

Отборочное тестирование на 2 поток Школы аналитиков данных МТС 2023 г.

Блок 1. Теория вероятности и математическая статистика

1. Бросают два кубика: белый и черный. Определи вероятность того, что на белом и на черном кубиках выпадут разные цифры.

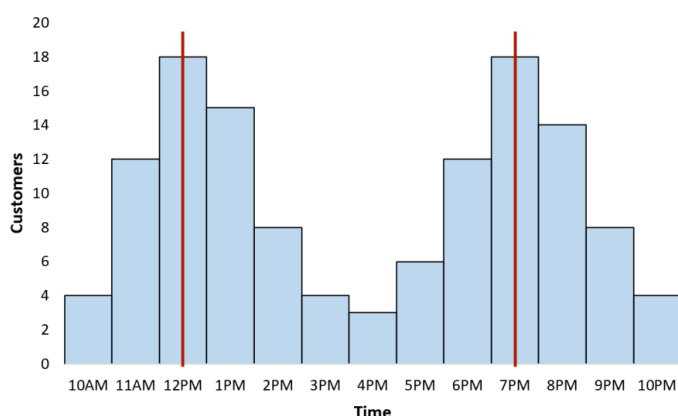
- 0.5
- 0.83
- 0.16
- 0.08

2. Два друга (1 и 2) бросают честную монетку (р выпадения решки – 0.5). Друг 1 бросает n кол-во раз, а друг 2 бросает $n+1$ кол-во раз. Найди вероятность того, что у друга 2 выпадет «решка» больше, чем у друга 1

- 0.5
- 1
- 0.33
- 0.66

3. Ты можешь кинуть кубик от 1 до 2 раз, твой выигрыш – это то значение, которое у тебя выпало, т.е. если кинул первый раз и выпало 2, а кинув второй кубик, выпадает 3, твой выигрыш – 5. Но если на втором кубике меньше, чем на первом – выигрыш равен 0. Сколько должно выпасть на первом кубике, чтобы ты точно решился бросать второй раз? Ответ укажи целыми числами через запятую без пробелов (например, 9,9,9)

4. Выбери верное утверждение:



- Красные линии - две медианы
- Красные линии – две моды

5. Какое значение из следующей выборки наиболее близко к 0.7-й квантили выборки?

[1, 3, 9, 2, 4, 0, 1, 5, -2, 12]

- 5
- 0
- 9
- 3

6. В коробке лежат 5 черных кубов и 5 белых кубов. Из коробки без возвращения вытаскивают 5 кубов. Известно, что из них 2 куба белые. Найди вероятность того, что другие три куба черные. Округли до 3 знаков после запятой

7. Среди клиентов банка 25% имеют задолженность по потребительскому кредиту и 30% имеют задолженность по ипотеке. Доля клиентов, у кого есть обе задолженности – 12%. Какова доля клиентов, имеющих задолженность по потребительскому кредиту, имеет и задолженность по ипотеке?

- 0,4
- 0,48
- 0,17
- 0,21

8. Отклонение нулевой гипотезы позволяет нам сделать следующий вывод:

- Нет значимых различий
- Все группы значимо отличаются
- Как минимум две группы значимо отличаются

9. Есть три группы исследуемых, которые следует сравнивать попарно.

Исследователям удалось отклонить основную нулевую гипотезу о равенстве средних при $p < 0,05$. Для попарного сравнения трех групп использовалась поправка Бонферрони, и различия между группами признавались статистически достоверными, если уровень значимости (p-value) был ниже:

- 0.0001
- 0.03
- 0.05
- 0.016

10. О чем говорит ЦПТ?

- Сумма независимых одинаково распределённых случайных величин распределена нормально с некоторыми параметрами
- Сумма нормальных случайных величин распределена как нормальное стандартное распределение
- Любую непрерывную случайную величину можно модифицировать в нормальную случайную величину

11. Какую проблему решают подходы к множественной проверке гипотез?

- Для того, чтобы проверять множество гипотез нужны оптимизации. Это становится сложнее в плане вычислений
- Уровень ошибки первого рода возрастает

12. Какова вероятность выявить коэффициент корреляции не меньше 0.6 при объеме выборки 39 и уровне значимости 0.05?

Блок 2. Машинное обучение и математика

1. Сколько ошибок второго рода совершила модель?

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

- 10
- 50
- 5
- 100
- 15

2. Пользуясь картинкой выше, посчитай метрику Precision. Округли до 2 знаков после запятой

3. Дан вектор [6, 0, 6, 9]. Найди L1-норму этого вектора

4. Дан вектор [6, 0, 6, 9]. Найди L2-норму этого вектора. Округли до 2 знаков после запятой

5. Найди косинус угла между векторами a и b, где $a=(4;3;7)$ и $b(1;3;2)$. Округли до 2 знаков после запятой

6. Найди градиент функции в точке $M(0,1)$. Округли до 1 знака после запятой, раздели значения в ответе пробелом (например, 3.3 3.2)

$$Z = \frac{xy}{x+y+1}$$

7. Вычисли интеграл. Ответ округли до 1 знака после запятой (например, 3.3)

$$\int_0^{2\pi} \sin(x)^8 dx$$

8. Дана функция плотности распределения случайной величины X .
Найди мат. ожидание $M(X)$, округли до 2 знаков после запятой (например, 1.72)

$$f(x) = \begin{cases} 0, & \text{при } x \leq 0, \\ 3x^2, & \text{при } 0 < x \leq 1, \\ 0, & \text{при } x > 1. \end{cases}$$

9. Найди дисперсию $D(X)$ для функции распределения, данной в 19 задании, ответ округли до 2 знаков после запятой (например, 1.72)

10. Найди среднее квадратичное отклонение для функции распределения, данной в 19 задании, ответ округли до 2 знаков после запятой (например, 1.72)

Блок 3. Python и Computer Science

1. Есть две переменные a и b , где $a = \text{"mts"}$ и $b = 2$. Что будет результатом выражения: $a+b$ и каким типом данных будет результат?

