



3101679375411

## Титульный лист

Направление ☒ анализ данных ☐ информатика ☐ история  
☐ математика ☐ обществознание ☐ русский язык  
☐ физика ☐ химия

Класс ☐ 8 ☐ 9 ☐ 10 ☒ 11

Фамилия

КУТЬИНА

Имя

ВИКТОРИЯ

Отчество

ВИКТОРОВНА

Дата рождения

20 12 2007

Город участия

ЕКАТЕРИНБУРГ

Аудитория

И5278

Телефон

+79530583910

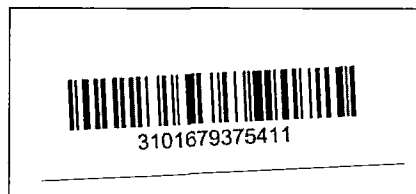
Дата

01 02 2025

Подпись

Пример  
заполнения

А Б В Г Д Е Ж З И Й К Л М Н О П Р С Т У Ф  
Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я 1 2 3 4 5 6 7 8 9 0



## Проверочный лист

Заполняется участниками

Направление ☒ анализ данных ☐ информатика ☐ история  
☐ математика ☐ обществознание ☐ русский язык  
☐ физика ☐ химия

Класс ☐ 8 ☐ 9 ☐ 10 ☒ 11

Город участия Е К А Т Е Р И Н Б У Р Г

Заполняется организаторами

Количество доп. листов 0 4 Количество черновиков к проверке 0 0

Время выхода с до

## Протокол проверки

Заполняется жюри

| Номер задания      | 1 | 2  | 3  | 4  | 5 | 6  | 7 | 8 | 9 | 10 |
|--------------------|---|----|----|----|---|----|---|---|---|----|
| Балл члена жюри №1 | 0 | 12 | 16 | 15 | 0 | 25 |   |   |   |    |
| Балл члена жюри №2 | 0 | 12 | 16 | 15 | 0 | 25 |   |   |   |    |

Итоговый балл 68

Подпись  
члена жюри №1

Подпись  
члена жюри №2

Пример  
заполнения

А Б В Г Д Е Ж З И Й К Л М Н О П Р С Т У Ф  
Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я 1 2 3 4 5 6 7 8 9 0



Задание 1 0,5

1 При плохой погоде жители Свердловской области используют общественный транспорт чаще, чем при хорошей погоде

2 Неблагоприятные погодные условия на улице (низкая температура, дождь и т.п.) → ~~не~~ человеку не комфортно в таких условиях добираться пешком → возрастают альтернативные издержки перемещения не на <sup>общественном</sup> транспорте (т.е. пешком) → выгоднее оплатить билет на один из видов общественного транспорта → больше людей отдаёт предпочтение общественному транспорту

3 Противоположная гипотеза при плохой погоде жители Свердловской области реже пользуются общественным транспортом, чем при хорошей погоде

Механизм неблагоприятных погодных условий на улице → человек ~~выходит~~ из дома ~~только по неотложным делам, поскольку~~ ~~возрастают~~ и издержки нахождения на улице (мокрая одежда, ощущение холода и т.п.) → человек <sup>(моду)</sup> выходит из дома только по неотложным делам → меньше людей реже и в меньшем количестве выходят на улицу → ~~меньше~~ меньше людей используют обществен- 1

вещный транспорт (из-за отсутствия  
необходимости куда-либо ехать)

Альтернативный механизм

на улице неблагоприятные погодные условия

→ растут издержки нахождение на улице →  
модн у которых есть много транспортное  
средство (автомобиль) предпочитают его об-  
щественному транспорту, поскольку при пере-  
звешении на авто минимизируется время  
нахождение на улице, поскольку есть воз-  
можность добраться сразу "от точки до  
точки"

