

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):
09.04.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

Образовательная программа
Искусственный интеллект в корпоративных информационных системах

очная форма обучения

Составитель:
Спиридонова Е.М., д.э.н., доцент

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.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

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/>

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

Очная форма обучения

Лекции

№	Тема лекции	Количество часов
1	Статистические показатели, характеризующие совокупность и распределение	2
2	Критерии согласия и проверка статистических гипотез	2
3	Корреляционно-регрессионный анализ: закономерности, тенденции и прогнозы. Проблемы регрессионного анализа	2
4	Многомерные методы: факторный, кластерный и дискриминантный анализ	3
5	Примеры анализа и моделирования реальных социально-экономических процессов	3
	Итого:	12

Практические занятия

№	Тема практического занятия	Количество часов
1	Статистические показатели, характеризующие совокупность и распределение	4
2	Критерии согласия и проверка статистических гипотез	5
3	Корреляционно-регрессионный анализ: закономерности, тенденции и прогнозы. Проблемы регрессионного анализа	5
4	Многомерные методы: факторный, кластерный и дискриминантный анализ	5
5	Примеры анализа и моделирования реальных социально-экономических процессов	5
	Итого:	24

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	285	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	Хорова 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	Ромашова Владислава Игоревна	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	Туганов Александр Терентьевич	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?

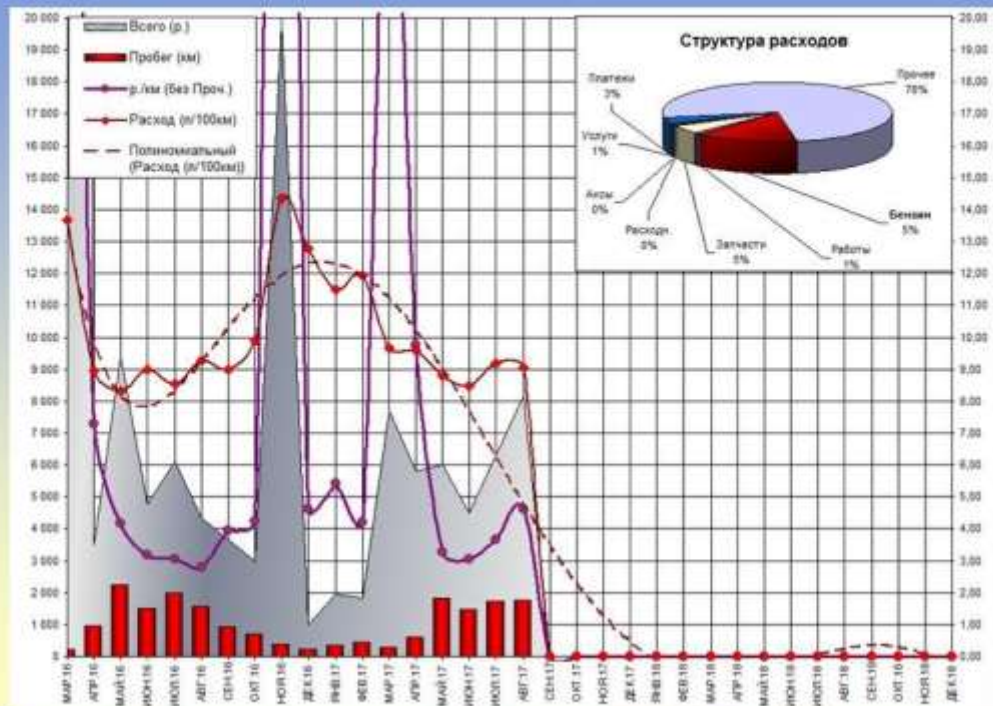
+ для ВКР (и статьи): Собрать «свои» данные!

