

Решение и критерии заключительного этапа Изумруд Анализ Данных 2025-2026

11 класс

Задача 1. (20 баллов) Вы хотите исследовать, связаны ли экологическая ситуация в регионах Германии и уровень здоровья населения. Данные — кросс-секционные по регионам за один и тот же год. Ниже приведён список возможных пар показателей.

1. (10 баллов) Отметьте все пары, которые подходят для исследования взаимосвязи.
2. (10 баллов) Для каждой выбранной и невыбранной пары дайте краткое объяснение (1–2 предложения): почему именно эта комбинация релевантна или, наоборот, проблемна (например, «абсолютное число вместо нормированного», «не отражает здоровье», «разные уровни агрегации» и т. п.).

Варианты пар показателей:

- a) Общее количество больничных дней в регионе и среднегодовая концентрация PM2.5 (мкг/м³).
- b) Число госпитализаций из-за астмы на 100 000 жителей и среднегодовой уровень загрязнения воздуха NO₂ (мкг/м³).
- c) Количество автомобилей в регионе и число заболевших раком лёгких.
- d) Смертность от сердечно-сосудистых заболеваний на 100 000 жителей и средний уровень шума в дБА.
- e) Число случаев COVID-19 и число солнечных дней в году.
- f) Средняя продолжительность жизни и доля зелёных зон (%) в регионе.
- g) Доля населения с ожирением (%) и среднее число парков на 10 000 жителей.
- h) Количество поликлиник на 100 000 жителей и уровень загрязнения воды по индексу качества (0–100).
- i) Расходы региональных властей на здравоохранение на душу населения и среднегодовой уровень загрязнения воздуха PM10 (мкг/м³).
- j) Смертность от хронических заболеваний дыхательной системы на 100 000 жителей и уровень загрязнения воздуха бенз(а)пиреном (нг/м³).

Пояснения к терминам

- *PM2.5 / PM10 — концентрация мелких твёрдых частиц в воздухе диаметром до 2.5 или 10 микрометров. Чем выше значение (мкг/м³), тем хуже качество воздуха; частицы проникают в дыхательную систему.*
- *NO₂ (диоксид азота) — газ, выделяющийся в основном при работе автомобильных двигателей и сжигании топлива. Вредит дыхательной системе.*
- *Бенз(а)пирен — канцерогенное вещество из промышленных выбросов, дыма и выхлопных газов.*
- *Индекс качества воды (0–100) — сводный показатель загрязнённости воды: чем выше значение, тем чище вода.*
- *Зелёные зоны (%) — доля территории региона, занятая парками, скверами и лесопарками.*
- *Уровень шума (дБА) — средний уровень шумового загрязнения; дБА — условная шкала, учитывающая чувствительность человеческого уха к разным частотам.*

Решение и критерии оценки

Подходящие пары: b, d, f, g, h, i, j.

Неподходящие пары: a, c, e.

Пояснения:

- a) Абсолютное число больничных дней зависит от размера региона → не нормировано.
- b) Отличная пара: нормированный показатель + конкретный загрязнитель, связанный с астмой.
- c) Абсолютное число автомобилей не учитывает население и не напрямую связано с

заболеваемостью.

d) Нормированная смертность и шумовое загрязнение — правдоподобная связь.

e) COVID и солнечные дни — скорее случайная корреляция, биологического механизма нет.

f) Продолжительность жизни и зелёные зоны — экологическая гипотеза, нормировано.

g) Ожирение и количество парков на душу населения — связь через физическую активность.

h) Кол-во поликлиник (нормировано) и качество воды — отражает доступ к медицине и фактор риска.

i) Затраты на медицину и PM10 — может отражать реакцию на загрязнение.

j) Заболеваемость дыхательных путей и канцерогенное загрязнение воздуха — ожидаемая связь.

Участник получает по 1 баллу за верный вывод (подходит/не подходит) для каждой пары. Участник получает по 1 баллу за верное обоснование для каждой пары (только при условии верного вывода).

Задача 2. (20 баллов) В городе есть свободный фонд муниципального жилья (пустующие комнаты и малосемейные общежития). Администрация решила уменьшить неравенство в обеспеченности площадью и распределить часть фонда между жителями с наименьшей площадью на человека. Забирать у кого-то метры нельзя; допускается только добавлять площади отдельным жителям из муниципального фонда.

