

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
Ярославский государственный университет им. П.Г.Демидова

УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

Направление подготовки (специальность):

02.04.02 Фундаментальная информатика и информационные технологии

Образовательная программа

Искусственный интеллект и компьютерные науки

очная форма обучения

Составитель:

СПИРИДОНОВА ЕЛЕНА МИХАЙЛОВНА,
Д.Э.Н., ДОЦЕНТ Ф-ТА ИВТ
ЯРГУ ИМ. П.Г. ДЕМИДОВА

г. Ярославль

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

Основная литература:

1. Ильшев, А. М., Общая теория статистики [Электронный ресурс] : учебник / А. М. Ильшев, М., Юнити-Дана, 2015, 535с
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

Дополнительная литература:

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.

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине (модулю)

Для самостоятельной работы особенно рекомендуется использовать учебную литературу.

Также для подбора учебной литературы рекомендуется использовать широкий спектр интернет-ресурсов:

1. Электронно-библиотечная система «Университетская библиотека online» (www.biblioclub.ru) - электронная библиотека, обеспечивающая доступ к наиболее востребованным материалам-первоисточникам, учебной, научной и художественной литературе ведущих издательств (*регистрация в электронной библиотеке – только в сети университета. После регистрации работа с системой возможна с любой точки доступа в Internet.).
2. Информационная система "Единое окно доступа к образовательным ресурсам" (<http://window.edu.ru/library>).

Целью создания информационной системы "Единое окно доступа к образовательным ресурсам" (ИС "Единое окно ") является обеспечение свободного доступа к интегральному каталогу образовательных интернет-ресурсов и к электронной библиотеке учебно-методических материалов для общего и профессионального образования.

Информационная система "Единое окно доступа к образовательным ресурсам" создана по заказу Федерального агентства по образованию в 2005-2008 гг. Головной разработчик проекта - Федеральное государственное автономное учреждение Государственный научно-исследовательский институт информационных технологий и телекоммуникаций (ФГАУ ГНИИ ИТТ "Информика") www.informika.ru.

ИС "Единое окно" объединяет в единое информационное пространство электронные ресурсы свободного доступа для всех уровней образования в России. Разделы этой системы:

- Электронная библиотека – является крупнейшим в российском сегменте Интернета хранилищем полнотекстовых версий учебных, учебно-методических и научных материалов с открытым доступом. Библиотека содержит более 30 000 материалов, источниками которых являются более трехсот российских вузов и других образовательных и научных учреждений. Основу наполнения библиотеки составляют электронные версии учебно-методических материалов, подготовленные в вузах, прошедшие рецензирование и рекомендованные к использованию советами факультетов, учебно-методическими комиссиями и другими вузовскими структурами, осуществляющими контроль учебно-методической деятельности.

- Интегральный каталог образовательных интернет-ресурсов содержит представленные в стандартизированной форме метаданные внешних ресурсов, а также содержит описания полнотекстовых публикаций электронной библиотеки. Общий объем каталога превышает 56 000 метаописаний (из них около 25 000 - внешние ресурсы). Расширенный поиск в "Каталоге" осуществляется по названию, автору, аннотации, ключевым словам с возможной фильтрацией по тематике, предмету, типу материала, уровню образования и аудитории.

- Избранное. В разделе представлены подборки наиболее содержательных и полезных, по мнению редакции, интернет-ресурсов для общего и профессионального образования.

- Библиотеки вузов. Раздел содержит подборки сайтов вузовских библиотек, электронных каталогов библиотек вузов и полнотекстовых электронных библиотек вузов.

Для самостоятельного подбора литературы в библиотеке ЯрГУ рекомендуется использовать:

1. Личный кабинет (http://lib.uniyar.ac.ru/opac/bk_login.php) дает возможность получения on-line доступа к списку выданной в автоматизированном режиме литературы, просмотра и копирования электронных версий изданий сотрудников университета (учеб. и метод. пособия, тексты лекций и т.д.) Для работы в «Личном кабинете» необходимо зайти на сайт Научной библиотеки ЯрГУ с любой точки, имеющей доступ в Internet, в пункт меню «Электронный каталог»; пройти процедуру авторизации, выбрав вкладку «Авторизация», и заполнить представленные поля информации.