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

3. Расчет параметров: 05.08.2011
Авто: VW Touareg MP 3.6 FSI 2011

Учитывать в стоимости одного километра:

- (X) Начальная стоимость машины: 3200000
- (X) Бензин (топливо): 397188
- (X) Стоимость стоянки (парковка): 16300
- (X) Расходные материалы: 31382
- (X) Замасло: 7000
- (X) Услуги сервиса: 300000
- (X) Штрафы: 400
- (X) Аксессуары: 47000
- (X) Прочие: 20488

Средний расход топлива: 18 л/100 км, средняя цена 1 километра: 64,28 руб/км
Пробег: 91596 км, зарплата 8181 руб/мес, затраты: 3191780 рублей

Дата	Тип	Сумма	Заправка / Оборот	Комментарий
11.09.2017	Бензин (топливо)	1000 - 40	61086	86
09.08.2017	Услуги сервиса	300	51430	Ремонт тормозных дисков
20.08.2017	Бензин (топливо)	2000 - 51 (18,89 л)	11270	86
07.08.2017	Бензин (топливо)	1000 - 25 (7,14 л)	51080	86

3. Расчет параметров: 05.08.2011
Авто: VW Touareg MP 3.6 FSI 2011

Добавить запись

Дата (ДД.ММ.ГГГГ):

Тип записи:

- ☐ Приобретение машины
- ☐ Бензин (топливо)
- ☐ Стоянка (парковка)
- ☐ Расходные материалы
- ☐ Замасло
- ☐ Услуги сервиса
- ☐ Штрафы
- ☐ Аксессуары
- ☐ Прочие

Сумма:

Заправка (литры):

Ремонт (стоимость):

Комментарий:

Добавить

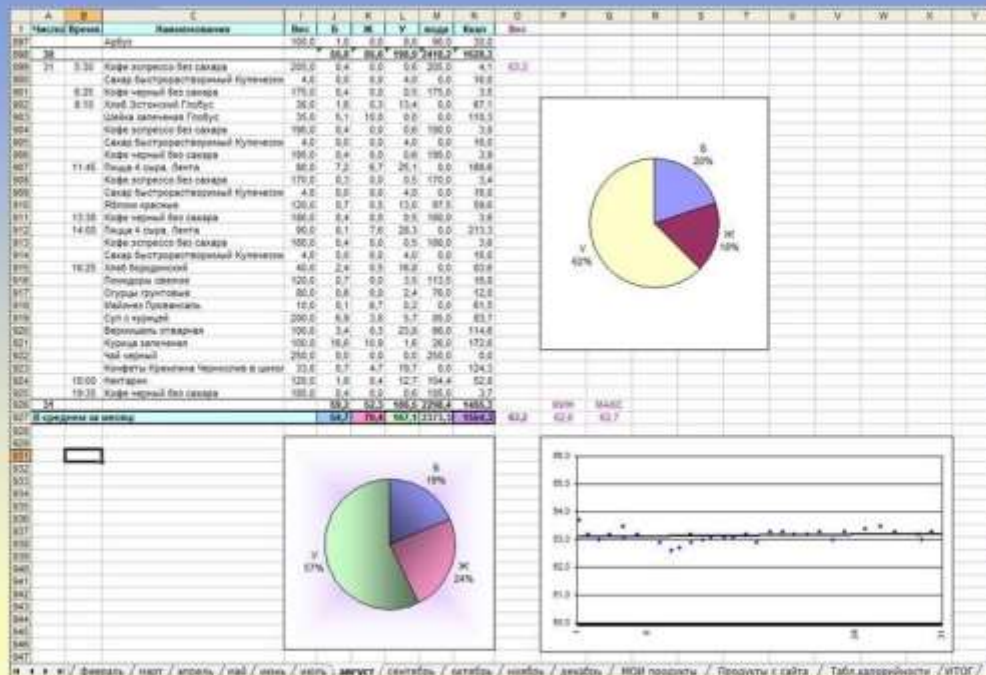
Учитывать в стоимости одного километра:

6. Калькулятор калорий

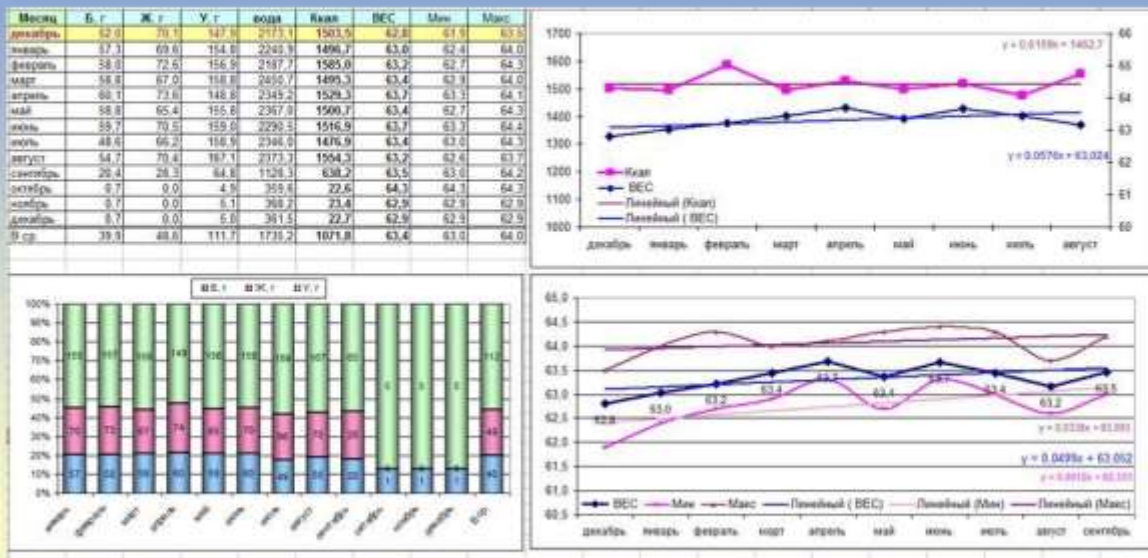
Задача: вести «учет» потребленных продуктов питания, в разрезе килокалорий, белков–жиров–углеводов, воды (а при необходимости – и разл.микроэлементов) с целью контроля веса (для спортсменов и/или последователей ЗОЖ). 😊

