	4	40,42	51	45,20	39	42,05	47	48,05	29	39,41	61	53,22	22	56,77	64	57,20	59	56,42	62	60,61	38	58,65	60	63,96	38
МОУ СОШ № 34	5	32,58	70	43,58	49			35,85	60	30,14	67			61,11	78	53,04	68			53,43	67	57,69	63		
МОУ СОШ № 39	4	30,25	82	43,93	45	39,50	66			40,85	54	42,44	58	58,83	50	60,30	43	56,34	61			59,91	52	60,97	53
МОУ СОШ № 55	4	45,43	29	43,14	61	42,85	44	48,09	27	43,02	41	48,49	40	63,89	21	60,20	34	62,27	29	63,32	28	63,92	36	67,20	29
МОУ СОШ № 66	4	50,44	15	53,13	14	48,06	22	50,35	19	48,94	22	64,46	2	60,41	40	62,63	23	59,64	40	60,49	40	63,64	39	69,96	17
МОУ СОШ № 68	1	52,34	8	55,42	7	54,31	7	57,28	6	57,17	6	57,71	12	63,56	22	64,93	14	67,71	8	65,98	16	70,93	14	71,00	18
МОУ СОШ № 69	5	34,64	65	33,77	72	37,73	63	38,06	63	39,50	63	39,71	62	56,78	63	49,78	74	53,69	87	50,67	80	57,05	64	63,85	60
МОУ СОШ № 82	4	37,15	57	39,17	61	43,10	42	40,20	53	41,61	48	51,39	28	49,21	52	58,03	61	61,64	31	61,21	36	60,00	51	62,11	52
МОУ СОШ № 72	4	32,75	69	37,64	68	34,25	69			44,71	35			54,25	73	60,24	35	57,81	47			59,69	54		
МОУ СОШ № 80	1	49,34	19	50,20	21	48,98	25	50,29	20	51,82	14	67,90	11	67,79	8	69,40	6	66,12	14	67,13	13	74,43	6	71,20	14
МОУ СОШ № 81	4	44,82	32	41,30	68	37,93	62	49,13	23	38,00	63	47,72	44	62,20	30	60,15	27	59,19	45	58,81	48	59,54	53	68,84	21
МОУ СОШ № 87	4	44,21	35	47,81	29	50,19	20	42,80	46	51,43	15	60,89	7	69,52	47	69,88	37	67,04	11	66,81	41	66,78	26	67,79	26
МОУ СОШ № 90	4	43,09	38	54,39	8	56,38		53,72	13	53,39	12	56,22	16	62,04	33	65,00	19	66,30	13	64,29	21	65,11	31	69,24	23
МОУ СОШ № 95	2	19,10	82	27,74	77	35,77	67	33,85	61	32,40	74	39,79	61	68,00	5	48,50	77	49,64	70	49,41	81	50,25	74	53,89	68
МОУ гимназия № 2	1	60,21	2	59,82	2	59,11	2	59,67	3	62,48	2	62,31	5	68,81	3	71,08	3	70,74	3	73,45	4	78,65	2	76,84	5
МОУ СОШ № 2	4	47,11	24	48,24	27	44,40	38	61,95	16	50,93	17	52,29	24	59,87	42	62,50	25	60,43	33	63,03	29	68,67	17	64,71	38
МОУ СОШ № 19		40,78	47											57,13	60										
МОУ СОШ № 41		29,00	79											62,67	28										
МОУ СОШ № 46		25,00	88											48,11	83										
МОУ СОШ № 47	5	32,13	72	39,21	60	42,88	43	31,95	65				44,64	61	56,00	68	54,64	65	67,68	50	48,91	83		58,00	62
МОУ СОШ № 48	4	36,24	63	37,03	67	40,25	55	33,58	60	39,48	60	45,12	60	58,93	49	56,08	64	60,10	30	53,79	66	58,13	62	62,92	45
МОУ СОШ № 50		32,66	71											49,22	81										
МОУ СОШ № 51	5	34,10	67	41,80	55	36,10	66			40,89	52	42,84	67	55,19	51	59,45	42	66,93	50			59,60	56	62,71	47
