

ВТОРОЙ ОТБОРОЧНЫЙ ЭТАП, 2 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 треть — пользователи Android, четверть — пользователи браузерной версии, а остальные — пользователи iOS. При этом после теста выяснилось, что из пользователей Android кнопкой чата воспользовались 70%, среди пользователей браузерной версии — 90%, а среди пользователей iOS — всего 55%. (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 кинотеатра. Заняться проведением наблюдений было решено утром в будни. Он тщательно пересчитал всех посетителей кинотеатра в понедельник и среду. За время наблюдений получилось, что в среднем его 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) Один маркетплейс в конце 2023 года решил объявить список из 8 своих самых продаваемых товаров за год. Каждый из аналитиков маркетплейса, 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 3 до 7 товаров, глава аналитики угадал все товары, стажер угадал лишь 1 товар. А остальные угадали ровно по 2 товара, и среди них не оказалось тех, кто угадал один и тот же набор из 2 товаров, а также тех, кто угадал бы товар, угаданный стажером.

Какое максимальное число аналитиков может работать в этом маркетплейсе?

- a. 46 (+)
 - b. 48
 - c. 60
 - d. 62
4. (5 баллов) Выборка состоит из 3047 различных чисел, причем все числа положительные и представимы в виде конечной десятичной дроби. Все числа из выборки упорядочили по возрастанию и разделили на 11 равных по величине групп, которые пронумеровали от 1 до 11. То есть первые 277 чисел упорядоченной выборки оказались в группе 1, следующие 277 чисел — в группе 2 и так далее. Выберите все верные утверждения:
- a. Среднее арифметическое чисел в группе 9 меньше среднего арифметического чисел в группе 10 (+)
 - b. Медиана выборки может оказаться не в группе 6
 - c. Сумма двух чисел из группы 1 не может оказаться больше числа из группы 11
 - d. Если одно число меньше другого в 11 раз, то они могут оказаться в одной группе (+)
 - 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	12
Немного улучшилась	8
Не изменилась	16
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) Даня устал от подсчета среднего арифметического и медианы — и решил придумать свое среднее, которое назвал «среднее Данино». Оно считается так: если имеется выборка из N различных чисел, то она сначала упорядочивается либо по возрастанию, либо 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:

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

Даня решил протестировать работу придуманного им среднего и взял какую-то выборку из 16 различных натуральных чисел a_1, a_2, \dots, a_{16} , каждое от 1 до 100. Он упорядочил выборку либо по возрастанию, либо по убыванию, затем посчитал среднее Данино и медиану выборки — и сравнил их. Медиана в случае четного количества чисел равна полусумме чисел в середине

упорядоченной выборки.

Какие результаты могли получиться у Дани?

- a. Если Даня упорядочил выборку по возрастанию, то медиана могла оказаться больше среднего Даниного
 - b. Если Даня упорядочил выборку по возрастанию, то медиана могла оказаться меньше среднего Даниного (+)
 - c. Если Даня упорядочил выборку по убыванию, то медиана могла оказаться больше среднего Даниного (+)
 - d. Если Даня упорядочил выборку по убыванию, то медиана могла оказаться меньше среднего Даниного
 - 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.)	8	10	10	20	30	60	100	170	600

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 размеры субсидий через запятую в порядке, соответствующем порядку жителей, которые их получают. Если ответ представлен в виде дроби, округлите его до целого. Единицы измерения указывать не нужно.

Ответ: 90, 60, 20

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: 170

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:

