Необходимо сдать: Rmd и сгенерированный по нему html/pdf-файл с подробным отчётом по проведённому исследованию, содержащий визуализацию исходных данных, описания и выводы каждого этапа анализа — используемые методы, обоснование их применимости, графики.

Работа с реальными данными

Требуется подобрать и применить наилучший статистический метод, позволяющий ответить на вопрос прикладной задачи; обосновать выбор метода, его применимость и оптимальность. Помимо выводов, касающихся математических особенностей решения, необходимо в терминах предметной области сформулировать количественные выводы, которые могли бы быть понятны гипотетическому заказчику-нематематику.

Данные доступны по ссылке `https://yadi.sk/d/Y0F3614a3ExSYe`.

1. Трахов Роман

Рак лёгких в Китае. Для участников исследования, проживающих в одном из восьми городов Китая, известно, курят ли они и больны ли раком лёгких.

Задача. Как связаны риск заболевания раком лёгких, курение и город проживания участников исследования?

Данные. china_smoking.xls.

2. Федоряка Дмитрий

Годовой заработок. Опрос US Bureau of Labor Statistics 2002 года содержит данные о годовом заработке 55729 участников; известны также их пол (1 = male, 2 = female), возраст, уровень образования (1 = no high school, 2 = some high school, 3 = high school diploma, 4 = some college, 5 = bachelor's degree, 6 = postgraduate degree) и тип работы (5 = private sector, 6 = government, 7 = self-employed).

Задача. Оценить влияние образования и пола на годовой заработок.

Данные. workers.xls.

3. Чуйкова Екатерина

Линька крабов. У 472 самок крабов metacarcinus magister измерена ширина панциря до и после линьки. Измерения были получены двумя способами: 1) 12000 крабов измеряли, помечали сигнальными маячками и выпускали обратно в естественную среду перед периодом линьки, затем часть крабов за вознаграждение возвращалась в лабораторию рыбаками, выловившими их с помощью стандартных ловушек; 2) крабов, выловленных на суше во время спаривания непосредственно перед линькой, приносили в лабораторию, измеряли, затем, через несколько дней после линьки, измеряли снова. Для второй категории известен год вылова.

Задача. Исследовать различия между изменениями размеров панциря особей, линька которых проходила в лабораторных условиях и в естественных. Для последних оценить влияние года вылова.

Данные. crabs.csv.

4. Сноровихина Виктория

Белки в коре мозга мышей. В 1080 образцах коры мозга мышей измерен уровень экспрессии 77 белков. Часть образцов взята от трисомных мышей (лабораторная модель синдрома Дауна), часть — от здоровых; в эксперименте перед получением образцов некоторые мыши получали стимул к обучению, а некоторые — нет; наконец, части мышей

вводился Мемантин, а части — физраствор. Цель эксперимента — проверить, восстанавливает ли Мемантин способность к обучению у трисомных мышей.

Задача. Влияет ли Мемантин на экспрессию белков здоровых мышей?

Данные. memantine.xls.

5. Погодин Роман

Прочность промышленных вентиляторов. Измерен разрушающий крутящий момент 64 промышленных вентиляторов; для каждого известны тип отверстия, форма барабана и метод соединения.

Задача. Связан ли разрушающий крутящий момент с характеристиками вентилятора?

Данные. fans.txt.

6. Фокина Дарья

Биомаркеры рака груди. В эксперименте принимали участие 24 человек, у которых не было рака груди (normal), 25 человек, у которых это заболевание было диагностировано на ранней стадии (early neoplasia), и 23 человека с сильно выраженными симптомами (cancer). Секвенирование — это определение степени активности генов в анализируемом образце с помощью подсчёта количества соответствующей каждому гену РНК; именно эта количественная мера активности каждого из 15748 генов для каждого из 72 человек записана в данных.

