



3101760120700

Титульный лист

Направление

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

Класс

- | | | | |
|----------------------------|----------------------------|-----------------------------|--|
| ☐ 8 | ☐ 9 | ☐ 10 | ☒ 11 |
|----------------------------|----------------------------|-----------------------------|--|

Фамилия

ЛАПТЕВА

Имя

МИЛАНА

Отчество

ВИТАЛИЕВНА

Дата рождения

17 05 2007

Город участия

КРАСНОЯРСК

Аудитория

1-20

Телефон

9969145822

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



3101760120700

Проверочный лист

Заполняется участниками

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

Класс ☐ 8 ☐ 9 ☐ 10 ☒ 11

Город участия К Р А С Н О Я Р С К

Заполняется организаторами

Количество доп. листов 0 1 Количество черновиков к проверке

Время выхода с до

Протокол проверки

Заполняется жюри

Номер задания	1	2	3	4	5	6	7	8	9	10
Балл члена жюри №1	14	9	6	0	0	25				
Балл члена жюри №2	14	9	6	0	0	25				

Итоговый балл 54

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

- 1) Гипотеза В ясную погоду чаще пользуются общественным транспортом (характеристики ясной погоды: солнечно, ветер до 4 км/час (безветренно тоже уместно), облаков нет) +
(Другая гипотеза в дождливую погоду реже пользуются общественным транспортом, предпочитая передвижение на такси)
- 2) Механизм гипотезы 1) В ясную погоду легче дойти до остановки; 2) В ясную погоду легче догнать нужный маршрут; 3) В ясную погоду больше хочется выйти на природу, что находится за городом; 4) В ясную погоду не обязательно брать такси, можно выбрать менее дорогой вариант передвижения (данные аргументы уместны для той части населения, что имеет какой-либо основный вид деятельности (учеба, работа и т.д.) и при этом не имеет своего транспортного средства
- 3) Противоположная гипотеза В ясную погоду реже пользуются общественным транспортом
Механизм противоположной гипотезы: 1) В ясную погоду люди чаще выбирают прогулки пешком; 2) В ясную погоду люди чаще предпочитают заниматься домашними делами, такими как прополка огорода, уход за цветами, посадка овощей, цветов и фруктов и тому подобное; 3) В ясную погоду чаще занимаются спортом в придомовых территориях; 4) В ясную погоду люди гуляют с домашними питомцами на улице, вокруг своего района

Задание 2

1) Давайте укажем какой столбец за что отвечает

A - порядковый номер абитуриента

B - имя

C - фамилия

D - город

E - возраст

F - пол

G - кол-во баллов ЕГЭ по математике

H - кол-во баллов ЕГЭ по русскому языку

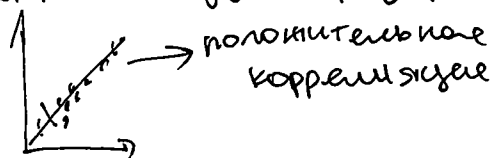
I - кол-во баллов ЕГЭ по биологии

Продолжение задания 2

1) кол-во баллов ЕГЭ по математике и кол-во баллов по биологии +

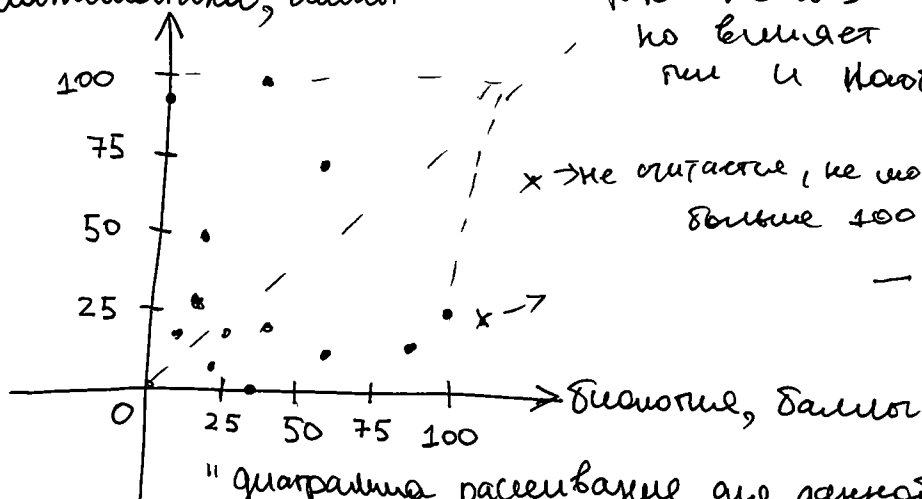
2) Учитывая, что первые столбцы содержат информацию про переменные, обозначая их, то 1-ый студент шел со второй строки выборка состояла из $(1200 - 2 + 4)^+ 1199$ абитуриентов

