

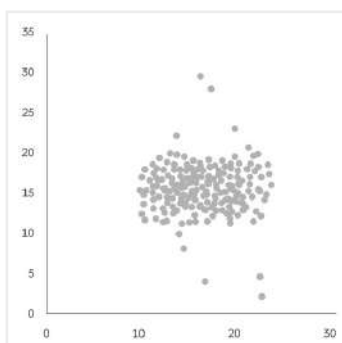
Вариант 1

- По выборке из 100 значений посчитали выборочную дисперсию. После этого все значения увеличили в 2 раза. Как изменится выборочная дисперсия, если вычислить ее по новым значениям выборки?

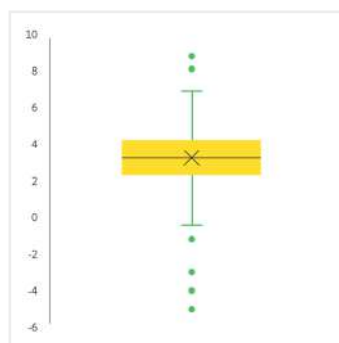
- A. Не изменится
- B. Увеличится в 2 раза
- C. Увеличится в 4 раза
- D. Уменьшится в 2 раза

- Сопоставьте названия типов визуализации с графиками:

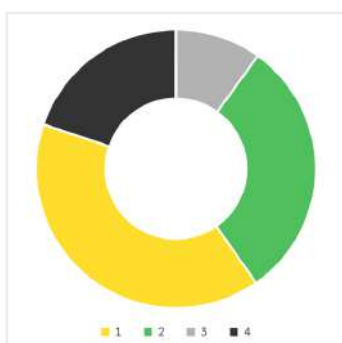
1.



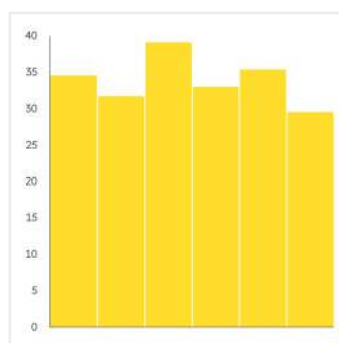
2.



3.



4.



- 1. Диаграмма рассеяния
- 2. Диаграмма размаха
- 3. Секторная диаграмма
- 4. Гистограмма

3. Исследователи проверили гипотезу о среднем значении с использованием Т-теста. Уровень значимости установлен равным 0,05. Как интерпретировать такое значение уровня значимости?

A. Вероятность того, что альтернативная гипотеза верна, равна 5%

B. Допустимая вероятность отвержения нулевой гипотезы, если она на самом деле верна, равна 5%

C. Если нулевая гипотеза верна, в 5% случаев мы допускаем, что можем ошибочно отвергнуть ее при многократном повторении теста на семплированных данных из одной генеральной совокупности

D. Вероятность ошибки второго рода равна 5%

4. Сотрудники отдела прогнозирования и анализа данных изучили ежемесячные данные по ценам на гибискусы и гортензии за 2024 год. Средние цены оказались одинаковыми. Значения дисперсий: 144 P^2 для гибискуса, и 225 P^2 для гортензии. Какое из утверждений верное?

A. Цена гибискуса в 2024 году была более изменчива, чем цена гортензии

B. Цена гортензии в 2024 году была более изменчива, чем цена гибискуса

C. Ежемесячные цены на гортензию и гибискус всегда одинаковы

D. Стандартные отклонения цен на гибискус и гортензию одинаковы

5. Студент построил 95%-й доверительный интервал для средней цены на паровой утюг и получил оценку от 1 500 до 7 500 P . Какая гипотеза не будет отвергнута на уровне значимости 5% при использовании двусторонней альтернативной гипотезы?

A. Дисперсия цены на паровой утюг равна 3 000 P

B. Средняя цена парового утюга равна 2 700 P

C. Средняя цена парового утюга равна 1 300 P

D. Стандартное отклонение цены на паровой утюг равно 3 000 P

6. 15 учеников прошли тест. Среднее значение набранных баллов у 10 учеников получилось 65. Оказалось, что 5 работ преподаватель забыл внести в ведомость. После их внесения среднее по всем 15 работам стало равно 90 баллам. Максимум за тест можно набрать 100 баллов. Какие выводы можно сделать о добавленных пяти работах?

A. Преподаватель забыл добавить работы отличников, набравших максимум

B. Преподаватель неправильно рассчитал среднее

C. Среднее число баллов в добавленных пяти работах равно 90

D. Стандартное отклонение набранных баллов одним учеником в среднем равно 25 баллам

7. Сопоставьте следующие данные о студентах, собранные в рамках опроса, и типы признаков, которыми могут быть описаны эти данные.

Пол (мужской/женский) — категориальный

Средний балл (от 0 до 10) — непрерывный

Количество пропущенных занятий (целое число) — дискретный

Уровень удовлетворенности (цифра от 1 до 5) — порядковый

8. Оля из города N опросила своих друзей в социальных сетях, готовы ли они купить лабubu. У Оли 700 друзей в соцсетях, из них 643 проживают с ней в одном городе. Результаты опроса такие: 441 человек готовы купить лабubu, 59 человек — нет, 100 затруднились ответить. Какие выводы можно сделать по результатам опроса?

A. В среднем 73,5% населения города N готовы купить лабubu

B. В среднем 63% населения города N готовы купить лабubu

C. Полученной информации недостаточно для открытия магазина на всероссийском маркетплейсе

D. В городе N примерно 68% готовы купить лабubu

9. Ваш друг рассматривает инвестиционные проекты для вложения сбережений в размере 1 000 коинов. Он рассматривает банковский депозит на 6 месяцев под годовую ставку доходности 12% и частный проект с ежемесячной доходностью в ближайшие 6 месяцев +2%, +3%, -5%, +1%, +0,5%, +5%. Куда вы посоветуете другу вложить свои сбережения? Предполагается, что процентный доход ежемесячно выплачивается сразу после начисления (капитализация процентного дохода отсутствует).

A. Вложиться в частный проект: доход в нем за полгода будет больше, чем на депозитном вкладе

B. Открыть депозит: он принесет больший доход, чем частный проект

C. Оба варианта равнозначны

D. Недостаточно информации для принятия решения, так как нужно учитывать капитализацию

10. Заводы автомобильной краски сговорились и последние 10 лет производили преимущественно черную краску. Доля объема поставленной автоконцернам черной краски составила 80%. На остальные 20 цветов пришлось оставшиеся 20%. Конверсия в покупку автомобиля каждого цвета на рынке одинаковая. Маркетологи опубликовали исследование, в котором привели набор тезисов.

Выберите один или несколько верных тезисов маркетологов.

A. Потребители предпочитают черный цвет автомобиля остальной палитре

B. Доля красных машин составляет 1% или более

C. Потребителям не нравятся яркие цвета автомобилей

D. Большинство продающихся автомобилей черного цвета

E. Нет верных ответов среди вышеперечисленных вариантов

11. В колл-центре желтого банка работают только два сотрудника: Маша и Степа. Маша тратит на одного клиента 1 минуту, Степа — 3 минуты. Время рабочей смены — 3 часа. Сколько в среднем длится время обслуживания одного клиента в колл-центре?

A. 2 минуты

B. 45 секунд

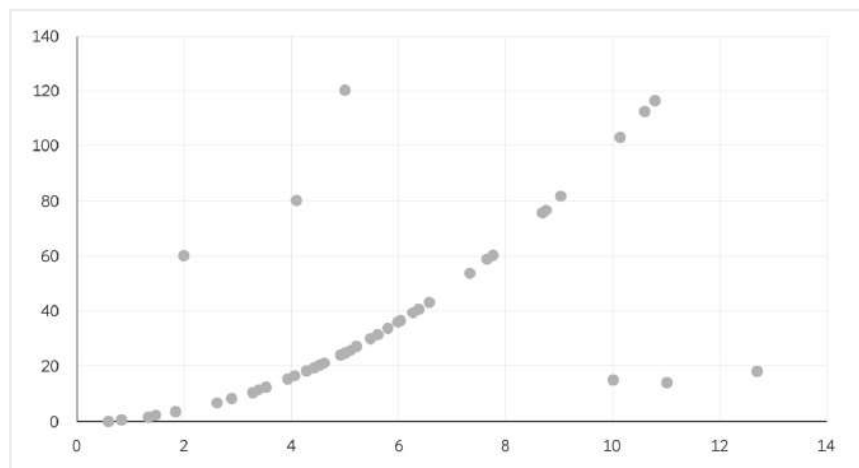
C. 1 минуту

D. 1 минуту 30 секунд

12. В лавке у разбойника лежит 15 острых ножей, расположенных так, что каждый следующий нож на 1 см длиннее предыдущего. Разбойник рассказал, что среднее арифметическое длин первых 11 ножей равно 12, а среднее арифметическое последних 11 ножей — 16. Какова длина наиболее длинного ножа у разбойника?

Ответ: 21

13. Директор цеха составил график зависимости оплаты труда сотрудников всех профессий в цеху (в тысячах рублей по вертикальной оси) от длины рабочей смены (в часах по горизонтальной оси). Выберите верные утверждения, которые соответствуют построенному графику.



A. Между длиной рабочей смены и оплатой существует линейная положительная связь

B. Коэффициент корреляции Спирмена между длиной рабочей смены и оплатой отрицательный

C. В цеху трое перерабатывают и им недоплачивают, а трое недорабатывают и им переплачивают

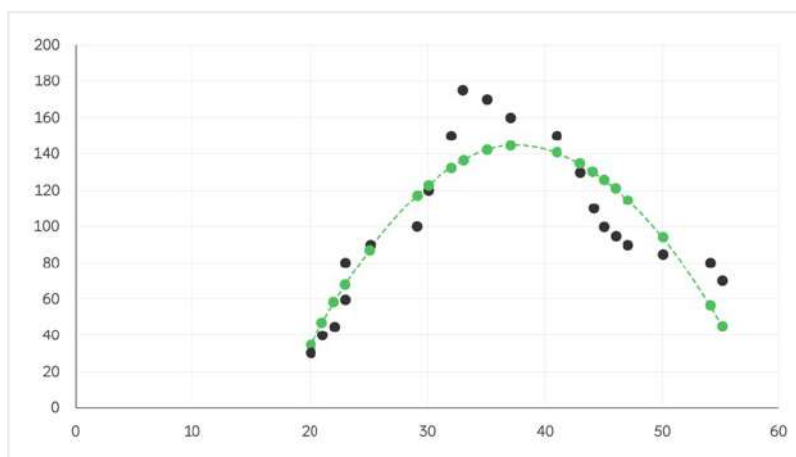
D. Зависимость между оплатой труда и длительностью рабочей смены можно описать нелинейной функцией, исключив выбросы

14. Аналитики компании решили оценить зависимость заработной платы сотрудников от их возраста. Для этого они оценили регрессионную модель:

$$\text{Зарплата} = -344.78 + 25.76 * \text{возраст} - 0.34 * \text{возраст}^2$$

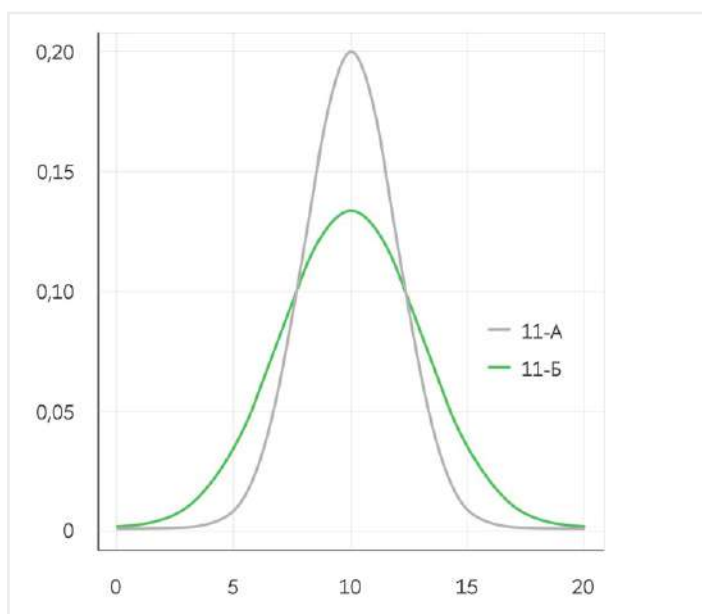
Коэффициент детерминации R^2 модели оказался равен 0,79.

Результаты аналитики представили на графике: черным цветом — фактические данные, зеленым цветом — оцененную модель.



Ответ: 38

15. Саша построил распределения количества игр в библиотеке магазина «Стем» для учащихся 11 «А» и 11 «Б». В результате он получил такой график распределений.



Какие выводы может сделать Саша из этого графика?

А. В среднем в обоих классах по 10 игр у каждого учащегося

В. В 11 «Б» разброс числа игр у учащихся больше, чем в 11 «А»

C. Медиана в 11 «А» не равна моде в 11 «Б»

D. Больше половины учащихся 11 «А» имеют в своей библиотеке 7 и более игр

16. Фанатам творчества Михаила Афанасьевича Булгакова задали вопрос: сколько литров подсолнечного масла нужно разлить Аннушке в зависимости от роста повстречавшегося ей незнакомца? Результаты опроса представили в таблице с усредненными значениями.

Рост	Средний объем масла (по данным опроса)
Низкий	1,5 л
Средний	2 л
Высокий	1 л
Очень высокий	3 л

При помощи каких статистических инструментов можно подтвердить или опровергнуть наличие зависимости между объемом пролитого масла и ростом незнакомца?

A. Коэффициент корреляции Спирмена

B. F-тест на равенство дисперсий

C. Коэффициент корреляции R^2

D. T-тест

17. Аналитик получил датасет со следующими переменными: пол, возраст, количество братьев и (или) сестер, стиль вождения. Какие из этих переменных имеет смысл использовать в качестве зависимых, а какие — в качестве регрессоров?

A. Пол — зависимая переменная, а остальные в качестве регрессоров

B. Возраст — зависимая переменная, а остальные в качестве регрессоров

C. Количество братьев и (или) сестер — зависимая переменная, а остальные в качестве регрессоров

D. Стиль вождения в качестве зависимой переменной, а остальные в качестве регрессоров

18. Вам нужно визуализировать распределение брендов смартфонов у целевой аудитории вашего блога. Какой (-ие) тип (-ы) визуального представления вы выберете?

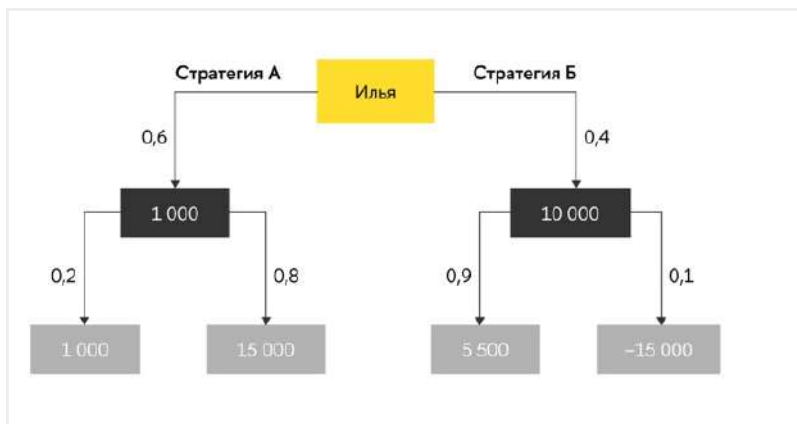
A. Диаграмма рассеяния

B. Столбчатая диаграмма

C. Гистограмма

D. Круговая диаграмма

19. Илья руководит стартапом по анализу данных. Ему нужно выбрать одну из дальнейших стратегий развития стартапа: А или Б. Обычно руководители стартапов с вероятностью 0,6 выбирают стратегию А и с вероятностью 0,4 — стратегию Б.



Каждая стратегия состоит из двух этапов. Стратегия А на первом этапе принесет 1 000 бонусов, а на втором — с вероятностью 0,2 еще 1 000 бонусов или с вероятностью 0,8 еще 15 000 бонусов. Стратегия Б на первом этапе принесет 10 000 бонусов, а на втором — с вероятностью 0,9 еще 5 500 бонусов или с вероятностью 0,1 Илья потеряет 15 000 бонусов.

Какую стратегию стоит выбрать Илье?

- A. А
 - B. Стратегии равноценны
 - C. Б
 - D. Недостаточно информации для принятия решения
20. Зарплата аналитика за последние 4 года менялась так: 150 000, 200 000, 230 000, 300 000 ₽. Чему равен средний темп прироста зарплаты аналитика в год в процентах? Укажите ответ одним числом без дополнительных знаков и символов, округлив до целого значения.

Ответ: 26

21. Перед вами таблица сопряженности двух признаков: пол и сфера работы.

	Образование	Медицина	Финансы
Муж.	350	215	460
Жен.	500	210	370

Какие выводы об имеющейся выборке можно сделать на основе этой таблицы?