## Задание 2

1 Фокусом исследования стали столбцы

G - кол-во баллов ЕГЭ по математике

I - кол-во баллов ЕГЭ по биологии + +

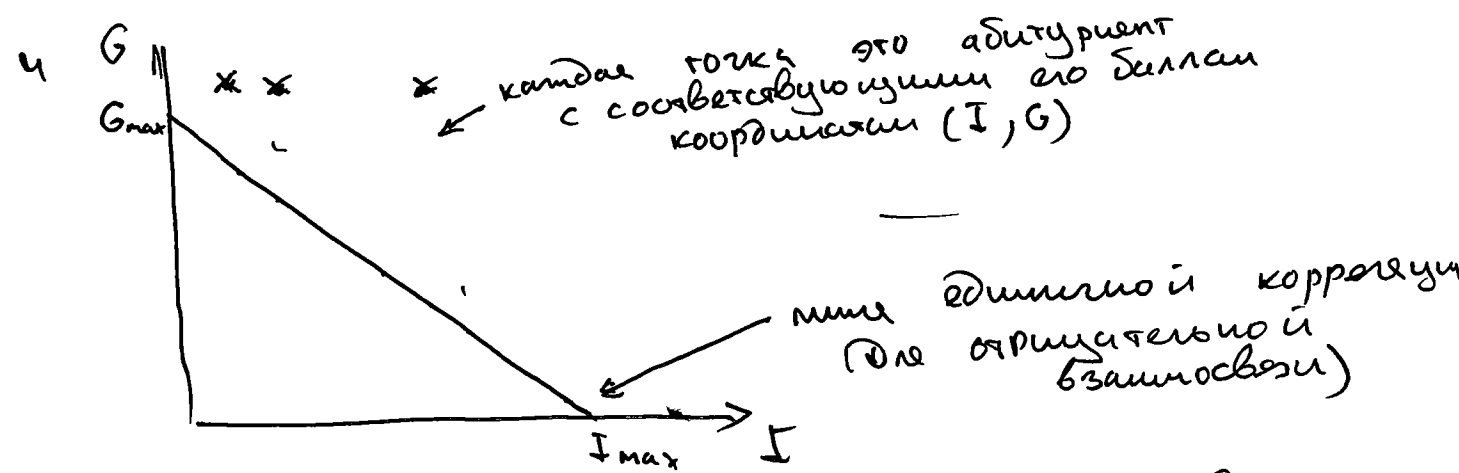
2 В выборку исследования попали 1199  
абитуриентов (поскольку корреляция проверяется  
только для строк с 201200 => 1199 абитуриентов)

3 взаимосвязь отрицательная, поскольку у  
значения корреляции стоит знак "-" +

взаимосвязь отсутствует, т.е. она не статисти-  
чески значима, поскольку ~~результат~~ получен-  
ный результат по модулю ниже статистически  
значимого значения  
о нем можно сказать, что она очень  
слабая

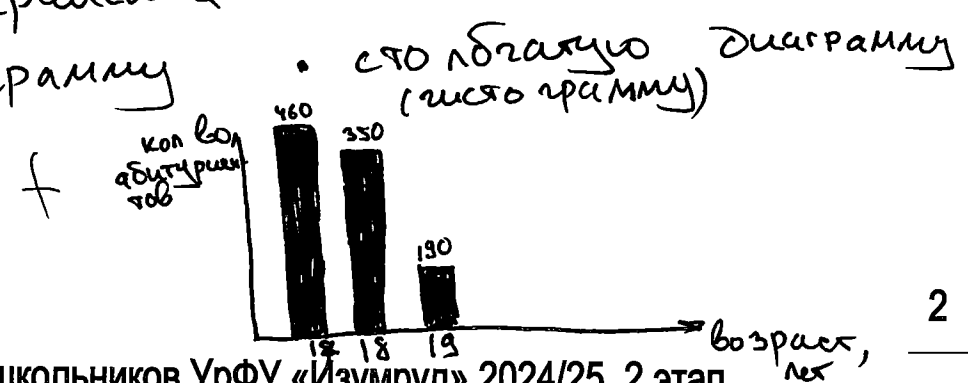
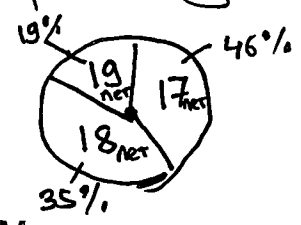
линия отреза

4



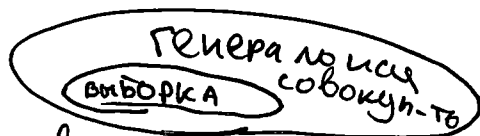
$I_{max}, G_{max}$  - максимально возможное кол-во баллов за экзамен

5 Поскольку абитуриенты, поступающие в один год обычно имеют возрастной разбег максимум 1-2 года (безусловно бывают и исключения в виде зачисляющихся школу экстерном или взрослые, решившие получить высшее образование не сразу после школы), поэтому наиболее подходящими диаграммами я назвала бы - **круговую диаграмму**



Эти диаграммы позволяют наглядно показать распределение данных для переменных с небольшой вариативностью (в нашем случае это 3 различных значения)