В таблице указана текущая площадь на человека (m^2) у 10 жителей:

Житель	A	B	C	D	E	F	G	H	I	J
Площадь ($m^2/чел$)	9	12	14	16	20	24	28	35	60	80

Администрация ставит цели:

1. (6 баллов) *Сделать медиану площади равной 25 m^2 .* Найдите минимально возможный общий объём площади, которую нужно добавить (в m^2).

2. (6 баллов) *Дополнительно обеспечить минимальный стандарт 15 m^2 на человека для каждого жителя.* Также найдите минимально возможный общий объём площади, которую нужно добавить (в m^2), сохранив медиану 25 m^2 .

3. (8 баллов) *(Бюджетный сценарий)* Если суммарный доступный фонд ограничен 12 m^2 , какова максимально достижимая медиана? Укажите оптимальное распределение этих 12 m^2 по жителям, которое её достигает.

Решение и критерии оценки

1. Для 10 человек медиана — среднее из 5-го и 6-го значений в отсортированном ряду. Сейчас это $E=20$ и $F=24$, медиана = $(20+24)/2 = 22$. Чтобы сделать медиану ровно 25, достаточно поднять оба эти значения до 25 и ничего больше не трогать. Это даёт минимальный суммарный объём, потому что любые добавки другим людям не влияют на медиану, пока 5-е и 6-е остаются ниже 25.

Поднять E: $20 \rightarrow 25$ (+5 m^2)

Поднять F: $24 \rightarrow 25$ (+1 m^2)

Итого минимум: $5 + 1 = 6 m^2$. Получившийся упорядоченный ряд: 9, 12, 14, 16, 25, 25, 28, 35, 60, 80 \rightarrow медиана = $(25+25)/2 = 25$.

Участник получает 6 баллов за верный ответ.

2. Нужно довести все значения ниже 15 до 15, не испортив медиану из п.1.

A: $9 \rightarrow 15 (+6 \text{ м}^2)$

B: $12 \rightarrow 15 (+3 \text{ м}^2)$

C: $14 \rightarrow 15 (+1 \text{ м}^2)$

Из п.1 оставляем E: +5 и F: +1 (медиана 25 сохраняется).

Дополнительно к п.1: $+6 + 3 + 1 = +10 \text{ м}^2$.

Суммарно (п.1 + п.2): $6 + 10 = 16 \text{ м}^2$.

Проверка порядка после всех добавок: 15, 15, 15, 16, 25, 25, 28, 35, 60, 80 \rightarrow медиана = 25, у всех ≥ 15 .

Участник получает 6 баллов за верный ответ.

3. Хотим максимизировать медиану при суммарном ресурсе 12 м^2 . Для чётного n медиана — среднее 5-го и 6-го значений: E и F. Пусть целевая медиана = M. Тогда нужно сделать $E \geq M$ и $F \geq M$ при минимальных добавках. Исходно $E=20$, $F=24$.

Суммарная прибавка для достижения уровня M:

$(E_{\text{надбавка}}) + (F_{\text{надбавка}}) = (M-20) + (M-24) = 2M - 44$.

Ограничение по фонду: $2M - 44 \leq 12 \rightarrow 2M \leq 56 \rightarrow M \leq 28$.

Вывод: максимально достижимая медиана = 28 м^2 , оптимально:

E: $20 \rightarrow 28 (+8 \text{ м}^2)$

F: $24 \rightarrow 28 (+4 \text{ м}^2)$

Итого как раз 12 м^2 . Порядок сохраняется (... 16, 28, 28, 28, 35, 60, 80).

Участник получает 8 баллов за верный ответ.

Задача 3. (20 баллов) На производственной линии оценивают качество работы смен по 10-балльной шкале (чем выше, тем лучше). За неделю получены баллы:

Смена А: 3, 5, 5, 6, 8

Смена В: 2, 4, 6, 7, 10

Смена С: 1, 4, 5, 8, 9

1) (3 балла) Для каждой смены найдите: среднее, медиану, размах. Сравните смены: у какой среднее выше? У какой разброс больше?

2) (3 балла) Сравните дисперсии смен А, В и С. У какой смены результаты наиболее «разнообразны»?

На другой площадке за 6 смен получили баллы качества по 100-балльной шкале (все баллы — целые числа):

Смена D: 65, 72, 80, 95, x , y

Известно, что среднее = 80, а дисперсия = 50.

