

ВТОРОЙ ОТБОРОЧНЫЙ ЭТАП, 3 ВАРИАНТ 2023/2024

1. (2 балла) Один интернет-магазин решил добавить для пользователей мобильных устройств кнопку чата в углу экрана. Но перед этим он выбрал большую репрезентативную группу своих постоянных клиентов, чтобы посмотреть, насколько востребованной будет эта функция. Оказалось, что в этой группе половина — пользователи Android, пятая часть — пользователи браузерной версии, а остальные — пользователи iOS. При этом после теста выяснилось, что из пользователей Android кнопкой чата воспользовались 35%, среди пользователей браузерной версии — 85%, а среди пользователей iOS — 60%. (Известно, что каждый клиент в группе пользуется только одной из платформ.)

Исходя из этих данных, если постоянный клиент воспользовался кнопкой чата, то пользователем какой платформы из этих трех он вероятнее всего является?

- a. Android
 - b. Браузерная версия
 - c. iOS (+)
 - d. Из этих условий невозможно однозначно определить ответ
2. (2 балла) Юный исследователь захотел оценить количество посетителей ближайшего торгового центра. Он тщательно пересчитал всех посетителей на протяжении недели. За время наблюдений получилось, что в среднем ТЦ посещает 1548 человек в день, или 46 440 человек за месяц. Как вы можете прокомментировать полученную оценку числа посетителей?
- a. Эта оценка близка к реальному количеству посетителей (+)
 - b. Эта оценка будет занижать количество посетителей

- c. Эта оценка будет завышать количество посетителей
 - d. Недостаточно данных, чтобы ответить на вопрос
3. (2 балла) Одна компания в конце 2023 года решила объявить список из 8 своих самых продаваемых товаров за год. Каждый из аналитиков компании, включая главу аналитики и стажера, попробовал угадать этот список — и записал заранее 8 товаров, которые, как он предположил, туда попадут. После оглашения списка оказалось, что половина аналитиков угадали от 4 до 7 товаров, глава аналитики угадал все товары, стажер угадал лишь 2 товара. А остальные угадали ровно по 3 товара, и среди них не оказалось тех, кто угадал один и тот же набор из 3 товаров, а также тех, кто угадал хотя бы один из товаров, угаданных стажером.

Какое максимальное число аналитиков может работать в этом магазине?

- a. 34
 - b. 44 (+)
 - c. 54
 - d. 74
4. (5 баллов) Выборка состоит из 3510 различных чисел, причем все числа положительные и представимы в виде конечной десятичной дроби. Все числа из выборки упорядочили по возрастанию и разделили на 13 равных по величине групп, которые пронумеровали от 1 до 13. То есть первые 270 чисел упорядоченной выборки оказались в группе 1, следующие 270 чисел — в группе 2 и так далее. Выберите все верные утверждения:
- a. Среднее арифметическое чисел в группе 3 может оказаться больше среднего арифметического чисел в группе 4
 - b. Сумма двух чисел из группы 1 не может оказаться больше числа из группы 13
 - c. Если одно число меньше другого в 13 раз, то они могут оказаться в одной группе (+)
 - d. Медиана выборки обязательно окажется в группе 7 (+)
 - e. Среди перечисленного выше нет верных ответов
5. (5 баллов) В таблице ниже представлено распределение ответов школьников на вопрос «Как изменилась ваша успеваемость в этом учебном году

по сравнению с предыдущим?». Все ученики отвечали на вопрос правдиво, таким образом представленные числа отражают реальное изменение успеваемости.

Ответ	Частота
Значительно улучшилась	16
Немного улучшилась	8
Не изменилась	12
Немного ухудшилась	4
Сильно ухудшилась	3

Выберите все верные варианты ответа:

- a. Успеваемость улучшилась более чем у половины школьников (+)
 - b. Успеваемость не улучшилась более чем у половины школьников
 - c. У медианного школьника успеваемость не изменилась
 - d. У медианного школьника успеваемость немного улучшилась (+)
 - e. Среди перечисленного выше нет верных ответов
6. (5 баллов) Диана устала от подсчета среднего арифметического и медианы — и решила придумать свое среднее, которое назвала «среднее Дианино». Оно считается так: если имеется выборка из N различных чисел, то она сначала упорядочивается по убыванию, а затем считается среднее арифметическое первых 2 чисел, потом — среднее арифметическое получившегося среднего и следующего числа из выборки, затем — получившегося теперь числа и следующего числа из выборки и так далее.