- TypeError
- "mtsmts"
- "mts" + "mts"
- "2mts"

2. Что вернет выражение $a * b$, где $a = \text{'mts'}$ и $b = \text{False}$?

- TypeError
- Пустая строка
- NameError

3. Выбери верные утверждения о dict в Python

- Обновление ключа не влияет на порядок расположения элементов словаря
- Обновление ключа влияет на порядок расположения элементов словаря
- Ключи, добавленные после удаления, вставляются в конец словаря
- Ключи, добавленные после удаления, вставляются в начало словаря
- Ключом словаря может быть list и set
- Ключом словаря может быть int, str

4. Какие виды баз данных существуют? Выбери варианты:

- Иерархические
- Объектно-ориентированные
- Объектно-реляционные
- Реляционные
- Сетевые
- Функциональные

5. Дана таблица Trip. Какие из запросов подсчитают количество полётов из города "Москва" в город "Ростов"?

id	company	plane	town_from	town_to	time_out	time_in
1100	4	Boeing	Rostov	Paris	1900-01-01T14:30:00.000Z	1900-01-01T17:50:00.000Z
1101	4	Boeing	Paris	Rostov	1900-01-01T08:12:00.000Z	1900-01-01T11:45:00.000Z
1123	3	TU-154	Rostov	Vladivostok	1900-01-01T16:20:00.000Z	1900-01-02T03:40:00.000Z
1124	3	TU-154	Vladivostok	Rostov	1900-01-01T09:00:00.000Z	1900-01-01T19:50:00.000Z
1145	2	IL-86	Moscow	Rostov	1900-01-01T09:35:00.000Z	1900-01-01T11:23:00.000Z
1146	2	IL-86	Rostov	Moscow	1900-01-01T17:55:00.000Z	1900-01-01T20:01:00.000Z
1181	1	TU-134	Rostov	Moscow	1900-01-01T06:12:00.000Z	1900-01-01T08:01:00.000Z
1182	1	TU-134	Moscow	Rostov	1900-01-01T12:35:00.000Z	1900-01-01T14:30:00.000Z
1187	1	TU-134	Rostov	Moscow	1900-01-01T15:42:00.000Z	1900-01-01T17:39:00.000Z

- SELECT DISTINCT company, COUNT(*) OVER (PARTITION BY company ORDER BY town_from) AS flights FROM Trip WHERE town_from = 'Moscow' AND town_to = 'Rostov'
- SELECT company, COUNT(*) as flights FROM Trip WHERE town_from = 'Moscow' AND town_to = 'Rostov' GROUP BY company ORDER BY company ASC
- SELECT company, COUNT(*) OVER (PARTITION BY company ORDER BY town_from DESC ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING) AS flights FROM Trip WHERE town_from = 'Moscow' AND town_to = 'Rostov'
- SELECT DISTINCT company, COUNT(*) OVER (PARTITION BY company ORDER BY town_from ASC ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS flights FROM Trip WHERE town_from = 'Moscow' AND town_to = 'Rostov' GROUP BY 1

6. Имеется отсортированный список из 64 наименований товаров магазина. Ты ищешь в нем значение методом бинарного поиска. Какое максимальное кол-во проверок для этого может понадобиться?

7. Что будет, если список наименований товаров магазина из прошлого задания увеличится вдвое?

8. Какая сложность алгоритма будет при создании таблицы умножения для всех элементов массива? Например, массив состоит из элементов [2, 3, 7, 8, 10], сначала каждый элемент умножается на 2, потом на 3, потом на 7 и т.д.

- $O(n^2)$

- $O(n)$
- $O(\log n)$
- $O(1)$

9. Какой результат будет выведен в результате выполнения кода ниже?

```
class A:
    def foo(self):
        print("A")

class B(A):
    pass

class C(A):
    def foo(self):
        print("C")

class D:
    def foo(self):
        print("D")

class E(B, C, D):
    pass

E().foo()
```

- A
- D
- C

10. Какие фрагменты кода написаны в соответствии с PEP8?

- `func(x = 7)`
- `x,y = 2,2`
- `a = [1,2,3]`
- `x, y = 2, 2`
- `a = [1, 2, 3]`

11. С помощью какого алгоритма решается задача, представленная ниже? Напиши ответ в именительном падеже

```
def func_(n: int) -> int:
    if (n == 0 or n == 1):
        return n
    else:
        return func_(n-1) + func_(n-2)
```

12. Какие алгоритмы, представленные ниже, являются жадными?

- Быстрая сортировка
- Поиск в ширину
- Алгоритм Дейкстры