3.а) (7 баллов) Найдите целые x и y , которые удовлетворяют этим условиям.

3.б) (7 баллов) При фиксированном среднем 80 и целых x , y найдите минимально возможную и максимально возможную дисперсию этой шестёрки и укажите соответствующие x и y .

Решение и критерии оценки

1. Смена А: сумма $27 \rightarrow$ среднее $=27/5=5,4$; медиана $=5$; размах $=8-3=5$.

Смена В: сумма $29 \rightarrow$ среднее $=29/5=5,8$; медиана $=6$; размах $=10-2=8$.

Смена С: сумма $27 \rightarrow$ среднее $=27/5=5,4$; медиана $=5$; размах $=9-1=8$.

Сравнение: самое высокое среднее у В (5.8). Наибольший разброс (размах 8) у В и С (больше, чем у А).

Участник получает 2 балла за верно посчитанные средние значения, медианы и размахи. Участник получает 1 балл за верные сравнения (только при верно посчитанных значениях).

2. А: отклонения от 5.4: $-2.4, -0.4, -0.4, 0.6, 2.6 \rightarrow$ квадраты: $5.76, 0.16, 0.16, 0.36, 6.76$.
Сумма квадратов $=13.20 \rightarrow$ дисперсия $=13.20/5=2.64$.

В: от 5.8: $-3.8, -1.8, 0.2, 1.2, 4.2 \rightarrow$ квадраты: $14.44, 3.24, 0.04, 1.44, 17.64$.

Сумма $=36.80 \rightarrow$ дисперсия $=36.80/5=7.36$.

С: от 5.4: $-4.4, -1.4, -0.4, 2.6, 3.6 \rightarrow$ квадраты: $19.36, 1.96, 0.16, 6.76, 12.96$.

Сумма $=41.20 \rightarrow$ дисперсия $=41.20/5=8.24$.

Вывод: самая большая дисперсия у С (8.24), затем В (7.36), затем А (2.64). Это согласуется с наличием более крайних значений у В и С.

Участник получает 2 балла за верно посчитанные дисперсии. Участник получает 1 балл за верные сравнения (только при верно посчитанных значениях).

3.а. Среднее: $(65+72+80+95+x+y)/6=80 \Rightarrow x+y=480-(65+72+80+95)=168$.

Обозначим $a=x-80, b=y-80$. Тогда $a+b=(x+y)-160=8$.

Сумма квадратов для известных четырёх относительно 80:

$$(65-80)^2+(72-80)^2+(80-80)^2+(95-80)^2=225+64+0+225=514.$$

Популяционная дисперсия 50 означает, что суммарная сумма квадратов отклонений всей шестёрки $=50 \cdot 6=300$.

Но уже у первых четырёх сумма квадратов $514 > 300$. Противоречие. Таких целых x, y не существует: условия « $\bar{x} = 80$ » и « $D=50$ » одновременно невыполнимы (даже без целочисленности), так как требуемая общая сумма квадратов $=300$ меньше суммы квадратов уже известных четырёх значений $=514$.

Участник получает 7 баллов за верный вывод.

3.б. Минимальная дисперсия. При фиксированной сумме $a+b$ величина a^2+b^2 минимальна, когда a и b максимально близки. Для целых: $a=4, b=4 \rightarrow x=84, y=84$.

Тогда $a^2+b^2=16+16=32$; полная сумма квадратов $=514+32=546$;

$$D_{min} = \frac{546}{6} = 91.$$

Максимальная дисперсия. Увеличиваем «расстояние» от среднего, соблюдая $x+y=168$ и границы $0-100$. Край: $x=100 \Rightarrow a=20, y=68 \Rightarrow b=-12$ (или наоборот).

Тогда $a^2+b^2=400+144=544$; полная сумма квадратов $=514+544=1058$;

$$D_{max} = \frac{1058}{6} \approx 176.33.$$

Участник получает по 3.5 балла за верно найденные максимальное и минимальное значение дисперсии.

Задача 4. (20 баллов) В последние годы активно обсуждается влияние смартфонов на сон подростков.

Исследователи хотят проверить: связана ли **продолжительность использования смартфона вечером с продолжительностью сна школьников**.