3) Заметим, что значение отрицательное когда это могло произойти и как можно сделать следующий вывод о взаимосвязи между баллами ЕГЭ по математике и по биологии корреляция этих двух предметов отрицательная и при этом значение $-0,02$ — это отрицательная корреляция. Данные идут в разброс, максимально далеко друг от друга.



→ отрицательная связь

4) математика, баллы



увеличение баллов по математике не обяз-о обозначает положительное влияние на баллы по биологии и наоборот

x не считается, не может быть больше 100 в

"диаграмма рассеивание для данных взаимосвязи"

5) Я считаю, что круговая и столбчатая диаграммы подойдут для того, чтобы показать распределение переменной "возраст" +

• Почему круговая диаграмма? 1) Можно в процентном соотношении показать возраст абитуриентов, 2) Данные, отраженные в круговой диаграмме, легче воспринимать, 3) Подходящая инфографика проста в понимании для широкого спектра людей,

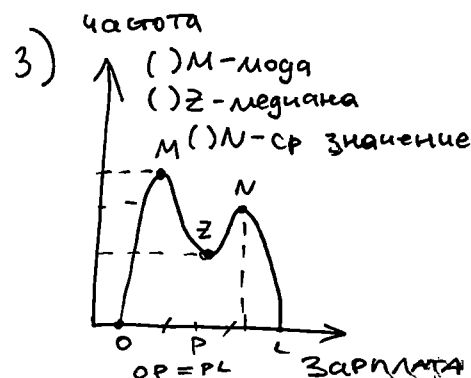
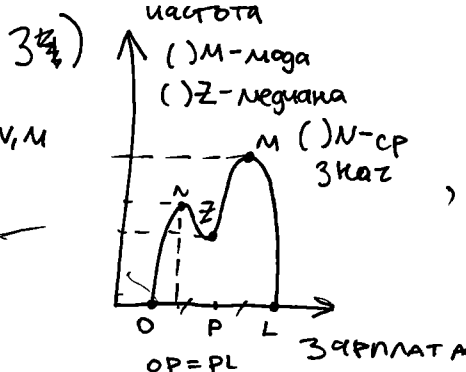
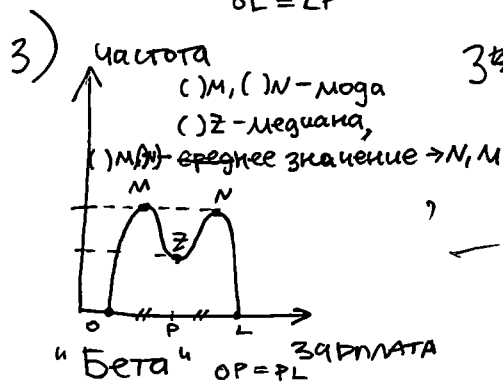
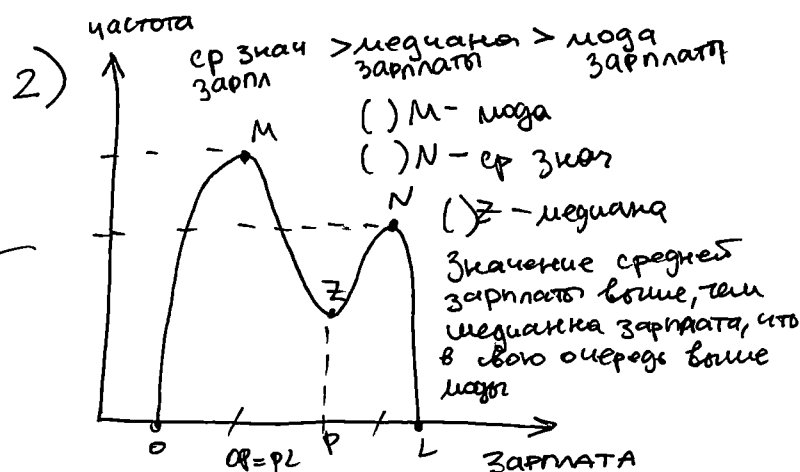
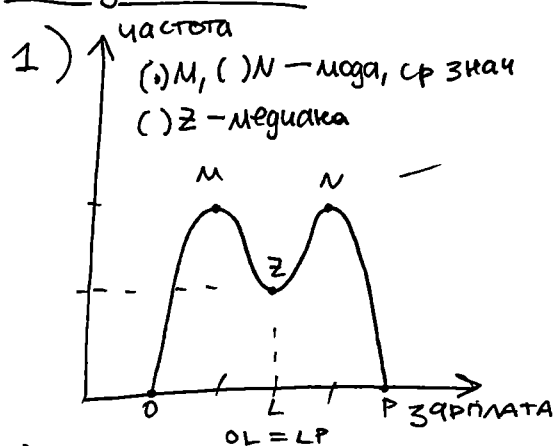
• Почему столбчатая диаграмма? 1) Можно в численном количестве указать какой возраст сколько раз попадается, 2) Простая и понятная инфографика, данные которой легко усваиваются чем выше столбик, тем чаще вариант попадается, тем ниже — тем реже.