Например, если выборка состоит из 5 упорядоченных по возрастанию/убыванию чисел a_1, a_2, a_3, a_4, a_5 , то:

$$m_D = \frac{\frac{a_1 + a_2}{2} + a_3}{2} + a_4 \over 2} + a_5$$

Диана решила протестировать работу придуманного ею среднего и взяла какую-то выборку из 16 различных натуральных чисел a_1, a_2, \dots, a_{16} , каждое от 1 до 100. Она упорядочила выборку по убыванию, затем посчитала среднее Дианино, среднее арифметическое и медиану выборки — и сравнила их.

Медиана в случае четного количества чисел равна полусумме чисел в середине упорядоченной выборки.

Какие результаты могли получиться у Дианы?

- a. Среднее Дианино могло оказаться больше среднего арифметического
 - b. Среднее Дианино могло оказаться больше медианы
 - c. Среднее Дианино могло оказаться меньше среднего арифметического (+)
 - d. Среднее Дианино могло оказаться меньше медианы (+)
 - e. Среди перечисленного выше нет верных ответов
7. (8 баллов, по 2 на каждый вопрос) Предположим, в стране проживает 9 человек, доходы которых представлены в таблице ниже. Государству не нравится, что средний доход отличается от медианного, и оно хочет это исправить. Для этого предлагается ввести в стране налоги на богатых и раздать эти деньги другим гражданам. При этом государство также хочет собирать как можно меньше налогов с богатых.

Вам необходимо посчитать, сколько налогов и субсидий необходимо собрать и раздать, чтобы медиана и среднее оказались равны. Если подходящих способов несколько, используйте любой из них для ответа на все вопросы ниже

Житель	A	B	C	D	E	F	G	H	I
Доход (y. e.)	10	15	20	30	50	80	110	220	500

Вопрос 1. Кто из жителей будет получать субсидию? Перечислите их в алфавитном порядке через запятую.

Ответ: E, F, G

Вопрос 2. Какую субсидию будет получать каждый из тех жителей, которых вы перечислили выше? Перечислите их через запятую в порядке, соответствующем порядку жителей. Если ответ представлен в виде дроби, округлите его до целого. Единицы измерения указывать не нужно.

Ответ: 65, 35, 5

Вопрос 3. Кто из жителей будет платить налог? Перечислите их в алфавитном

порядке через запятую.

Ответ: либо I; либо H, I

Вопрос 4. Какой объем налогов суммарно заплатят все, кого вы перечислили выше? Если ответ представлен в виде дроби, округлите его до целого. Единицы измерения указывать не нужно.

Ответ: 105

8. (5 баллов) В таблице ниже приведены данные об общей кредитной задолженности компаний и индивидуальных предпринимателей по видам экономической деятельности и направлениям использования средств (в млн рублей).