- а. За апрель 2019 общий объем кредитов вырос примерно на 1,25%

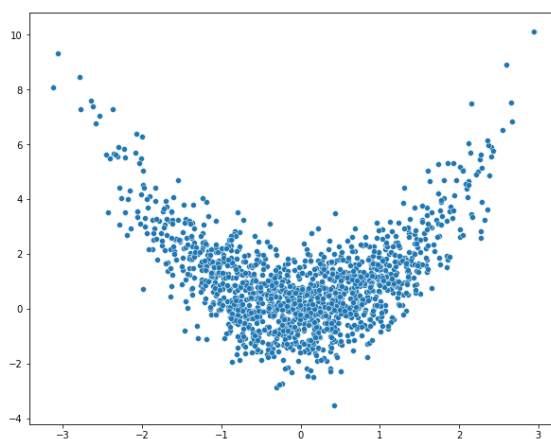
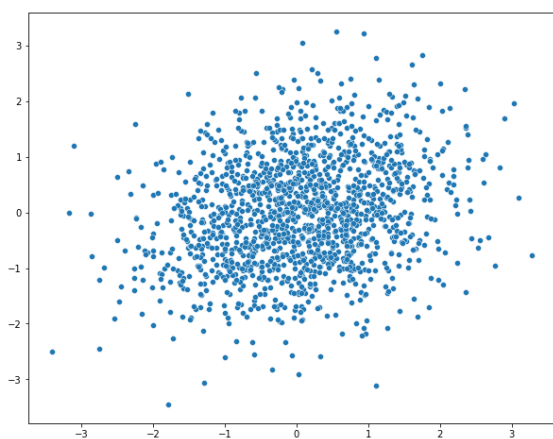
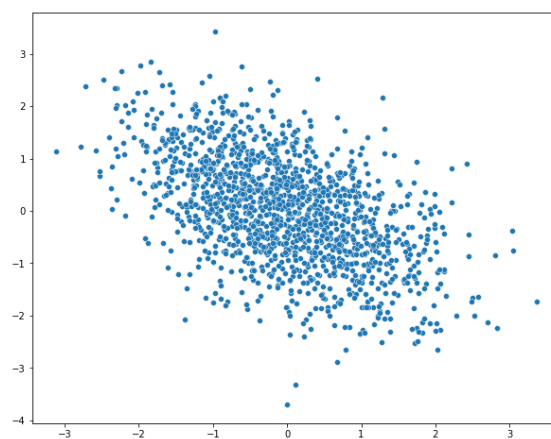
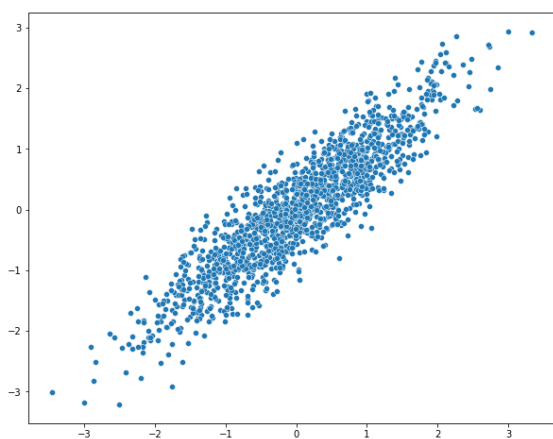
b. Кредиты в прочих видах деятельности в среднем растут в феврале — марте 2019 (+)

c. На два вида экономической деятельности с самым большим объемом задолженности в указанный период приходилось более 40% всей задолженности в экономике (+)

d. Доля задолженности компаний транспорта и связи в общем объеме снизилась 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.