Бланк ответов

Задание 3 "Что влияет на выбор аптеки среди пенсионеров Тюмени"

- 1) (Относительно) репрезентативна
- 2) Репрезентативна (мы можем понять какие пункты стоит улучшить, чтобы они выбрали нас)
- 3) Полностью репрезентативна
- 4) Полностью не репрезентативна Никто из Свердловской области не сможет совершить покупку в Тюменской аптеке (очень маловероятно)
- 5) Репрезентативна
- 6) Репрезентативна
- 7) Частично репрезентативна, ведь врачи могут рекомендовать аптеку, но это все равно не целевая аудитория, они такие знают где покупать лекарства
- 8) Не репрезентативна, даже если список покупок составлен пенсионерами, ведь выбор за аптекой все еще на этих детей

Задание 4



- 4) "Не бывает такого распределения, для которого медиана была бы строго меньше моды, а мода - строго меньше среднего значения"
Медиана - среднее значение, потерявшие про сть
среднее значение - среднее арифметическое

$$7 \left(X_{\frac{n}{2}} < X_c < \frac{X_1 + X_2 + \dots + X_n}{n} \right) = ?$$

или продолжение задание 4 (отдельный листок)

Задание 5

1) Ошибка состоит в том, что Паша мог не учесть тех людей, кто вообще не покупает авиабилеты, предпочитая поезд, машину, ^{многие самолеты} и т.д.

Решение называется ошибкой выжившего. Такой феномен имеет место быть, поскольку мы учитываем доход тех, кто уже купил авиабилет, хотя какое-то положительное кол-во раз, совершенно не учитывая тех, кто не покупал авиабилет.

2) Корректный вывод среди тех компаний города, что выбирают авиaperелеты, те, кто имеет более высокую зарплату, может позволить себе большее количество авиaperелетов к покупке.

3) Добавить к исследованию дополнительные переменные для анализа, ~~(кроме)~~ помимо основной переменной интереса.

Можно учитывать туристические, деловые и иные виды поездок, наличие или отсутствие обратного билета, пребывание в точке Б (ее длительность). Можно включить в исследование тот самый феномен, что не покупает авиабилеты, поскольку сам владеет самолетом, вертолетом и т.д.

Можно также учитывать тех людей, кто покупает авиабилеты со скидками.