1. (2 балла) Сформулируйте направленную гипотезу о связи «вечерний смартфон \rightarrow продолжительность сна».

2. (3 балла) Назовите 3 механизма, которые могут объяснять такую связь.

3. (6 баллов) Опишите дизайн эксперимента, который позволит проверять причинность.

Укажите:

- единицу наблюдения,
- интервенцию/экзогенную вариацию,
- способ сбора данных,
- переменные для анализа,
- план анализа,
- отметьте этические аспекты

4. (3 балла) Укажите не менее 3 конфаундеров и предложите способы их учёта.

Конфаундер — это переменная, которая одновременно связана и с «причиной», и с «результатом», искажающая реальную связь между ними.

5. (3 балла) Опишите, какими конкретно переменными вы будете измерять «вечерний смартфон» и «сон».

6. (3 балла) Назовите 3 угрозы валидности и меры для их снижения.

Решение и критерии оценки

1. Пример:

Чем больше времени школьник проводит со смартфоном вечером, тем меньше он спит / тем позже засыпает.

Участник получает 2 балла за корректную и направленную гипотезу (0 баллов, если не указано направление связи «при увеличении/уменьшении», в гипотезе есть неточности, или она не по теме).

2. Примеры:

Смещение режима: прокрастинация откладывает отход ко сну.

Возбуждение/стресс: контент и уведомления мешают заснуть.

Синее излучение: подавляет выработку мелатонина → труднее заснуть.

Участник получает 3 балла за каждый корректный механизм.

3. Пример эксперимента:

- Единица: класс (кластерная рандомизация).
- Интервенция: «цифровой закат» — ограничение смартфона после 21:00 на 4 недели.
- Измерения исходов: фитнес-браслет/актиметр + дневник сна.
- Экспозиция: отчёты «Экранное время» (минуты после 20:00, типы приложений).
- План анализа: ИТТ и difference-in-differences; учёт несоблюдения.
- Этика: согласие родителей и школьников, обезличивание данных.

Участник получает 6 баллов за каждый корректный компонент. Участник получает 0 баллов, если интервенция не позволяет установить причинно-следственную связь.

4. Примеры: учебная нагрузка, стресс, кофеин/тренировки, хронотип, семейные правила, время подъёма, здоровье.

Участник получает по 1 баллу за каждый верный конфаундер.

5. Пример: Вечерний смартфон = минуты после 20:00 по отчётам устройства. Сон = часы сна за ночь, время засыпания, число пробуждений.

Участник получает 2 балла за верный замер вечернего смартфона и 1 балл за мерный замер сна.

6. Примеры:

- Ошибка измерения → использовать объективные трекеры.
- Эффект Хотгорна → предусмотреть адаптационный период.
- «Перетекание» поведения между группами → кластерная рандомизация.
- Сезонность/экзамены → фиксированные эффекты по неделям.

Участник получает по 1 баллу за каждую корректную угрозу. Участник получает 0 баллов за все задание если текст является нечитаемым и/или не разделен на подпункты в соответствии с условием задания.

Задача 5. (20 баллов) В группе из 10 сотрудников провели два теста — X и Y (баллы от 0 до 100). Известно, что

$$\sum_{i=1}^{10} X_i = 700 \text{ (средний балл по } X \text{ равен } \bar{X}=70)$$

$$\sum_{i=1}^{10} Y_i = 600 \text{ (средний балл по } Y \text{ равен } \bar{Y}=60)$$

$$\frac{1}{10} \sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = 0$$

1. (8 баллов) Чему равны ковариация $\text{Cov}(X, Y)$ и корреляция ρ_{XY} ?
2. (6 баллов) Всем участникам прибавили по 5 баллов к результату X , получив новый тест $X' = X+5$. Как изменятся $\text{Cov}(X', Y)$ и $\rho_{X'Y}$?
3. (6 баллов) Затем вместо прибавления применили масштабирование: $X'' = 3X$. Как изменятся $\text{Cov}(X'', Y)$ и $\rho_{X''Y}$ по сравнению с исходными?

Подсказка: используйте свойства ковариации:

$$\text{Cov}(X + c, Y) = \text{Cov}(X, Y)$$

$$\text{Cov}(aX, Y) = a \text{Cov}(X, Y)$$

Корреляция — это ковариация, делённая на произведение стандартных отклонений.