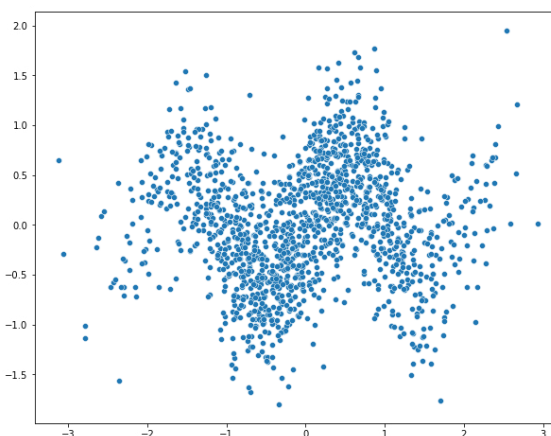
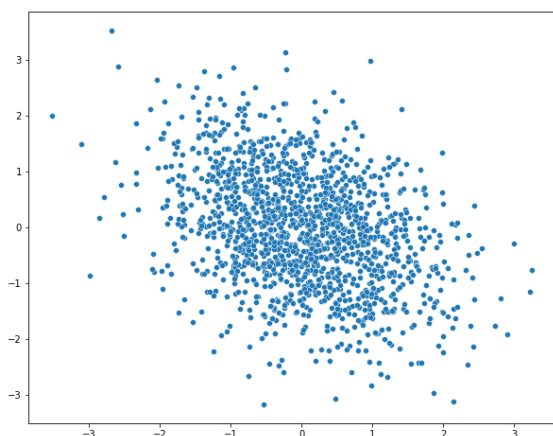
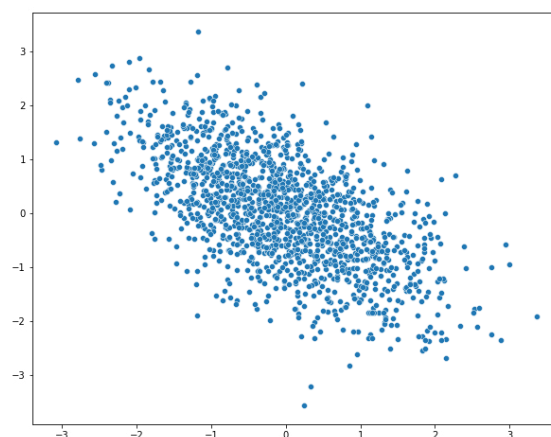
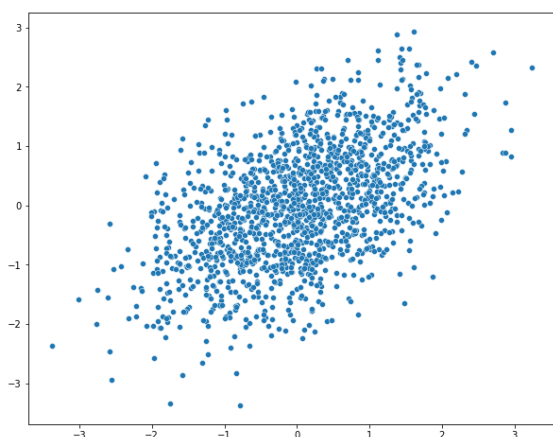
Виды экономической деятельности и отдельные направления использования средств	01.02.2019	01.03.2019	01.04.2019
ВСЕГО	25 728 662	25 885 238	26 209 480
добыча полезных ископаемых	959 863	973 512	986 700
обрабатывающие производства	6 082 270	6 137 454	6 138 053
производство и распределение электроэнергии, газа и воды	1 218 916	1 207 717	1 180 787
сельское хозяйство, охота и лесное хозяйство	1 725 294	1 746 871	1 800 908
строительство	1 063 041	1 087 213	1 108 528
транспорт и связь	1 771 028	1 805 037	1 841 668
оптовая и розничная торговля	3 681 859	3 678 621	3 747 668
операции с недвижимым имуществом, аренда и предоставление услуг	3 874 280	3 915 409	3 936 643
прочие виды деятельности	5 078 084	5 040 035	5 163 325
на завершение расчетов	274 027	293 369	305 200

Выберите все верные утверждения из представленных ниже:

- a. За март 2019 общий объем кредитов вырос на 1,25% (+)
- b. Кредиты в секторе производства и распределения электроэнергии, газа и воды в среднем растут в феврале — марте 2019