- 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. 😊

City / Region	Elevation (m)	Month											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
City A	120	10.5	11.2	12.1	13.0	14.5	15.8	16.5	15.2	14.0	12.8	11.5	10.2
City B	150	11.0	11.8	12.5	13.5	15.0	16.2	17.0	15.5	14.2	13.0	11.8	10.5
City C	180	11.5	12.2	13.0	14.0	15.5	16.8	17.5	16.0	14.8	13.5	12.2	11.0
City D	210	12.0	12.8	13.5	14.5	16.0	17.2	18.0	16.5	15.2	14.0	12.8	11.5
City E	240	12.5	13.2	14.0	15.0	16.5	17.8	18.5	17.0	15.8	14.5	13.2	12.0
City F	270	13.0	13.8	14.5	15.5	17.0	18.2	19.0	17.5	16.2	15.0	13.8	12.5
City G	300	13.5	14.2	15.0	16.0	17.5	18.8	19.5	18.0	16.8	15.5	14.2	13.0
City H	330	14.0	14.8	15.5	16.5	18.0	19.2	20.0	18.5	17.2	16.0	14.8	13.5
City I	360	14.5	15.2	16.0	17.0	18.5	19.8	20.5	19.0	17.8	16.5	15.2	14.0
City J	390	15.0	15.8	16.5	17.5	19.0	20.2	21.0	19.5	18.2	17.0	15.8	14.5
City K	420	15.5	16.2	17.0	18.0	19.5	20.8	21.5	20.0	18.8	17.5	16.2	15.0
City L	450	16.0	16.8	17.5	18.5	20.0	21.2	22.0	20.5	19.2	18.0	16.8	15.5
City M	480	16.5	17.2	18.0	19.0	20.5	21.8	22.5	21.0	19.8	18.5	17.2	16.0
City N	510	17.0	17.8	18.5	19.5	21.0	22.2	23.0	21.5	20.2	19.0	17.8	16.5
City O	540	17.5	18.2	19.0	20.0	21.5	22.8	23.5	22.0	20.8	19.5	18.2	17.0
City P	570	18.0	18.8	19.5	20.5	22.0	23.2	24.0	22.5	21.2	20.0	18.8	17.5
City Q	600	18.5	19.2	20.0	21.0	22.5	23.8	24.5	23.0	21.8	20.5	19.2	18.0
City R	630	19.0	19.8	20.5	21.5	23.0	24.2	25.0	23.5	22.2	21.0	19.8	18.5
City S	660	19.5	20.2	21.0	22.0	23.5	24.8	25.5	24.0	22.8	21.5	20.2	19.0
City T	690	20.0	20.8	21.5	22.5	24.0	25.2	26.0	24.5	23.2	22.0	20.8	19.5
City U	720	20.5	21.2	22.0	23.0	24.5	25.8	26.5	25.0	23.8	22.5	21.2	20.0
City V	750	21.0	21.8	22.5	23.5	25.0	26.2	27.0	25.5	24.2	23.0	21.8	20.5
City W	780	21.5	22.2	23.0	24.0	25.5	26.8	27.5	26.0	24.8	23.5	22.2	21.0
City X	810	22.0	22.8	23.5	24.5	26.0	27.2	28.0	26.5	25.2	24.0	22.8	21.5
City Y	840	22.5	23.2	24.0	25.0	26.5	27.8	28.5	27.0	25.8	24.5	23.2	22.0
City Z	870	23.0	23.8	24.5	25.5	27.0	28.2	29.0	27.5	26.2	25.0	23.8	22.5
City AA	900	23.5	24.2	25.0	26.0	27.5	28.8	29.5	28.0	26.8	25.5	24.2	23.0
City AB	930	24.0	24.8	25.5	26.5	28.0	29.2	30.0	28.5	27.2	26.0	24.8	23.5
City AC	960	24.5	25.2	26.0	27.0	28.5	29.8	30.5	29.0	27.8	26.5	25.2	24.0
City AD	990	25.0	25.8	26.5	27.5	29.0	30.2	31.0	29.5	28.2	27.0	25.8	24.5
City AE	1020	25.5	26.2	27.0	28.0	29.5	30.8	31.5	30.0	28.8	27.5	26.2	25.0
City AF	1050	26.0	26.8	27.5	28.5	30.0	31.2	32.0	30.5	29.2	28.0	26.8	25.5