Решение и критерии оценки:

1. $\text{Cov}(X, Y) = 0$ по условию $\Rightarrow \rho_{XY} = 0$ (при ненулевых стандартных отклонениях).
Вывод: линейной связи нет (корреляция ноль), хотя нелинейная зависимость теоретически возможна.

Участник получает 4 балла за верную ковариацию и 4 балла за верную корреляцию.

2. $X' = X+5$. $\text{Cov}(X', Y) = \text{Cov}(X + 5, Y) = \text{Cov}(X, Y) = 0$

Стандартное отклонение X не меняется при сдвиге $\Rightarrow \rho_{X'Y} = \rho_{XY} = 0$. Вывод: ничего не изменилось.

Участник получает 6 баллов за верный вывод.

3. $X'' = 3X$. $\text{Cov}(X'', Y) = \text{Cov}(3X, Y) = 3 \text{Cov}(X, Y) = 3 \cdot 0 = 0$

Стандартное отклонение X масштабировалось втрое, но и ковариация умножилась на 3 — отношение остаётся тем же $\Rightarrow \rho_{X''Y} = \rho_{XY} = 0$. Вывод: ковариация остаётся 0, корреляция остаётся 0.

Участник получает 6 баллов за верный вывод.

10 класс

Задача 1. (20 баллов) Ниже приведены результаты двух небольших групп студентов, написавших тест по 10-балльной шкале.

Группа А: 4, 5, 5, 6, 7

Группа В: 3, 4, 6, 8, 9

- (5 баллов) Найдите для каждой группы: среднее, медиану, размах. Сравните группы: в какой среднее выше, а в какой разброс оценок больше?
- (5 баллов) Сравните дисперсии групп А и В. У какой группы результаты более «разнообразные»? Объясните, почему так вышло, исходя из самих рядов.
- (10 баллов) В другой группе из 6 студентов известны оценки (оценки являются целыми числами по 100-балльной шкале):

Группа С: 70, 80, 85, 90, x , y

Известно, что среднее = 85, а дисперсия = 50. Найдите x и y .

Решение и критерии оценки

1. Группа А: среднее = $(4+5+5+6+7)/5 = 27/5 = 5.4$; медиана = 5; размах = $7-4 = 3$.

Группа В: среднее = $(3+4+6+8+9)/5 = 30/5 = 6$; медиана = 6; размах = $9-3 = 6$.

Сравнение: у группы В среднее выше ($6 > 5.4$) и разброс больше ($6 > 3$).

Участник получает по 1 баллу за верно рассчитанные средние значения, медианы, размахи для обеих групп, соответственно. Участник получает по 1 баллу за верное сравнение средних значений и размахов, соответственно (при неверно посчитанных значениях средних и размахов ставится 0 баллов даже при верном сравнении).

2. Дисперсия А: $((4-5.4)^2 + (5-5.4)^2 + (5-5.4)^2 + (6-5.4)^2 + (7-5.4)^2) / 5 = 1.04$.

Дисперсия В: $((3-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (9-6)^2) / 5 = 5.2$.

Вывод: у группы В разброс существенно больше, что видно и по большему размаху. Участник получает по 2 балла за каждую верно посчитанную дисперсию и 1 балл за вывод и обоснование (при неверно посчитанных дисперсиях за вывод и обоснование ставится 0 баллов).

3. Среднее: $(70+80+85+90+x+y)/6 = 85 \rightarrow x+y=185$.

Дисперсия: $(70-85)^2+(80-85)^2+(85-85)^2+(90-85)^2+(x-85)^2+(y-85)^2/6 = 50$.

$\rightarrow (225+25+0+25)+(x-85)^2+(y-85)^2 = 300 \rightarrow (x-85)^2+(y-85)^2=25$.

Система:

- $x+y=185$

- $(x-85)^2+(y-85)^2=25$

Подстановка: $y=185-x \rightarrow (x-85)^2+(100-x)^2=25 \rightarrow 2x^2-370x+17225=25 \rightarrow 2x^2-$

$370x+17200=0 \rightarrow$ дискриминант отрицательный. Вывод: система не имеет

действительных решений \rightarrow таких данных не может быть.

Участник получает 10 баллов за верный вывод.

Задача 2. (20 баллов) Вы хотите проверить, связаны ли учебные и бытовые привычки учеников в школах города и их учебные результаты. Данные собраны по школам за один и тот же учебный год (одна строка = одна школа). Ниже дан список возможных пар показателей.