- A. Вероятность, что случайно выбранный человек из выборки работает в сфере финансов, выше, чем вероятность, что он работает в образовании
- B. Мужчин меньше в выборке, чем женщин

С. Сфера образования и финансы одинаково представлены в выборке

Д. Больше трети мужчин выборки работают в сфере образования

Е. Вероятность работать в медицине выше среди мужчин, чем среди женщин

22. Ангелина работает врачом и сталкивается с пациентами, от которых можно заразиться тяжелыми заболеваниями. Вероятность остаться здоровой при контакте с зараженным равна 0,95. Раз в 6 дней в больницу, где работает Ангелина, попадает пациент с такой болезнью, и Ангелина с ним контактирует. Определите, через сколько дней вероятность, что Ангелина заразится болезнью, станет выше, чем вероятность не заразиться.

Укажите ответ одним числом без дополнительных знаков и символов, округлив в случае необходимости до целого значения.

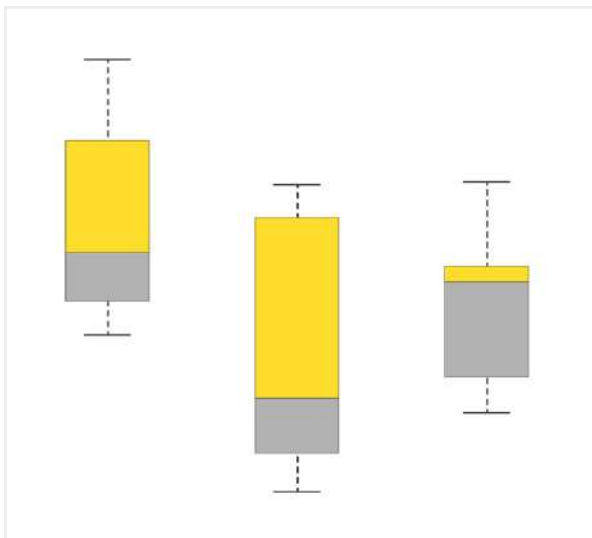
Ответ: 84

23. В соревнованиях по программированию участвовали 120 человек. Медиана набранных баллов среди участников — 50 баллов, а среднее значение баллов — 45. Организаторы решили добавить 20 дополнительных баллов самым лучшим участникам, чтобы поощрить их.

Какое минимальное число участников нужно наградить дополнительными баллами, чтобы средний балл стал равен медиане (50 баллов)?

Ответ: 30

24. Перед вами графики Boxplot для трех выборок с параметром длины усов = $1,5 \times \text{IQR}$.



Выберите верные утверждения, опираясь на эти графики:

— Дисперсия третьей выборки меньше, чем первых двух

— Медианное значение во второй выборке наименьшее из представленных

— В выборках нет выбросов, превышающих $1,5 \times \text{IQR}$

— Максимальное из всех значений по трем выборкам находится в первой из них

25. В школьном кружке по анализу данных провели опрос среди 100 учащихся. Каждый ученик должен был ответить на два вопроса:

- 1) Любите ли вы математику?
- 2) Любите ли вы информатику?

Анкеты обработали и получили следующие результаты:

- 60 человек ответили, что любят математику;
- 55 человек ответили, что любят информатику;
- 20 человек ответили, что не любят ни математику, ни информатику.

Сколько учеников любят и математику, и информатику?

Ответ: 35

26. Компания занимается онлайн-продажами. Каждый клиент проходит несколько этапов:

1. Посещение сайта.
2. Добавление товара в корзину.
3. Оформление заказа.
4. Оплата заказа.

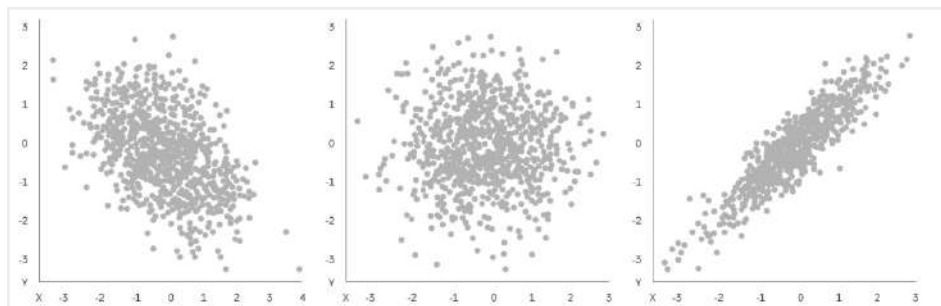
За один день аналитики собрали информацию:

- На сайт зашло 10 000 пользователей.
- 30% из всех, кто заходил на сайт, оформили заказ.
- 2 500 клиентов оплатили заказ.
- Только 50% тех, кто добавил товар в корзину, оплатили его.

Рассчитайте долю оформивших заказ клиентов среди пользователей, добавивших товар в корзину. Ответ округлите до второго знака после запятой.

Ответ: 0,6 или 60

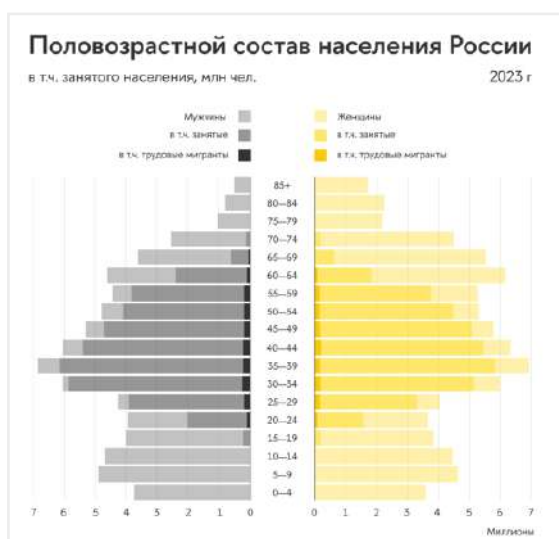
27. Перед вами три диаграммы рассеяния. Сопоставьте их с коэффициентами корреляции, которые получены на этих выборках. В ответ запишите буквы, в порядке расположения картинок.



- A. 0,5
- B. 0
- B. 0,8
- Г. -0,5

Ответ: 1Г, 2Б, 3В, 4А

28. Перед вами иллюстрация половозрастной пирамиды России с указанием работающего (занятого) населения. Выберите верные утверждения для этой картинки.



- Доля занятых женщин в возрасте 30—34 года выше, чем у мужчин в том же возрасте
- Среди мужчин трудовых мигрантов больше, чем среди женщин
- Максимальная доля работающего населения среди мужчин приходится на возрастную группу от 25 до 29 лет
- В возрасте 75—79 лет количество женщин примерно вдвое больше, чем мужчин
- После 75 лет население состоит из женщин менее чем на 50%

29. Выберите из списка все корректные методы борьбы с выбросами в данных

- A. Бутстрап
- B. Удаление 1% значений
- C. Удаление значений, которые отклоняются сильнее чем $3 \times \text{IQR}$
- D. Стандартизация исходной метрики

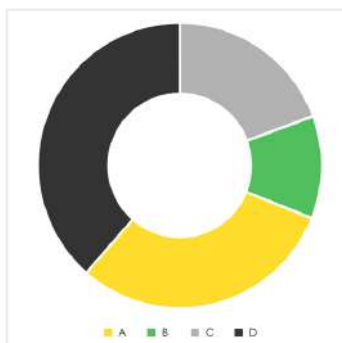
Вариант 2

- По выборке из 144 значений посчитали выборочную дисперсию. После этого все значения увеличили в 10 раз. Как изменится выборочная дисперсия, если вычислить ее по новым значениям выборки?

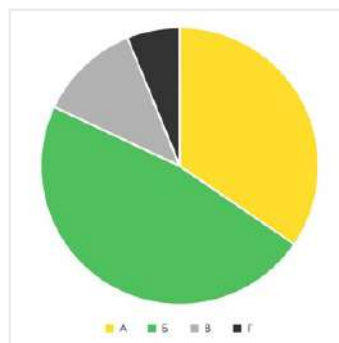
- Не изменится
- Увеличится в 10 раз
- Увеличится в 100 раз
- Изменится в 0,1 раза

- Сопоставьте названия типов визуализации с графиками:

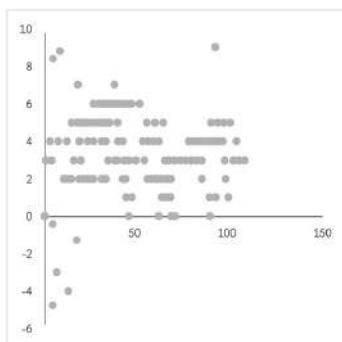
1.



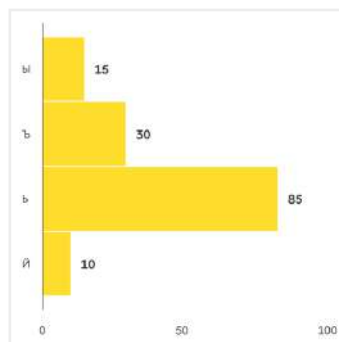
2.



3.



4.



- Кольцевая диаграмма
- Секторная диаграмма
- Диаграмма рассеяния
- Гистограмма

3. Исследователи проверили гипотезу о среднем значении с использованием Z-теста. Уровень значимости установлен равным 0,01. Как интерпретировать такое значение уровня значимости?
- A. Вероятность того, что альтернативная гипотеза верна, равна 1%
 - B. Допустимая вероятность отвержения альтернативной гипотезы, если она на самом деле верна, равна 1%
 - C. Если нулевая гипотеза не верна, в 1% случаев мы допускаем, что можем ошибочно принять ее при многократном повторении теста на семплированных данных из одной генеральной совокупности
 - D. Вероятность ошибки первого рода равна 1%
 - E. Если тест показал $p\text{-value} = 0,1$, вероятность того, что нулевая гипотеза верна, равна 1%
4. Сотрудники отдела прогнозирования и анализа данных изучили ежемесячные данные по ценам на рис и гречку за 2021 год. Средние цены оказались одинаковыми. Значения дисперсий: 64 ₺^2 для риса и 82 ₺^2 для гречки. Какое из утверждений верное?
- A. Цена риса в 2021 году была более изменчива, чем цена гречки
 - B. Цена гречки в 2021 году была более изменчива, чем цена риса
 - C. Ежемесячные цены на рис и гречку всегда одинаковы
 - D. Стандартные отклонения цен на рис и гречку одинаковы
5. Влад построил 80%-й доверительный интервал для средней цены пары кроссовок «Майк» и получил оценку от 7 000 до 15 000 ₺. Какая гипотеза не отвергнется на 20%-м уровне значимости при использовании двусторонней альтернативной гипотезы?
- A. Дисперсия цены пары кроссовок «Майк» равна 4 000 ₺
 - B. Средняя цена пары кроссовок «Майк» равна 2 700 ₺
 - C. Средняя цена пары кроссовок «Майк» равна 9 535 ₺
 - D. Стандартное отклонение цены пары кроссовок «Майк» равно 4 000 ₺
6. В автошколе после тестирования инструктор оценил среднее значение баллов за тест, набранных 10 сдавшими, и получил 85 баллов. Оказалось, что 5 работ инструктор забыл внести в ведомость. После их внесения среднее по всем 15 работам оказалось равно 90 баллам. Максимум за тест можно было набрать 100 баллов. Что можно сказать о добавленных пяти работах?
- A. Инструктор забыл добавить работы учеников, набравших максимум баллов
 - B. Инструктор неправильно рассчитал среднее
 - C. Среднее число баллов в добавленных пяти работах равно 99
 - D. Стандартное отклонение набранных баллов одним учеником в среднем равно 10 баллам

7. Сопоставьте следующие данные о сотрудниках, собранные в рамках опроса, и типы признаков, которыми могут быть описаны эти данные.

Должность (менеджер/разработчик/аналитик) — категориальный

Процент надбавки от фиксированной заработной платы — непрерывный

Количество рабочих дней в командировках — дискретный

Уровень стресса (целое число от 1 до 10) — порядковый

8. Алексей из Москвы провел опрос среди своих подписчиков в соцсети, чтобы узнать, будут ли они покупать новый кофейный напиток. Всего у Алексея 1 000 подписчиков, 850 из которых живут в Москве. Результаты опроса такие:

— 620 человек ответили «Да, куплю»

— 150 человек сказали «Нет»

— 230 человек затруднились с ответом

Какие выводы можно сделать по результатам этого опроса?

A. В среднем в России 62% потребителей готовы купить новый кофейный напиток

B. В среднем в России 15% потребителей готовы купить новый кофейный напиток

C. В Москве примерно 73% потребителей готовы купить новый кофейный напиток

D. Полученных данных недостаточно для запуска продаж по всей стране

9. У Кати есть 10 000 ₽, и она думает об открытии депозита на 6 месяцев под годовую ставку 20%. Сотрудники банка предложили ей инновационный продукт, ежемесячная доходность которого будет такая: +10%, +1%, −3%, +12%, +1%, −2%. Куда стоит вложить средства Кате? Предполагается, что процентный доход ежемесячно выплачивается после начисления (капитализация процентного дохода отсутствует).

A. Стоит выбрать инновационный продукт: так Катя сможет заработать через полгода больше, чем на вкладе в депозит

B. Стоит открыть депозит: он принесет больший доход, чем инновационный продукт

C. Все равно, куда вкладываться

D. Недостаточно информации для принятия решения, так как нужно учитывать капитализацию

10. Производители чехлов для смартфонов последние 8 лет по негласному соглашению выпускали преимущественно черные чехлы. В результате доля черных чехлов в поставках ретейлерам составила 85%, а на все остальные цвета (белый, синий, красный и др.) пришлось только 15%. Конверсия в покупку чехлов каждого цвета одинаковая. Аналитики рынка опубликовали исследование с несколькими утверждениями. Какие из этих утверждений можно считать достоверными на основе приведенных данных?

A. Покупатели предпочитают черные чехлы для смартфонов всем другим цветам

B. Доля синих чехлов составляет 2% или более

С. Потребителям не нравятся яркие чехлы для телефонов

D. Большинство проданных чехлов черного цвета

E. Нет верных ответов среди вышеперечисленных вариантов

11. На космической станции связи работают два робота-оператора: R2 и D2. R2 тратит 2 минуты, чтобы обработать один запрос. D2 тратит на один запрос 6 минут. Их рабочая смена длится 4 часа. Сколько в среднем длится обработка запроса на космической станции?

A. 3 минуты

B. 3 минуты 30 секунд

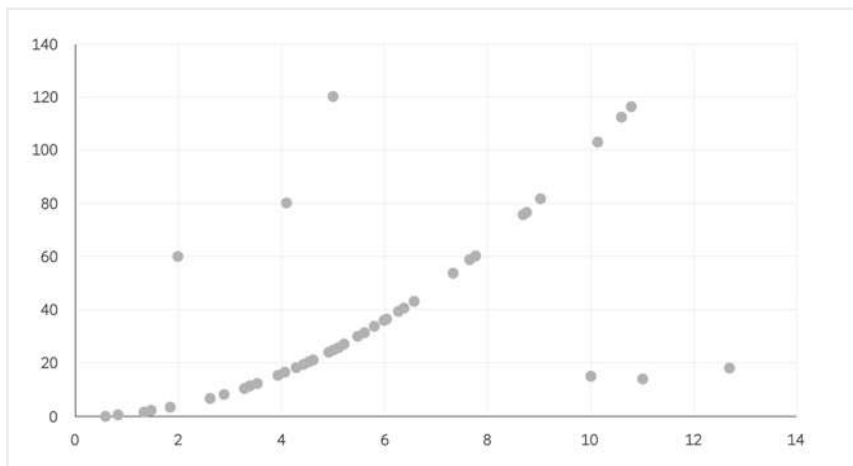
C. 5 минут

D. 1 минуту 30 секунд

12. В сказочной стране Конфетляндии у волшебника есть 15 волшебных конфет, выстроенных в ряд так, что каждая следующая конфета весит на 1 г больше предыдущей. Известно, что средний вес первых 11 конфет равен 10 г, а средний вес последних 11 конфет — 14 г. Найдите вес самой тяжелой конфеты в этом ряду.

Ответ: 19

13. Директор цеха по производству шпuleк составил график зависимости оплаты труда сотрудников всех профессий в цеху (в тысячах рублей по вертикальной оси) от длины рабочей смены (в часах по горизонтальной оси). Выберите верные утверждения, которые соответствуют построенному графику.