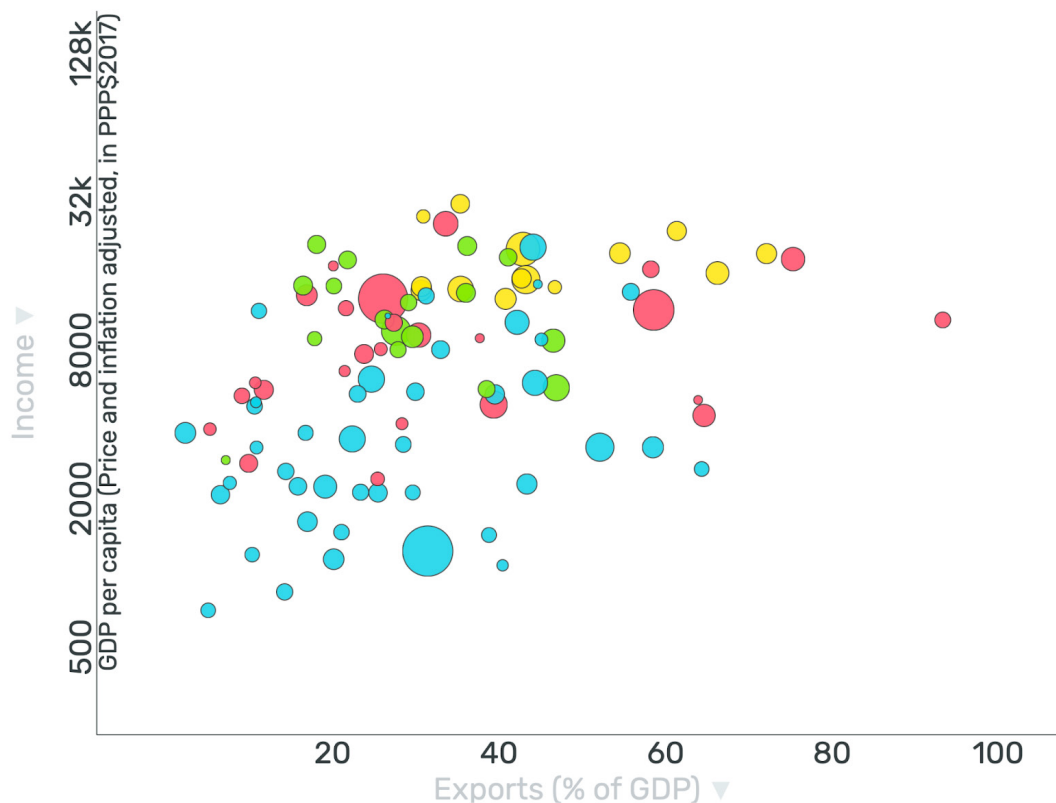
- c. На четыре вида экономической деятельности с самым большим объемом задолженности в указанный период приходилось более 75% всей задолженности в экономике
- d. Доля задолженности строительных компаний выросла за рассматриваемый промежуток времени (+)
- e. Среди перечисленного выше нет верных ответов

9. (4 балла, по 1 за каждый график) Установите взаимосвязь между данными на графике и значениями коэффициента корреляции.



- a. 0,5
- b. 0,6
- c. 0,33
- d. 0,02

10. (5 баллов) На рисунке изображена диаграмма рассеяния, показывающая взаимосвязь между доходами населения и размером экспорта по состоянию на 2021 год. На графике представлены в основном развивающиеся страны.



По вертикальной оси отражен реальный ВВП на душу населения в долларах (показатель скорректирован на инфляцию и пересчитан по паритету покупательной способности для сопоставимости между странами). Для наглядности графика значения ВВП преобразованы, при этом способ качественно не влияет на выводы.

По горизонтальной оси отражено отношение экспорта страны к ее ВВП.

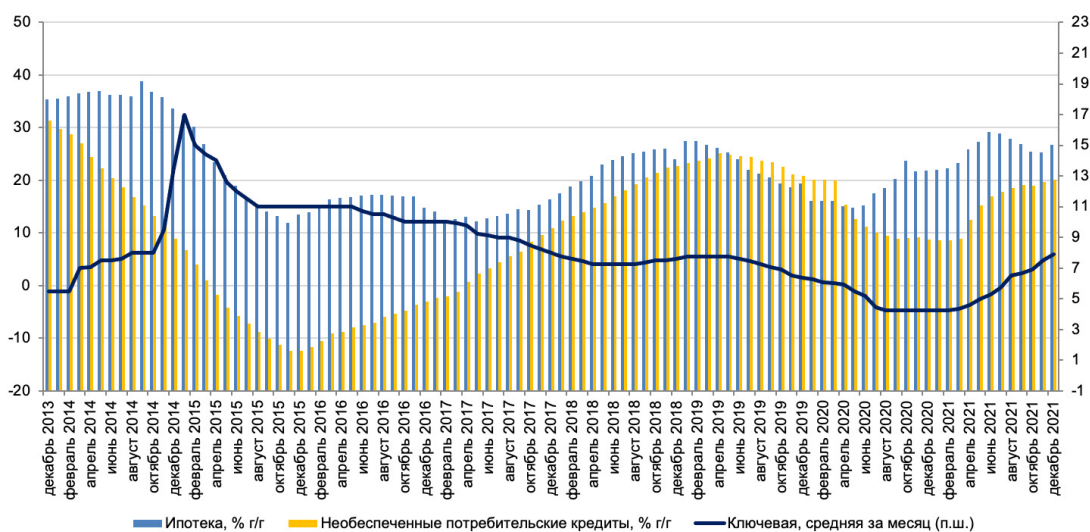
Размер кругов отображает третий показатель — отношение совокупного внешнего долга страны к валовому национальному доходу — и отражает задолженность страны (государства и частных лиц) перед нерезидентами (иностранными государствами, фирмами и людьми). Чем больше размер круга, тем больше задолженность страны.