City AG	1080	26.5	27.2	28.0	29.0	30.5	31.8	32.5	31.0	29.8	28.5	27.2	26.0
City AH	1110	27.0	27.8	28.5	29.5	31.0	32.2	33.0	31.5	30.2	29.0	27.8	26.5
City AI	1140	27.5	28.2	29.0	30.0	31.5	32.8	33.5	32.0	30.8	29.5	28.2	27.0
City AJ	1170	28.0	28.8	29.5	30.5	32.0	33.2	34.0	32.5	31.2	30.0	28.8	27.5
City AK	1200	28.5	29.2	30.0	31.0	32.5	33.8	34.5	33.0	31.8	30.5	29.2	28.0
City AL	1230	29.0	29.8	30.5	31.5	33.0	34.2	35.0	33.5	32.2	31.0	29.8	28.5
City AM	1260	29.5	30.2	31.0	32.0	33.5	34.8	35.5	34.0	32.8	31.5	30.2	29.0
City AN	1290	30.0	30.8	31.5	32.5	34.0	35.2	36.0	34.5	33.2	32.0	30.8	29.5
City AO	1320	30.5	31.2	32.0	33.0	34.5	35.8	36.5	35.0	33.8	32.5	31.2	30.0
City AP	1350	31.0	31.8	32.5	33.5	35.0	36.2	37.0	35.5	34.2	33.0	31.8	30.5
City AQ	1380	31.5	32.2	33.0	34.0	35.5	36.8	37.5	36.0	34.8	33.5	32.2	31.0
City AR	1410	32.0	32.8	33.5	34.5	36.0	37.2	38.0	36.5	35.2	34.0	32.8	31.5
City AS	1440	32.5	33.2	34.0	35.0	36.5	37.8	38.5	37.0	35.8	34.5	33.2	32.0
City AT	1470	33.0	33.8	34.5	35.5	37.0	38.2	39.0	37.5	36.2	35.0	33.8	32.5
City AU	1500	33.5	34.2	35.0	36.0	37.5	38.8	39.5	38.0	36.8	35.5	34.2	33.0
City AV	1530	34.0	34.8	35.5	36.5	38.0	39.2	40.0	38.5	37.2	36.0	34.8	33.5
City AW	1560	34.5	35.2	36.0	37.0	38.5	39.8	40.5	39.0	37.8	36.5	35.2	34.0
City AX	1590	35.0	35.8	36.5	37.5	39.0	40.2	41.0	39.5	38.2	37.0	35.8	34.5
City AY	1620	35.5	36.2	37.0	38.0	39.5	40.8	41.5	40.0	38.8	37.5	36.2	35.0
City AZ	1650	36.0	36.8	37.5	38.5	40.0	41.2	42.0	40.5	39.2	38.0	36.8	35.5
City BA	1680	36.5	37.2	38.0	39.0	40.5	41.8	42.5	41.0	39.8	38.5	37.2	36.0
City BB	1710	37.0	37.8	38.5	39.5	41.0	42.2	43.0	41.5	40.2	39.0	37.8	36.5
City BC	1740	37.5	38.2	39.0	40.0	41.5	42.8	43.5	42.0	40.8	39.5	38.2	37.0
City BD	1770	38.0	38.8	39.5	40.5	42.0	43.2	44.0	42.5	41.2	40.0	38.8	37.5
City BE	1800	38.5	39.2	40.0	41.0	42.5	43.8	44.5	43.0	41.8	40.5	39.2	38.0
City BF	1830	39.0	39.8	40.5	41.5	43.0	44.2	45.0	43.5	42.2	41.0	39.8	38.5
City BG	1860	39.5	40.2	41.0	42.0	43.5	44.8	45.5	44.0	42.8	41.5	40.2	39.0
City BH	1890	40.0	40.8	41.5	42.5	44.0	45.2	46.0	44.5	43.2	42.0	40.8	39.5
City BI	1920	40.5	41.2	42.0	43.0	44.5	45.8	46.5	45.0	43.8	42.5	41.2	40.0
City BJ	1950	41.0	41.8	42.5	43.5	45.0	46.2	47.0	45.5	44.2	43.0	41.8	40.5
City BK	1980	41.5	42.2	43.0	44.0	45.5	46.8	47.5	46.0	44.8	43.5	42.2	41.0
City BL	2010	42.0	42.8	43.5	44.5	46.0	47.2	48.0	46.5	45.2	44.0	42.8	41.5
City BM	2040	42.5	43.2	44.0	45.0	46.5	47.8	48.5	47.0	45.8	44.5	43.2	42.0
City BN	2070	43.0	43.8	44.5	45.5	47.0	48.2	49.0	47.5	46.2	45.0	43.8	42.5