### Задание 3

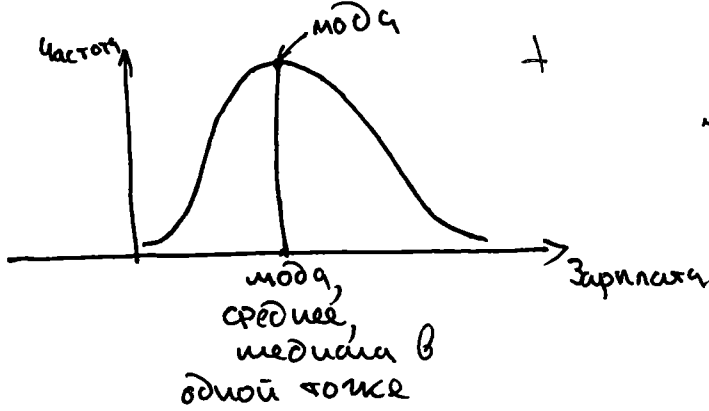


- 1 эта выборка нерепрезентативна, она будет смещена из-за уже сформированной лояльности к аптеке "ТюменьЗдрав" +
- 2 эта выборка не нерепрезентативна, поскольку она будет смещена. Так у нас не будет тех, кто являлся посетителем "ТюменьЗдрав" => мы не сможем в полной мере проанализировать замечаемые модели различий в нашей и других аптеках +
- 3 эта выборка будет репрезентативна, поскольку опрашиваемые выбраны из генеральной совокупности случайным образом +
- 4 выборка будет нерепрезентативна, поскольку генеральная совокупность - пенсионеры Тюмени, а в этой выборке будут присутствовать люди, выходящие за рамки генеральной совокупности, то есть подростки +
- 5 выборка будет нерепрезентативна, поскольку здесь будет смещение из-за того, что для всех опрашиваемых будет единым критерий местоположения, это делает невозможность выводов по всей генеральной совокупности +
- 6 выборка будет нерепрезентативна, поскольку люди, имеющие проблемы со здоровьем имеют повышенный спрос на аптеки, это даст смещение выборки, так будут представлены не все подгруппы, входящие в генеральную совокупность +
- 7 выборка будет нерепрезентативна, поскольку в ней не будут представлены генеральной совокупности, поэтому нельзя будет по ней сделать корректных и практически применимых выводов +
- 8 выборка будет не репрезентативна, поскольку в ней не будут представлены генеральной совокупности (пенсионеров Тюмени), что делает ее неподходящей для качественного анализа.

Задача

Задача 4

1

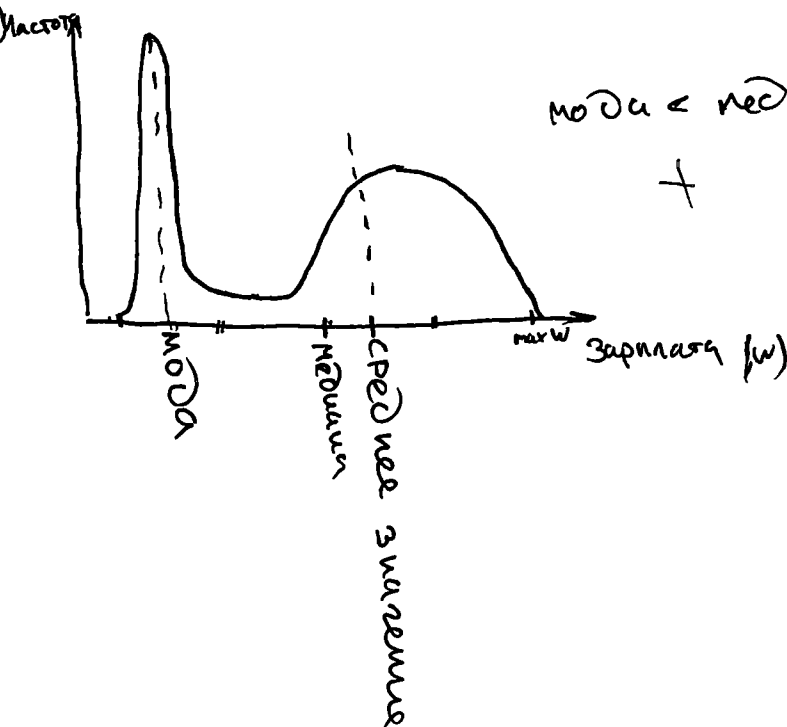


мода - наиболее часто встречающийся вариант  $\Rightarrow$  с наибольшей частотой