1. (10 баллов) Отметьте все пары, которые подходят для исследования взаимосвязи.

2. (10 баллов) Для каждой выбранной и невыбранной пары кратко объясните (1–2 предложения), почему эта комбинация релевантна или проблемна (например: «абсолютное число вместо показателя на ученика», «не отражает учебные результаты», «сомнительная логическая связь», «разные уровни измерения» и т. п.).

Варианты пар показателей:

а) Общее число двоек по итоговой контрольной в школе и среднее время на домашние

задания (минут в день).

- b) Доля двоек по итоговой контрольной (%) и среднее время на домашние задания (минут в день).
- с) Средний балл по математике и число подписчиков школьного аккаунта в соцсетях (человек).
- d) Доля пропусков по болезни (%) и средняя продолжительность сна учеников (часов в сутки).
- e) Средний балл по русскому языку и число кабинетов информатики (штук).
- f) Доля учеников, занимающихся спортом ≥ 3 раза в неделю (%) и средний балл по итоговой контрольной.
- g) Общее число смартфонов у учеников и среднее время использования телефона (часов в день).
- h) Среднее время использования телефона (часов в день) и средний балл по итоговой контрольной.
- i) Средний размер класса (учеников на одного учителя) и средний балл по итоговой контрольной.
- j) Доля учеников, получающих бесплатные обеды (%) и средний балл по итоговой контрольной.

Решение и критерии оценки

Подходящие пары: b, d, f, h, i, j.

Неподходящие пары: a, c, e, g.

Короткие обоснования:

- a) Общее число двоек vs время на ДЗ — плохо: абсолют зависит от размера школы; лучше нормировать (есть вариант b).
 - b) Доля двоек (%) vs время на ДЗ — хорошо: оба показателя сопоставимы между школами; ожидаемая логика («больше времени — меньше двоек/выше качество»), хотя причинность не гарантируется.
 - с) Средний балл vs подписчики — плохо: нет прямой учебной логики; популярность аккаунта не отражает качество обучения.
 - d) Пропуски по болезни (%) vs сон (часы) — хорошо: оба нормированы; понятный механизм (недосып \leftrightarrow чаще болеют/пропуски).
 - e) Средний балл vs число кабинетов — скорее плохо: абсолют без учёта размера школы; и сам по себе «число кабинетов» не гарантирует качество.
 - f) Доля регулярно занимающихся спортом (%) vs средний балл — приемлемо: возможная связь через дисциплину/здоровье; оба нормированы.
 - g) Общее число смартфонов vs среднее время на телефоне — плохо: первое — абсолют; к тому же сама по себе «численность смартфонов» мало что говорит о привычках.
 - h) Время на телефоне (часы) vs средний балл — хорошо: сопоставимые средние по школе; реалистичная гипотеза (больше экранного времени \leftrightarrow ниже учебные результаты или наоборот).
 - i) Средний размер класса vs средний балл — хорошо: понятный механизм (нагрузка на учителя/внимание ученикам).
 - j) Доля на бесплатных обедах (%) vs средний балл — хорошо: индикатор социально-экономического фона; уместно изучать связь с результатами.
- Участник получает по 1 баллу за верный вывод (подходит/не подходит) для каждой пары. Участник получает по 1 баллу за верное обоснование для каждой пары (только при условии верного вывода).

Задача 3. (20 баллов) В городе есть свободный фонд муниципального жилья (пустующие комнаты и малосемейные общежития). Администрация решила уменьшить неравенство в обеспеченности площадью и распределить часть фонда между жителями с наименьшей площадью на человека. Забирать у кого-то метры нельзя; допускается только добавлять площади отдельным жителям из муниципального фонда.

В таблице указана текущая площадь на человека (m^2) у 10 жителей:

Житель	A	B	C	D	E	F	G	H	I	J
Площадь (m^2 /чел)	9	12	14	16	20	24	28	35	60	80

Администрация ставит цели:

- (6 баллов) *Сделать медиану площади равной $25 m^2$.* Найдите минимально возможный общий объём площади, которую нужно добавить (в m^2).
- (6 баллов) *Дополнительно обеспечить минимальный стандарт $15 m^2$ на человека для каждого жителя.* Также найдите минимально возможный общий объём площади, которую нужно добавить (в m^2), сохранив медиану $25 m^2$.
- (8 баллов) *(Бюджетный сценарий)* Если суммарный доступный фонд ограничен $12 m^2$, какова максимально достижимая медиана? Укажите оптимальное распределение этих $12 m^2$ по жителям, которое её достигает.