МОУ СОШ № 62	4	54,10	6	57,61	3	53,83	9	55,29	7	59,18	3	63,02	4	63,41	23	64,39	16	65,83	12	71,34	13	67,65	27		
МОУ СОШ № 69	4	44,25	34	47,73	30	51,31	15	40,90	60	53,95	8	53,46	21	68,47	52	59,82	39	66,93	12	58,92	47	65,56	29	67,92	25
МОУ СОШ № 67	4	46,21	28	43,90	46	41,74	51	38,93	68	47,70	26	39,64	68	61,69	37	61,76	28	58,58	44	61,71	69	66,43	22	68,45	61
МОУ СОШ № 69	4	41,43	43	45,83	35			44,40	43	44,80	34	38,79	64	69,66	40	55,30	62			61,92	30	64,00	35	62,21	50
МОУ СОШ № 77	5	38,42	64	36,85	68					41,33	45	38,83	66	60,35	70	53,36	67					59,83	54	55,67	60
МОУ СОШ № 79	5	43,77	38	46,73	28	41,07	54			39,33	73			61,60	36	57,90	53	63,07	22			65,20	30		
МОУ СОШ № 83	4	52,95	7	44,72	40	50,97	16	47,96	30	39,65	57	47,90	42	66,56	12	66,17	13	65,73	10	69,74	10	65,61	27	68,64	19
МОУ СОШ № 84	1	40,75	47	30,19	71	50,44	19	38,38	67	37,49	65	55,90	16	63,95	20	58,81	46	64,44	21	66,83	14	66,38	29	74,21	8
МОУ СОШ № 84	2	19,85	86	22,05	78	27,00	71	31,17	64	31,22	76	31,65	71	67,33	57	43,78	76	45,53	78	49,88	65	44,82	77	45,12	23
МОУ СОШ № 97	1	57,24	4	56,97	4	58,02	4	63,04	2	63,72	9	65,90	18	69,14	4	67,83	8	74,84	1	76,73	2	73,28	7	74,52	6
МОУ СОШ № 1	4	43,78	39	52,05	15	53,98	8	46,55	35	54,52	7	47,82	43	65,35	69	61,02	31	63,39	23	61,45	33	64,96	32	66,11	35
МОУ СОШ № 4	1	44,89	31	49,62	23			54,61	6	53,71	10	56,55	19	67,83	27	71,57	2			71,67	3	78,60	1	73,30	2
МОУ СОШ № 25	4	44,37	33	48,79	32	44,47	35	43,43	46	37,48	60	48,92	48	68,28	54	66,09	60	60,40	34	64,24	54	69,20	59	63,08	44
МОУ СОШ № 33	1	70,41	1	70,42	1	70,39	1	71,44	1	71,25	1	71,80	1	69,05	2	71,99	1	72,03	1	81,36	1	74,69	1	81,33	1
МОУ СОШ № 42	1	51,50	13	51,66	19	51,38	14			53,49	11	51,97	26	68,00	5	69,13	3	69,81	6			76,05	4	76,68	3
МОУ СОШ № 43	1	43,66	40	48,89	24	49,69	21	61,57	14	50,40	21	64,89	20	66,32	14	62,93	22	67,35	10	66,79	18	71,83	9	73,96	7
МОУ СОШ № 49	4	43,16	45	55,39	8	50,30	19	63,87	11	51,29	16	62,11	8	65,11	16	65,85	12	67,70	9	63,91	24	73,18	5	73,84	8
МОУ СОШ № 70	4	38,30	61	44,46	41	45,05	32	40,78	61	59,74	58	62,59	23	67,70	61	58,21	69	60,41	60	61,33	34	68,83	60	68,89	59
МОУ СОШ № 83	2	18,91	83	31,72	74	34,77	68	31,44	66					48,25	84	45,81	76	49,87	69	45,74	66				
Прованский	1	61,65	12	64,34	10	69,13	3	68,00	4	60,83	20	68,44	9	66,92	11	67,95	7	70,01	5	70,90	7	71,39	12	72,26	12
МОУ СОШ № 8	4	37,50	73	32,27	73	38,90	59	41,52	47	35,05	69	28,85	70	62,27	77	60,73	72	50,62	86	57,82	90	63,79	70	49,85	70
СШ № 10	5	19,90	88	27,54	74	32,04	71	22,58	75	27,55	75	27,55	75	62,68	61							55,21	21	55,21	21