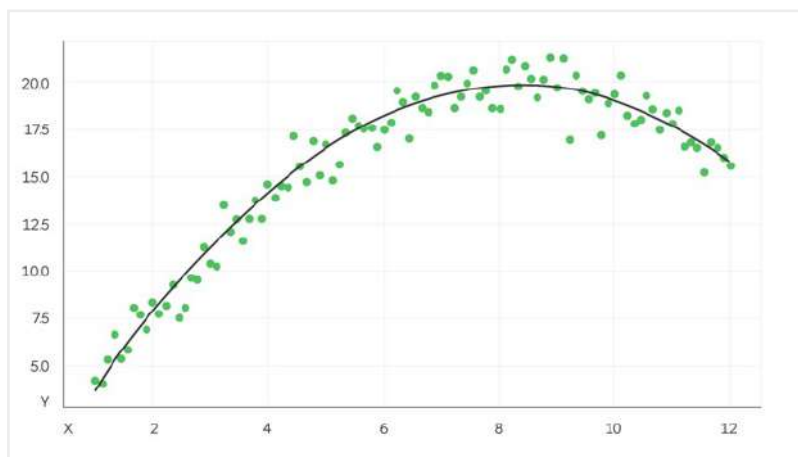


- A. Между длиной рабочей смены и оплатой есть линейная отрицательная связь
- B. Коэффициент корреляции Спирмена между длиной рабочей смены и оплатой нулевой
- C. В цеху трое перерабатывают и им недоплачивают
- D. Зависимость между оплатой труда и длительностью рабочей смены можно описать нелинейной функцией, исключив выбросы

14. Аграрная компания провела исследование, чтобы определить, как количество внесенных удобрений (килограмм на гектар) влияет на урожайность винограда (килограмм с гектара). Для анализа данных оценили квадратичную регрессионную модель:

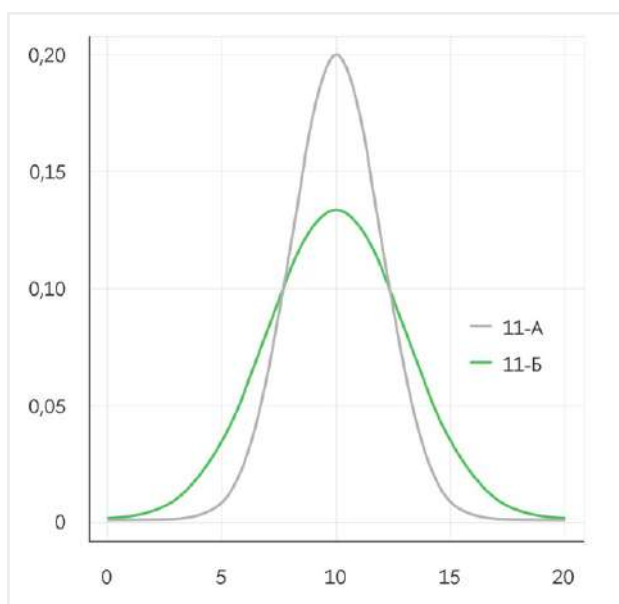
$$\text{Урожайность} = -0,3 \times \text{Удобрения}^2 + 5 \times \text{Удобрения} - 1$$

Коэффициент детерминации $R^2 = 0,82$. Определите оптимальное количество удобрений, при котором урожайность достигает максимума, согласно оцененной модели. На графике зеленым цветом показаны фактические данные, черным — оцененная зависимость.



Ответ: 8

15. Дима построил распределения количества игр в библиотеке магазина «Стем» для учащихся 11 «А» и 11 «Б». Получился такой график распределений.



- A. В обоих классах среднее число игр у одного учащегося равно 10
B. В 11 «Б» разброс числа игр у учащихся меньше, чем в 11 «А»

C. Медиана в 11 «А» равна моде в 11 «Б»

D. Больше половины учащихся 11 «А» обладают 8 и более играми

16. Любителей творчества Чарльза Диккенса спросили: сколько чашек чая выпивает мисс Хэвишем в зависимости от количества гостей на ее знаменитых чаепитиях? В результате получилась следующая таблица с усредненными значениями

Количество гостей	Среднее количество выпитых чашек чая (оцененное по данным опроса)
Мало	5 чашек
Средне	4 чашки
Много	3 чашки
Очень много	2 чашки

При помощи каких статистических инструментов можно подтвердить или опровергнуть наличие зависимости между количеством выпитых чашек чая и количеством гостей на чаепитии у мисс Хэвишем?

A. Коэффициент корреляции Спирмена

B. F-тест на равенство дисперсий

C. Z-тест

D. U-тест Манна — Уитни

17. Исследователь получил набор данных о вкусовых предпочтениях мороженого анкетированных со следующими переменными: цвет упаковки, возраст потребителя, уровень дохода и предпочитаемый вкус мороженого. Какие из этих переменных имеет смысл использовать в качестве зависимых, а какие в качестве независимых?

A. Цвет упаковки как независимая переменная, а остальные как зависимые

B. Возраст потребителя как зависимая переменная, а остальные как независимые

C. Уровень дохода как независимая переменная, а остальные как зависимые

D. Предпочитаемый вкус мороженого как зависимая переменная, а остальные как независимые

18. Специалист по флоре тропического леса хочет визуализировать распределение различных видов растений в исследуемой области. Какой тип визуализации лучше всего подходит для этой задачи?

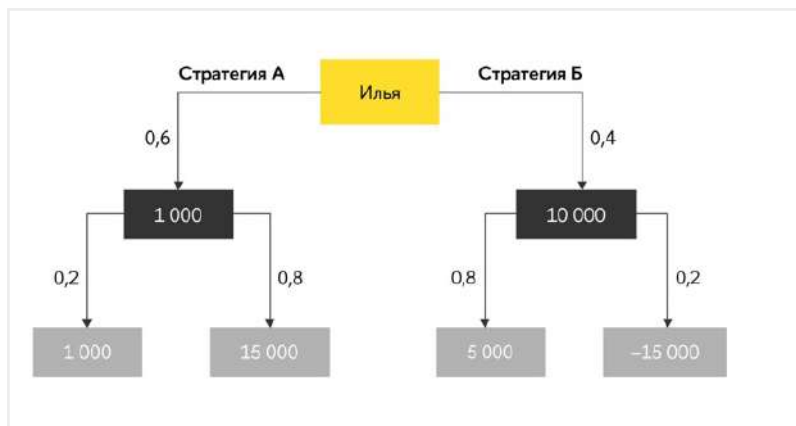
A. Тепловая карта

B. Линейный график

C. Диаграмма Венна

D. Столбчатая диаграмма

19. Илья руководит стартапом по анализу данных. Ему нужно выбрать одну из дальнейших стратегий развития стартапа: А или Б. Обычно руководители стартапов с вероятностью 0,6 выбирают стратегию А и с вероятностью 0,4 — стратегию Б.



Каждая стратегия состоит из двух этапов. Стратегия А на первом этапе принесет 1 000 бонусов, а на втором — с вероятностью 0,2 еще 1 000 бонусов или с вероятностью 0,8 еще 15 000 бонусов. Стратегия Б на первом этапе принесет 10 000 бонусов, а на втором — с вероятностью 0,8 еще 5 000 бонусов или с вероятностью 0,2 Илья потеряет 15 000 бонусов.

Какую стратегию стоит выбрать Илье?

- A. А
- B. Стратегии равноценны
- C. Б
- D. Недостаточно информации для принятия решения
20. Владелец небольшого кафе фиксировал количество посетителей каждый год в течение последних 4 лет и получил такие данные по годам: 1 200 посетителей, 1 500, 1 800, 2 250. Чему равен средний темп прироста числа посетителей кафе за год в процентах? Укажите ответ одним числом без дополнительных знаков и символов, округлите ответ до целого.
- Ответ: 23
21. Аналитики исследовали, каких домашних животных предпочитают жители города. Получилась такая таблица сопряженности, отражающая связь между типом жилья и предпочтением вида питомца.

Тип жилья	Кошки	Собаки	Рыбки
Квартира	250	180	120
Частный дом	150	320	80

Известно также, что все жители города держат какого-нибудь питомца!

A. Вероятность, что случайно выбранный житель держит собаку, выше, чем вероятность, что он держит кошку

B. Среди жителей частных домов больше тех, кто держит рыбок, чем среди жителей квартир

C. Кошки и рыбки одинаково популярны среди всех жителей

D. Больше половины жителей квартир предпочитают держать кошек

E. Вероятность держать собаку выше среди жителей частных домов, чем среди жителей квартир

22. Леша — профессиональный киберспортсмен и ежедневно участвует в онлайн-турнирах. Вероятность того, что его аккаунт не будет взломан при попытке взлома, составляет 0,98. Раз в 4 дня фиксируется попытка взлома аккаунтов всех участников турнира. Через сколько дней вероятность того, что аккаунт Алексея будет взломан, станет выше, чем вероятность, что этого не произойдет?

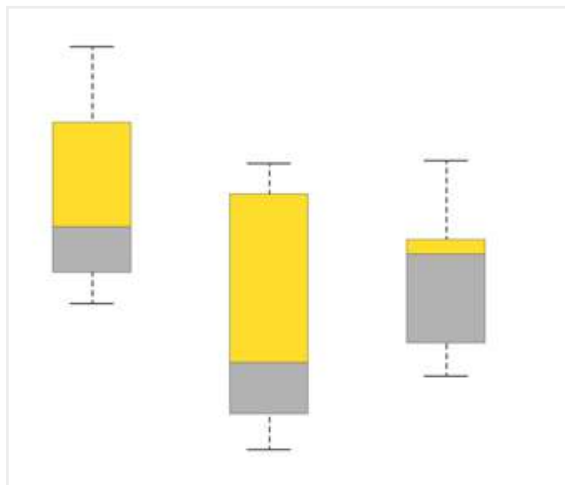
Укажите ответ одним числом без дополнительных знаков и символов, округлив в случае необходимости до целого значения.

Ответ: 140

23. В выборке из 100 значений медиана равна 50, а среднее — 45. Вася может увеличить X чисел в выборке на 50. Какого минимального X достаточно, чтобы среднее стало равно медиане, равной 50?

Ответ: 10

24. Перед вами графики Boxplot для трех выборок с параметром длины усов = $1,5 \times \text{IQR}$. Выберите верные утверждения.



— Дисперсия второй выборки больше, чем двух других

— Медианное значение в третьей выборке наименьшее из представленных графиков

— В выборках нет выбросов, превышающих $1,5 \times \text{IQR}$

— Максимальное значение из всех значений по трем выборкам находится в первой выборке

25. 100 ученикам школы задали два вопроса:

- 1) Занимаетесь ли вы спортом?
- 2) Играете ли вы на музыкальных инструментах?

По результатам опроса 50 учеников занимаются спортом; 45 учеников играют на музыкальных инструментах; 25 учеников не занимаются ни спортом, ни музыкой.

Сколько учеников и занимаются спортом, и играют на музыкальных инструментах?

Ответ: 20

26. Компания проводит онлайн-продажи, и каждый клиент проходит несколько этапов:

1. Посещение сайта.
2. Добавление товара в корзину.
3. Оформление заказа.
4. Оплата заказа.

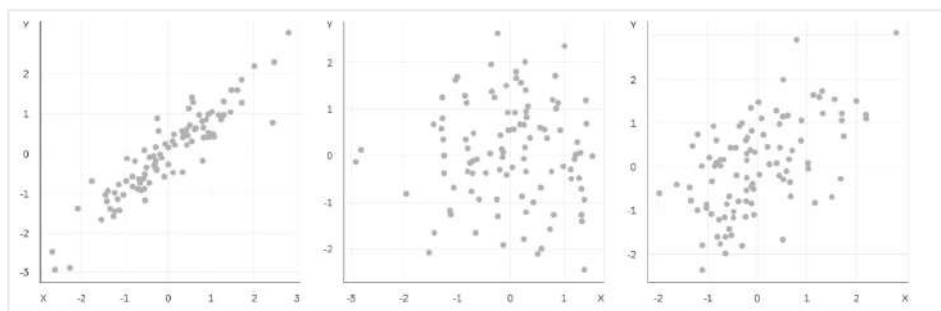
Аналитики собрали следующую информацию за один день:

- На сайт зашло 5 000 пользователей.
- Из всех, кто заходил на сайт, 20% пользователей оформили заказ.
- 500 клиентов оплатили заказ.
- Из всех, кто добавил товар в корзину, только 50% его оплатили.

Рассчитайте долю оформивших заказ клиентов среди пользователей, добавивших товар в корзину. Ответ округлите до второго знака после запятой.

Ответ: 1 или 100

27. Даны три диаграммы рассеяния. Сопоставьте их с коэффициентами корреляции, которые получены на этих выборках. В ответ запишите буквы в порядке расположения картинок.



- A. -0,4
- B. 0
- B. 0,9
- Г. 0,4

Ответ: 1В, 2Б, 3Г, 4А

28. Перед вами иллюстрация половозрастной пирамиды России с указанием работающего (занятого) населения. Выберите верные утверждения для этой картинки.



- Доля занятых женщин в возрасте 30—34 года ниже, чем у мужчин в том же возрасте
- Среди мужчин трудовых мигрантов меньше, чем среди женщин
- Максимальная доля работающего населения среди мужчин приходится на возрастную группу от 30 до 34 лет
- В возрасте 75—79 лет количество женщин примерно вдвое больше, чем мужчин
- После 75 лет население состоит из женщин более чем на 50%

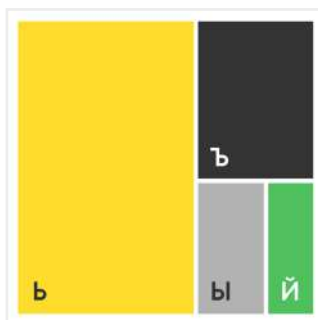
29. Выберите все варианты ответов, в которых перечислены корректные методы борьбы с выбросами в данных.

- A. Бутстрап
- B. Использование Т-тестов вместо Z-тестов
- C. Удаление значений, которые отклоняются сильнее чем $1,5 \times \text{IQR}$
- D. Нормализация исходных данных

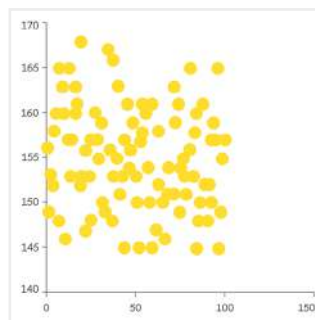
Вариант 3

- По выборке из 225 значений посчитали выборочную дисперсию. После этого все значения увеличили в 3 раза. Как изменится выборочная дисперсия, если вычислить ее по новым значениям выборки?
 - Не изменится
 - Увеличится в 3 раза
 - Увеличится в 6 раз
 - Увеличится в 9 раз
 - Изменится примерно в 1,732 раза
- Сопоставьте названия типов визуализации с графиками:

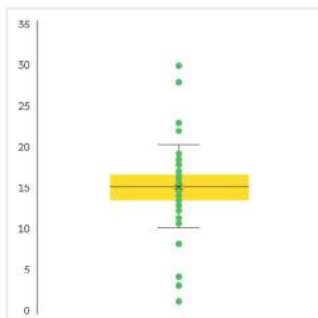
1.



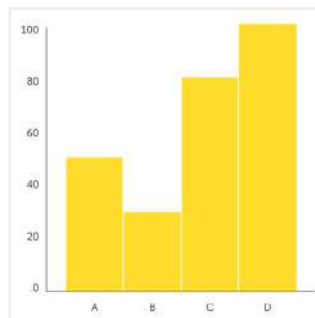
2.



3.



4.



1. Древовидная диаграмма

2. Диаграмма рассеяния

3. Ящик с усами

4. Гистограмма

3. Исследователи проверили гипотезу о среднем значении с использованием Т-теста. Уровень статистической значимости установлен равным 0,05. Выберите верные варианты ответов.

A. Вероятность того, что альтернативная гипотеза верна, равна 5%

B. Допустимая вероятность принятия нулевой гипотезы, если она на самом деле не верна, равна 95%

C. Если нулевая гипотеза верна, в 5% случаев мы допускаем, что можем ошибочно отвергнуть ее при многократном повторении теста на семплированных данных из одной генеральной совокупности

D. Вероятность ошибки первого рода равна 5%

E. Если тест показал $p\text{-value} = 0,054$, вероятность того, что альтернативная гипотеза верна, равна 5%

4. Сотрудники отдела прогнозирования и анализа данных изучили ежемесячные данные по ценам на пшеницу и ячмень за 2022 год. Средние цены оказались одинаковыми. Значения дисперсий: 100 ₺^2 для пшеницы и 121 ₺^2 для ячменя. Какое утверждение верное?

A. Цена пшеницы в 2022 году была более изменчива, чем цена ячменя

B. Цена ячменя в 2022 году была более изменчива, чем цена пшеницы

C. Ежемесячные цены на пшеницу и ячмень всегда одинаковы

D. Стандартные отклонения цен на пшеницу и ячмень одинаковы

5. Маша построила 99%-й доверительный интервал для средней цены на электрический чайник и получила оценку от 1 000 до 5 000 ₺. Какая гипотеза не будет отвергнута на уровне значимости 1% при использовании двусторонней альтернативной гипотезы?

A. Средняя цена электрического чайника равна 4 500 ₺

B. Средняя цена электрического чайника равна 9 000 ₺

C. Стандартное отклонение цены на электрический чайник равно 2 000 ₺

D. Нет верного варианта ответа

6. В автошколе после тестирования инструктор оценил среднее значение баллов за тест, набранных 10 сдавшими, и получил 80 баллов. Оказалось, что 5 работ инструктор забыл внести в ведомость. После их внесения среднее по всем 15 работам оказалось равно 90 баллам. Максимум за тест можно было набрать 100 баллов. Что можно сказать о добавленных пяти работах?

A. Инструктор забыл добавить работы учеников, набравших максимум баллов

B. Инструктор неправильно рассчитал среднее

C. Среднее число баллов в добавленных пяти работах равно 99

D. Стандартное отклонение набранных баллов одним учеником в среднем равно 10 баллам

7. Сопоставьте следующие данные о квартирах, собранные в рамках опроса собственников, с типами признаков, которыми они могут быть описаны.