Цвет кругов обозначает разные регионы мира: голубой — Африка, красный — Азия (включая Австралию и Новую Зеландию), зеленый — Америка (Северная и Южная вместе), желтый — Европа.

Выберите все верные утверждения, которые можно сделать на основании этого графика:

- a. В среднем доход на душу населения в странах Северной и Южной Америки выше, чем в странах Африки (+)
- b. В среднем отношение экспорта к ВВП в странах Европы ниже, чем в странах Африки
- c. Разброс доходов на душу населения в странах Европы ниже, чем в странах Африки (+)
- d. Три страны с наибольшим отношением задолженности иностранному сектору к доходу — страны Африки и Европы
- e. Среди перечисленного выше нет верных ответов

11. (5 баллов) На графике (взято из Обзора российского финансового сектора и финансовых инструментов Банка России за 2021 год) показана динамика ипотечного и необеспеченного потребительского кредитования в России за 2014–2021 годы, а также ключевая ставка за тот же промежуток времени. Темпы прироста кредитования отложены по левой шкале, ключевая ставка — по правой.



Выберите все утверждения, которые можно сделать на основе приведенного выше графика:

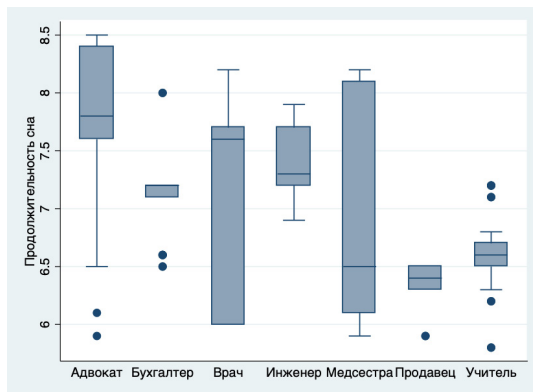
- a. Объемы ипотечного кредитования в среднем растут меньше, чем объемы необеспеченного кредитования
- b. С апреля 2017 по апрель 2018 объем ипотечного кредитования снижался
- c. В течение 2017 года темп роста обоих видов кредитования в среднем увеличивался (+)

d. На протяжении 2016–2018 годов наблюдается отрицательная взаимосвязь между темпом роста ипотечного кредитования и ключевой ставкой (+)

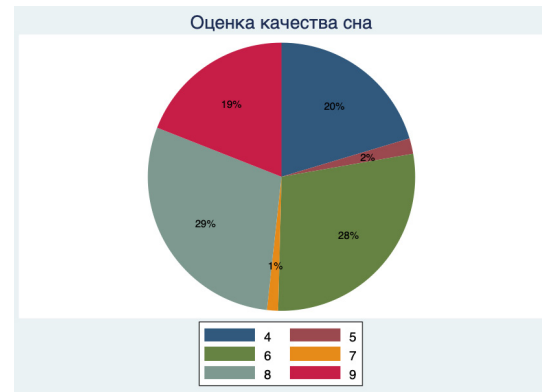
e. Среди перечисленного выше нет верных ответов

12. (10 баллов) (Файл sleep.xlsx) В представленном файле вы можете найти информацию о роде деятельности человека, различных индикаторах качества его сна (субъективная оценка качества, продолжительность, наличие проблем со сном) и состоянии его здоровья. Какие из приведенных ниже графиков могли бы быть получены при проведении анализа взаимосвязи различных показателей качества сна с различными персональными характеристиками человека?