2. Электронная библиотека учебных материалов ЯрГУ

(http://www.lib.uniyar.ac.ru/opac/bk_cat_find.php) содержит более 2500 полных текстов учебных и учебно-методических материалов по основным изучаемым дисциплинам, изданных в университете. Доступ в сети университета, либо по логину/паролю.

3. Электронная картотека «Книгообеспеченность»

(http://www.lib.uniyar.ac.ru/opac/bk_bookreq_find.php) раскрывает учебный фонд научной библиотеки ЯрГУ, предоставляет оперативную информацию о состоянии книгообеспеченности дисциплин основной и дополнительной литературой, а также цикла дисциплин и специальностей. Электронная картотека «Книгообеспеченность» доступна в сети университета и через Личный кабинет.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая перечень информационных справочных систем (при необходимости)

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/>

Перечень информационных технологий, используемых при изучении дисциплины, включая программное обеспечение

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);
- 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.

Учебно-методические указания и рекомендации к изучению тем лекционных и практических занятий, самостоятельной работе студентов

Содержание дисциплины

Наименование раздела дисциплины	Название темы с кратким содержанием
---------------------------------	-------------------------------------

Наименование раздела	Название темы с кратким содержанием
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
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
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
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
Примеры анализа и моделирования реальных социально-экономических процессов	5. Примеры анализа и моделирования реальных социально-экономических процессов: □ анализ результатов ЕГЭ, □ моделирование стоимости жилья, □ анализ факторов развития региона и др.

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.)

Список заданий

будет пополняться...

- 1 Анализ абитуриентов ф-та (ЕГЭ)
- 2 Анализ студентов (успеваемость)
- 3 Анализ поведения клиентов банка
- 4 Анализ цен на недвижимость
- 5 Калькулятор расходов на автомобиль
- 6 Калькулятор калорий (КБЖУ, вес)
- 7 Анализ школ по результатам ЕГЭ



❖ 1. Анализ абитуриентов

А. Рассчитать основные характеристики распределения баллов ЕГЭ абитуриентов ф-та ИВТ на момент окончания подачи документов (26.07.17):

- 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	Суздальева 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	185	Приоритет на другом направлении, нет оригиналов документов
4	46873	Тасинов Илдан Магерамович	Нет	Нет	275	Приоритет на другом направлении, нет оригиналов документов
5	43542	Мельникова Татьяна 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	Аробейди Жуман Таназовна	Да	Нет	266	Проходит по конкурсу в первую волну на другое направление
8	43376	Хорош 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	Прокопцкая 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	45421	Ткачев 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	Широква Анна Владимировна	Да	Да	251	Проходит по конкурсу в первую волну
20	44694	Тухватов Александр Термикович	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). оч.старые! (1997 г.) по квартирам Москвы, \$

Б). по нежилым помещениям Челябинска, 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?

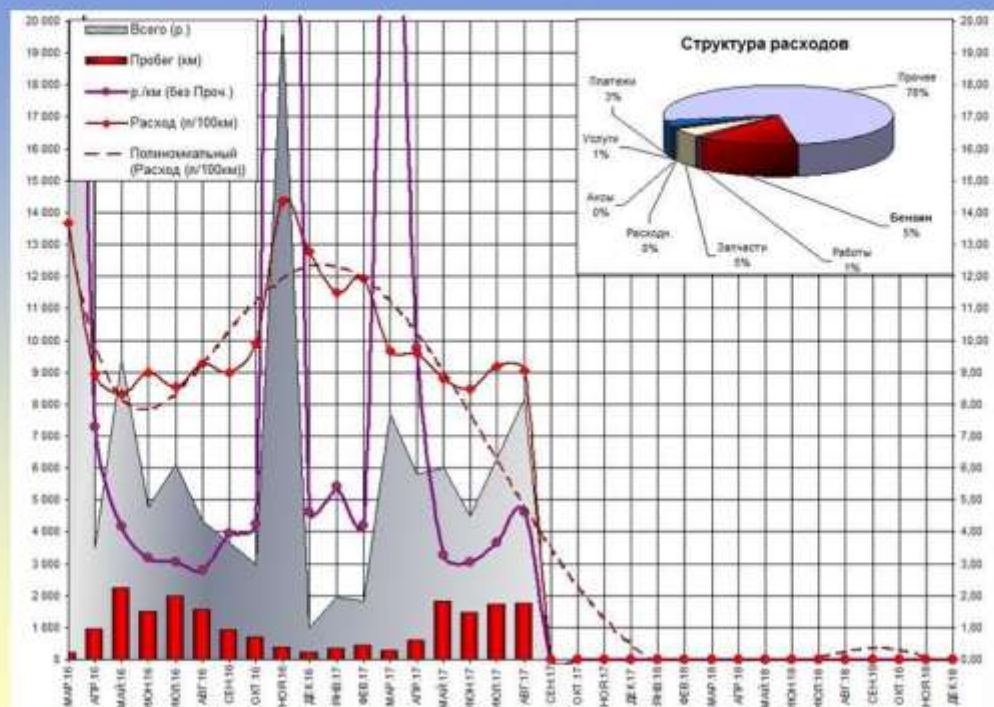
+ для ВКР (и статьи): Собрать «свои» данные!