Количество балконов — дискретный

Тип дома (кирпичный или не кирпичный) — категориальный

Степень удаленности от метро (по шкале от 1 до 3) — порядковый

Высота потолков (в см) — непрерывный

8. Машенька решила узнать у своих друзей в социальной сети, хотят ли они купить эликсир невидимости. Всего у Машеньки 256 онлайн-друзей, из которых 225 живут с ней в одном городе. По результатам опроса 113 друзей готовы купить эликсир, 19 друзей отказались, а 10 затруднились ответить. Какие выводы можно сделать из этого опроса?

A. В среднем в России 76,3% людей готовы купить эликсир невидимости

B. В среднем в России 44% людей готовы купить эликсир невидимости

C. В Машинном городе примерно 50% людей готовы купить эликсир невидимости

D. Нет верного ответа

9. Пирожок-Мастер рассматривает инвестиционные проекты для вложения своих сбережений в размере 5 000 золотых монет. Ему предложили два варианта: банковский депозит на 6 месяцев под годовую ставку доходности 10% и проект по выращиванию волшебных бобов, ежемесячная доходность которого в ближайшие 6 месяцев оценивается как +1%, +2%, -3%, +0,5%, +0,2%, +4%. Куда стоит вложить свои сбережения Пирожку-Мастеру? Предполагается, что процентный доход ежемесячно выплачивается сразу Пирожку-Мастеру и не вкладывается им в проект (капитализация процентного дохода отсутствует).

A. Стоит поддержать проект по выращиванию волшебных бобов: там можно больше заработать через полгода, чем на депозите

B. Стоит открыть депозит: он принесет больше, чем проект по выращиванию волшебных бобов

C. Все равно, куда вкладываться: доходность вариантов одинаковая

D. Недостаточно информации для принятия решения, так как нужно учитывать капитализацию

10. В сказочном королевстве заговор: последние 10 лет все гномы-мастера выпускают в основном только зонтики с узором в горошек. В результате доля зонтиков в горошек на полках в магазинах составила 90%, а на все остальные узоры пришлось только 10%. Конверсия в покупку зонтиков с разными расцветками одинаковая. Королевские аналитики опубликовали исследование, в котором привели набор тезисов. Выберите один или несколько верных тезисов аналитиков.

A. Жители королевства предпочитают зонтики в горошек всем остальным узорам

B. Доля зонтиков в полоску составляет не менее 1%

C. Жителям королевства не нравятся яркие узоры на зонтиках

D. Большинство продающихся зонтиков в королевстве имеют узор в горошек

E. Нет верных ответов среди вышеперечисленных вариантов

11. В колл-центре работают два оператора: Алина и Борис. Алина тратит на одного клиента 10 минут, а Борис, который любит подробно отвечать на каждый вопрос, — 40 минут. Их рабочая смена длится 4 часа. Сколько в среднем длится обслуживание одного клиента в колл-центре?

A. 20 минут

B. 16 минут

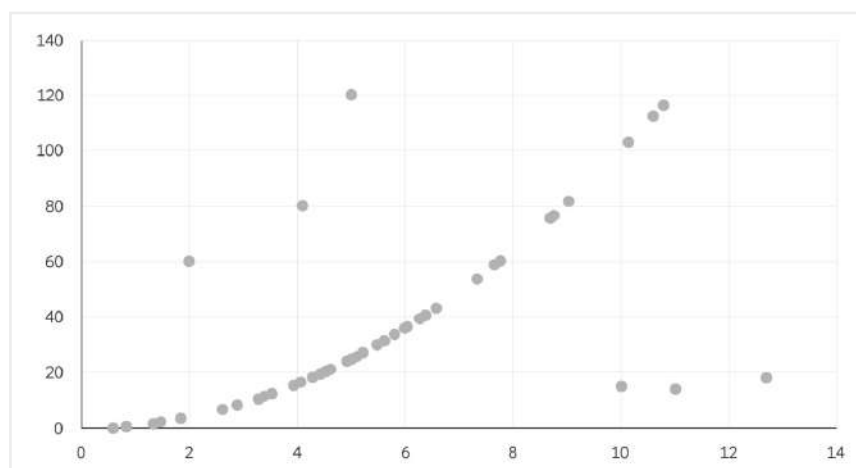
C. 8 минут

D. 30 минут

12. В личной библиотеке Даши есть 17 книг, выстроенных на полке так, что каждая следующая книга на 1 страницу толще предыдущей. Даша рассказала, что среднее арифметическое чисел страниц первых 7 книг из них равно 47, а среднее арифметическое следующих 7 — 54. Сколько страниц в самой толстой книге в библиотеке Даши?

Ответ: 60

13. Директор цеха составил график зависимости оплаты труда сотрудников всех профессий в цеху (в тысячах рублей по вертикальной оси) от длины рабочей смены (в часах по горизонтальной оси). Выберите верные утверждения, которые соответствуют построенному графику.



A. Между длиной рабочей смены и оплатой есть нелинейная положительная связь

B. Коэффициент корреляции Спирмена между длиной рабочей смены и оплатой равен нулю

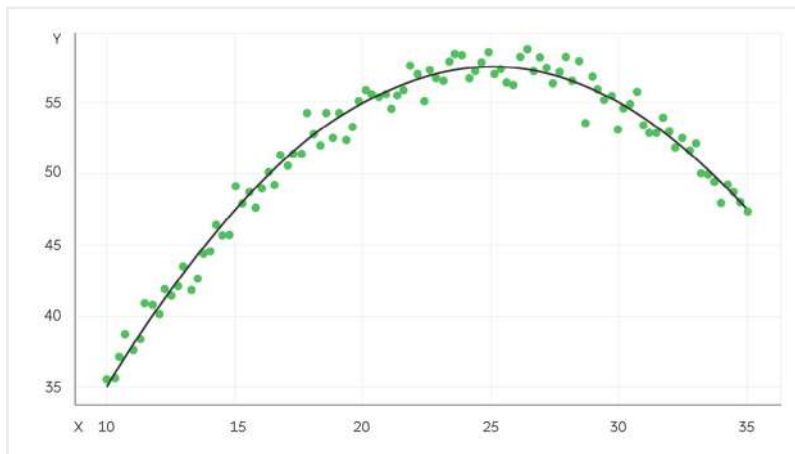
C. В цеху трое недорабатывают и им переплачивают

D. Зависимость между оплатой труда и длительностью рабочей смены нельзя описать нелинейной функцией, исключив выбросы

14. Спортивные аналитики изучили, как температура воздуха во время забега влияет на среднюю скорость марафонцев (в километрах в час). Для этого они оценили квадратичную модель:

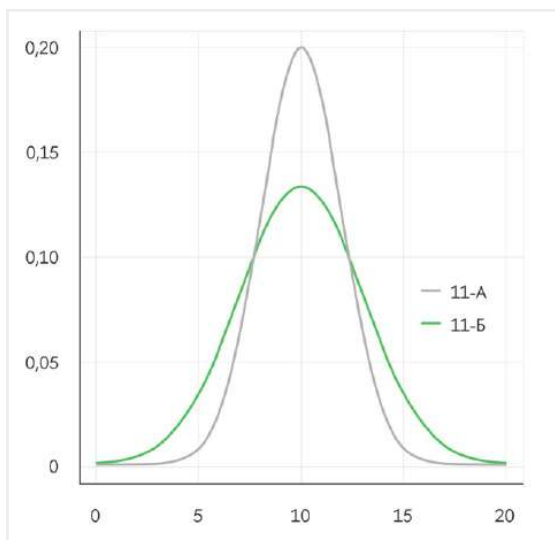
$$\text{Средняя скорость} = -0,1 \times \text{Температура}^2 + 5 \times \text{Температура} - 5$$

Коэффициент детерминации $R^2 = 0,76$. Определите температуру, при которой средняя скорость марафонцев достигает максимума. На графике: зеленым — наблюдаемые значения, черным — оцененная модель.



Ответ: 25

15. Денис построил распределения количества игр в библиотеке магазина «Стем» для учащихся 11 «А» и 11 «Б». В результате он получил такой график распределений.



- A. Среднее число игр у одного учащегося равно 5 в обоих классах
- B. В 11 «Б» разброс числа игр у учащихся больше, чем в 11 «А»
- C. Мода в 11 «А» равна медиане в 11 «Б»
- D. Больше половины учащихся 11 «А» обладают 15 и более играми

16. В сказочном лесу провели исследование и обнаружили: число ягод, собранных эльфами, зависит от времени суток. Данные были собраны с помощью волшебных часов, которые автоматически записывают количество ягод, собранных в разное время суток. В результате получили такое распределение средних значений.

Время суток	Среднее количество собранных ягод
Утро	15
День	12
Вечер	10
Ночь	8

При помощи каких статистических инструментов можно подтвердить или опровергнуть наличие зависимости между количеством собранных ягод и временем суток?

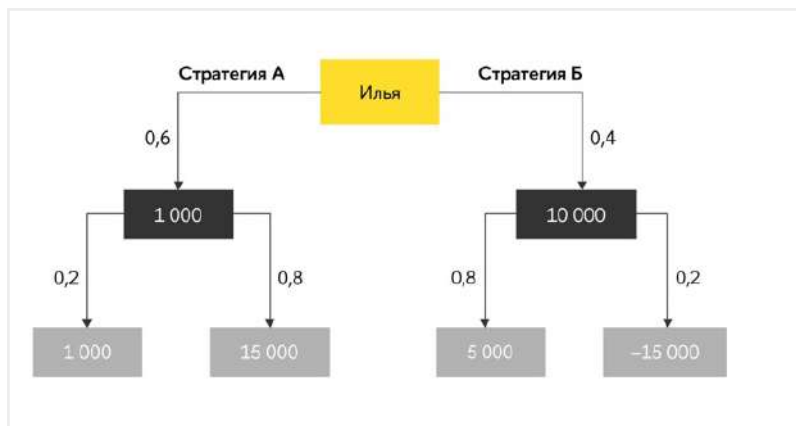
- A. Коэффициент корреляции Спирмена
- B. Тест Стьюдента
- C. Тест Манна-Уитни
- D. Логистическая регрессия
17. Ученый получил набор данных со следующими переменными: тип тренировки, возраст спортсмена, спортивные достижения и количество тренировок в неделю. Какие из этих переменных имеет смысл использовать в качестве зависимых, а какие — в качестве независимых?

- A. Тип тренировки как независимая переменная, а остальные как зависимые
- B. Возраст спортсмена как зависимая переменная, а остальные как независимые
- C. Спортивные достижения как зависимая переменная, а остальные как независимые
- D. Количество тренировок в неделю как независимая переменная, а остальные как зависимые

18. Ученый-эколог изучает зависимость уровня загрязнения воздуха от количества автомобилей на дорогах города. Он собрал данные о концентрации вредных веществ в воздухе и количестве автомобилей в разные дни недели. Какой тип визуализации лучше остальных подходит для анализа этой зависимости?

- A. Диаграмма рассеяния
- B. Столбчатая диаграмма
- C. Гистограмма
- D. Круговая диаграмма

19. Илья руководит стартапом по анализу данных. Ему нужно выбрать одну из дальнейших стратегий развития стартапа: А или Б. Обычно руководители стартапов с вероятностью 0,6 выбирают стратегию А и с вероятностью 0,4 — стратегию Б.



Каждая стратегия состоит из двух этапов. Стратегия А на первом этапе принесет 1 000 бонусов, а на втором — с вероятностью 0,2 еще 1 000 бонусов или с вероятностью 0,8 еще 15 000 бонусов. Стратегия Б на первом этапе принесет 10 000 бонусов, а на втором — с вероятностью 0,8 еще 5 500 бонусов или с вероятностью 0,2 Илья потеряет 15 000 бонусов.

Какую стратегию стоит выбрать Илье?

- A. А
- B. Стратегии равноценны
- C. Б
- D. Недостаточно информации для принятия решения
20. В Зеленограде последние 4 года фиксировали количество посаженных деревьев в рамках городской программы озеленения:
- 2021 год — 500 деревьев;
- 2022 год — 650 деревьев;
- 2023 год — 780 деревьев;
- 2024 год — 975 деревьев.
- Чему равен средний темп прироста количества посаженных деревьев за год в процентах? Укажите ответ одним числом без дополнительных знаков и символов, округлите его до целого.
- Ответ: 25
21. Аналитики изучали предпочтения транспорта среди жителей города. По результатам они составили таблицу сопряженности, которая отражает связь между возрастной группой и предпочитаемым видом транспорта.

Возрастная группа	Метро	Наземный транспорт	Личный автомобиль
Молодежь (18—25)	300	150	100
Взрослые (26—50)	250	200	300
Пожилые (51+)	100	250	50

Какие выводы можно сделать на основе данной таблицы?

- A. Вероятность, что случайно выбранный житель предпочитает наземный транспорт, выше, чем вероятность, что он предпочитает метро
- B. Среди молодежи меньше тех, кто пользуется личным автомобилем, чем среди взрослых
- C. Метро и личный автомобиль одинаково популярны среди всех жителей
- D. Больше половины пожилых людей предпочитают наземный транспорт
- E. Вероятность пользоваться личным автомобилем выше среди взрослых, чем среди молодежи

22. Иван — фермер, который выращивает редкий сорт яблок. Иногда в сад прилетают вредители и пытаются уничтожить урожай. Вероятность того, что попытка уничтожения урожая окажется неуспешной, составляет 0,92. Раз в 5 дней в его сад прилетают вредители, способные нанести ущерб. Через сколько дней вероятность того, что урожай Ивана будет уничтожен, превысит вероятность, что он останется невредимым?

Укажите ответ одним числом без дополнительных знаков и символов, округлив в случае необходимости до целого значения.

Ответ: 45

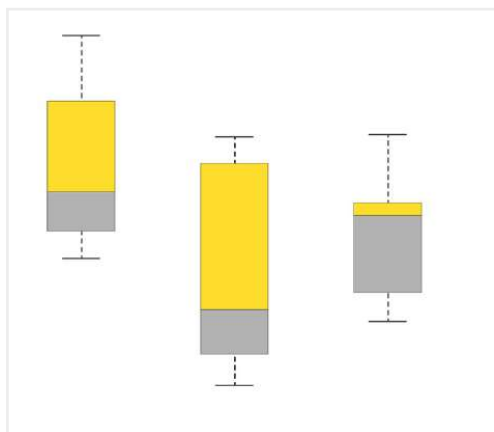
23. В компании работает 60 сотрудников. Медиана заработных плат составляет 60 тысяч рублей, а средняя зарплата — 58 тысяч рублей. Руководство решило повысить зарплату X сотрудникам с самой низкой зарплатой на 10 тысяч рублей, чтобы улучшить средние показатели по компании.

Какому минимальному количеству сотрудников (X) нужно повысить зарплату, чтобы средняя стала равна медиане (60 тысяч рублей)?

Ответ: 12

24. Перед вами графики Boxplot для трех выборок с параметром длины усов = $1,5 \times \text{IQR}$. Выберите верные утверждения.

- Среднее значение сильнее всего отклоняется от медианы в третьей выборке
- Медианное значение в третьей выборке наибольшее из представленных графиков
- В выборках есть выбросы, превышающие $1,5 \times \text{IQR}$
- Минимальное значение из всех значений по трем выборкам находится во второй выборке



25. В библиотеке провели опрос среди 120 посетителей. Им задали два вопроса:

- 1) Любите ли вы читать книги?
- 2) Увлекаетесь ли вы рисованием?

По результатам опроса 95 человек любят читать книги; 40 человек увлекаются рисованием; 15 человек не любят ни читать, ни рисовать.

Сколько посетителей любят и читать, и рисовать?

Ответ: 30

26. Компания проводит онлайн-продажи, и каждый клиент проходит несколько этапов:

1. Посещение сайта.
2. Добавление товара в корзину.
3. Оформление заказа.
4. Оплата заказа.

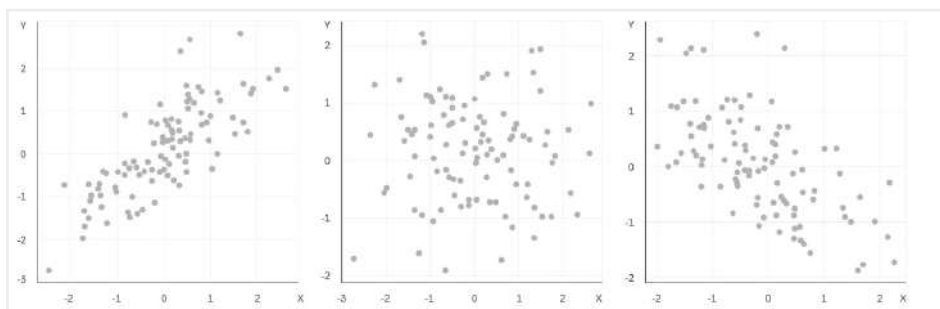
Аналитики собрали следующую информацию за один день:

- На сайт зашло 10 000 пользователей.
- Из всех, кто заходил на сайт, 80% пользователей оформили заказ.
- 4 500 клиентов оплатили заказ.
- Из всех, кто добавил товар в корзину, только 50% его оплатили.