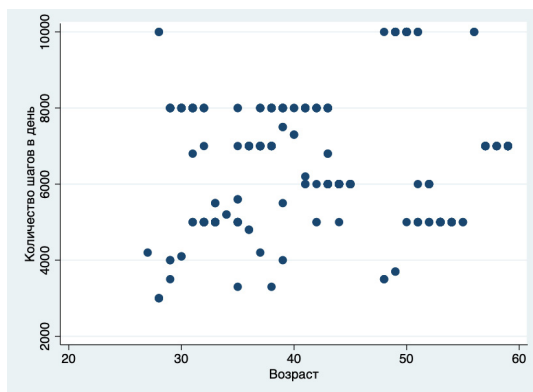
a)



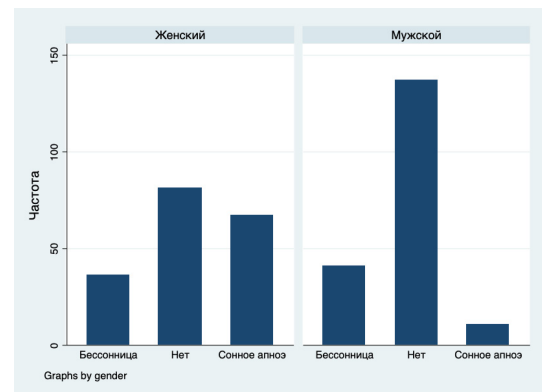
b)



c)



d) +



e) Среди перечисленного выше нет верных ответов

13. (6 баллов) Вы провели исследование потребления сладких газированных напитков в Англии и получили следующее уравнение регрессии:

$$\text{soda}_i = 2.5 \text{ stress}_i + 0.5 \text{ age}_i - 0.01 \text{ age}_i^2 + 0.3 \text{ divorce}_i - 0.05 \times \text{child}_i + 0.3 \text{ divorce}_i \times \text{child}_i + 8.3$$

где soda_i — потребление сладких газированных напитков в литрах в год;

stress_i — уровень стресса индивида от 0 до 10;

age_i — возраст индивида;

divorce_i — 1, если индивид в разводе, иначе 0;

child_i — 1, если у индивида есть дети, иначе 0.

Вас интересует разница в потреблении сладких газированных напитков у двух групп людей одного возраста, находящихся в разводе, однако у людей в группе 1 есть дети и уровень стресса выше на 2 балла по сравнению с людьми из группы 2. У людей в группе 2 детей нет.

В ответе укажите разницу в среднем объеме потребляемых за год напитков между первой и второй группами (в литрах). Если ответ представлен в виде десятичной дроби, запишите его через точку или запятую с округлением до двух знаков после запятой. Например, если средний объем потребления напитков в первой группе — 15, а во второй — 20, в ответе необходимо указать -5.

Ответ: 5,25

14. (5 баллов) Вы хотите исследовать взаимосвязь между неравенством доходов и количеством краж в населенных пунктах Ленинградской области в 2021 году. Предположим, что вы можете получить любые из перечисленных показателей моментально и в полном объеме. Какие наборы показателей могли бы использоваться? Выберите все подходящие комбинации.

(Индекс Робин Гуда (Robin Hood index), также известный как индекс Гувера (Hoover index), — это показатель неравенства по доходам, равен доле дохода общества, которую необходимо перераспределить для достижения равенства.

Коэффициент Джини (индекс Джини, или «индекс справедливости») — показатель, отражающий степень неравенства в распределении доходов внутри различных групп населения. Рассчитывается на основе информации о том, какую долю от совокупного дохода получает каждая группа населения — от самых бедных до самых богатых. Чем ближе к 1, тем выше уровень неравенства.)

- а. Размах в официальных (белых) зарплатах по населенным пунктам И количество краж на 100 человек за 2021 год по населенным пунктам

- b. Индекс Робин Гуда по населенным пунктам И количество краж за 2021 год по населенным пунктам
- c. Индекс Джини по населенным пунктам И количество краж на 1000 человек по населенным пунктам (+)
- d. Разность между совокупными доходами 10% самых богатых и 10% самых бедных граждан по населенным пунктам И процент нераскрытых краж по населенным пунктам
- e. Среди перечисленного выше нет верных ответов

15. (8 баллов) (Файл «Результаты 2022.xls») В этом файле представлены результаты финала олимпиады DANO в 2022 году. Этот этап разделен на два тура. Задачный тур является индивидуальным (то есть каждому участнику засчитываются набранные им баллы). Он был написан в онлайн-формате. Проектный тур является командным (то есть все участники одной команды, защищавшие проект, получают одинаковое количество баллов). Он проводился очно. Туры проводились в разное время.