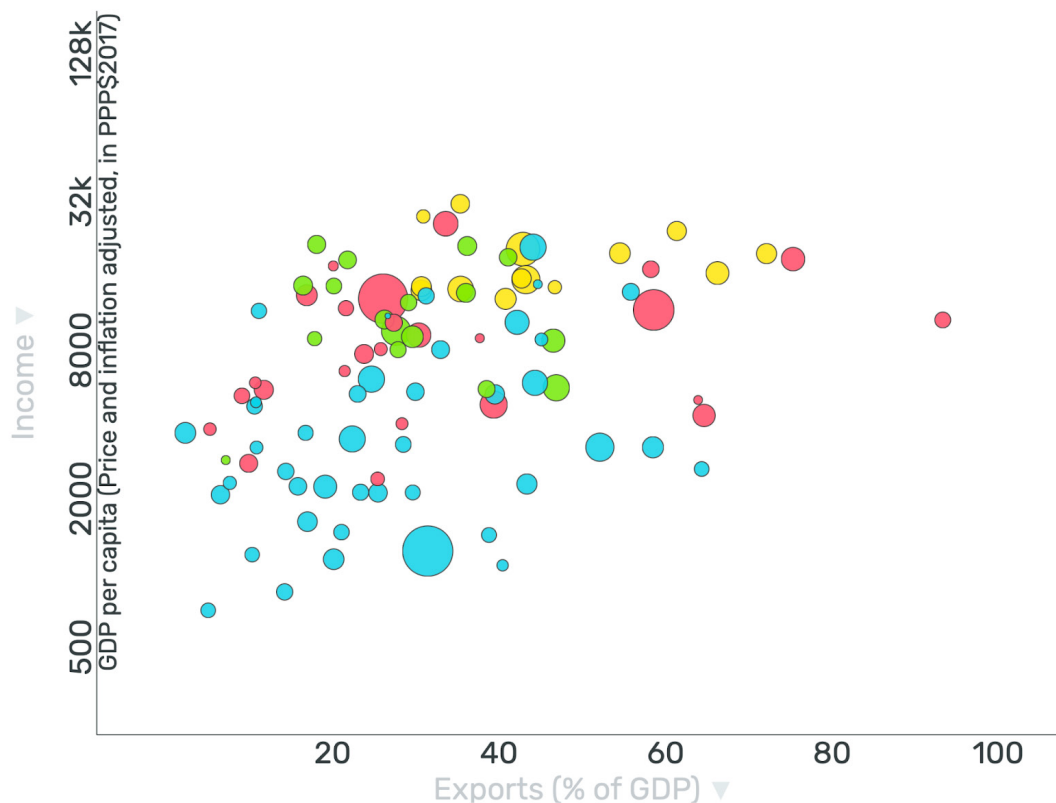
a. 0,9

b. -0,5

c. 0,2

d. 0

10. (5 баллов) На рисунке изображена диаграмма рассеяния, показывающая взаимосвязь между доходами населения и долей экспорта в ВВП 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.

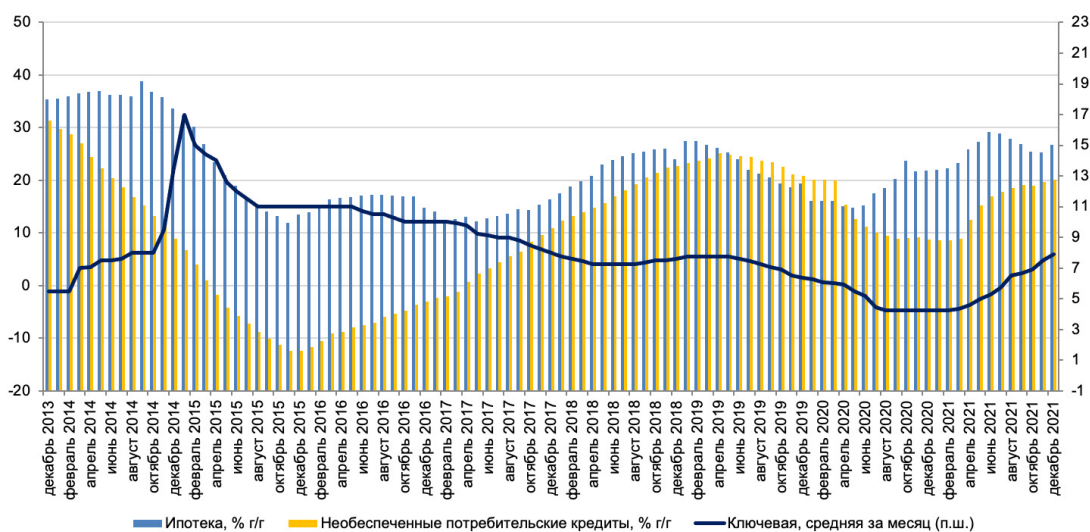
Размер кругов отображает третий показатель — отношение совокупного внешнего долга страны к валовому национальному доходу — и отражает задолженность страны (государства и частных лиц) перед нерезидентами (иностранными государствами, фирмами и людьми). Чем больше размер круга, тем больше задолженность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

- a. В среднем доход на душу населения в странах Европы ниже, чем в странах Африки
- b. В среднем отношение экспорта к ВВП в странах Европы выше, чем в странах Африки (+)
- c. Разброс доходов на душу населения в странах Европы ниже, чем в странах Азии (+)
- d. Три страны с наибольшим отношением задолженности иностранному сектору к доходу — страны Европы и Азии
- 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:

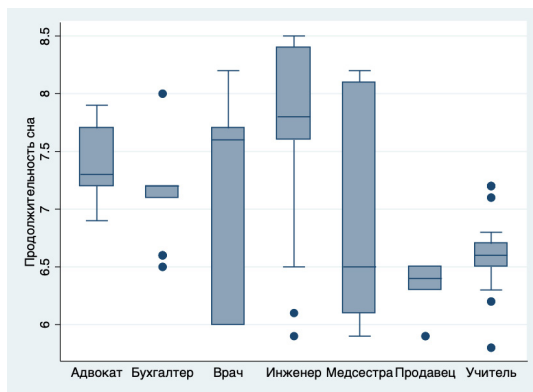
- a. Темпы роста ипотечного кредитования в среднем более стабильны, чем темпы роста необеспеченного кредитования (+)
- b. С октября 2014 по октябрь 2015 объем ипотечного кредитования снижался
- c. В течение 2017 года темп роста необеспеченного потребительского кредитования увеличивался (+)

d. В первой половине 2015 года наблюдается отрицательная взаимосвязь между темпами 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?