Рассчитайте долю оформивших заказ клиентов среди пользователей, добавивших товар в корзину. Ответ округлите до второго знака после запятой.

Ответ: 0,89 или 88,88 или 88,89

27. Перед вами три диаграммы рассеяния. Сопоставьте их с коэффициентами корреляции, которые получены на этих выборках. В ответ запишите буквы в порядке расположения картинок.



A. 0,26

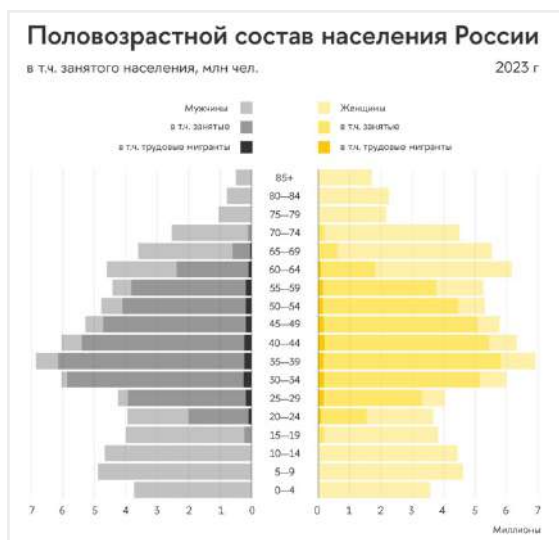
B. 0

B. 0,75

Г. -0,56

Ответ: 1B, 2Б, 3Г, 4А

28. Перед вами иллюстрация половозрастной пирамиды России с указанием работающего (занятого) населения. Выберите верные утверждения для этой картинки.



— Доля занятых женщин в возрасте 55—59 лет ниже, чем у мужчин в том же возрасте

— Среди мужчин трудовых мигрантов больше, чем среди женщин

— Максимальная доля работающего населения среди мужчин приходится на возрастную группу от 25 до 29 лет

— В возрасте 60—64 лет количество женщин примерно вдвое больше, чем мужчин

— После 75 лет население состоит из женщин менее чем на 50%

29. Выберите все варианты ответов, в которых перечислены корректные методы борьбы с выбросами в данных.

A. Бутстрап

B. Стандартизация данных

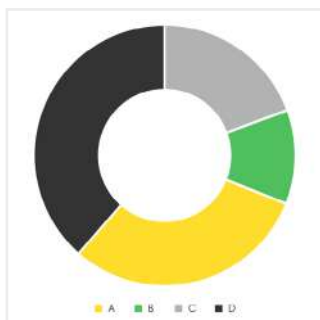
C. Удаление значений, которые отклоняются сильнее чем $1,5 \times \text{IQR}$

D. Дисперсионный анализ (ANOVA)

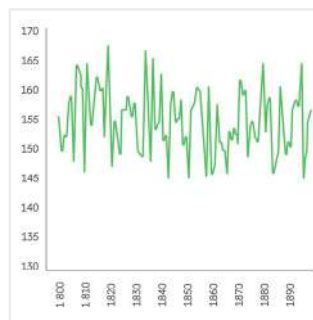
Вариант 4

- По выборке из 256 значений посчитали выборочную дисперсию. После этого все значения увеличили в 9 раз. Как изменится выборочная дисперсия, если вычислить ее по новым значениям выборки?
 - Не изменится
 - Увеличится в 81 раз
 - Увеличится в 9 раз
 - Уменьшится в 18 раз
 - Нет верного варианта ответа среди вышеперечисленных
- Сопоставьте названия типов визуализации с графиками:

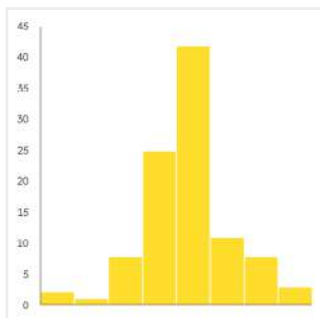
1.



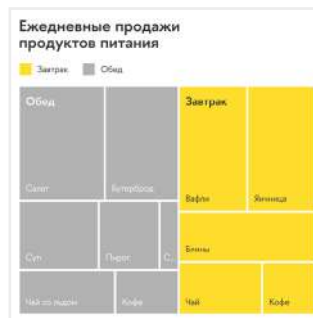
2.



3.



4.



- Кольцевая диаграмма
- График временного ряда
- Гистограмма
- Древовидная диаграмма

3. Исследователи проверили гипотезу о среднем значении с использованием Z-теста. Уровень значимости установлен равным 0,10. Как интерпретировать такое значение уровня значимости?

A. Вероятность того, что альтернативная гипотеза верна, равна 10%

B. Допустимая вероятность отвержения нулевой гипотезы, если она на самом деле верна, равна 10%

C. Если нулевая гипотеза верна, в 10% случаев мы можем ошибочно отвергнуть ее при многократном повторении теста на семплированных данных из одной генеральной совокупности

D. Вероятность ошибки второго рода равна 10%

E. Если тест показал $p\text{-value} = 0,97$, вероятность того, что нулевая гипотеза верна, равна 10%

4. Сотрудники отдела прогнозирования и анализа данных изучили ежемесячные данные по ценам на рис и гречку за 2023 год. Средние цены оказались одинаковыми. Значения дисперсий: 49 P^2 для сахара и 36 P^2 для соли. Какое утверждение верное?

A. Цена сахара в 2023 году была более изменчива, чем цена соли

B. Цена соли в 2023 году была более изменчива, чем цена сахара

C. Ежемесячные цены на сахар и соль всегда одинаковы

D. Стандартные отклонения цен на сахар и соль одинаковы

5. Коля построил 97%-й доверительный интервал для средней цены на микроволновую печь и получил оценку от 3 000 до 9 000 P . Какая гипотеза не отвергается на 3%-м уровне значимости при использовании двусторонней альтернативной гипотезы?

A. Дисперсия цены на микроволновую печь равна 4 000 P

B. Средняя цена микроволновой печи равна 5 500 P

C. Средняя цена микроволновой печи равна 2 500 P

D. Нет верного варианта ответа

6. В автошколе после тестирования инструктор оценил среднее значение баллов за тест, набранных 10 сдавшими, и получил 85,5 балла. Оказалось, что 5 работ инструктор забыл внести в ведомость. После их внесения среднее по всем 15 работам оказалось равно 90 баллам. Максимум за тест можно было набрать 100 баллов. Что можно сказать о добавленных пяти работах?

A. Инструктор забыл добавить работы учеников, набравших максимум

B. Инструктор неправильно рассчитал среднее

C. Среднее число баллов в добавленных пяти работах равно 99

D. Стандартное отклонение набранных баллов одним учеником в среднем равно 10 баллам

7. Сопоставьте данные о питомцах, собранные в рамках опроса в ветклинике, с типами признаков, которыми они могут быть описаны.

Пол питомца (самец/самка) — бинарный

Вид животного (кошка/собака/...) — категориальный

Размер (для определения переноски: крупные/средние/мелкие — порядковый

Вес питомца — непрерывный

8. Петя из Солнечногорска провел опрос среди своих друзей в социальных сетях, чтобы узнать, хотят ли они купить новый VR-гаджет. У Пети 515 онлайн-друзей, 127 из которых живут с ним в одном городе. По результатам опроса 44 друга готовы купить гаджет, 59 друзей ответили «нет», а 24 затруднились ответить. Какие выводы можно сделать из этого опроса?

A. В среднем в России 8,5% людей готовы купить новый VR-гаджет

B. В среднем в России 34% людей готовы купить новый VR-гаджет

C. В Солнечногорске примерно 34% людей готовы купить новый VR-гаджет

D. Петя может выдвинуть гипотезу, что в среднем большинство людей не готовы покупать новый VR-гаджет

9. Волшебник Мерлин задумался, куда бы ему спрятать свои 5 000 золотых драконовых монет, чтобы они приумножились. У него есть два варианта: волшебный сундук, который за 6 месяцев увеличивает содержимое на 8% годовых, и проект по выращиванию говорящих тыкв с ежемесячной доходностью +1%, +2%, −3%, +0,5%, +0,2%, +4% за каждый из следующих шести месяцев. Куда стоит вложить свои монеты Мерлину, если доход он будет получать ежемесячно и сразу тратить на волшебные зелья (капитализация процентного дохода отсутствует)?

A. Стоит поддержать проект по выращиванию говорящих тыкв: на нем можно заработать больше, чем с волшебным сундуком

B. Стоит выбрать волшебный сундук: он принесет больше монет, чем проект с тыквами

C. Все равно, куда вкладывать монеты: доходность вариантов одинаковая

D. Недостаточно информации для принятия решения, так как нужно учитывать капитализацию

10. В Грибном царстве последние 8 лет все эльфы-производители шапок выпускали преимущественно красные шапки. В результате доля красных шапок в поставках магазинам составила 83%, а на все остальные цвета (зеленый, синий, желтый и др.) пришлось только 17%. Конверсия в покупку шапок каждого цвета одинаковая. Лесные мудрецы опубликовали исследование с несколькими утверждениями. Какие из этих утверждений можно считать достоверными на основе приведенных данных?

A. Жители царства предпочитают красные шапки всем другим цветам

B. Доля зеленых шапок составляет 2% или более

C. Большинство шапок в царстве красные

D. Большинство продающихся шапок красного цвета

E. Нет верных ответов среди вышеперечисленных вариантов

11. В подводном колл-центре работают два сотрудника: осьминог Олли и рыба-клоун Немо. Олли благодаря своим восьми рукам справляется с одним звонком за 3 минуты. Немо, который любит рассказывать анекдоты, тратит на один звонок 9 минут. Их рабочая смена длится 6 часов. Сколько в среднем длится обслуживание одного клиента в колл-центре?

A. 2 минуты

B. 3 минуты

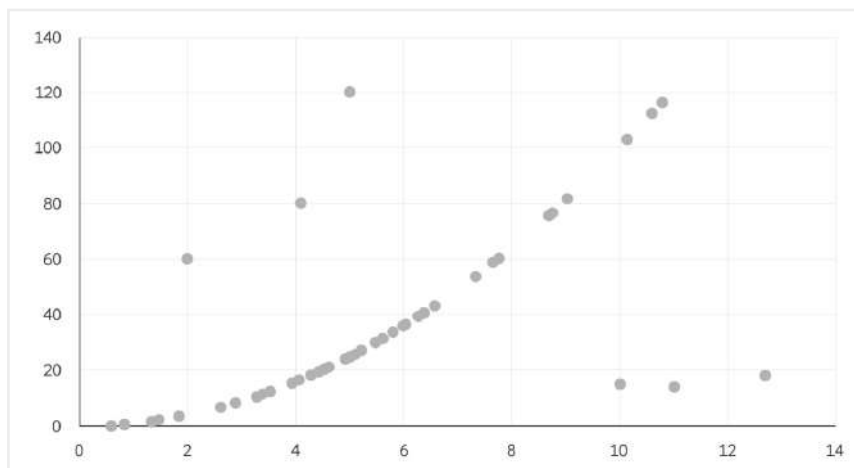
C. 4 минуты 30 секунд

D. 6 минут

12. В тайном хранилище у археолога найдено 17 загадочных артефактов, выстроенных в ряд так, что возраст каждого следующего артефакта на 1 год больше предыдущего. Известно, что средний возраст первых 11 артефактов равен 149 годам, а средний возраст последних 11 артефактов равен 155 годам. Найдите возраст самого старого артефакта в этом ряду.

Ответ: 160

13. Директор цеха составил график зависимости оплаты труда сотрудников (в тысячах рублей по вертикальной оси) от длины рабочей смены (в часах по горизонтальной оси). Выберите верные утверждения, которые соответствуют построенному графику.



A. Между длиной рабочей смены и оплатой есть линейная отрицательная связь

B. Коэффициент корреляции Спирмена между длиной рабочей смены и оплатой отрицательный

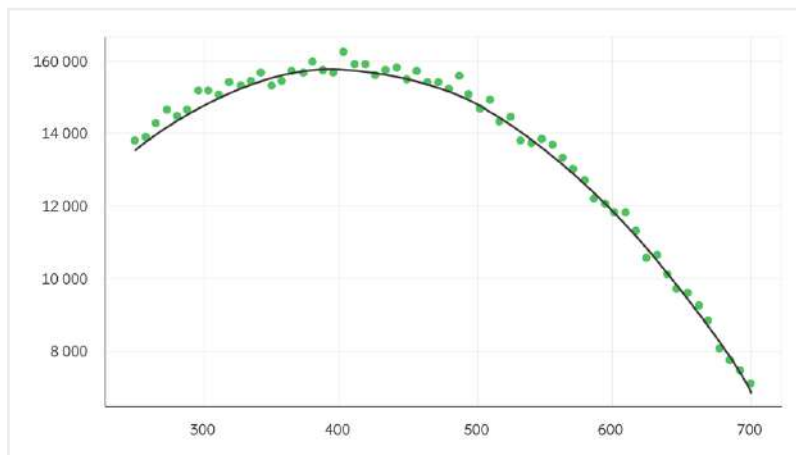
C. Меньше 25% сотрудников получают больше 100 тысяч рублей

D. Зависимость между оплатой труда и длительностью рабочей смены можно описать нелинейной функцией, исключив выбросы

14. Сеть кофеен решила оценить, как цена чашки капучино (в рублях) влияет на ежемесячную прибыль одной точки (в тысячах рублей). Для этого была построена квадратичная регрессионная модель:

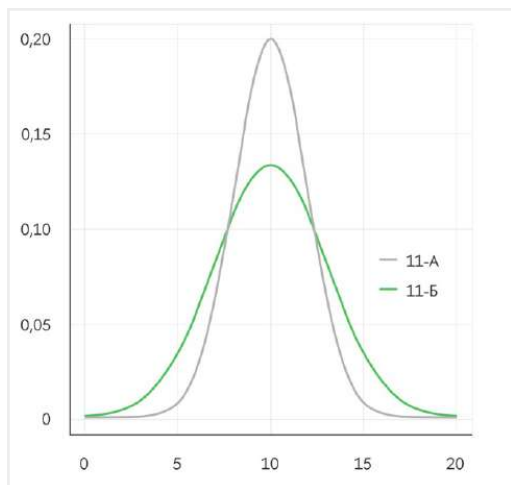
$$\text{Прибыль} = -0,1 \times \text{Стоимость чашки}^2 + 80 \times \text{Стоимость чашки} - 50$$

Коэффициент детерминации $R^2 = 0,81$. Определите цену на кофе, при которой прибыль будет максимальной, согласно оцененной модели. На графике зеленым цветом представлены фактические данные, черным — результат модели. Ответ округлите до целого.



Ответ: 400

15. Слава построил распределения количества игр в библиотеке магазина «Стем» для учащихся 11 «А» и 11 «Б». В результате он получил такой график распределений.



Что из этого графика может заключить Слава об учащихся?

- A. В обоих классах среднее число игр у одного учащегося равно 5
- B. В 11 «Б» разброс числа игр у учащихся меньше, чем в 11 «А»
- C. Мода в 11 «А» равна медиане в 11 «Б»
- D. Меньше половины учащихся 11 «А» обладают 11 и менее играми

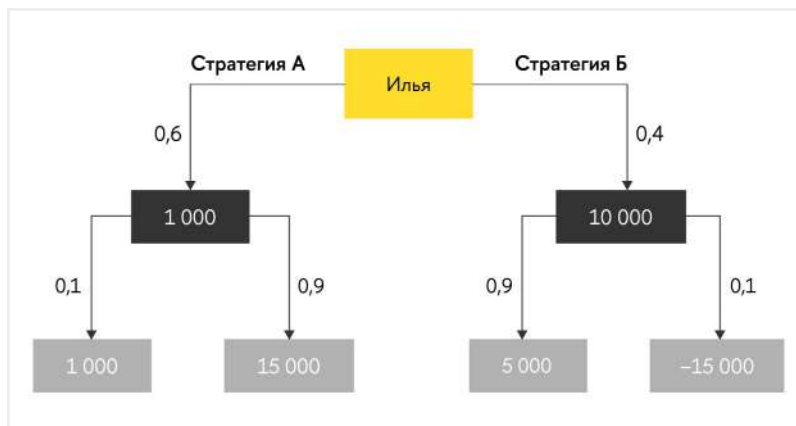
16. В ходе исследования в волшебной лаборатории обнаружили: количество искр, создаваемых волшебной палочкой, зависит от типа используемого камня. Данные были собраны с помощью магических датчиков, которые автоматически фиксируют количество искр для каждого типа камня. В результате получили такое распределение средних значений.

Тип камня	Среднее количество искр
Простой булыжник	50
Полудрагоценный камень	60
Драгоценный камень	70

При помощи каких статистических инструментов можно подтвердить или опровергнуть наличие зависимости между числом искр и типом драгоценного камня?