7. Анализ школ по ЕГЭ

Задача: провести кластерный анализ по результатам ЕГЭ по математике и русскому (рекомендуется использовать простое Евклидово расстояние) за каждый год с разными мерами близости:

- по принципам ближнего и дальнего соседа,
- по принципу средней связи,
- по центрам тяжести

? Выделились ли «устойчивые» кластеры?
(предположительно: «спецшколы» с высокими результатами, «вечерние» – с низкими...)



7. Анализ школ по ЕГЭ

? Совпадают ли выделенные кластеры с «кластерами» Департамента образования? (предположительно может быть хорошее совпадение по вечерним и спецшколам...) 😊

☑ Добавьте данные:

- ✓ за последние годы (с 2014 до наст. времени)
- ✓ и/или по другим предметам (*).

Вероятно, следует использовать взвешенное Евклидово расстояние (для придания разной значимости тем или иным предметам)?

Обосновать!

7. Анализ школ по ЕГЭ

☑ Проранжируйте школы по:

- ✓ Математике & Русскому
- ✓ Другим предметам (если есть данные!)

Выделите сколько-то (по 10) «сильнейших» и «слабейших» школ. Эти группы должны «пересчитываться» автоматически!

+ для ВКР: Предложите методику расчета «интегральной оценки» по двум/нескольким/всем предметам (возможно, и с учетом других важных «достижений»?) для рейтингования школ.

См.: Спиридонова Е.М. Интегральная оценка образовательных достижений и рейтингование ОУ: статистический подход. Тезисы. Евразийский образовательный диалог: материалы международного форума. Международный форум: 17-19 апреля 2013 г. – Часть 3, с. 126-128.

Критерии оценивания Индивидуального расчетного задания:

Если задание выполнено полностью, обоснованно получены ответы на все вопросы, при этом использованы нужные методики и обобщающие показатели, все расчеты выполнены верно, таблицы и диаграммы с итоговыми показателями построены без ошибок, наглядны и «читабельны», отчет логичен и аккуратно оформлен, то:

- за «стандартную» задачу (например, №1, №3 или №4, где используются предоставленные преподавателям базы данных) ставится **10** баллов;
- за задачу, где предполагается самостоятельный сбор информации (например, в задаче №4 собрана «свежая» и достаточная по объему база по продаваемым квартирам на Авито или в других открытых источниках информации) - **15** баллов;
- за задачу с программной реализацией (например, если в № 1 создан модуль для формирования рейтинговых списков со статусами абитуриентов, или в №№ 5, 6 - соответствующие, реально работающие онлайн-калькуляторы) - **20** баллов.

В общем случае, при выборе «своей» темы для исследования, задание в целом – со «статистической точки зрения» – оценивается исходя из максимума в 10 баллов.

При этом баллы добавляются (но не более 10 баллов в сумме):

- за «свою» интересную и/или имеющую практический смысл задачу - 1-3 балла;
- за «свои» данные (в зависимости от их объема и полноты) - 3-5 баллов;
- за «свои» инструменты расчетов и построения «оригинальных» диаграмм - 2-4 балла;
- за хорошо оформленный отчет (логично выстроенный, с «красивой» графикой и обоснованными выводами) - 1-2 балла;
- за выложенные в «общий доступ» данные и результаты анализа - 1 балл.

Балл снижается:

- за вопросы, оставшиеся без обоснованного ответа, - «-2» - «-3» балла за каждый;
- за неверно выбранную методику обоснования и/или расчета (например, вместо регрессионного анализа используется анализ средних) - «-2» балла (в каждом случае);
- за грубые ошибки в расчетах, в результате которых получены неверные выводы, - «-1» балл за каждую;
- за несущественные ошибки, не повлиявшие на выводы, - «-1» - «-2» балла за все;
- за отсутствие таблиц и/или диаграмм в отчете (нет наглядности!) - «-1» - «-2» балла;
- отчет оформлен небрежно, диаграммы и таблицы не «читабельны» - «-1» балл;
- автор допускает стилистические и орфографические ошибки - «-0,5» балла.

1.2. Контрольные задания и иные материалы, используемые для промежуточной аттестации

Вопросы к Экзамену:

1. Статистические показатели центра распределения. Выбор формулы для расчета средней; средние степенные. Средние порядковые.
2. Показатели вариации. Критерии однородности совокупности.
3. Показатели формы распределения. Метод моментов.
4. Задачи выборочного наблюдения. Ошибки выборки. Расчет численности выборки.
5. Показатели взаимосвязи: ковариация и корреляция; их свойства и формулы связи между ними.
6. Понятие функции и плотности распределения. Мат.ожидание и дисперсия основных видов распределений.
7. Критерии согласия для проверки гипотез: назначение и примеры использования.
8. Критерий «Хи-квадрат» для проверки гипотезы о соответствии эмпирического распределения определенному виду.
9. Критерий Уилкоксона (Манна-Уитни) для сравнения двух независимых выборок
10. Корреляционный анализ: частные, парные и множественные коэффициенты корреляции; проверка их значимости
11. Регрессионный анализ: парная и множественная регрессия, их оценка по МНК. Содержательная интерпретация моделей.
12. Оценка качества модели, ее надежности и предсказательной способности. Критерии выбора модели.
13. Использование фиктивных переменных в регрессионных моделях.
14. Проблемы регрессионного анализа: спецификация модели. Последствия невключения важных переменных. Замещающие переменные.
15. Проблема мультиколлинеарности. Причины, последствия и методы смягчения.
16. Проблема гетероскедастичности: причины и последствия; методы обнаружения и устранения.
17. Проблема автокорреляции 1-го порядка: причины и последствия, методы обнаружения и устранения. Автокорреляция более высоких порядков.
18. Кластерный анализ: сущность, расстояния и методы кластеризации
19. Дискриминантный анализ: цели и различные подходы; вычисление дискриминантной функции.
20. Факторный анализ: задачи и возможности; метод главных компонент .

