

## Задача 1.

I. Для решения этой задачи проанализируем математические свойства данных функций  
*Функция А соответствует боксплоту 4.* Данная функция является параболой ветвями вниз, значит у нее есть максимум, который и будет является верхней границей наших усиков. Находим его:  $-b/2a = 0.5$  и подставляем в функцию, получаем 0,75. Следовательно наш боксплот либо первый, либо четвертый. Для дальнейшего анализа создадим программу на Python (приведена в текстовом формате в конце решения), которая будет создавать боксплот на основе массива чисел от 0 до 1 с шагом 0,05, и построим наш боксплот. Заметим, что медиана находится между 0.65 и 0.7, следовательно боксплот для этой функции - четвертый.

*Функция В соответствует боксплоту 5,* так как по сути эта функция “отзеркаливает” значения относительно 0.5, исходя из чего боксплот данной функции будет таким же, как и боксплот исходной функции с медианой 0.5

*Функция С соответствует боксплоту 2.* Данная функция возрастает на данном отрезке, поэтому значения функции при 0 и 1 будут границами усиков. Подставляем 0 и 1 и получаем 0.5 и 1 соответственно. Заметим, что таким усикам соответствуют 2 графика - 2 и 3. Функция Е также имеет похожие характеристики, поэтому построим их боксплоты и сравним между собой, используя код, представленный ниже, но подставляя другие уравнения в 8 строчку кода. Получаем, что медиана выше у боксплота функции Е, значит ей соответствует 3 боксплот, а функции С - второй.

*Функция D соответствует боксплоту 6.* Квадратичная функция является строго возрастающей на нашем отрезке, исходя из чего наименьшими и наибольшими значениями будут значения 0 и 1 соответственно, а медиана останется медианой. Таким образом, мы должны получить боксплот с усиками, тянущимися от 0 до 1 и медианой в 0.25

*Функция Е соответствует боксплоту 3.* Данная функция также является строго возрастающей, потому границы усиков нового боксплота - значения функции при 0 и 1, то есть 0.5 и 1. Медиана также будет достигаться в точке 0.5, подставляем ее в функцию и получаем 0.75. Далее проводим сравнения с функцией С, описанные ранее, и определяем, что функции соответствует 3 боксплот.

```
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
import numpy as np
arr = [0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95,1]
arr1 = []
for item in arr:
    arr1.append(-item**2+item+0.5)
df = pd.DataFrame(arr1)
plt.boxplot(df)
plt.show()
```

## II.

Оценка средней зарплаты завышена, так как почти все респонденты с высоким уровнем дохода прошли опрос, а с низким уровнем дохода прошли только половина. Если бы опрос прошли все, то к нашим данным добавилось бы много значений зарплаты ниже 33 тысяч и из-за этого средняя зарплата бы снизилась.

Медиана может быть как завышена, так и занижена. Возможно такое, что доходы подавляющего большинства опрошенных принадлежат к интервалу от 30,622 тысяч до 33 тысяч, и тогда новая медиана будет выше старой. Однако возможно и такое, что большинство опрошенных имеют доходы ниже 30 622 рублей и тогда медианный доход окажется ниже. Также медиана может не измениться.

Значение 75 персентилля преувеличено. У нас добавится много людей с маленькими доходами, исходя из чего граница 75-персентилля гарантированно сместится вниз

Значение 25 персентилля также может быть как преувеличено, так и преуменьшено. Если большинство людей, не прошедших опрос имеют зарплату ниже 19,979 рублей, то значение 25-персентилля может уменьшиться, однако если большинство опрошенных имеет зарплаты ближе к 33 тысячам, то 25 персентиль увеличится из-за увеличения людей.

Максимальная зарплата, скорее всего, является верной, так как почти все сотрудники с высокой зарплатой прошли опрос, а значит данные о наибольшем значении верны. Тем не

менее, нельзя сказать точно, является ли она верной, так как в задании сказано, что “почти все” сотрудники с высокой зарплатой прошли опрос, но не все, следовательно максимальная зарплата может быть у кого-то, кто опрос все-таки не прошел. Сделать однозначные выводы относительно настоящего распределения невозможно. На представленной диаграмме можно заметить 2 явных пика: при порядка 20000 и при 45000 рублях. Можно предположить, что пик около 20000 станет еще более явным и ярко выраженным. Значения меньше 30 тысяч сильно вырастут и из-за этого график распределения может выглядеть как нормальное распределение в диапазоне 10-30 тысяч и маленький “бугорок” от 30 тысяч с пиком, где будет значительно меньше значений.