- A. Коэффициент корреляции Спирмена
- B. Линейная регрессия
- C. Логистическая регрессия
- D. Т-тест
17. Аналитик получил набор данных со следующими переменными: пол слушателя, возраст слушателя, предпочтение музыкального стиля и уровень образования. Какие из этих переменных имеет смысл использовать в качестве зависимых, а какие — в качестве независимых?
- A. Пол слушателя как независимая переменная, а остальные как зависимые
- B. Возраст слушателя как зависимая переменная, а остальные как независимые
- C. Предпочтение музыкального стиля как зависимая переменная, а остальные как независимые
- D. Уровень образования как независимая переменная, а остальные как зависимые
18. Директор школы хочет наглядно представить распределение бюджета между различными направлениями: закупкой учебников, ремонтом классов, зарплатой учителей и внеклассными мероприятиями. Какой тип визуализации лучше всего подойдет для демонстрации доли каждого направления в общем бюджете?
- A. Диаграмма рассеяния
- B. Линейный график
- C. Пузырьковая диаграмма
- D. Круговая диаграмма

19. Илья руководит стартапом по анализу данных. Ему нужно выбрать одну из дальнейших стратегий развития стартапа: А или Б. Обычно руководители стартапов с вероятностью 0,6 выбирают стратегию А и с вероятностью 0,4 — стратегию Б.



Каждая стратегия состоит из двух этапов. Стратегия А на первом этапе принесет 1 000 бонусов, а на втором — с вероятностью 0,1 еще 1 000 бонусов или с вероятностью 0,9 еще 15 000 бонусов. Стратегия Б на первом этапе принесет 10 000 бонусов, а на втором — с вероятностью 0,9 еще 5 000 бонусов или с вероятностью 0,1 Илья потеряет 15 000 бонусов.

Какую стратегию стоит выбрать Илье?

- A. А
- B. Стратегии равноценны
- C. Б
- D. Недостаточно информации для принятия решения
20. В ходе тренировок спортсмен показывал следующие результаты в метании копья за последние 4 сезона:
- в первом сезоне — 41 м;
- во втором сезоне — 45 м;
- в третьем сезоне — 51 м;
- в четвертом сезоне — 52 м.

Чему равен средний темп прироста дальности полета копья в процентах?
Укажите ответ одним числом без дополнительных знаков и символов, округлите до целого.

Ответ: 8

21. В туристическом агентстве провели анализ предпочтений клиентов по видам отдыха и сезонам, когда они предпочитают отдыхать. Составили такую таблицу сопряженности.

Сезон	Пляжный	Горнолыжный	Экскурсионный
Лето	450	20	130
Зима	50	350	100
Межсезонье	100	50	250

Какие выводы можно сделать на основе данной таблицы?

- A. Вероятность, что случайно выбранный клиент предпочитает пляжный отдых, выше, чем вероятность, что он выберет горнолыжный отдых
- B. В межсезонье меньше клиентов, чем зимой
- C. Горнолыжный и экскурсионный отдых одинаково популярны среди всех клиентов
- D. Больше половины клиентов, выбирающих отдых летом, предпочитают пляжный отдых
- E. Вероятность выбрать горнолыжный отдых выше зимой, чем летом

22. Инженер Наталья работает на орбитальной станции, где существует риск поломки оборудования из-за космического излучения. Вероятность того, что оборудование не выйдет из строя во время периода критической нагрузки, составляет 0,97. Критическая нагрузка на системы станции фиксируется раз в 7 дней. Через сколько дней вероятность того, что оборудование выйдет из строя, станет выше, чем вероятность, что оно останется работоспособным?

Укажите ответ одним числом без дополнительных знаков и символов, округлив в случае необходимости до целого значения.

Ответ: 161

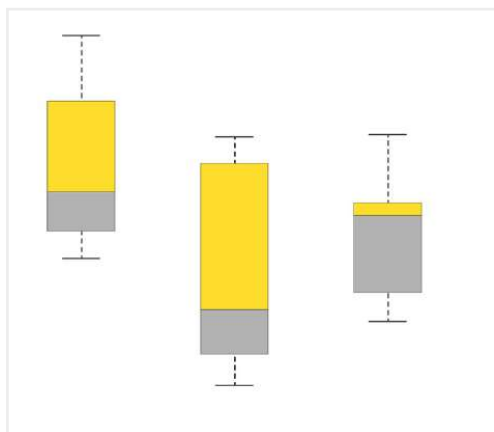
23. На производственной линии за день упаковывается 100 коробок с продукцией. Медиана веса коробок составляет 500 г, а средний вес — 497 г. Инженер по качеству может добавить по 10 г продукции в X коробок, чтобы привести средний вес к медиане.

Какое минимальное количество коробок (X) нужно дополнить, чтобы средний вес стал равен медиане (500 г)?

Ответ: 30

24. Перед вами графики Boxplot для трех выборок с параметром длины усов = $1,5 \times \text{IQR}$. Выберите верные утверждения.

- Среднее значение сильнее всего отклоняется от медианы во второй выборке
- Медианное значение в первой выборке наибольшее из представленных графиков
- В выборках есть выбросы, превышающие $1,5 \times \text{IQR}$
- Минимальное значение из всех значений по трем выборкам находится в третьей выборке



25. В туристическом клубе провели опрос среди 100 участников. Им задали два вопроса:

- 1) Любите ли вы путешествовать?
- 2) Увлекаетесь ли вы фотографией?

По результатам опроса 70 человек любят путешествовать; 60 человек увлекаются фотографией; 10 человек не любят ни путешествовать, ни фотографировать.

Сколько участников любят и путешествовать, и фотографировать?

Ответ: 40

26. Компания проводит онлайн-продажи, и каждый клиент проходит несколько этапов:

1. Посещение сайта.
2. Добавление товара в корзину.
3. Оформление заказа.
4. Оплата заказа.

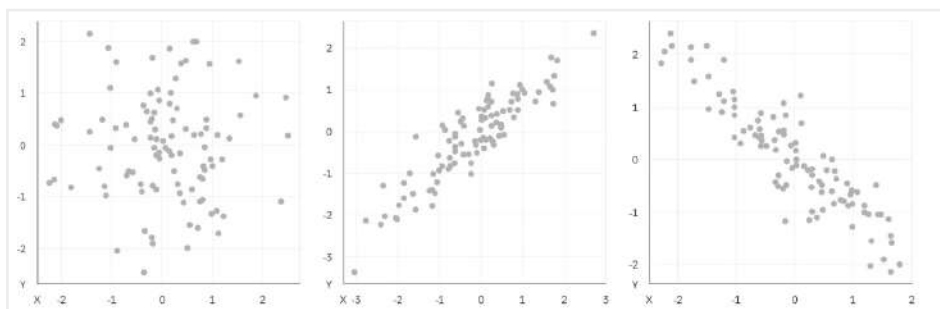
Аналитики собрали следующую информацию за один день:

- На сайт зашло 7 000 пользователей.
- Из всех, кто заходил на сайт, 12% пользователей оформили заказ.
- 351 клиент оплатил заказ.
- Из всех, кто добавил товар в корзину, только 30% его оплатили.

Рассчитайте долю оформивших заказ клиентов среди пользователей, добавивших товар в корзину. Ответ округлите до второго знака после запятой.

Ответ: 0,72 или 71,79

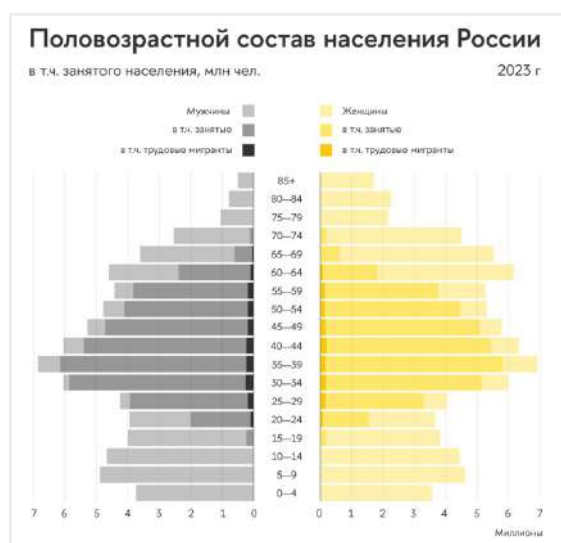
27. Даны три диаграммы рассеяния. Сопоставьте их с коэффициентами корреляции, которые получены на этих выборках. В ответ запишите буквы в порядке расположения картинок.



- А. 0
- Б. 0,5
- В. 0,9
- Г. -0,9

Ответ: 1А, 2Б, 3Г, 4Б

28. Перед вами иллюстрация половозрастной пирамиды России с указанием работающего (занятого) населения. Выберите верные утверждения для этой картинки.



- Доля занятых женщин в возрасте 55—59 лет ниже, чем у мужчин в том же возрасте
- Среди мужчин трудовых мигрантов больше, чем среди женщин
- Максимальная доля работающего населения среди мужчин приходится на возрастную группу от 25 до 29 лет
- В возрасте 60—64 лет количество женщин примерно вдвое больше, чем мужчин
- После 75 лет население состоит из женщин менее чем на 50%

29. Выберите все варианты ответов, в которых перечислены корректные методы борьбы с выбросами в данных.

A. Удаление 20% значений

B. Стандартизация данных

C. Удаление значений, которые отклоняются сильнее чем $1,5 \times IQR$

D. Использование T-тестов вместо Z-тестов

Разбор вопросов

1. По выборке из 144 значений посчитали выборочную дисперсию. После этого все значения увеличили в 10 раз. Как изменится выборочная дисперсия, если вычислить ее по новым значениям выборки?

- А. Не изменится
- Б. Увеличится в 10 раз
- В. Увеличится в 100 раз
- Г. Изменится в 0,1 раза
- Д. Нет верного варианта ответа среди вышеперечисленных

Решение:

Выборочная дисперсия — это мера разброса данных относительно выборочного среднего. Она вычисляется по формуле:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Рассмотрим, что произойдет, если все значения выборки увеличить в 10 раз, то есть заменить x_i на $10x_i$. Новое среднее будет:

$$10\bar{x} = 10\bar{x}$$

Тогда новая дисперсия:

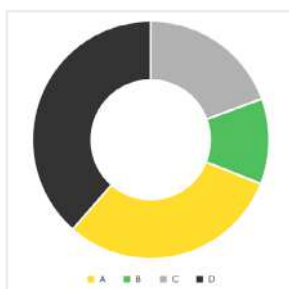
$$s_{\text{нов}}^2 = \frac{1}{n-1} \sum_{i=1}^n (10x_i - 10\bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n [10(x_i - \bar{x})]^2 = \frac{1}{n-1} \sum_{i=1}^n 100(x_i - \bar{x})^2 = 100 \cdot s^2$$

Таким образом, выборочная дисперсия увеличится в 100 раз.

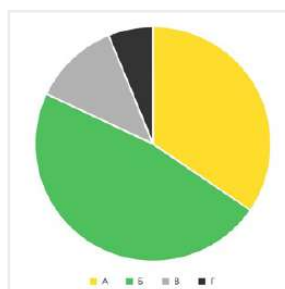
Ответ: В — Увеличится в 100 раз

2. Сопоставьте названия типов визуализации с графиками:

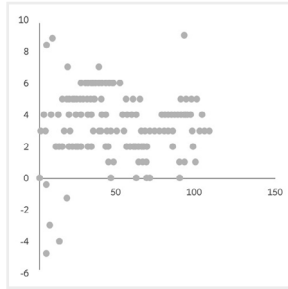
1.



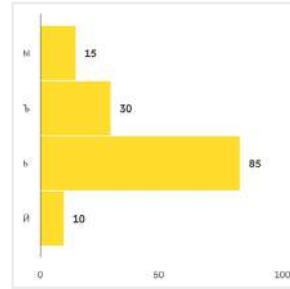
2.



3.



4.



1. Кольцевая диаграмма
2. Секторная диаграмма
3. Диаграмма рассеяния
4. Гистограмма

Решение:

Диаграммы в данной задаче можно сопоставить по визуальным признакам:

Кольцевая диаграмма — изображает, например, доли в виде кольца (1)

Секторная диаграмма — представляет собой круг, разделенный на сектора. В англоязычной литературе и на слэнге известна как pie chart (2)

Диаграмма рассеяния — обычно по координатной плоскости рассеяны точки, соответствующие значениям двух признаков (3)

Гистограмма — иногда можно встретить название столбчатой диаграммы. Представляет собой столбцы чаще вертикальные, но иногда горизонтальные как в варианте (4)

3. Исследователи проверили гипотезу о среднем значении с использованием Z-теста. Уровень значимости установлен равным 0,01. Как интерпретировать такое значение уровня значимости?
 - А. Вероятность того, что альтернативная гипотеза верна, равна 1%
 - Б. Допустимая вероятность отвержения альтернативной гипотезы, если она на самом деле верна, равна 1%.
 - В. Если нулевая гипотеза не верна, в 1% случаев мы допускаем, что можем ошибочно принять ее при многократном повторении теста на семплированных данных из одной генеральной совокупности
 - Г. Вероятность ошибки первого рода равна 1%
 - Д. Если тест показал $p\text{-value} = 0,1$, вероятность того, что нулевая гипотеза верна, равна 1%

Решение:

В задании указан выбранный уровень значимости, который в статистике показывает вероятность совершения ошибки 1 рода (то есть принятие нулевой гипотезы, когда она не верна). Таким образом, вариант Г должен быть выбран верным.

Далее рассмотрим каждый из вариантов ответа последовательно.

В варианте ответа А делается предположение об априорной вероятности верности альтернативной гипотезы – об этом мы никакой информации не имеем, соответственно, вариант А не верен.

В варианте Б говорится об ошибке 2 рода, о чем в условии задания также не сказано, следовательно, вариант Б неверен.

В варианте В также говорится об ошибке 2 рода, следовательно, вариант В неверен.

В варианте Д говорится об априорной вероятности верности нулевой гипотезы, о чем информации в условии задания нет.

Ответ: Г — Вероятность ошибки первого рода равна 1%

4. Сотрудники отдела прогнозирования и анализа данных изучили ежемесячные данные по ценам на рис и гречку за 2021 год. Средние цены оказались одинаковыми. Значения дисперсий: 64 P^2 для риса и 82 P^2 для гречки. Какое из утверждений верное?

- А. Цена риса в 2021 году была более изменчива, чем цена гречки
- В. Цена гречки в 2021 году была более изменчива, чем цена риса
- С. Ежемесячные цены на рис и гречку всегда одинаковы
- Д. Стандартные отклонения цен на рис и гречку одинаковы

Решение:

Задача на интерпретацию второго центрального момента случайной величины — дисперсии (в теории вероятностей) и её выборочной оценки (математическая статистика). Дисперсия отражает разброс показателя (накопленное квадратичное отклонение от среднего) и всегда относительна. У какой цены значение выборочной дисперсии больше, у той больше разброс, а значит и изменчивость. Информация о средних здесь не обязательно, хотя и полезна на случай, если хочется проверить ответ через определение выборочной дисперсии.

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2, \text{ где } n - \text{объем выборки}$$

Ответ: В — Цена гречки в 2021 году была более изменчива, чем цена риса

5. Влад построил 80%-й доверительный интервал для средней цены пары кроссовок «Майк» и получил оценку от 7 000 до 15 000 P . Какая гипотеза не отвергнется на 20%-м уровне значимости при использовании двусторонней альтернативной гипотезы?

- А. Дисперсия цены пары кроссовок «Майк» равна 4 000 P
- В. Средняя цена пары кроссовок «Майк» равна 2 700 P
- С. Средняя цена пары кроссовок «Майк» равна 9 535 P
- Д. Стандартное отклонение цены пары кроссовок «Майк» равно 4 000 P

Решение:

В условии задания не указана к какому распределению принадлежит средняя цена кроссовок, поэтому формула, по которой получен доверительный интервал не может быть однозначно идентифицирована и вычислена дисперсию/стандартное отклонение. Ответы А и Г не верны.

На 20% уровне значимости не отвергнется гипотеза, согласно которой истинное значение средней цены принадлежит указанному интервалу, в вариантах ответа только одна цена входит в этот интервал: $7\ 000 < 9\ 535 < 15\ 000$.

Ответ: В — Средняя цена пары кроссовок «Майк» равна 9 535 ₽

6. В автошколе после тестирования инструктор оценил среднее значение баллов за тест, набранных 10 сдавшими, и получил 85 баллов. Оказалось, что 5 работ инструктор забыл внести в ведомость. После их внесения среднее по всем 15 работам оказалось равно 90 баллам. Максимум за тест можно было набрать 100 баллов. Что можно сказать о добавленных пяти работах?

- A. Инструктор забыл добавить работы учеников, набравших максимум баллов
B. Инструктор неправильно рассчитал среднее
C. Среднее число баллов в добавленных пяти работах равно 99
D. Стандартное отклонение набранных баллов одним учеником в среднем равно 10 баллам

Решение:

Задача на знание определения выборочного среднего и умение работать с ним.

$$\frac{1}{10} \sum_{i=1}^{10} x_i = 85, \text{ следовательно сумма по десяти работам равна } \sum_{i=1}^{10} x_i = 850$$

$$\frac{1}{15} \sum_{i=1}^{15} x_i = 90, \text{ следовательно сумма по 15 работам равна } \sum_{i=1}^{15} x_i = 1350$$

Разность сумм равно 500, что является суммой набранных баллов по 5 работам. Так как максимум равен 100 баллам, то очевидно, что работы забытые инструктором набрали максимум баллов.

