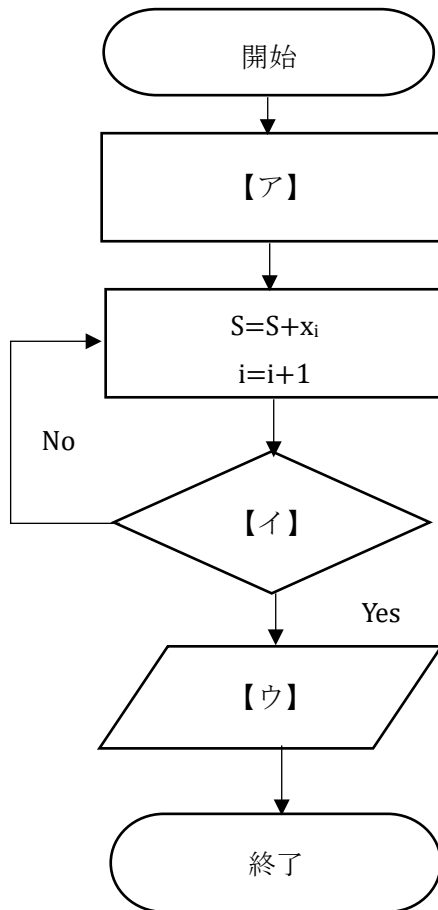




問 4

データ  $(x_1, x_2, \dots, x_{50})$  から標本平均を計算し、その値を出力する基本的なアルゴリズムをフローチャートで描いた。次のフローチャートの【ア】、【イ】、【ウ】に記入するものの組合せとして、下の①～⑤のうちから最も適切なものを一つ選べ。



- |                  |            |                   |
|------------------|------------|-------------------|
| ① 【ア】 $S=1, i=0$ | 【イ】 $i>50$ | 【ウ】 $i$ を出力       |
| ② 【ア】 $S=0, i=1$ | 【イ】 $i<50$ | 【ウ】 $S$ を出力       |
| ③ 【ア】 $S=0, i=1$ | 【イ】 $i>50$ | 【ウ】 $S/i$ を出力     |
| ④ 【ア】 $S=1, i=0$ | 【イ】 $i<50$ | 【ウ】 $S/(i-1)$ を出力 |
| ⑤ 【ア】 $S=0, i=1$ | 【イ】 $i>50$ | 【ウ】 $S/(i-1)$ を出力 |

問 5

国勢調査は日本に住んでいるすべての人と世帯を対象とする国の最も重要な統計調査であり、基本単位区と呼ばれる地域の集計も行われている。日本の人口は国勢調査にもとづいて確定される。

基本単位区は全国で  $J$  箇所あるとし、基本単位区ごとに集計された 5 歳年齢階級別人口について数式を用いて表す。年齢階級として  $K=21$  個の階級（5 歳未満，5 歳以上～10 歳未満，10 歳以上～15 歳未満，…，100 歳以上）を考え、 $j$  番目の基本単位区( $j=1, \dots, J$ )の  $k$  番目( $k=1, \dots, K$ )の年齢階級の人口を  $(j, k)$  成分とする行列  $\mathbf{A}=(A_{jk})$  を考える。日本の  $k$  番目の年齢階級別総人口  $S_k$  と、1 基本単位区当たりの人口  $M$  を表す式の組合せとして、次の①～⑤のうちから最も適切なものを一つ選べ。

$$\textcircled{1} \quad S_k = \sum_{j=1}^K A_{kj}, \quad M = \frac{1}{J} \sum_{k=1}^J \sum_{j=1}^K A_{kj}$$

$$\textcircled{2} \quad S_k = \sum_{j=1}^J A_{jk}, \quad M = \frac{1}{KJ} \sum_{k=1}^K \sum_{j=1}^J A_{jk}$$

$$\textcircled{3} \quad S_k = \sum_{j=1}^J A_{jk}, \quad M = \frac{1}{J} \sum_{j=1}^J \sum_{k=1}^K A_{jk}$$

$$\textcircled{4} \quad S_k = \sum_{j=1}^K A_{kj}, \quad M = \frac{1}{K} \sum_{k=1}^J \sum_{j=1}^K A_{kj}$$

$$\textcircled{5} \quad S_k = \sum_{j=1}^J A_{jk}, \quad M = \frac{1}{K} \sum_{k=1}^K \sum_{j=1}^J A_{jk}$$