Разница в уровнях экспрессии гена между группами считается практически значимой, если средние уровни в группах отличаются более, чем в полтора раза; таким образом, необходимо посчитать величину fold change:

$$F_c(C,T) = \begin{cases} \frac{T}{C}, & T > C, \\ -\frac{C}{T}, & T < C, \end{cases}$$

где C, T — средние значения экспрессии гена в control и treatment группах соответственно, и считать практически значимыми те отличия, для которых $|F_c(C,T)| > 1.5$.

Задача. По каким генам имеются статистически и практически значимые отличия в уровнях экспрессии между здоровыми испытуемыми и испытуемыми с ранней стадией рака? Между здоровыми и испытуемыми с сильно выраженными симптомами?

Данные. gene_high_throughput_sequencing.csv

7. Молибог Игорь

Дома престарелых Нью-Мексико. Для 52 лицензированных домов престарелых Нью-Мексико известны: число коек, суммарное годовое число дней в стационаре и койко-дней (в сотнях), суммарные годовые расходы на уход за пациентами, зарплату медсестёр и инфраструктуру (в сотнях долларов).

Задача. Есть ли различия между сельскими и городскими домами престарелых? По каким признакам?

Данные. nursing_homes.txt.

8. Чигринский Виктор

Размеры черепа древних египтян. Было измерено 150 черепов, найденных при раскопках в Египте. Находки относятся к пяти различным временным периодам. Для каждого черепа известны: максимальная ширина, базибрегматическая высота, базиальвеолярная длина, высота носа, примерная дата формирования. Была выдвинута гипотеза о том, что изменение этих параметров со временем может свидетельствовать о скрещивании египтян с другими популяциями.

Задача. Проверить, есть ли различия между размерами черепов различных временных периодов, если есть, то какие периоды отличаются друг от друга.

Данные. skulls.txt.

9. Катугина Татьяна

Риск сердечно-сосудистых заболеваний. В рамках исследования риска сердечно-сосудистых заболеваний изучалось влияние регулярных занятий спортом на частоту сердечных сокращений при выполнении шестиминутного упражнения на беговой дорожке. Выборка состоит из 800 испытуемых двух групп. Половина из них пробегала не меньше 15 миль в неделю, половина вела малоподвижный образ жизни. Известен пол испытуемых и их пульс по окончании упражнения.

Задача. Оценить влияние регулярных занятий спортом на частоту сердечных сокращений после нагрузки с учётом пола.

Данные. heartrate.txt.

10. Шульгина Евгения

Массовая доля жира в организме. Массовая доля жира, важная характеристика здоровья, рассчитывается через плотность тела, измеряемую при помощи взвешивания в воде. Для 252 мужчин проведены такие расчёты. Имеются также данные антропометрии (возраст, рост, вес, обхват грудной клетки и т.д.).

Задача. Как связаны простые антропометрические показатели (возраст, рост, вес, ИМТ) с обхватами?

Данные. fat.xls.

11. Багаев Рамазан

Оптимальные условия размножения штаммов золотистого стафилококка. При подозрении на инфекционное заболевание для правильной постановки диагноза часто бывает важно из взятых у пациентов образцов вырастить как можно более многочисленную колонию бактерий, чтобы её было удобнее исследовать. Считается, что оптимальные параметры для размножения штаммов стафилококка в лабораторных условиях следующие: температура 35 градусов, концентрация триптона в питательном растворе 1.0%, время выдержки 24 часа. Для проверки оптимальности этих условий было проведено 30 экспериментов над пятью различными штаммами стафилококка. Для каждого из экспериментов известны время выдержки, температура, концентрация триптона, а также измеренное по окончании выдержки число колониеобразующих единиц (КОЕ) бактерий каждого штамма.

Задача. Оценить зависимость итогового числа КОЕ каждого штаммов стафилококка от внешних условий; одинакова ли эта зависимость?

Данные. Staphylococcus aureus.txt.

12. Горишний Юрий

Одеяла с электрообогревом. Одеяла с электрообогревом применяются в хирургии для восстановления температуры тела пациента после операции. Имеются четыре вида одеяла: стандартный, b0, и три экспериментальных — b1, b2, b3. Для 41 пациента известно время, за которое нормальная температура тела восстанавливается при использовании одеяла одного из видов.