Предлагая в целом учесть ценную дискриминацию всех типов, тогда лучше описать, а перед этим определить поведение определенного человека.

Ценовая дискриминация 1-ого, 2-ого, 3-его типа + дискриминационные данные с данными о профессии и заработной плате.

(1) Дополнение к 1-ому пункту

Ошибка выжившего заключается в следующем:

- 1) Супербогатые люди могут путешествовать по миру и на бизнес;
 - 2) Как положительно может увеличиться кол-во авиaperелетов;
 - 3) Если человек еще смог достичь МРОТ'а? Никак.
- Следует учитывать к какой категории новаторов а также относит авиaperелеты.

Это явно не первое исследование

В своем исследовании нужно также учитывать как взброс, так и их противоположность.

Задание 6

1) Я хочу провести исследование на тему школьных опозданий. Я предполагаю, что те учащиеся, кто живет ближе к учебному заведению опаздывают реже, чем те, кто живет дальше (гипотеза). Аргументирую это тем, что школьники, кто живет ближе знают, что у них больше в запасе времени перед школой, ведь путь у них короче, значит и быстрее, соответственно, они могут себе позволить подольше поспать. Осознавая то, что у тебя больше времени, чем просто достаточно, можешь позволить себе проблем с тайм-менеджментом. Те же учащиеся, кто живет дальше точно определяют запас своего времени и эффективнее им руководствуются, поскольку у таких учащихся чувство времени и его хода развито сильнее, чем у тех, кто живет ближе.

Результаты исследования могут быть полезны школьникам учащимся и методистам, чтобы можно было выявить наборы собственного распределения времени у школьников, а также самим школьникам, чтобы лучше следить за своей успеваемостью. В этом исследовании стоит учитывать демографические данные пол, возраст, состояние здоровья, регион проживания и т.д., а также стоит учитывать нагрузку в учебном заведении, профань учащегося, расстояние от дома до школы, вид транспорта, которым пользуется ученик, а также то, какой класс он обучается. Поколение со здоровым чувством времени и умениями в распоряжении личного времени будет успешным/более успешным.

2) Генеральная совокупность выборка должна быть обширной, репрезентативной и случайной. Я заинтересована в результатах исследования для 14 регионов Республики Саха (Якутия), поэтому и выберу случайным образом 6000 школьников, чтобы данные исследование были каденными. Я хочу взять 4 возрастные группы на четверти, чтобы после эти данные применить для анализа и сравнение результатов в общем виде.

43) Доступные данные 1) потенциальные источники;
сайт росстат, научные журналы, достоверные источники
сети и т.д., Переменные же данные могут быть целочис-
ленными, дробными, могут быть потоки данных за определен-
ный промежуток времени, а могут также использоваться
запасы мер в опр. промежуток времени

Также эффективными способами собрать данные будет
анкетирование

Вопросы анкеты (примерно)

1) Ваш пол, возраст, пол, возраст, школа,
лицей, гимназия,

2) Расстояние от вашего дома до школы,

3) Время, которое вы тратите на дорогу,

4) Как часто вы опаздываете 1) редко (1 раз в месяц)

5) Высокая ли у вас нагрузка 2) часто (2 раза - 9 раз в неделю)

и т.д. в школе? 1) да

3) очень часто (10+ раз в неделю)

2) затруд. ответить, 4) никогда (один раз в месяц)

3) нет

Также дополнительными параметрами можно считать
мотивацию учащегося посещать учебное заведение и его
интерес к учебе в целом, его будущая профессия и т.д.