Ответ: А — Инструктор забыл добавить работы учеников, набравших максимум баллов

7. Сопоставьте следующие данные о сотрудниках, собранные в рамках опроса, и типы признаков, которыми могут быть описаны эти данные.

Переменная	Тип признака
Должность (менеджер/разработчик/аналитик)	Дискретный
Процент надбавки от фиксированной заработной платы	Категориальный
Количество рабочих дней в командировках	Порядковый
Уровень стресса (целое число от 1 до 10)	Непрерывный

Решение:

Должность является текстовой переменной, которая характеризует «класс» или категорию наблюдения, которая не может быть описана числом, поэтому должность — это категориальная переменная.

Количество рабочих дней в командировках может быть измерено только целым числом, то есть дискретно.

Уровень стресса задается целыми числами, которые характеризуют силу стресса, то есть является порядковой переменной.

Следовательно, оставшийся «Процент надбавки от фиксированной заработной платы» — это непрерывная переменная.

Ответ: Должность: категориальный. Процент: непрерывный.

Количество рабочих дней: дискретный. Уровень стресса: порядковый.

8. Алексей из Москвы провел опрос среди своих подписчиков в соцсети, чтобы узнать, будут ли они покупать новый кофейный напиток. Всего у Алексея 1 000 подписчиков, 850 из которых живут в Москве. Результаты опроса такие:

- 620 человек ответили «Да, куплю»
- 150 человек сказали «Нет»
- 230 человек затруднились с ответом

Какие выводы можно сделать по результатам этого опроса?

- A. В среднем в России 62% потребителей готовы купить новый кофейный напиток
- B. В среднем в России 15% потребителей готовы купить новый кофейный напиток
- C. В Москве примерно 73% потребителей готовы купить новый кофейный напиток
- D. Полученных данных недостаточно для запуска продаж по всей стране

Решение:

Алексей опросил преимущественно жителей Москвы (из 1000 опрошенных 850 из Москвы). Результаты опроса, которые получил Алексей смещены в сторону предпочтений москвичей. В вариантах ответа предлагается проецировать результаты на всю страну, что некорректно, так как выборка относительно всей страны оказалась нерепрезентативной. Вариант C отражает долю опрошенных ответивших «Да, куплю» среди числа подписчиков из Москвы (850). Это некорректно, так как для принятия решения по Москве нас интересовали бы только жители Москвы, а среди ответивших «Да, куплю» вполне могли оказаться жители других регионов.

Ответ: D — Полученных данных недостаточно для запуска продаж по всей стране

9. У Кати есть 10 000 ₽, и она думает об открытии депозита на 6 месяцев под годовую ставку 20%. Сотрудники банка предложили ей инновационный продукт, ежемесячная доходность которого будет такая: +10%, +1%, −3%, +12%, +1%, −2%. Куда стоит вложить средства Кате? Предполагается, что процентный доход ежемесячно выплачивается после начисления (капитализация процентного дохода отсутствует).

- A. Стоит выбрать инновационный продукт: так Катя сможет заработать через полгода больше, чем на вкладе в депозит
- B. Стоит открыть депозит: он принесет больший доход, чем инновационный продукт
- C. Все равно, куда вкладываться
- D. Недостаточно информации для принятия решения, так как нужно учитывать капитализацию

Решение:

Рассчитаем какой доход получит Катя от двух этих вариантов.

Инновационный продукт: $10+1-3+12+1-2 = 19\%$

Вклад: $20/2=10\%$ (так как 20% это годовая ставка, а Катя положит деньги на 6 месяцев, то есть на полгода)

При этом в варианте Г не стоит принимать во внимание капитализацию, так как и в инновационном продукте, и во вкладе указано, что капитализации нет.

Ответ: А — Стоит выбрать инновационный продукт, так Катя сможет заработать через полгода больше, чем на вкладе в депозит.

10. Производители чехлов для смартфонов последние 8 лет по негласному соглашению выпускали преимущественно черные чехлы. В результате доля черных чехлов в поставках ретейлерам составила 85%, а на все остальные цвета (белый, синий, красный и др.) пришлось только 15%. Конверсия в покупку чехлов каждого цвета одинаковая. Аналитики рынка опубликовали исследование с несколькими утверждениями. Какие из этих утверждений можно считать достоверными на основе приведенных данных?

- A. Покупатели предпочитают черные чехлы для смартфонов всем другим цветам
- B. Доля синих чехлов составляет 2% или более
- C. Потребителям не нравятся яркие чехлы для телефонов
- D. Большинство проданных чехлов черного цвета
- E. Нет верных ответов среди вышеперечисленных вариантов

Решение:

Утверждения А и С предлагают гипотезы относительно потребительских предпочтений покупателей. Они неверны, так как в самой задаче ничего о предпочтениях не сказано. Мы знаем лишь факт, связанный с производителем, что длительное время по каким-то причинам выпускались преимущественно черные чехлы. Для варианта В у нас нет никакой подтверждающей информации о доле синих чехлов. Вариант D оказывается верным, так как это факт, следующий из условий сложившихся на рынке чехлов вследствие ограничений производителя.

Ответ: D

11. На космической станции связи работают два робота-оператора: R2 и D2. R2 тратит 2 минуты, чтобы обработать один запрос. D2 тратит на один запрос 6 минут. Их рабочая смена длится 4 часа. Сколько в среднем длится обработка запроса на космической станции?

- А. 3 минуты
- Б. 3 минуты 30 секунд
- В. 3 минуты
- Г. 5 минут

Решение:

4 часа = $4 \cdot 60$ мин = 240 мин - длительность смены каждого из роботов в минутах

Суммарно робота обрабатывают за одну смену 480 мин, по 240 мин каждый

Так как R2 обрабатывает один запрос за 6 мин, то за свое время смены он обработает $240/2 = 120$ клиентов. А робот D2 соответственно $240/6=40$ клиентов.

Суммарно за смену роботы выполняют запросы $120+40 = 160$ клиентов за суммарно 480 минут. То есть в среднем по $480/160 = 3$ мин / клиента.

Ответ: В — 3 минуты

12. В сказочной стране Конфетляндии у волшебника есть 15 волшебных конфет, выстроенных в ряд так, что каждая следующая конфета весит на 1 г больше предыдущей. Известно, что средний вес первых 11 конфет равен 10 г, а средний вес последних 11 конфет — 14 г. Найдите вес самой тяжелой конфеты в этом ряду.

Решение:

Пусть вес первой конфеты a грамм. Тогда вес второй $a+1$, третьей $a+2$, и так до 15-й: $a+14$.

Для первых 11 конфет: $a, a+1, a+2, \dots, a+10$

Известно, что их средний вес 10 г, а значит:

$$\frac{a+(a+1)+\dots+(a+10)}{11} = 10$$

Откуда:

$$\frac{11a+55}{11} = 10$$

$$a=5$$

Значит, первая конфета весит 5 г.

Для последних 11 конфет (с номерами с 5-й по 15-ю): $a+4, a+5, \dots, a+14$

Подставим $a=5$:

9, 10, 11, ..., 19

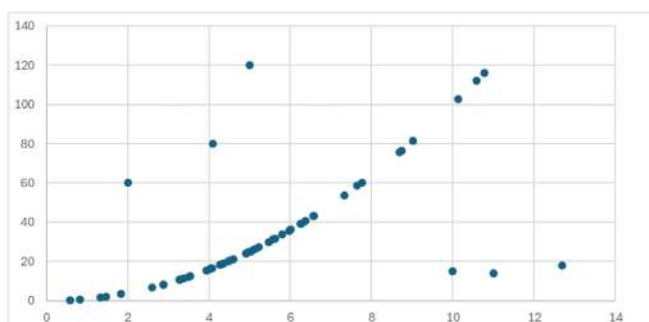
Средний вес этих 11 конфет по условию 14 г. Проверим:

Среднее: $154/11=14$ — верно.

Тогда самая тяжелая конфета — это 15-я конфета: $\alpha+14=5+14=19$ грамм.

Ответ: 19

13. Директор цеха по производству шпuleк составил график зависимости оплаты труда сотрудников всех профессий в цеху (в тысячах рублей по вертикальной оси) от длины рабочей смены (в часах по горизонтальной оси). Выберите верные утверждения, которые соответствуют построенному графику.



- А. Между длиной рабочей смены и оплатой есть линейная отрицательная связь.
- Б. Коэффициент корреляции Спирмена между длиной рабочей смены и оплатой нулевой.
- В. В цеху трое перерабатывают и им недоплачивают.
- Г. Зависимость между оплатой труда и длительностью рабочей смены можно описать нелинейной функцией, исключив выбросы

Решение:

Рассмотрим последовательно варианты ответов.

Между длиной рабочей смены и оплатой на графике видна положительная взаимосвязь, которая к тому же нелинейна. Поэтому вариант А неверен.

Коэффициент корреляции Спирмена будет положительным, поскольку на графике явно видна положительная взаимосвязь. Вариант Б также неверен.

В условии указано, что на графике отображены представители всех профессий. Мы можем предположить, что верхние 3 выбивающиеся наблюдения — это представители более высокооплачиваемых профессий (директора, ведущие мастера и тд), а оставшиеся 3 наблюдения — представители профессий с более низкой квалификацией (например, стажеры, подмастерье и тд). Поэтому нельзя сделать вывод, что они недорабатывают или перерабатывают, поскольку иначе в этом выводе не учитывается квалификация сотрудников. Вариант В неверен.

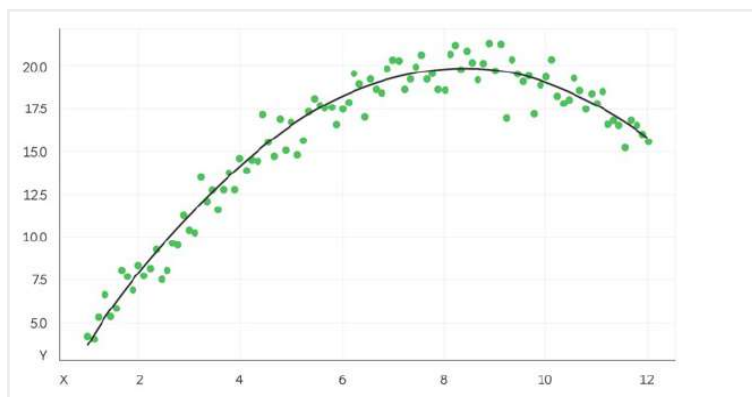
Вариант Г верен, так как исключив 6 выбивающихся наблюдений, функция действительно имеет нелинейный вид.

Ответ: Г — Зависимость между оплатой труда и длительностью рабочей смены можно описать нелинейной функцией, исключив выбросы.

14. Аграрная компания провела исследование, чтобы определить, как количество внесенных удобрений (килограмм на гектар) влияет на урожайность винограда (килограмм с гектара). Для анализа данных оценили квадратичную регрессионную модель:

$$\text{Урожайность} = -0,3 \times \text{Удобрения}^2 + 5 \times \text{Удобрения} - 1$$

Коэффициент детерминации $R^2 = 0,82$. Определите оптимальное количество удобрений, при котором урожайность достигает максимума, согласно оцененной модели. На графике зеленым цветом показаны фактические данные, черным — оцененная зависимость.



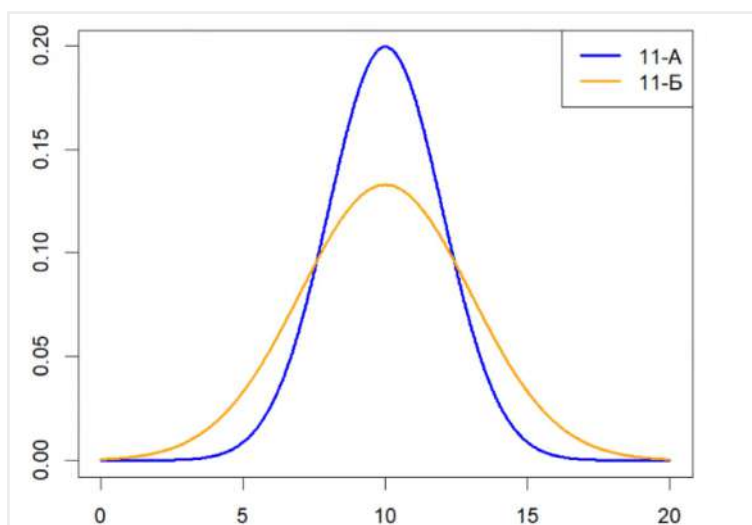
Решение:

Из условия следует, что зависимость между урожайностью и удобрениями квадратичная: Урожайность = $-0,3 \times \text{Удобрения}^2 + 5 \times \text{Удобрения} - 1$.

Найденная вершина параболы по абсциссе и будет ответом: $-5 / (-2 \cdot 0.3) = 8.33$

Ответ: 8

15. Дима построил распределения количества игр в библиотеке магазина «Стем» для учащихся 11 «А» и 11 «Б». Получился такой график распределений.



Что из этого графика может заключить Дима об учащихся?

А. В обоих классах среднее число игр у одного учащегося равно 10

Б. В 11 «Б» разброс числа игр у учащихся меньше, чем в 11 «А»

В. Медиана в 11 «А» равна моде в 11 «Б»

Г. Больше половины учащихся 11 «А» обладают 8 и более играми.

Решение:

Рассмотрим все варианты ответов по порядку.

Вариант А верен, поскольку на графиках показаны два нормальных распределения со средним значением в 10.

Вариант Б не верен, поскольку распределение в классе А более сконцентрированное, а значит, в нем дисперсия (разброс) меньше, чем у класса Б.

Вариант В верен, поскольку представленные распределения являются симметричными относительно среднего значения, а значит, медиана в обоих распределениях равна 10. Также по графикам видно, что мода (самое популярное значение в выборке) приходится также на 10.

Вариант Г верен, поскольку медиана распределения приходится на 10, значит, что половина учеников имеет 10 игр и больше. Так как $8 < 10$, то больше половины учащихся имеют 8 и больше.

Ответ: А — В обоих классах среднее число игр у одного учащегося равно 10.

В — Медиана в 11 «А» равна моде в 11 «Б». Г — Больше половины учащихся 11 «А» обладают 8 и более играми.

16. Любителей творчества Чарльза Диккенса спросили: сколько чашек чая выпивает мисс Хэвишем в зависимости от количества гостей на ее знаменитых чаепитиях? В результате получилась следующая таблица с усредненными значениями

Количество гостей	Среднее количество выпитых чашек чая (оцененное по данным опроса)
Мало	5 чашек
Средне	4 чашки
Много	3 чашки
Очень много	2 чашки

При помощи каких статистических инструментов можно подтвердить или опровергнуть наличие зависимости между количеством выпитых чашек чая и количеством гостей на чаепитии у мисс Хэвишем?

А. Коэффициент корреляции Спирмена

Б. F-тест на равенство дисперсий

В. Z-тест

Г. U-тест Манна — Уитни

Решение:

В таблице приведены два признака: порядковый (количество гостей) и количественный (среднее количество выпитых чашек). Из всех предложенных вариантов только коэффициент корреляции Спирмена предлагает рецепт по расчету корреляции между порядковыми и количественными признаками с дальнейшей проверкой гипотезы о независимости признаков.

17. Исследователь получил набор данных о вкусовых предпочтениях мороженого анкетированных со следующими переменными: цвет упаковки, возраст потребителя, уровень дохода и предпочитаемый вкус мороженого. Какие из этих переменных имеет смысл использовать в качестве зависимых, а какие в качестве независимых?

- А. Цвет упаковки как независимая переменная, а остальные как зависимые
- Б. Возраст потребителя как зависимая переменная, а остальные как независимые
- В. Уровень дохода как независимая переменная, а остальные как зависимые
- Г. Предпочитаемый вкус мороженого как зависимая переменная, а остальные как независимые

Решение:

В качестве зависимой переменной (таргета/целевой переменной/...) в исследовании стоит выбирать те переменные, которые не оказывают влияние на другие, но при этом должен существовать механизм, по которому на них будут оказывать влияние независимые переменные.

В данном случае логичнее предположить, что на предпочтение определенного мороженого могут влиять цвет упаковки, возраст потребителя и его уровень дохода.

Ответ: Г — Предпочитаемый вкус мороженого как зависимая переменная, а остальные как независимые.

18. Специалист по флоре тропического леса хочет визуализировать распределение различных видов растений в исследуемой области. Какой тип визуализации лучше всего подходит для этой задачи?

- А. Тепловая карта
- Б. Линейный график
- С. Диаграмма Венна
- Д. Столбчатая диаграмма

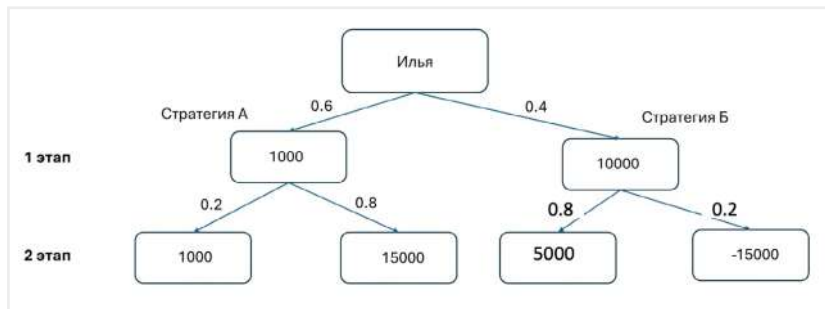
Решение:

Ключевым словосочетанием является «хочет визуализировать распределение». Распределение хорошо представлять при помощи гистограммы (столбчатой диаграммы), но если вчитаться в задание, то мы имеем дело с пространственными данными о произрастании растений в географической области. Тепловая карта как нельзя лучше подходит для такого и широко используется географами и биологами.