5. Автокалькулятор

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

- Рассч. 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?

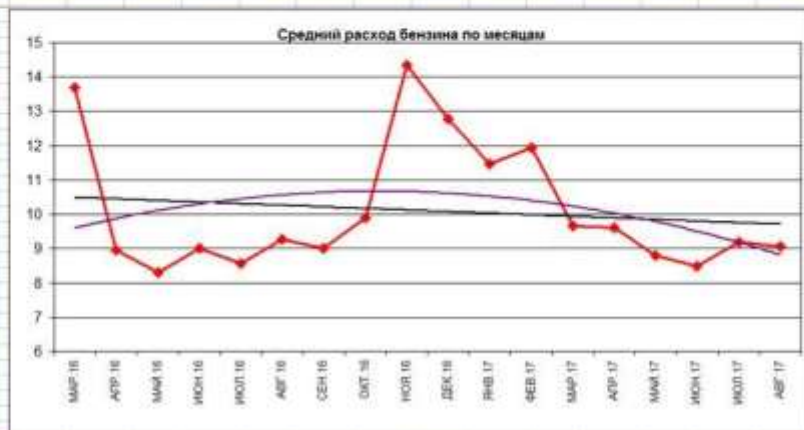
5. Автокалькулятор



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	2100
АПР 16	960
МАЙ 16	2 247
ИЮН 16	1 518
ИЮЛ 16	1 995
АВГ 16	1 642
СЕН 16	923
ОКТ 16	700
НОЯ 16	386
ДЕК 16	217
ЯНВ 17	362
ФЕВ 17	438
МАР 17	262
АПР 17	595
МАЙ 17	1 840
ИЮН 17	1 489
ИЮЛ 17	1 726
АВГ 17	1 763
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.

CarAutoNet
на auto.yar.net.ru

Настройка параметров

Зарегистрирован: 13.07.2011
Авто: VW Touareg MP 3.6 FSI 2011

Учитывать в стоимости одного километра:

- ☒ Начальная стоимость машины: 3200000
- ☒ Бонус (процент): 397188
- ☒ Стоимость (стоимость (параметр)): 16300
- ☒ Расходные материалы: 31382
- ☒ Замасло: 7000
- ☒ Услуги сервиса: 300000
- ☒ Страховка: 480
- ☒ Аксессуары: 47000
- ☒ Прочие: 204880

Средний расход топлива: 18,7 л/100 км, средняя цена 1 километра: 64,28 руб/км
Пробег: 91596 км, зашло 8381 литра топлива, потрачено 3296780 рублей

Дата	Тип	Сумма	Заправка /	Одометр	Комментарий
11.09.2017	Бонус (процент)	1000 - 40	61386	86	
09.08.2017	Услуги сервиса	300	51430		Ремонт тормозов, сцепления
20.08.2017	Бонус (процент)	2000 - 51 (10,99 А)	51270	86	
07.05.2017	Бонус (процент)	1000 - 25 (7,14 А)	51080	86	

3. Добавить запись

Дата (ДД.ММ.ГГГГ):

Тип записи:

- ☐ Приобретенные материалы
- ☐ Бонус (процент)
- ☐ Стоимость (параметр)
- ☐ Расходные материалы
- ☐ Замасло
- ☐ Услуги сервиса
- ☐ Страховка
- ☐ Аксессуары
- ☐ Прочие

Сумма:

Заправка (литры):

Расходные материалы:

Комментарий:

Добавить

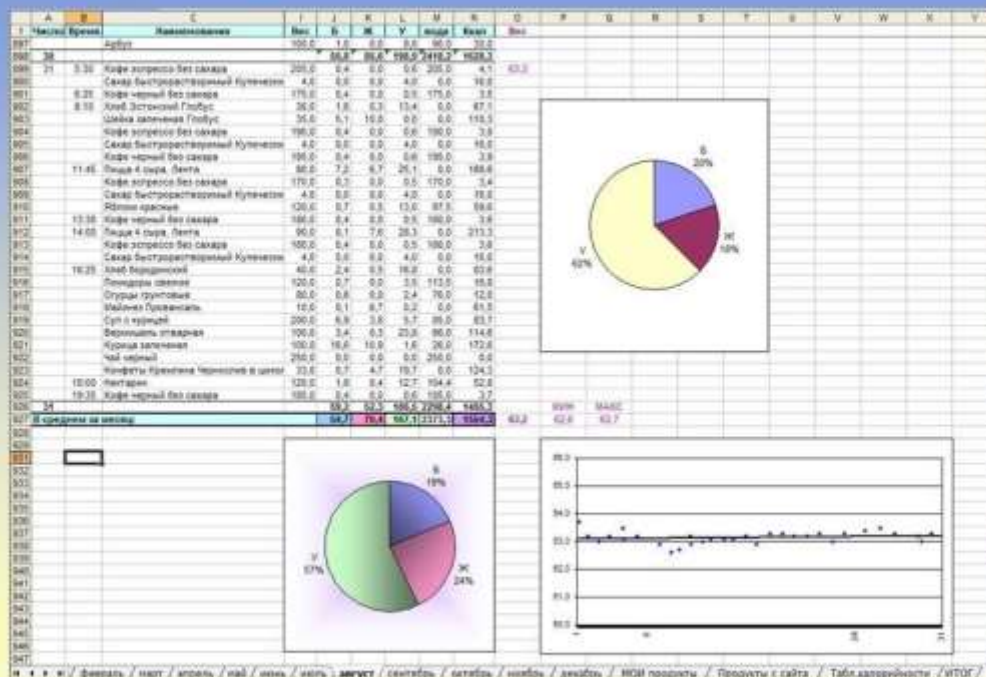
Учитывать в стоимости одного километра:

6. Калькулятор калорий

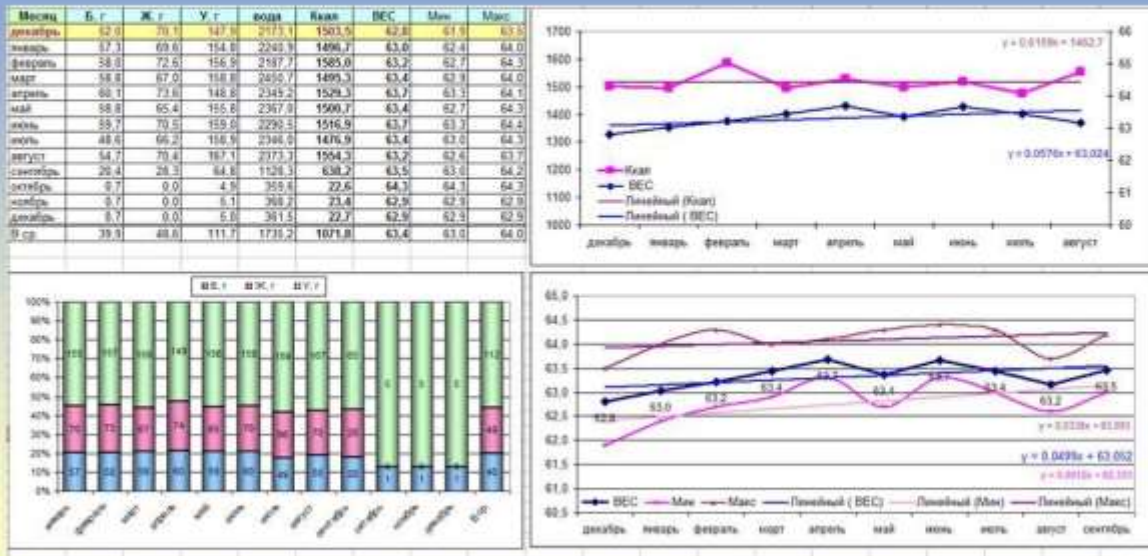
Задача: вести «учет» потребленных продуктов питания, в разрезе килокалорий, белков–жиров–углеводов, воды (а при необходимости – и разл.микроэлементов) с целью контроля веса (для спортсменов и/или последователей ЗОЖ). 😊