4) Типа переменных 1) дифференцированные (линейные шкалы)
2) порядковые шкалы (интервальные)

Типы используемых графиков

1) круговые диаграммы

2) столбчатые диаграммы

3) диаграмма рассеивания

4) линейные графики

5) гистограммы

6) диаграмма вращений

7) диаграмма вращений

8) диаграмма вращений

9) диаграмма вращений

10) диаграмма вращений

11) диаграмма вращений

12) диаграмма вращений

13) диаграмма вращений

14) диаграмма вращений

15) диаграмма вращений

16) диаграмма вращений

17) диаграмма вращений

18) диаграмма вращений

19) диаграмма вращений

20) диаграмма вращений

Будет посчитано расстояние
от школы до дома, время,
затраченное на путь, будут
использоваться средства транспорта,
мотивация и желание учиться,

качество сна учащегося, также кол-во часов сна (в
зависимости от пола, оно разное), так что учитываем
демографические данные, нагрузка в учебном заведении

5) Я могу справиться с 1) заданными /возможными (ответы)
данными,

2) нескрепками,

3) Возвращение от ответа,

4) Сложности в сборе, анализе больших данных (более 1000 тыс.);

5) Недостаточность количественных данных (низкая польза),

6) ~~Невозможность~~ сложности отображения качественных данных,

7) Сложности в выявлении связи между данными,

8) Проблемы с "чувствительными" вопросами,

9) Сложности с получением средств на подобное исследование

Задание 4 (Продолжение)

дополнительный лист
лист 1

Давайте обозначим некоторый конечный ряд чисел $- [X_n]$

В котором справедливо следующее $X_1, X_2, \dots, X_n$ — конечный ряд
Есть первые и последние значения

Есть ~~среднее~~ $X_{\frac{n}{2}}$, если n — четное (Если n — нечетное, то $\frac{n}{2}$ округляют в большую или меньшую сторону)

Каково X_c — который встречается чаще остальных значений
 X_c — мода

m — среднее значение $\frac{X_1 + X_2 + \dots + X_n}{n} = m$

Что мы должны опровергнуть или доказать

$\neg$ — отрицание, "НЕ",

1 — истина, 0 — ложь,

$$\neg \left(X_{\frac{n}{2}} < X_c < m \right) = \neg \left(X_{\frac{n}{2}} < X_c < \frac{X_1 + X_2 + \dots + X_n}{n} \right) = F$$

$F = 0$ или 1 ?

Если мы предполагаем, что $F = 1$, то

$$\neg \left(X_{\frac{n}{2}} < X_c < \frac{X_1 + X_2 + \dots + X_n}{n} \right) = 1, \text{ то есть такого распределения не бывает}$$

Если мы опровергаем, что $F = 1$, то

$$\neg \left(X_{\frac{n}{2}} < X_c < \frac{X_1 + X_2 + \dots + X_n}{n} \right) = 0, \text{ то есть такое распределение бывает}$$

Разбиваем нашу гипотезу

1) Если $X_{\frac{n}{2}} = X_c$, то да действительно такого распределения не бывает,

2) Если все элементы послед-сти одинаковые, то тоже не бывает

3) Предположим, что $X_{\frac{n}{2}}$, где n — нечетное, то у нас две медианы, соответственно X_x и X_{x+1} , если $X_{x+1} = X_c$, то тоже не бывает

4) $X_{\frac{n}{2}}$, где n — четное и $X_{\frac{n}{2}} \neq X_c$, элементы различные

Задача 4 (Продолжение) дополнительный лист

$$7 \left(X_{\frac{n}{2}} < X_c < \frac{X_1 + X_2 + \dots + X_{\frac{n}{2}} + \dots + X_c + X_c + \dots + X_n}{n} \right) \text{ лист}$$

НЕ
✓ ∇ опровергаю, т.е. соглашусь с данным утвержде-
нием

Если $X_{\frac{n}{2}}$ — это медиана, а $X_c > \frac{X_n}{2}$, то каждый раз медиана сдвигается с каждым новым повторением. Чтобы от этого избавиться можно будем увеличивать разницу между модой и медианой, тогда будем уменьшаться ее значение. Если все идет по возрастанию, если нет, то можно наоборот.

2, 3, 2, 1, 5, 5, 5
 ↓
 медична

Er zwar $\frac{27 + \overset{10}{3} + \overset{12}{2} + \overset{13}{7} + \overset{15}{5} + 5 + 5}{7} = \frac{42}{7} = 6$

te 21, 3, 2, ~~1~~, 5, 5, 5
 ↓
 uguale
 1 - 5 - 0

$$1 < 5 < 6$$

Поэтому окончательный ответ Бывает такое распределение