Задача. Отличаются ли экспериментальные одеяла от стандартного?

Данные. blanket.txt.

13. Володин Сергей

Засеивание облаков и уровень осадков Исследовалось воздействие засеивания облаков на обилие дождей. Измерения проводились в течение 108 периодов на пяти участках земли в Тасмании – участки обозначены в файле как западный, восточный, южный, северный и северо-восточный. В выборке содержатся данные об уровне осадков (в миллиметрах) на каждом из пяти участков, о времени года, к которому относится период, и о том, проводилось ли засеивание. Необходимо проверить, как засеивание облаков повлияло на уровень осадков отдельно по каждому из пяти экспериментальных участков.

Задача. Проверить, одинаково ли проявляется эффект засеивания на каждом из них, или, возможно, он как-то зависит от исходного уровня осадков на участке?

Данные. cloudseeding.txt.

14. Рязанов Андрей

Заживление ран. На 26 пациентах было испытано экспериментальное лекарство, способствующее заживлению ран; для сравнения ещё к 26 пациентам применялась стандартная терапия. Измерялась площадь раны до начала терапии, после курса лечения и на заключительном визите через длительное время после завершения лечения. Кроме того, приведена субъективная оценка изменения состояния раны пациентом и врачом.

Задача. Отличается ли эффективность экспериментального лекарства от эффективности стандартного?

Данные. wounds.csv.

15. Водопьян Даниил

Данные антропометрии. Для 247 мужчин и 260 женщин измерены две группы антропометрических показателей – легко измеримые характеристики скелета и обхваты, всего 21 признак. Указаны возраст, пол, вес и рост.

Задача. Как связаны друг с другом характеристики скелета и обхваты?

Данные. body.xlsx.

16. Гончаренко Владислав

Комментарии в блогах. Для 60021 постов в блогах, опубликованных не более, чем за 72 часа до базового времени, собрана информация о количестве комментариев, времени публикации, длине и количестве каждого из 200 часто встречающихся слов.

Задача. Чем отличаются сообщения, не получившие ни одного комментария, от тех, которые хотя бы кто-нибудь прокомментировал?

Данные. blog_feedback.xlsx.

17. Рындин Максим

Пожертвования на благотворительность. Благотворительная организация разослала 4268 писем с предложением сделать пожертвование и получила отклик с пожертвованиями от 1707 адресатов. Для каждого адресата известны: индикатор ответа на предыдущее письмо, число недель, прошедших с момента предыдущего пожертвования, размеры текущего, предыдущего и среднего по всем предыдущим пожертвованиям в голландских гульденах, число писем, отправляемых адресату в год, доля писем, в ответ на которые приходят пожертвования.

Задача. Какие признаки отличают людей, совершающих пожертвования?

Данные. charity.xlsx.

18. Стельмах Иван

Задержка авиарейсов. Для 4029 рейсов, вылетавших из Нью-йоркского аэропорта Ла-Гуардия, известны название авиакомпании-перевозчика, аэропорт назначения, диапазон планируемого времени вылета, день недели, месяц, продолжительность полёта и время задержки вылета.

Задача. Есть ли закономерности в задержках вылетов?

Данные. FlightDelays.csv.

19. Баяндина Анастасия

Годовой заработок. Опрос US Bureau of Labor Statistics 2002 года содержит данные о годовом заработке 55729 участников; известны также их пол (1 = male, 2 = female), возраст, уровень образования (1 = no high school, 2 = some high school, 3 = high school diploma, 4 = some college, 5 = bachelor's degree, 6 = postgraduate degree) и тип работы (5 = private sector, 6 = government, 7 = self-employed).

Задача. Оценить влияние пола и типа работы на годовой заработок.

Данные. workers.xls.

20. Рябых Алексей

General Social Survey. General Social Survey – ежегодный социологический опрос нескольких тысяч граждан США; на сайте `https://gssdataexplorer.norc.org` доступны все данные с 1972 по 2014 год (GSS.xls).