- 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



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. 😊

[illegible]

7. Анализ школ по ЕГЭ

Задача: провести кластерный анализ по результатам ЕГЭ по математике и русскому (рекоменд. использ. 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 (2014 – 2017)
- ✓ и 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.

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.

ПРИМЕРЫ контрольной работ

Контрольная работа №1

Задача 1.

Известны следующие данные о результатах ЕГЭ по математике Ярославской области в 2010 г.:

Набранные баллы	Количество, чел.
0 – 10	32
11 – 20	300
21 – 30	1080
31 – 40	1048
41 – 50	1516
51 – 60	1170

61 – 70	772
71 – 80	226
81 – 90	31
91 – 100	10
Итого:	6185

По этим данным рассчитайте:

1. Показатели центра распределения: среднее, моду, медиану.
2. Показатели вариации: среднее линейное отклонение, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Сделайте вывод об однородности совокупности.
3. Показатели формы распределения: асимметрию и эксцесс. Проверьте их на значимость.

Задача 2.

Имеются данные среднем расходе бензина (л на 100 км) автомобиля по месяцам:

	2011	2012	2013	2014	2015
Январь		12,2	15	10,7	12,9
Февраль		11,3	11,9	9,9	11,0
Март		10,0	12,1	9,5	11,0
Апрель		10,0	8,6	8,5	
Май	9,1	8,4	8,1	8,7	
Июнь	8,0	8,5	8,6	8,7	
Июль	8,3	8,2	8,6	8,3	
Август	8,7	8,3	8,4	8,5	
Сентябрь	9,0	8,9	8,9	9,4	
Октябрь	10,1	10,0	9,9	10,9	
Ноябрь	11,0	10,9	9,8	9,9	
Декабрь	9,9	14,5	11,2	10,9	

По этим данным:

1. Выявите общую тенденцию (построив уравнение прямой за весь период).
2. Произведите аналитическое выравнивание с помощью функции Фурье по одной и двум гармоникам данных: а) отдельно за какой-нибудь «полный» год, б)* за весь период.
3. Изобразите фактические и выровненные данные с помощью секторной диаграммы. Сделайте вывод о возможных причинах сезонности и общей тенденции.

Ответы на задачи Контрольной работы №1:

Задача 1.

- $\bar{x} = 44,414$; $Mo = 46,749$; $Me = 45,172$
- $d = 12,839$; $D = 251,543$; $o = 15,860$ г; $v = 35,7\% \Rightarrow$ совокупность школьников по результатам ЕГЭ по математике - не однородна.
- $As = -0,147 \Rightarrow$ левосторонняя асимметрия, существенная;
 $Ex = -0,375 \Rightarrow$ распределение плосковершинное; эксцесс - существенный

Задача 2.

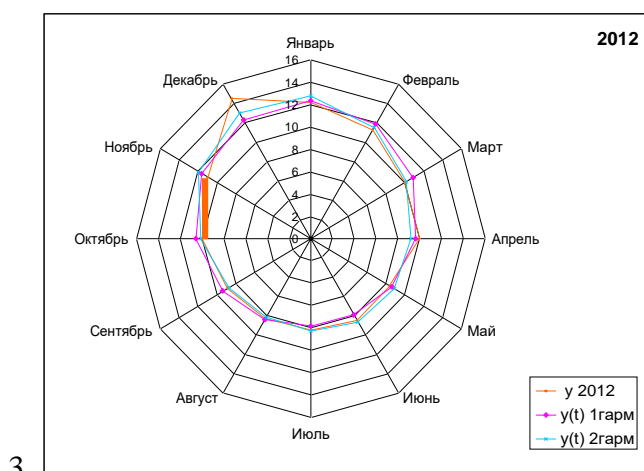
- $Y = 9,819 - 0,0008 \cdot t$



2 а. Для примера - за 2012 г.:

по одной гармонике: $Y = 10,100 + 2,266 \cdot \cos t - 0,452 \cdot \sin t$

по двум: $Y = 10,100 + 2,266 \cdot \cos t - 0,452 \cdot \sin t + 0,433 \cdot \cos 2t - 0,5438 \cdot \sin 2t$



3.

Контрольная работа №2

Задача 1.

Проведите классификацию учеников по результатам ЕГЭ:

Баллы ЕГЭ	Ученики					
	А	Б	В	Г	Д	Е
по Математике	55	58	55	70	73	70
по Русскому языку	80	86	89	77	77	71

В качестве расстояния между «объектами» примите обычное евклидово расстояние. Расстояние между кластерами определите: а) по принципу средней связи, б) по центрам тяжести.

Ответ:

- а) А, Б и В - «гуманитарии»; Г, Д и Е - «технари»; расстояние между кластерами = 18,557.
 б) А, Б и В - «гуманитарии»; Г, Д и Е - «технари»; расстояние между кластерами = 18,028.

Задача 2.

Сеть супермаркетов осуществляет торговую деятельность на территории 12 регионов. В планах – постепенное освоение новых рынков сбыта в других регионах. Аналитики компании выделили наиболее значимые показатели, характеризующие социально–экономическое развитие регионов (см. таблицу):

Регион	Группы регионов	Товарооборот на душу населения, тыс.р.	Инвестиции на душу населения, тыс.р..	Коэффициент плотности автомобильных дорог	Коэффициент покупательной способности
А	Группа X, где деятельность сети была успешной	28,94	8,64	32,06	2,29
Б		31,59	3,96	25,56	2,16
В		23,63	6,33	30,05	1,79
Г		23,62	8,22	29,69	1,62
Д		21,43	5,78	27,57	1,59
Е		17,62	4,62	24,62	1,57
Ж		86,02	20,37	61,69	5,09
З	Группа Y, где деятельность сети НЕ была успешной	17,97	2,45	28,41	1,41
И		14,07	3,94	25,86	1,22
К		11,33	2,06	21,73	0,84
Л		15,93	4,76	31,05	1,31
М		20,18	2,8	25,92	1,53
Н	Группа Z – регионы, входящие в круг интересов	17,47	5,97	28,17	1,29
О		14,88	6,28	15,78	1,32
П		16,27	7,80	29,91	1,32
Р		23,16	8,20	37,83	1,62
С		15,39	6,82	41,28	1,11
Т		19,28	9,68	27,79	1,82

В каких регионах из группы Z следует открыть супермаркеты?



Ответ: в регионах О, Т и П (рассчитано с помощью дискриминирующей функции).

Вопросы к зачету:

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 зачету

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.

ЗАЧЕТ ставится по итогам работы в семестре:

- если выполнены:
 - индивидуальное расчетное задание (можно выполнить одно более сложное или трудоемкое на 15-20 баллов, или пару более простых на 10 баллов),
 - обе контрольные работы,
 - и при этом набрано не менее 25 баллов,
- то оценка «зачтено» может быть поставлена «автоматом».

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

Критерии оценивания:

Оценка «**зачтено**» ставится, если

- ✓ задача решена верно (при этом допускаются несущественные арифметические ошибки, не повлиявшие на итоговые результаты и выводы);
- ✓ теоретический вопрос раскрыт в полной мере (при этом допускаются отдельные неточности, не носящие принципиального характера), на дополнительные вопросы даны правильные ответы.

Оценка «**не зачтено**» выставляется студенту, который:

- 1) не выполнил индивидуальное расчетное задание,
- 2) а на зачете:
 - не справился с задачей,
 - не может ответить на дополнительные («наводящие») вопросы,
 - не имеет целостного представления о взаимосвязях и компонентах дисциплины.