• среднее  $\frac{\sum_{i=1}^n x_i}{n}$  (сумма всех  $x_i$ , деленная на кол-во наблюдений), поскольку нормальное распределение симметрично среднее будет находиться в той же точке, что и мода, поскольку для нормального

• медиана - значение, находящееся в середине проранжированной в порядке возрастания выборки (если число наблюдений нечетное берется просто среднее в середине выборки число, если не число наблюдений четное, то берется среднее между двумя числами, стоящими в середине проранжированной выборки), поскольку нормальное распределение симметрично относительно своего центра медиана будет совпадать с модой и средним

2



мода < медиана < среднее





# Репозиторийный Бланк ответов 1

## Задание 5

1 Наша идея заключается в том, что на количество авиаперелетов в большей степени могут влиять другие факторы, помимо фактора дохода, например, должность на работе (подразумевающая большое количество командировок), к тому же мы можем говорить лишь о том, что "чем больше у человека доход, тем большее число авиаперелетов он совершит". Мы не можем говорить об изменении поведения <sup>(в сторону большего числа авиаперелетов)</sup> ~~наблюдается~~ <sup>наблюдается</sup> лишь тогда, когда при росте доходов, поскольку не только доход, как я уже и говорила, влияют на количество авиаперелетов.

К тому же если мы можем максимум сделать предположение о том, что при росте доходов у человека его спрос на авиаперевозки растет (при этом мы не можем утверждать, ~~опираясь~~ <sup>опираясь</sup> ~~на~~ <sup>на</sup> ~~этом~~ <sup>этом</sup> ~~основании~~ <sup>основании</sup> ~~вырастает~~ <sup>вырастает</sup> и предположение, ~~вырастает~~ <sup>удовлетворяющее</sup> этот возраст-~~интенсивность~~ <sup>интенсивность</sup> спроса, поскольку мы исследуем взаимосвязи в поведении отдельного человека, а не взаимосвязи ~~между~~ <sup>между</sup> в поведении нескольких агентов).

2 Размер дохода положительно влияет на количество авиаперелетов.

3 Можно было бы построить распределение, с целью изменения дохода для отдельного человека (например, изменение дохода в 2024 году в сравнении с 2023).

по одной оси и дельта кол-ва абисперен  
(разница между 2024 и 2023 годами соответс-  
но) И при наших статистически значи-  
мой положительной взаимосвязи между этими  
переменными поспорить с выводом Паши  
было бы нельзя (можно было бы найти аль-  
тернативные механизмы, но сам вывод будет  
в таком случае ~~он~~ оставаться корректным)

Задание 6 <sup>250</sup>

1 Исследование фактора плодородия <sup>(урожайности)</sup> почв в  
зависимости от культуры предшественника

Гипотеза: после выращивания зернобобовых ~~пло-~~  
~~дородие почв~~ урожайность культур увеличи-  
вается в большей степени, нежели после  
других предшественных культур

Механизм на поле выращиваются зернобо-  
бовые → они насыщают землю микроэlemen-  
тами (увеличение содержания гумуса, фосфора,  
кальция и т.п.) \* → урожайность последующих  
культур увеличивается

\* согласно учебникам по почвоведению и расте-  
ниеоводству

Это исследование может проводиться в  
рамках агропромышленных комплексов разны-  
х с/х компаний В моем случае, компани-  
ей, запросившей исследование, яв-ся „Эколива“-  
группа компаний в сфере сельского хозяйства  
(Продолжение на доп. бланке 2)

## Дополнительный бланк ответов к

Это исследование может быть полезно для выстраивания севооборота, увеличивающего урожайности отдельных культур за счет правильно высаженной культуры предшественника (а увеличение урожайности в свою очередь влечет рост прибыли компании). Помимо этого, поскольку Эконива является крупной холдинговой и одним из ее холдинговых это огромное производство, то увеличение урожайности культур также увеличивает производство кормов для коров (это при нацеленности на наращивание производства, яв-я хорошим фактором).

2. Генеральная совокупность - все с/х угодья принадлежащие компании "Эконива" и используемые для выращивания с/х культур (для полного анализа возможности компании в рамках принадлежа ей территории).