Если участник пропустил один из туров, он мог принять участие в другом. Все школьники, принимавшие участие в одном из туров, получили за них ненулевой балл (то есть итоговая оценка 0 за тот или иной тур означает, что школьник отказался от участия в этом туре или был отстранен). В проектном туре от участия могла отказаться как вся команда, так и отдельные ее участники.

Выберите все верные варианты ответа исходя из анализа представленных в файле данных:

- a. Средний балл за задачный тур у школьников из Санкт-Петербурга выше, чем у школьников не из Санкт-Петербурга
 - b. Коэффициент корреляции между оценками команд за постановку задачи и за презентацию — приблизительно 0,726 (+)
 - c. Средний балл за задачный тур у девятиклассников ниже, чем у одиннадцатиклассников (+)
 - d. Среди перечисленного выше нет верных ответов
16. (14 баллов, по 2 за каждое число) (файл defaults_v2.xlsx) По этому файлу делаются задания 16 и 17.

Вам представлен набор данных о наличии у клиента банка проблем с возвратом кредита, а также другая дополнительная информация о клиенте и взятом кредите. Банки часто используют такую информацию для того, чтобы строить модели для прогнозирования кредитоспособности людей, обращающихся за займом. В следующих двух заданиях вам предлагается рассмотреть несколько факторов на предмет их взаимосвязи с вероятностью допущения неплатежей по кредиту (вероятностью дефолта).

Рассчитайте некоторые значения показателей, необходимые вам для выполнения следующего задания. Если ответ представлен в виде десятичной дроби, запишите его через точку или запятую с округлением до двух знаков после запятой. Единицы измерения указывать не нужно.

Медианное значение LTI = 3,29 (Ответ: 328,93)

Доля дефолтов в группе с LTI ниже медианы = 8,05 (Ответ: 0,08)

Доля дефолтов в группе с LTI выше медианы = 7,82 (Ответ: 0,08)

Доля дефолтов в группе с автомобилями = 7,07 (Ответ: 0,07)

Доля дефолтов в группе без автомобилей = 8,38 (Ответ: 0,08)

Доля дефолтов в группе с недвижимостью = 7,89 (Ответ: 0,08)

Доля дефолтов в группе без недвижимости = 8,03 (Ответ: 0,08)

17. (9 баллов, по 1 за каждый выбор из вариантов) Продолжение задания 16.

Далее для выполнения задания считайте, что если доли дефолтов в двух интересующих вас группах различаются на 1 процентный пункт и более, то отличие считается значимым. Если разница меньше, то считается, что доли равны.

Вам необходимо выбрать подходящий вариант заполнения пропуска, чтобы получилось верное описание гипотез и выводов из расчетов, сделанных ранее.

Первая гипотеза состоит в том, что отношение объема взятого кредита к доходу (Loan to Income, LTI) (положительно / отрицательно / никак не) взаимосвязан с шансом допустить по нему дефолт. Поделив выборку по этому показателю на две равные части, мы можем посчитать долю дефолтов в группе с высоким LTI и с низким. В результате оказывается, что доля дефолтов в группе с высоким LTI (не отличается от / больше / меньше) доли дефолтов в группе с высоким LTI, то есть гипотеза (отвергается/не отвергается).

Вторая гипотеза состоит в том, что владение автомобилем, на который сейчас не оформлен кредит, (снижает / увеличивает / не влияет на) вероятность того, что человек будет иметь проблемы с возвратом кредитов. Доля дефолтов в группе обладателей автомобилей (не отличается от / больше / меньше) доли дефолтов в группе людей без машины. Таким образом, гипотеза (отвергается / не отвергается).

Третья гипотеза состоит в том, что владение недвижимостью, на которую сейчас не оформлен кредит, (снижает / увеличивает / не влияет на) вероятность того, что человек будет иметь проблемы с возвратом кредитов. Доля дефолтов в группе собственников недвижимости (не отличается от / больше / меньше) доли дефолтов в группе людей, которые не владеют квартирами. Таким образом, гипотеза (отвергается / не отвергается).