Задача. Исследовать связь уровня счастья (General happiness) опрошенных 2014 года с демографическими признаками (пол, возраст, раса, семейное положение, количество детей, образование, сексуальная ориентация, занятость, доход).

Данные. GSS.xls.

21. Колоскова Анастасия

Продолжительность жизни больных онкологическими заболеваниями. Выборка состоит из 64 пациентов, у которых был диагностирован неизлечимый рак какого-либо органа. Всем им в качестве поддерживающей терапии был назначен к приёму витамин C (считалось, что он может способствовать выздоровлению раковых больных). Приведены данные об остаточной продолжительности жизни пациентов в днях.

Задача. Исследовать связь между остаточной продолжительностью жизни и типом рака.

Данные. cancer.txt.

22. Луканин Артём

General Social Survey. General Social Survey – ежегодный социологический опрос нескольких тысяч граждан США; на сайте `https://gssdataexplorer.norc.org` доступны все данные с 1972 по 2014 год (GSS.xls)

Задача. Исследовать связь веры в научность астрологии опрошенных 2014 года с признаками, описывающими отношение к религии (религиозные предпочтения, уверенность в существовании бога, религиозная и духовная самоидентификация, участие в религиозных активностях).

Данные. GSS.xls.

23. Комаров Никита

Биоразлагаемость молекул. 1055 химических молекул описаны с помощью 41 признака (число атомов кислорода, нитратных групп, донорных связей с водородом, потенциал ионизации и т.д.); 355 из них биоразложимы.

Задача. Какие признаки отличают биоразложимые молекулы от устойчивых?

Данные. biodeg.xlsx.

24. Григорук Василий

Открытие депозита. Имеются результаты обзвона 4119 клиентов португальского банка, которым предлагалось завести депозит. Известны социально-демографические характеристики клиентов, история предыдущих коммуникаций, социально-экономические показатели на момент совершения звонка.

Задача. Какие признаки повышают вероятность открытия депозита клиентом по результатам обзвона?

Данные. deposit.xlsx.

25. Самсонов Никита

Белки в коре мозга мышей. В 1080 образцах коры мозга мышей измерен уровень экспрессии 77 белков. Часть образцов взята от трисомных мышей (лабораторная модель синдрома Дауна), часть — от здоровых; в эксперименте перед получением образцов некоторые мыши получали стимул к обучению, а некоторые — нет; наконец, части мышей вводился Мемантин, а части — физраствор. Цель эксперимента — проверить, восстанавливает ли Мемантин способность к обучению у трисомных мышей.

Задача. Отличается ли экспрессия белков у здоровых и трисомных мышей в каких-нибудь из экспериментальных подгрупп?

Данные. memantine.xls.

26. Зильберман Лев

Внешний вид и привлекательность самок мечехвостов. Изучалось влияние внешних характеристик самок морских ракообразных на их привлекательность для самцов. Выборка состоит из данных о наблюдениях над 173 особями и содержит закодированные данные о размере самок, их весе, цвете, состоянии панциря, а также о количестве спутников.

Задача. Сравнить по всем имеющимся признакам самок, имеющих хотя бы одного спутника, с самками, не имеющими ни одного.

Данные. horseshoe crab.txt.

27. Ковалёва Валерия

Обучение родителей воспитанию детей. 975 родителей участвовало в программе обучения воспитанию. Было проведено три опроса, в ходе которых родители отвечали на вопрос: «За последние несколько недель обращались ли дети к вам с проблемой или вопросом, который их беспокоил?» Первый опрос был проведён до начала обучения, второй – сразу после, и третий – по прошествии 6-8 недель после окончания обучения. Известен также уровень образования родителя.

Задача. Стали ли родители больше общаться с детьми в результате обучения? Проанализировать с учётом уровня образования родителей.

Данные. education.txt

28. Малышев Иван

Нарушения ПДД. В исследовании влияния обучения подростков вождению на число инцидентов с нарушениями ПДД контрольная группа состоит из 2409 человек. По каждому из них данные собираются на протяжении четырёх лет.