ПРИМЕРЫ задач к Экзамену

Для решения этих задач (в отличие от тех, что решались в течение семестра дома) специальных технических и/или программных средств расчета не требуется; достаточно обычного калькулятора.

Задача.

1. Известно, что сумма кредитов, выдаваемых одним менеджером ипотечного кредитования за месяц распределена нормально со средним 14 млн.руб. и стандартным отклонением 3,5 млн.руб.

Какова вероятность того, что общая сумма кредитов, выданных одним менеджером за год, окажется в интервале от 150 до 170 млн.руб.?

2. По результатам наблюдений за 36 менеджерами оказалось, что среднее количество кредитов в расчете на одного менеджера в месяц составило 32 кредита, а стандартное отклонение – 4,5 кредита.

На каком уровне значимости (α) можно утверждать, что один менеджер выдает больше 30 кредитов в месяц?

Ответ: 1. 0,497 ($\approx 0,5$); 2. 0,004

Задача.

Оценивается доля клиентов, которые воспользуются «новыми» услугами: возможностью отправки квитанции на e-mail и голосовым поиском при выборе услуги.

1. Сколько случайно выбранных клиентов надо опросить, чтобы с вероятностью 99% погрешность для доли клиентов была не более 0,1 (т.е., другими словами, чтобы ширина доверительного интервала не превышала 20%)?

2. Из 500 опрошенных 370 человек ответили, что будут пользоваться услугой «квитанция», и 280 – «поиском». Постройте 90%-ные интервалы для доли клиентов, которые воспользуются каждой из услуг.

3. Руководство считает, что внедрять услугу надо, только если есть полная уверенность, что ей воспользуются более 55% клиентов. Используя уровень доверия $\alpha=0,05$, определите какие услуги стОит внедрять?

Ответ: 1. 166 чел.; 2. 70,8% - 77,2%, 52,3% - 59,7%; 3. только «квитанцию».

Задача.

Доход от одного проекта имеет стандартное отклонение 100 млн.руб., а доход от другого проекта – 10 млн.руб.

1. Можно ли определить стандартное отклонение СОВОКУПНОГО дохода по этим двум проектам?

Если – да, то укажите возможные границы.

2. Известно, что корреляция между доходами проектов отсутствует (равна 0).

Что больше – стандартное отклонение суммы доходов или стандартное отклонение разности доходов этих двух проектов?

Ответ: 1. Да, 90 - 110; 2. Одинаково.

Экзамен проводится в традиционной форме: студент случайным образом выбирает билет, состоящий из одного теоретического вопроса и задачи (см. выше).

Допуском к экзамену является

- выполненное индивидуальное расчетное задание (можно выполнить одно более сложное или трудоемкое или пару более простых заданий),
- освоенный онлайн-курс, в том числе успешно пройденный итоговый тест

Критерии оценивания:

Оценка за Экзамен	Критерии оценивания уровня умений и владений
отлично	<ul style="list-style-type: none"> ● Студент правильно выполнил практическое задание билета. Показал отличные умения в рамках освоенного учебного материала. ● Студент правильно ответил на теоретические вопросы билета. Показал отличные знания в рамках усвоенного учебного материала. ● Ответил на все дополнительные вопросы.
хорошо	<ul style="list-style-type: none"> ● Студент выполнил практическое задание билета с небольшими неточностями. Показал хорошие умения в рамках освоенного учебного материала. ● Студент ответил на теоретические вопросы билета с небольшими неточностями. Показал хорошие знания в рамках усвоенного учебного материала. ● Ответил на большинство дополнительных вопросов.
удовлетворительно	<ul style="list-style-type: none"> ● Студент выполнил практическое задание билета с существенными неточностями. Показал удовлетворительные умения в рамках освоенного учебного материала. ● Студент ответил на теоретические вопросы билета с существенными неточностями. Показал удовлетворительные знания в рамках усвоенного учебного материала. ● При ответах на дополнительные вопросы было допущено много неточностей.
неудовлетворительно	<p>Студент:</p> <ul style="list-style-type: none"> ● не справился с задачей, ● не может ответить на дополнительные («наводящие») вопросы, ● не имеет целостного представления о взаимосвязях и компонентах дисциплины