Решение и критерии оценки

1. Для 10 человек медиана — среднее из 5-го и 6-го значений в отсортированном ряду. Сейчас это $E=20$ и $F=24$, медиана $= (20+24)/2 = 22$. Чтобы сделать медиану ровно 25 , достаточно поднять оба эти значения до 25 и ничего больше не трогать. Это даёт минимальный суммарный объём, потому что любые добавки другим людям не влияют на медиану, пока 5-е и 6-е остаются ниже 25 .

Поднять E: $20 \rightarrow 25$ (+5 m^2)

Поднять F: $24 \rightarrow 25$ (+1 m^2)

Итого минимум: $5 + 1 = 6 m^2$. Получившийся упорядоченный ряд: 9, 12, 14, 16, 25, 25, 28, 35, 60, 80 \rightarrow медиана $= (25+25)/2 = 25$.

Участник получает 6 баллов за верный ответ.

2. Нужно довести все значения ниже 15 до 15 , не испортив медиану из п.1.

A: $9 \rightarrow 15$ (+6 m^2)

B: $12 \rightarrow 15$ (+3 m^2)

C: $14 \rightarrow 15$ (+1 m^2)

Из п.1 оставляем E: +5 и F: +1 (медиана 25 сохраняется).

Дополнительно к п.1: $+6 + 3 + 1 = +10 m^2$.

Суммарно (п.1 + п.2): $6 + 10 = 16 m^2$.

Проверка порядка после всех добавок: 15, 15, 15, 16, 25, 25, 28, 35, 60, 80 \rightarrow медиана $= 25$, у всех ≥ 15 .

Участник получает 6 баллов за верный ответ.

3. Хотим максимизировать медиану при суммарном ресурсе $12 m^2$. Для чётного n медиана — среднее 5-го и 6-го значений: E и F. Пусть целевая медиана $= M$. Тогда нужно сделать $E \geq M$ и $F \geq M$ при минимальных добавках. Исходно $E=20$, $F=24$.

Суммарная прибавка для достижения уровня M:

$(E_{\text{надбавка}}) + (F_{\text{надбавка}}) = (M-20) + (M-24) = 2M - 44$.

Ограничение по фонду: $2M - 44 \leq 12 \rightarrow 2M \leq 56 \rightarrow M \leq 28$.

Вывод: максимально достижимая медиана = 28 м², оптимально:

E: 20 → 28 (+8 м²)

F: 24 → 28 (+4 м²)

Итого как раз 12 м². Порядок сохраняется (... 16, 28, 28, 28, 35, 60, 80).

Участник получает 8 баллов за верный ответ.

Задача 4. (20 баллов) В 2025 году команда исследователей УрФУ под руководством Евгения Максимовича Шмыкнёва отправилась в местные деревни для изучения фермерских практик. Фермерам было предложено пройти опрос, охватывающий как аспекты ведения хозяйства, так и социально-демографические вопросы о них и их семьях. Например, фермеров спрашивали: размер их поля в квадратных метрах, общую выручку от продажи урожая за последний сезон (в рублях), совокупные трудовые издержки за сезон (оплата рабочим, наёмный труд), совокупные нетрудовые издержки за сезон (семена, удобрения, вода для полива и т. д.), и многие другие показатели.

Ассистенты Шмыкнёва провели предварительный анализ данных и рассчитали корреляции между несколькими переменными:



1. (4 балла) Объясните корреляцию каждой переменной с площадью поля: направление, силу и возможный экономический механизм.

2. (4 балла) Чем может объясняться разница в значениях корреляций? Для каждого сравнения предложите экономический механизм.

3. (6 баллов) Предложите три переменные, которые могли бы иметь отрицательную корреляцию с площадью поля. Для каждой переменной объясните экономический механизм, обосновывающий такое направление связи.

4. (6 баллов) Какие факторы могли бы исказить рассчитанные корреляции, сделав их выше или ниже реальных? Предложите 3 фактора.