City BO	2100	43.5	44.2	45.0	46.0	47.5	48.8	49.5	48.0	46.8	45.5	44.2	43.0
City BP	2130	44.0	44.8	45.5	46.5	48.0	49.2	50.0	48.5	47.2	46.0	44.8	43.5
City BQ	2160	44.5	45.2	46.0	47.0	48.5	49.8	50.5	49.0	47.8	46.5	45.2	44.0
City BR	2190	45.0	45.8	46.5	47.5	49.0	50.2	51.0	49.5	48.2	47.0	45.8	44.5
City BS	2220	45.5	46.2	47.0	48.0	49.5	50.8	51.5	50.0	48.8	47.5	46.2	45.0
City BT	2250	46.0	46.8	47.5	48.5	50.0	51.2	52.0	50.5	49.2	48.0	46.8	45.5
City BU	2280	46.5	47.2	48.0	49.0	50.5	51.8	52.5	51.0	49.8	48.5	47.2	46.0
City BV	2310	47.0	47.8	48.5	49.5	51.0	52.2	53.0	51.5	50.2	49.0	47.8	46.5
City BW	2340	47.5	48.2	49.0	50.0	51.5	52.8	53.5	52.0	50.8	49.5	48.2	47.0
City BX	2370	48.0	48.8	49.5	50.5	52.0	53.2	54.0	52.5	51.2	50.0	48.8	47.5
City BY	2400	48.5	49.2	50.0	51.0	52.5	53.8	54.5	53.0	51.8	50.5	49.2	48.0
City BZ	2430	49.0	49.8	50.5	51.5	53.0	54.2	55.0	53.5	52.2	51.0	49.8	48.5
City CA	2460	49.5	50.2	51.0	52.0	53.5	54.8	55.5	54.0	52.8	51.5	50.2	49.0
City CB	2490	50.0	50.8	51.5	52.5	54.0	55.2	56.0	54.5	53.2	52.0	50.8	49.5
City CC	2520	50.5	51.2	52.0	53.0	54.5	55.8	56.5	55.0	53.8	52.5	51.2	50.0
City CD	2550	51.0	51.8	52.5	53.5	55.0	56.2	57.0	55.5	54.2	53.0	51.8	50.5
City CE	2580	51.5	52.2	53.0	54.0	55.5	56.8	57.5	56.0	54.8	53.5	52.2	51.0
City CF	2610	52.0	52.8	53.5	54.5	56.0	57.2	58.0	56.5	55.2	54.0	52.8	51.5
City CG	2640	52.5	53.2	54.0	55.0	56.5	57.8	58.5	57.0	55.8	54.5	53.2	52.0
City CH	2670	53.0	53.8	54.5	55.5	57.0	58.2	59.0	57.5	56.2	55.0	53.8	52.5
City CI	2700	53.5	54.2	55.0	56.0	57.5	58.8	59.5	58.0	56.8	55.5	54.2	53.0
City CJ	2730	54.0	54.8	55.5	56.5	58.0	59.2	60.0	58.5	57.2	56.0	54.8	53.5
City CK	2760	54.5	55.2	56.0	57.0	58.5	59.8	60.5	59.0	57.8	56.5	55.2	54.0
City CL	2790	55.0	55.8	56.5	57.5	59.0	60.2	61.0	59.5	58.2	57.0	55.8	54.5
City CM	2820	55.5	56.2	57.0	58.0	59.5	60.8	61.5	60.0	58.8	57.5	56.2	55.0
City CN	2850	56.0	56.8	57.5	58.5	60.0	61.2	62.0	60.5	59.2	58.0	56.8	55.5
City CO	2880	56.5	57.2	58.0	59.0	60.5	61.8	62.5	61.0	59.8	58.5	57.2	56.0
City CP	2910	57.0	57.8	58.5	59.5	61.0	62.2	63.0	61.5	60.2	59.0	57.8	56.5
City CQ	2940	57.5	58.2	59.0	60.0	61.5	62.8	63.5	62.0	60.8	59.5	58.2	57.0
City CR	2970	58.0	58.8	59.5	60.5	62.0	63.2	64.0	62.5	61.2	60.0	58.8	57.5
City CS													

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.

Отвѣты 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.

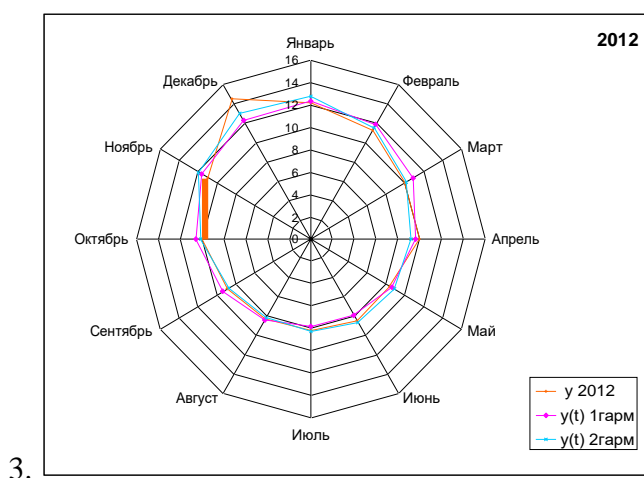
- $Y = 9,819 - 0,0008 \cdot t$



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.