Задача. Меняется ли в контрольной группе число инцидентов с годами? Если да, то как?

Данные. traffic_violation.txt.

29. Малыгин Виталий

Продажи платьев. Имеются данные по продажам 479 платьев на сайте aliexpress.com за полтора месяца осени 2013 года. Для каждого из платьев известны также стиль, ценовая категория, рейтинг, размер, сезон, ряд характеристик внешнего вида и индикатор участия в программе рекомендаций.

Задача. Исследовать, как каждый из признаков по отдельности влияет на уровень продаж.

Данные. aliexpress_dress_data.csv.

30. Цветкова Ольга

Годовой заработок. Опрос US Bureau of Labor Statistics 2002 года содержит данные о годовом заработке 55729 участников; известны также их пол (1 = male, 2 = female), возраст, уровень образования (1 = no high school, 2 = some high school, 3 = high school diploma, 4 = some college, 5 = bachelor's degree, 6 = postgraduate degree) и тип работы (5 = private sector, 6 = government, 7 = self-employed).

Задача Оценить влияние образования, пола и типа работы на годовой заработок.

Данные. workers.xls.

31. Биктайров Роман

Maryland's Pick-3 Lottery. Даны результаты розыгрыша лотереи Maryland's Pick-3 Lottery за 218 подряд идущих дней. Результатом является трёхзначное число.

Задача. Можно ли считать розыгрыш случайным?

Данные. lottery.txt.

32. Малькова Александра

Качество воды в Миннесоте. Для 895 источников воды в Миннесоте известны водоносный горизонт, водоём, уровень и химические свойства воды (pH, щёлочность, содержание алюминия, мышьяка, хлора и свинца) .

Задача. Сравнить свойства воды из разных водоёмов.

Данные. MnGroundwater.csv.

33. Гарипов Ильяс

Эффективность тромболитической терапии. Собраны данные по 206 пациентам второго кардиологического отделения московской городской клинической больницы №25. Имеются результаты 14 анализов, а также 8 дополнительных признаков, описывающих пациента (пол, возраст, курение, наличие диабета и т.д.).

Задача. Оценить влияние курения на вероятности выздоровления и возникновения осложнений, а также на результаты 14 анализов.

Данные. cardio.xls.

34. Ратько Мария

Цифры числа пи. Даны первые десять тысяч цифр числи пи.

Задача. Можно ли сказать, что все цифры встречаются с одинаковой частотой? Есть ли зависимость между подряд идущими цифрами?

Данные. pi10000.txt.

35. Писов Максим

Содержание азота в кормовых растениях. Четыре вида растений (Leucaena leucocephala, Acacia saligna, Prosopis juliflora, Eucalyptus citriodora), которые могут расти в долине реки Иордан, где климат достаточно засушливый, выращивались в лаборатории при разных условиях доступа к воде. Количество воды, получаемой растением в сутки, менялось в разных группах от 50 до 650 мм с шагом 100 мм. Исследователей интересовала возможность их использования для кормления сельскохозяйственных животных, для чего необходимо высокое содержание азота. Для 9 растений в каждой группе известно содержание азота.

Задача. Как содержание азота меняется для разных видов растений при разных условиях выращивания? Какие растения лучше всего подходят для сельскохозяйственного использования?

Данные. plants.xlsx

36. Змеев Максим

Допустимость наказаний. Известно мнение двенадцати родителей о допустимости наказания их детей по результатам оценки в психогенном эксперименте; допустимость выражается в баллах. Чем ниже балл, тем менее допустимым участник исследования считает наказание. Имеются результаты о наказании самим родителем, бабушкой и учителем ребёнка.

Задача. Как зависит оценка допустимости наказания от наказывающего?