Решение и критерии оценки

1. Площадь поля и выручка ($r = +0.34$) – связь положительная, средней силы; чем больше поле, тем выше объём урожая и доход.

Площадь поля и трудовые издержки ($r = +0.43$) – положительная и более сильная; большие поля требуют больше рабочих рук и затрат на труд.

Площадь поля и нетрудовые издержки ($r = +0.28$) – положительная, но слабее; хотя большие поля требуют больше семян и удобрений, эффект масштаба снижает рост

затрат.

Участник получает 4 балла за пункт, если представлены все 3 корреляции с их значениями и экономическими механизмами.

2. $r(\text{площадь} - \text{трудовые издержки}) > r(\text{площадь} - \text{выручка})$: потребность в труде растёт быстрее доходов, цены на урожай могут сдерживать выручку.

$r(\text{площадь} - \text{выручка}) > r(\text{площадь} - \text{нетрудовые издержки})$: доходы растут почти пропорционально, а закупки ресурсов дешевеют при больших объёмах.

$r(\text{площадь} - \text{трудовые издержки}) > r(\text{площадь} - \text{нетрудовые издержки})$: труд сложнее оптимизировать, чем ресурсы; экономия от масштаба действует слабее.

Участник получает 4 балла, если корректно объяснены все 3 сравнения.

3. Урожайность с гектара – на больших полях меньше внимания каждой части → урожайность снижается. Доля современных технологий – крупные поля сложнее оборудовать полностью → инновации внедряются частично. Возраст главы хозяйства – крупные поля чаще у старших фермеров, а молодые начинают с меньших участков.

Участник получает по 2 балла за каждую подходящую переменную и механизм.

4. Региональные различия в урожайности, рыночные колебания цен, государственные субсидии для крупных фермеров, разная структура затрат у хозяйств.

Участник получает по 2 балла за каждый подходящий фактор.

Задача 5. (20 баллов) В группе из 10 сотрудников провели два теста — X и Y (баллы от 0 до 100). Известно, что

$$\sum_{i=1}^{10} X_i = 450 \text{ (средний балл по X равен } \bar{X}=45)$$

$$\sum_{i=1}^{10} Y_i = 530 \text{ (средний балл по Y равен } \bar{Y}=53)$$

$$\frac{1}{10} \sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = 0$$

1. (8 баллов) Чему равны ковариация $\text{Cov}(X, Y)$ и корреляция ρ_{XY} ?

2. (6 баллов) Всем участникам прибавили по 5 баллов к результату X, получив новый тест $X' = X + 5$. Как изменятся $\text{Cov}(X', Y)$ и $\rho_{X'Y}$?

3. (6 баллов) Затем вместо прибавления применили масштабирование: $X'' = 3X$. Как изменятся $\text{Cov}(X'', Y)$ и $\rho_{X''Y}$ по сравнению с исходными?

Подсказка: используйте свойства ковариации:

$$\text{Cov}(X + c, Y) = \text{Cov}(X, Y)$$

$$\text{Cov}(aX, Y) = a \text{Cov}(X, Y)$$

Корреляция — это ковариация, делённая на произведение стандартных отклонений.

Решение и критерии оценки:

1. $\text{Cov}(X, Y) = 0$ по условию $\Rightarrow \rho_{XY} = 0$ (при ненулевых стандартных отклонениях). Вывод: линейной связи нет (корреляция ноль), хотя нелинейная зависимость теоретически возможна.

Участник получает 4 балла за верную ковариацию и 4 балла за верную корреляцию.

2. $X' = X + 5$. $\text{Cov}(X', Y) = \text{Cov}(X + 5, Y) = \text{Cov}(X, Y) = 0$

Стандартное отклонение X не меняется при сдвиге $\Rightarrow \rho_{X'Y} = \rho_{XY} = 0$. Вывод: ничего не изменилось.

Участник получает 6 баллов за верный вывод.

3. $X'' = 3X$. $\text{Cov}(X'', Y) = \text{Cov}(3X, Y) = 3 \text{Cov}(X, Y) = 3 \cdot 0 = 0$

Стандартное отклонение X масштабировалось втрое, но и ковариация умножилась на 3 — отношение остаётся тем же $\Rightarrow \rho_{X''Y} = \rho_{XY} = 0$. Вывод: ковариация остаётся 0,

корреляция остаётся 0.
Участник получает 6 баллов за верный вывод.