問 6

次の文章の空欄【ア】、【イ】に入る式および語句の組合せとして、下の①～⑥のうちから最も適切なものを一つ選べ。

所与の実数値のデータ  $x_1, \dots, x_n$  に対して、各  $x_i$  からの差の二乗和  $V = \sum_{i=1}^n (x_i - c)^2$  を最小にする  $c$  を求める。 $V$  を  $c$  で微分すると【ア】であり、これを 0 とおくことにより  $V$  を最小にする  $c$  は  $x_1, \dots, x_n$  の【イ】であることがわかる。

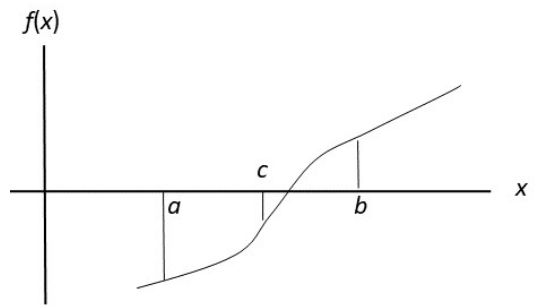
- |                                     |           |
|-------------------------------------|-----------|
| ① 【ア】 : $\sum_{i=1}^n x_i$          | 【イ】 : 中央値 |
| ② 【ア】 : $\sum_{i=1}^n x_i$          | 【イ】 : 平均値 |
| ③ 【ア】 : $-2 \sum_{i=1}^n (x_i - c)$ | 【イ】 : 中央値 |
| ④ 【ア】 : $-2 \sum_{i=1}^n (x_i - c)$ | 【イ】 : 平均値 |
| ⑤ 【ア】 : $\sum_{i=1}^n (x_i - c)$    | 【イ】 : 中央値 |
| ⑥ 【ア】 : $-\sum_{i=1}^n (x_i - c)$   | 【イ】 : 平均値 |

問 7

次の Python プログラムは、ある方程式  $f(x) = 0$  の解を二分法により求めている。二分法とは図にあるように、はさみうちの考え方によって、解を含む区間の長さを順次半分にしていく方法である。このプログラムから出力される方程式の解と、その値を求めるためにプログラム上で設定すべき  $a$  と  $b$  の値【ア】、【イ】の組合せとして、下の①～⑤のうちから最も適切なものを一つ選べ。

```
import math
epsilon = 1e-25
def f(x):
    return(math.log(x/2.0-1.0)/math.log(0.25)-0.5)

a = 【ア】
b = 【イ】
while (a - b) * (a - b) > epsilon:
    c = (a + b) * 0.5
    if f(c)==0:
        break;
    if f(a)*f(c) > 0.0:
        a = c
    else:
        b = c
print('{:.1f}'.format(c))
```



- ① 解: 1.5      【ア】: 0.0      【イ】: 3.0
- ② 解: 3.0      【ア】: 2.5      【イ】: 10.0
- ③ 解: 3.0      【ア】: -3.5      【イ】: 3.5
- ④ 解: 1.5      【ア】: -1.5      【イ】: 1.5
- ⑤ 解: 3.0      【ア】: 4.0      【イ】: 10.0

問 8

次の表は、各レコード（行）が個人に関するデータからなるデータベースである。どのレコードも他の少なくとも  $(k - 1)$  個のレコードと一致するとき、データベースは  $k$ -匿名性を持つという。この表では、例えば 3 行目のレコードは（年齢，性別，居住地）の組合せが他と異なるため、このデータベースは 2-匿名性を持たない。

年齢	性別	居住地
24	男性	東京
24	男性	東京
28	男性	東京
15	男性	大阪
17	男性	大阪
14	男性	大阪
23	女性	東京
26	女性	東京
28	女性	東京
18	女性	東京
15	女性	東京
12	女性	東京

そこで、個人情報の保護の観点から、年齢を 10 歳刻みに粗くして、0～9 歳，10～19 歳，20～29 歳，… のように 10 歳刻みの年齢階級に変換したデータベースを考える。たとえば 1 行目のレコードは

(24 歳，男性，東京) → (20 歳台，男性，東京)

と変換される。変換されたデータベースの  $k$ -匿名性，および変換されたデータベースにおいて年齢の階級値（15 歳および 25 歳とする）から計算した平均年齢に関して、次の文章の空欄【ア】，【イ】に適切な数値を記入せよ。数値は必要に応じて四捨五入することにより、半角，整数値で記入すること。

変換されたデータベースが  $k$ -匿名性を持つ  $k$  の最大値は【ア】である。また平均年齢は【イ】歳である。