Данные. punishment.txt

37. Парубченко Александр

Интеллект и размер головного мозга. Исследование проводилось среди студентов психологического факультета крупного университета. Все испытуемые должны были быть правшами, а также не иметь повреждений мозга, эпилепсии, алкоголизма и сердечных заболеваний. Участники предварительного этапа эксперимента прошли несколько IQ-тестов, после чего для дальнейшего участия было отобрано 20 мужчин и 20 женщин, имевших коэффициент интеллекта либо ниже 103, либо выше 130 баллов. Для каждого из отобранных при помощи магнитно-резонансной томографии были получены 18 снимков срезов головного мозга, и общее количество пикселей на всех 18 снимках было принято в качестве меры объёма мозга. Помимо этого были собраны данные о росте и массе тела испытуемых.

Задача. Исследовать взаимосвязи между коэффициентами интеллекта и биологическими характеристиками испытуемых (пол, рост, вес, объём мозга).

Данные. brain.xlsx.

38. Иванов Алексей

Продолжительность жизни и активность размножения самцов дрозофилы. Для изучения влияния активности размножения самцов дрозофилы на продолжительность их жизни был организован следующий эксперимент. По 25 самцов в пяти группах содержались в одинаковых условиях, за исключением одного отличия: в первой группе

к каждому самцу ежедневно подсаживалась готовая к размножению самка, во второй – восемь готовых к размножению самок, в третьей и четвёртой – соответственно, одна и восемь беременных самок, не готовых к размножению, наконец, к самцам пятой группы не подсаживали никого. Для каждого самца измерена продолжительность жизни, длина грудной клетки и доля времени, проводимого во сне.

Задача. Исследовать связь между продолжительностью жизни самцов дрозофилы и наличием самок разного типа и количества.

Данные. fly.txt.

39. Сафин Камиль

Эффективность тромболитической терапии. Собраны данные по 206 пациентам второго кардиологического отделения московской городской клинической больницы №25. Имеются результаты 14 анализов, а также 8 дополнительных признаков, описывающих пациента (пол, возраст, курение, наличие диабета и т.д.).

Задача. Оценить влияние возраста на вероятности выздоровления и возникновения осложнений, а также на результаты 14 анализов.

Данные. cardio.xls.

40. Неганов Алексей

Продажи платьев. Имеются данные по продажам 479 платьев на сайте aliexpress.com за полтора месяца осени 2013 года. Для каждого из платьев известны также стиль, ценовая категория, рейтинг, размер, сезон, ряд характеристик внешнего вида и индикатор участия в программе рекомендаций.

Задача. Исследовать, как ценовая категория и стиль влияют на уровень продаж.

Данные. aliexpress_dress_data.csv.

41. Павлов Сергей

General Social Survey. General Social Survey – ежегодный социологический опрос нескольких тысяч граждан США. На сайте `https://gssdataexplorer.norc.org/` доступны все данные с 1972 по 2014 год (GSS.xls).

Задача. Для опрошенных 2014 года исследовать связь суммарного количества правильных ответов на 11 вопросов на знание базовых научных фактов с демографическими признаками (пол, возраст, раса, семейное положение, количество детей, образование, сексуальная ориентация, занятость, доход).

Данные. GSS.xls.

42. Леонтьев Семён

Эффективность тромболитической терапии. Собраны данные по 206 пациентам второго кардиологического отделения московской городской клинической больницы №25. Имеются результаты 14 анализов, а также 8 дополнительных признаков, описывающих пациента (пол, возраст, курение, наличие диабета и т.д.).

Задача. Оценить влияние наличия диабета на вероятности выздоровления и возникновения осложнений, а также на результаты 14 анализов.

Данные. cardio.xls.

43. Владимирова Мария

Курение и болезнь Альцгеймера. Ретроспективное исследование влияния курения на болезнь Альцгеймера включает пациентов с болезнью Альцгеймера, другими формами деменции и другими диагнозами; известны статус курения и пол.

Задача. Как курение и пол связаны с различными формами снижения умственной деятельности?

Данные. alzheimer.txt.