Географические рамки будут определяться географическими рамками компании. В моем случае это зона центрального Черноземья Воронежская обл. и др., зона критического земледелия Санкт Петербург, Московская область, также Новосибирская, Оренбургская область и оставшиеся регионы в которых у компании есть сельскохозяйственные угодья.

Временные рамки: данные компании за последние 3 года, поскольку только 3 года назад была введена единая система данных о полях и культурах (в идеале конечно нужно было бы данные за последние 5-10 лет для выстраивания более точной схемы севооборота).

по это в данном случае невозможно с  
силу технических причин, описанных ранее)

Взаимосвязь будет проверяться на основании  
всей генеральной совокупности в силу ее  
не столь больших размеров, путем деления  
на подгруппы ~~по~~ (зерновые, зернобобовые, масличные,  
технические и т.д.), а ~~установленного~~ факта  
такие подгруппы по культурам предшествен-  
никам (аналогичные подгруппы только по значению  
культуры предшественника)

Размеры подгрупп варьируются от 500 до 2000  
полей в подборке в зависимости от куль-  
туры (подвыборки не балансируются по раз-  
меру, поскольку исследование основано на средних  
значениях урожайности и к тому же это сде-  
лано для того, чтобы при проверке гипотезы  
на устойчивость внутри дополнительных под-  
групп (по регионам, например) размер подвы-  
борки не оказался слишком маленьким для  
проведения анализа (меньше 20 наблюдений, кол-  
во варьируется в зависимости от выбранного  
способа проверки на устойчивость)

В данном исследовании из-за небольшого  
числа наблюдений проведение исследова-  
ния на всей генеральной совокупности будет  
репрезентативным, поскольку мы хотим  
проанализировать различные факторы влия-  
ющие на севооборот именно внутри компа-  
нии

(см продолжение на 201 бланке 3)

3. Основной источник данных - дача с данными "экоива" предоставленный для олимпиады по анализу данных "DANO", содержащий в себе данные для всех полей компании за 3 года, содержащее такие переменные как урожайность на ~~закосе~~, название культуры, название культуры - предшественника, регион произрастания культуры и т.д. (Данные о полях и работах на них)

Получить этого будет полезно использовать данные о погоде ~~взвешивать~~ для каждого из полей (но это затрудняется тем, что в даче указаны лишь регион и нет данных о координатах, что не позволяет установить точные погодные условия). Также в даче не были указаны названия удобрений и их состав, что не дает нам возможность проанализировать влияние удобрений на урожайность (покупание этих данных делало бы исследование более точным и детальным, затрачивая большее число факторов, влияющих на урожайность)

На мой взгляд может влиять не только культура предшественник, но и химический состав почвы, погодные условия, внесенные в почву удобрения и т.д. Так например правильно внесение удобрений

не могут увеличить урожайность куль-  
турные природные условия могут по-  
низить урожайность с/х культуры (и это  
будет затеено как выброс при исследовании,  
поскольку условия аномальные)

4<sup>+</sup> В исследовании используются как  
количественные переменные урожайность из  
гектар и т.д., так и качественные пе-  
ременные название культуры, название куль-  
туры-предшественника, регион произрастания  
культуры

Для визуализации результатов <sup>были (будут)</sup> ~~будут~~ <sup>исполь-</sup>  
зованы столбчатые диаграммы, показывающие  
распределение культур по регионам и  
культурам предшественникам, также будут  
использованы "язычки с усами" для демонстра-  
ции структуры данных до и после  
очистки от выбросов

Для ответа на главный исследователь-  
ский вопрос будет построена линейная  
регрессия с объясняющими переменными  
были плодородие почв, кол-во внесенных удоб-  
рений и контрольной переменной  
флаг является культура зернобобовая или  
нет

Для проверки устойчивости гипотезы  
(см. продолжение на той же странице 4)

## Репозиторийный блок 4

будет также построена линейная регрессия для подвыборок по регионам и для каждой отдельной группы <sup>текущих</sup> культур (зерновые, зернобобовые и т.д.) с большим числом объясняющих переменных (проверенных на отсутствие мультиколлинеарности)

5<sup>1</sup> В процессе исследования (оно мной и моей командой уже было проведено в рамках олимпиады DANO) мы столкнулись с такими трудностями, как

- отсутствие названия и хим. состава удобрений
- отсутствие данных о погоде и координатах поля (которые могли бы нам позволить самостоятельно привязать данные к погоде)
- наблюдение данных только за 3 года, это затрудняет желание выводов касательно долгосрочного севооборота

→