19. Илья руководит стартапом по анализу данных. Ему нужно выбрать одну из дальнейших стратегий развития стартапа: А или Б. Обычно руководители стартапов с вероятностью 0,6 выбирают стратегию А и с вероятностью 0,4 — стратегию Б.

Каждая стратегия состоит из двух этапов. Стратегия А на первом этапе принесет 1 000 бонусов, а на втором — с вероятностью 0,2 еще 1 000 бонусов или с вероятностью 0,8 еще 15 000 бонусов. Стратегия Б на первом этапе принесет 10 000 бонусов, а на втором — с вероятностью 0,8 еще 5 000 бонусов или с вероятностью 0,2 Илья потеряет 15 000 бонусов.

Какую стратегию стоит выбрать Илье?



- А. А
- Б. Стратегии равноценны
- В. Б
- Г. Недостаточно информации для принятия решения

Решение:

Посчитаем математическое ожидание бонусов при выборе каждой из стратегий.

Для стратегии А: $1\,000 + (0.2 * 1\,000 + 0.8 * 15\,000) = 1\,000 + 200 + 12\,000 = 13\,200$

Для стратегии Б: $10\,000 + (0.8 * 5\,000 + 0.2 * (-15\,000)) = 10\,000 + 4\,000 - 3\,000 = 11\,000$

То, какие стратегии обычно выбирают руководители («Обычно руководители стартапов с вероятностью 0,6 выбирают стратегию А и с вероятностью 0,4 — стратегию Б»), никак не влияет на прибыльность указанных стратегий.

$13\,200 > 11\,000$, значит, стратегия А предпочтительнее, и следует выбрать ее

Ответ: А — А

20. Владелец небольшого кафе фиксировал количество посетителей каждый год в течение последних 4 лет и получил такие данные по годам: 1 200 посетителей, 1 500, 1 800, 2 250. Чему равен средний темп прироста числа посетителей кафе за год в процентах? Укажите ответ одним числом без дополнительных знаков и символов, округлите ответ до целого.

Решение:

Обозначим данные по годам:

$$\text{Год 1: } N_0 = 1\,200$$

$$\text{Год 2: } N_1 = 1\,500$$

$$\text{Год 3: } N_2 = 1\,800$$

$$\text{Год 4: } N_3 = 2\,250$$

Годовые темпы прироста равны:

$$k_1 = \frac{1\,500}{1\,200} = 1.25$$

$$k_2 = \frac{1\,800}{1\,500} = 1.2$$

$$k_3 = \frac{2\,250}{1\,800} = 1.25$$

Средний темп роста за период находится по формуле среднего геометрического годовых коэффициентов роста:

$$\bar{k} = \sqrt[3]{k_1 \cdot k_2 \cdot k_3} = \sqrt[3]{1.25 \cdot 1.2 \cdot 1.25} \approx 1.234$$

Средний годовой темп прироста в %

$$\bar{r} = (\bar{k} - 1) \times 100\% \approx (1.234 - 1) \times 100\% \approx 23.4\%$$

Округляем до целого: 23%.

Ответ: 23

21. Аналитики исследовали, каких домашних животных предпочитают жители города. Получилась такая таблица сопряженности, отражающая связь между типом жилья и предпочтением вида питомца.

Тип жилья	Кошки	Собаки	Рыбки
Квартира	250	180	120
Частный дом	150	320	80

Известно также, что все жители города держат какого-нибудь питомца!

Какие выводы можно сделать на основе этих данных?

- А. Вероятность, что случайно выбранный житель держит собаку, выше, чем вероятность, что он держит кошку
- Б. Среди жителей частных домов больше тех, кто держит рыбок, чем среди жителей квартир
- В. Кошки и рыбки одинаково популярны среди всех жителей

Г. Больше половины жителей квартир предпочитают держать кошек

Д. Вероятность держать собаку выше среди жителей частных домов, чем среди жителей квартир

Решение:

Проверим каждый вариант ответов.

А. Вероятность, что случайный житель держит собаку $= (180+320)/(250+150+180+320+120+80) = 500/1100 \approx 0.45$. Вероятность, что случайный житель держит кошку $= (250+150)/1100 \approx 0.36$. Вариант А верен.

Б. Доля тех кто держит рыбок в частных домах $= 80/(150+320+80) = 80/550 \approx 0.145$. Доля тех кто держит рыбок в квартирах $= 120/(250+180+120) = 120/550 \approx 0.218$. Вариант Б не верен.

В. Держат кошек $250+150 = 400$ человек, а держат рыбок $120+80 = 200$ человек. Вариант В не верен.

Г. Держат кошку среди жителей квартир 250 чел, всего в квартирах живет 550 человек; $250/550$ это меньше 50%, следовательно ответ Г неверен

Д. Вероятность держать собаку среди жителей частных домов $= 320/(150+320+80) = 320/550 \approx 0.58$. Вероятность держать собаку среди жителей квартир $= 180/550 \approx 0.32$. $0.58 > 0.32$, следовательно вариант Д верен.

Ответ: А — Вероятность, что случайно выбранный житель держит собаку, выше, чем вероятность, что он держит кошку. Д — Вероятность держать собаку выше среди жителей частных домов, чем среди жителей квартир.

22. Леша — профессиональный киберспортсмен и ежедневно участвует в онлайн-турнирах. Вероятность того, что его аккаунт не будет взломан при попытке взлома, составляет 0,98. Раз в 4 дня фиксируется попытка взлома аккаунтов всех участников турнира. Через сколько дней вероятность того, что аккаунт Алексея будет взломан, станет выше, чем вероятность, что этого не произойдет?

Укажите ответ одним числом без дополнительных знаков и символов, округлив в случае необходимости до целого значения.

Решение:

Вероятность не быть взломанным при одной попытке взлома:

$$p = 0.98$$

Вероятность быть взломанным при одной попытке:

$$q = 1 - p = 0.02$$

Попытка взлома происходит раз в 4 дня.

Значит, через n дней будет $[n/4]$ попыток взлома.

Пусть $k = [n/4]$ — количество попыток взлома.

Вероятность не быть взломанным ни разу за k попыток:

$$P_{\text{не взломан}} = p^k = 0.98^k$$

Вероятность быть взломанным хотя бы раз:

$$P_{\text{взломан}} = 1 - 0.98^k$$

Нужно найти минимальное n , при котором:

$$P_{\text{взломан}} > P_{\text{не взломан}}$$

То есть:

$$1 - 0.98^k > 0.98^k$$

$$1 > 2 \cdot 0.98^k$$

$$0.98^k < 0.5$$

$$k \cdot \ln(0.98) < \ln(0.5)$$

$$k > \frac{\ln(0.5)}{\ln(0.98)}$$

$$k > \frac{(-0.693147)}{(-0.0202027)} \approx 34.309$$

То есть $k \geq 35$.

$k = 35$ попыток взлома.

Если попытка в день 4, 8, 12, ..., то k -я попытка в день $4k$.

Для $k = 35$ день последней попытки $4 \times 35 = 140$.

В этот день вероятность взлома станет больше 0.5? Проверим:

После 34 попыток: $0.98^{34} \approx 0.503$, $P_{\text{взлом}} \approx 0.497$ — еще не больше.

После 35 попыток: $0.98^{35} \approx 0.493$, $P_{\text{взлом}} \approx 0.507$ — стало больше.

Ответ: 140

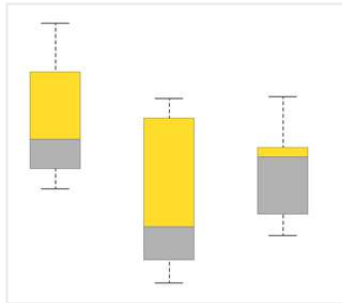
23. В выборке из 100 значений медиана равна 50, а среднее — 45. Вася может увеличить X чисел в выборке на 50. Какого минимального X достаточно, чтобы среднее стало равно медиане, равной 50?

Решение:

Запишем изначальное среднее как $M = 45$, а новое $M' = 50$. Тогда $M = 4500/100$, $M' = 5000/100$. $M' - M = 500$. Так как изменять каждое из значений можно только на 50, то мы должны изменить $X = 500/50 = 10$ значений, чтобы приравнять $M' = M = 5000/100$.

Ответ: 10

24. Перед вами графики Boxplot для трех выборок с параметром длины усов = $1,5 \times \text{IQR}$. Выберите верные утверждения.



- Дисперсия второй выборки больше, чем двух других
- Медианное значение в третьей выборке наименьшее из представленных графиков
- В выборках нет выбросов, превышающих $1,5 \times \text{IQR}$
- Максимальное значение из всех значений по трем выборкам находится в первой выборке

Ответ: Дисперсия второй выборки больше, чем двух других. В выборках нет выбросов, превышающих $1,5 \times \text{IQR}$. Максимальное значение из всех значений по трем выборкам находится в первой выборке

25. 100 ученикам школы задали два вопроса:

- 1) Занимаетесь ли вы спортом?
- 2) Играете ли вы на музыкальных инструментах?

По результатам опроса 50 учеников занимаются спортом; 45 учеников играют на музыкальных инструментах; 25 учеников не занимаются ни спортом, ни музыкой.

Сколько учеников и занимаются спортом, и играют на музыкальных инструментах?

Решение:

Если из 100 человек опрошенных 25 точно не занимаются ни спортом, ни музыкой, то $100 - 25 = 75$ человек занимаются хотя бы одним из этих занятий. При этом ответили хотя про одно из увлечений положительно $50 + 45 = 95$ опрошенных из 75 возможных, кто хотя бы чем-то занимается. Значит, что $95 - 75 = 20$ опрошенных занимаются одновременно и музыкой, и спортом.

Ответ: 20

26. Компания проводит онлайн-продажи, и каждый клиент проходит несколько этапов:

1. Посещение сайта.
2. Добавление товара в корзину.
3. Оформление заказа.
4. Оплата заказа.

Аналитики собрали следующую информацию за один день:

- На сайт зашло 5 000 пользователей.
- Из всех, кто заходил на сайт, 20% пользователей оформили заказ.
- 500 клиентов оплатили заказ.
- Из всех, кто добавил товар в корзину, только 50% его оплатили.

Рассчитайте долю оформивших заказ клиентов среди пользователей, добавивших товар в корзину. Ответ округлите до второго знака после запятой.

Решение:

Пусть x — число добавивших в корзину.

Из условия: «Из всех, кто добавил товар в корзину, только 50% его оплатили»

Значит, оплатившие = $0.5 * x$.

Но оплативших 500, поэтому:

$$0.5x = 500, \text{ следовательно } x = 1\,000$$

То есть 1 000 человек добавили товар в корзину.

Мы знаем, что оформили заказ 1 000 человек.

Все 1 000 оформивших заказ — это подмножество тех, кто добавил в корзину.

Тогда:

Добавили в корзину: 1 000 чел.

Оформили заказ: 1 000 чел.

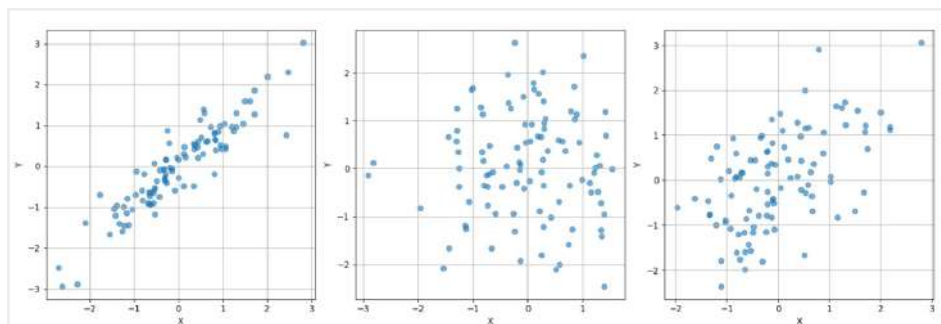
Доля:

$$1000/1000=1.00$$

То есть 100%.

Ответ: 1 или 100

27. Даны три диаграммы рассеяния. Сопоставьте их с коэффициентами корреляции, которые получены на этих выборках. В ответ запишите буквы в порядке расположения картинок.



А. $-0,4$

Б. 0

В. $0,9$

Г. $0,4$

Решение:

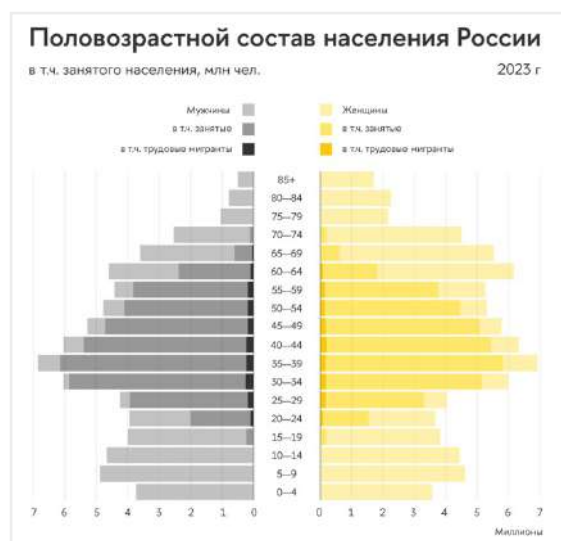
Первый график иллюстрирует ярко выраженную положительную взаимосвязь между переменными, которая почти складывается в линию. Из предложенных вариантов корреляция $0,9$ больше всего подходит для его описания.

Второй график не отображает вообще никакой направленности взаимосвязи: при изменении значений по X, значения по Y распределяются точно так же. Такой график иллюстрирует корреляцию $= 0$.

Третий график иллюстрирует средне выраженную положительную взаимосвязь, для такой взаимосвязи подходит коэффициент корреляции $= 0,4$.

Ответ: 1. В. $0,9$. 2. Б. 0. 3. Г. $0,4$.

28. Перед вами иллюстрация половозрастной пирамиды России с указанием работающего (занятого) населения. Выберите верные утверждения для этой картинки.



- Доля занятых женщин в возрасте 30—34 года ниже, чем у мужчин в том же возрасте
- Среди мужчин трудовых мигрантов меньше, чем среди женщин
- Максимальная доля работающего населения среди мужчин приходится на возрастную группу от 30 до 34 лет
- В возрасте 75—79 лет количество женщин примерно вдвое больше, чем мужчин
- После 75 лет население состоит из женщин более чем на 50%

Ответ: Доля занятых женщин в возрасте 30—34 года ниже, чем у мужчин в том же возрасте. Максимальная доля работающего населения среди мужчин приходится на возрастную группу от 30 до 34 лет. В возрасте 75—79 лет количество женщин примерно вдвое больше, чем мужчин. После 75 лет население состоит из женщин более чем на 50%

29. Выберите все варианты ответов, в которых перечислены корректные методы борьбы с выбросами в данных.

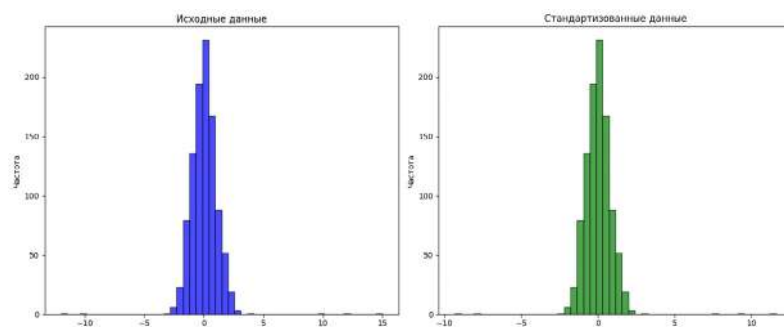
- А. Удаление 20% значений
- Б. Стандартизация данных
- В. Удаление значений, которые отклоняются сильнее чем $1,5 \times \text{IQR}$
- Г. Использование Т-тестов вместо Z-тестов

Решение:

Рассмотрим каждый вариант ответов.

А. Удаление 20% значений без уточнения конкретного способа выбора этих 20% никак не повлияет на распределение данных и не поможет бороться с выбросами. Этот вариант ответа не верен.

Б. Стандартизация данных влияет на масштаб данных, но не помогает бороться с выбросами, например как на картинке ниже:



Соответственно вариант Б не верен.

В. Удаление значений, которые отклоняются сильнее чем $1,5 \times \text{IQR}$ — это стандартный метод борьбы с выбросами, так как позволяет найти сильно отклоняющиеся значения от основной массы данных. Этот ответ верен.

Г. Z-тест, также как и t-тест чувствительны к выбросам, выбор между этими статистическими методами определяется знанием дисперсии генеральной совокупности, а не наличием выбросов в данных.

Ответ: В — Удаление значений, которые отклоняются сильнее чем $1,5 \times \text{IQR}$.