44. Стрельцов Фёдор

Продажи платьев. Имеются данные по продажам 479 платьев на сайте aliexpress.com за полтора месяца осени 2013 года. Для каждого из платьев известны также стиль, ценовая категория, рейтинг, размер, сезон, ряд характеристик внешнего вида и индикатор участия в программе рекомендаций.

Задача. Исследовать, как рейтинг и участие в программе рекомендаций влияют на уровень продаж.

Данные. aliexpress_dress_data.csv.

45. Копырин Денис

General Social Survey. General Social Survey – ежегодный социологический опрос нескольких тысяч граждан США; на сайте `https://gssdataexplorer.norc.org` доступны все данные с 1972 по 2014 год (GSS.xls).

Задача. Для опрошенных 2014 года исследовать связь отношения к аборту (abany) с демографическими признаками (пол, возраст, семейное положение, количество детей, сексуальная ориентация).

Данные. GSS.xls.

46. Склонин Илья

Годовой заработок. Опрос US Bureau of Labor Statistics 2002 года содержит данные о годовом заработке 55729 участников; известны также их пол (1 = male, 2 = female), возраст, уровень образования (1 = no high school, 2 = some high school, 3 = high school diploma, 4 = some college, 5 = bachelor's degree, 6 = postgraduate degree) и тип работы (5 = private sector, 6 = government, 7 = self-employed).

Задача. Оценить влияние образования и типа работы на годовой заработок.

Данные. workers.xls.

47. Шлёнский Владислав

Кассовые сборы кинофильмов. На сайте boxofficemojo.com имеются сведения о мировых кассовых сборах всех кинофильмов, выходивших в США, и студиях, их выпустивших. Рассмотрим данные о фильмах, вышедших в 2014 году. Будем считать крупными киностудии, выпустившие в этом году не менее 10 фильмов.

Задача. Сравнить средние кассовые сборы вышедших в 2014 году фильмов крупных киностудий в США, по всему миру и суммарно.

Данные. `http://www.boxofficemojo.com/yearly/chart/?view2=worldwide&yr=2014&p=.htm`

48. Борзов Артём

Краш-тест с манекенами. Имеются результаты 352 краш-тестов, при которых происходило лобовое столкновение автомобилей с бетонной стеной на скорости около 60 км/ч. Измерены показатели повреждения манекенов: критерий тяжести повреждений головы, замедление грудной клетки, нагрузка на левое и правое бедро.

Задача. Исследовать зависимость показателей повреждения от типа кузова, вида средств защиты, места манекена, веса автомобиля.

Данные. crush.xls.

49. Белозёрова Анастасия

Свойства грибов. Для 8416 грибов задано признаковое описание согласно справочнику The Audubon Society Field Guide to North American Mushrooms.

Задача. По каким признакам съедобные грибы отличаются от ядовитых, и каковы эти отличия? (замечание – стоит отфильтровать однозначные категории)

Данные. mushroom.csv.

50. Филиппенко Константин

Эхинацея и респираторные заболевания. Группа детей от 2 до 11 лет случайным образом была распределена по двум типам лечения ОРЗ — плацебо (329 испытуемых) и экстракт эхинацеи (367 испытуемых). Доза эхинацеи подбиралась соответственно возрасту ребёнка по рекомендации производителя. Для каждого заболевания известны тяжесть его протекания по мнению родителей и наличие сопутствующих симптомов.

Задача. Эффективно ли лечение ОРЗ с помощью экстракта эхинацеи?

Данные. echinacea.xlsx.

51. Малюков Евгений

Вакцина против папилломавируса. Собраны данные по 1413 пациенткам клиник при университете Джона Хопкинса, проходившим с 2006 по 2008 вакцинацию против папилломавируса человека препаратом Гардасил. Рекомендуемый курс – три укола в течение года – был пройдён только 469 пациентками. Производитель препарата исследует, в каких демографических группах и каком способе получения вакцины проведение полного курса наиболее вероятно.

Задача. Что отличает пациентов, прошедших полный курс вакцинации в течение года, от тех, кто его не прошёл?

Данные. gardasil.xls.