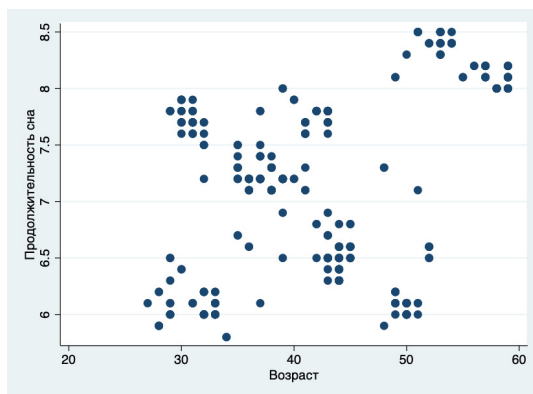
a) +



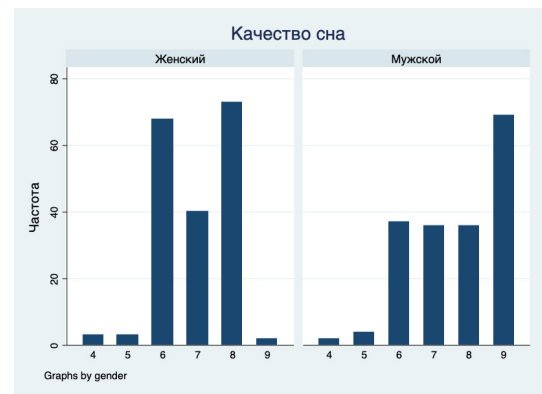
b)



c) +



d)



e) Среди перечисленного выше нет верных ответов

13. (6 баллов) Вы провели исследование успеваемости старшекурсников в вузах Франции и получили следующее уравнение регрессии:

$$gpa_i = 0.075 \text{ study}_i + 0.3 \text{ female}_i - 0.5 \text{ work}_i + 0.05 \text{ female}_i \times \text{study}_i - 0.03 \text{ work}_i \times \text{female}_i + 4.7$$

где gpa — средневзвешенный балл;

$study$ — среднее количество часов, посвященных учебе, в день;

$work$ — 1, если студент работает, иначе 0;

$female$ — 1, если студент женского пола, иначе 0.

Вас интересует разница в средних баллах у работающих студентов (группа 1) и неработающих студенток (группа 2). Из данных вы знаете, что и те и другие уделяют учебе в среднем 8 часов в день.

В ответе укажите разницу в среднем балле между первой и второй группами (в баллах). Если ответ представлен в виде десятичной дроби, запишите его через точку или запятую с округлением до двух знаков после запятой. Например, если средний балл первой группы — 15, а второй — 20, в ответе необходимо указать -5.

Ответ: -1,2

14. (5 баллов) Вы хотите исследовать взаимосвязь между количеством аварий на дорогах в регионах Франции и уровнем коррупции. Какие показатели могли бы использоваться для проведения исследования? Выберите все подходящие комбинации.

(Индекс восприятия коррупции - показатель, отражающий оценку уровня восприятий коррупции экспертами и предпринимателями по шкале от 0 (максимальный уровень коррупции) до 100 (минимальный уровень коррупции).)

- Количество ДТП в регионе И количество уголовных дел в регионе, возбужденных по статьям, связанным с коррупцией
- Количество уголовных дел в регионе, открытых вследствие ДТП, И средний размер взятки в регионе
- Количество ДТП на 1000 автомобилей в регионе И индекс восприятия коррупции в регионе (+)
- Количество погибших на 1000 ДТП в регионе И количество обвинительных приговоров в делах о коррупции в регионе
- Среди перечисленного выше нет верных ответов

15. (8 баллов) (Файл «Результаты 2022.xls») В файле с данными 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 него 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. А) Средний балл за задачный тур у школьников из Москвы выше, чем средний балл у школьников из других регионов (+)
 - b. Б) Коэффициент корреляции между оценками команд за постановку задачи и за аналитический блок — приблизительно 0,831 (+)
 - c. В) Средний балл за задачный тур у десятиклассников не ниже, чем у одиннадцатиклассников
 - d. Г) 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 = 3,26 (Ответ: 326,09)

Доля дефолтов в группе с LTI ниже медианы = 8,16 (Ответ: 0,08)

Доля дефолтов в группе с LTI выше медианы = 8,13 (Ответ: 0,08)

Доля дефолтов в группе с автомобилями = 7,36 (Ответ: 0,07)

Доля дефолтов в группе без автомобилей = 8,55 (Ответ: 0,09)

Доля дефолтов в группе с недвижимостью = 7,97 (Ответ: 0,08)

Доля дефолтов в группе без недвижимости = 8,55 (Ответ: 0,09)

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).