

Вибіркові характеристики

1. Знайти вибіркові середню та дисперсію напрямку та методом умовних варіант. Порівняти результати.
2. Знайти вибіркові моду та медіану.

Варіант 7

91	95	85	85	96	87	98	92	90	89
98	97	117	88	91	119	85	77	94	89
79	108	94	112	96	93	98	112	97	88
107	95	99	101	104	81	102	117	105	104
103	107	100	85	105	112	93	92	106	108
84	113	118	98	120	94	92	85	101	108
119	96	94	92	102	109	113	99	98	115
105	86	87	92	96	89	97	115	100	99
119	97	89	99	90	79	91	103	109	108
88	91	98	99	77	72	99	84	82	96

Рішення

1.1) Знайдемо середнє і дисперсію напрямку.

а) Для середнього знайдемо суму усіх чисел і поділимо на їх кількість.

Використаємо для цього програмне забезпечення (скрипт на мові пайтон)

```
def average(xs: Sequence[N]) -> float:
```

```
    length = len(xs)
    if not length:
        return 0
    else:
        return sum(xs) / length
```

Отримане значення: 97.52

б) Для дисперсії скористаємось формулою та вирахуємо її написаною функцією

$$D^*x = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$$

```
def dispersion(xs: List[N]) -> float:
```

```
    length = len(xs)
    avg = average(xs)
    res = sum([(x - avg)**2 for x in xs]) / length
    return res
```

Отримане значення: 114.54

1.2) Знайдемо середнє і дисперсію методом умовних варіант

а) Згрупуємо вибірку таким чином

$$u_i = \frac{1}{h} (z_i - C), i = 1, 2, \dots, k,$$

де h – довжина інтервалу групування, C – середина інтервалу з найбільшою частотою. Тоді вибіркові середнє \bar{x} та D^*x пов'язані з середнім \bar{u} та дисперсією D^*u наступним чином:

$$\bar{x} = h \cdot \bar{u} + C, \quad D^*x = h^2 \cdot D^*u.$$

б) h знайдемо за формулою Стерджеса

```
def distance_sturges(xs: Sequence[N]) -> int:
```

```
    """ Return h of selection """
    n = len(xs)
    m = 1 + 1.322 * math.log(n, 10)
    h = (max(xs) - min(xs)) / m
    return int(h)
```

в) вибірку з u знайдемо за формулами вище

```
def u(interval: List[N], C: float) -> float:
```

```
    diff = max(interval) - min(interval)
    if int(diff) == 0:
        h = 1
    else:
        h = int(diff)
    res = (average(interval) - C) / h
    return res
```

```
def list_u(intervals: Sequence[List[N]]) -> List[float]:
    C = median_longest(intervals)
    return [u(interval, C) for interval in intervals]
```

г) Знаючи це, можемо знаходити дисперсію і середнє значення

```
def dispersion_U(xs: List[N], h: Optional[int] = None) -> float:
    if h is None:
        h = distance_sturges(xs)
    list_intervals = intervals(xs, h)
    list_of_u = list_u(list_intervals)
    average_u_sqr = average([u**2 for u in list_of_u])
    average_u = average(list_of_u)
    return h**2 * average_u_sqr - average_u ** 2
```

```
def average_U(xs: Sequence[N], h: Optional[int] = None) -> float:
    if h is None:
        h = distance_sturges(xs)
    list_intervals = intervals(xs, h)
    C = median_longest(list_intervals)
    average_u = average(list_u(list_intervals))
    return h * average_u + C
```

Отримані таким чином значення:

x середнє = 93.70833333333334

(h за Стерджесом)

Dx* = 107.16040274591104

(h за Стерджесом)

Значення трохи відхиляються від прямих обрахунків, але не сильно. (Якщо зменшувати h — відстань в інтервалах – то середнє значення наближується до прямих обрахунків)

2) Медіана і мода

а) Медіана це значення у вибірці, що ділить вибірку на дві рівні частини. Якщо вибірка має непарну кількість значень, то ми беремо значення по середині, якщо парну – то беремо середнє від двох центральних значень.

Це зображено у цьому коді.

```
def median(xs: List[N]) -> float:
    length = len(xs)
    if length == 1:
        return xs[0]
    elif length % 2 != 0:
        center = int((length + 1) / 2) - 1
        return float(xs[center])
    else:
        first_idx = int(length / 2) - 1
        second_idx = int(length / 2 + 1) - 1
        return (xs[first_idx] + xs[second_idx]) / 2
```

Таке значення – 97.

б) Мода це значення, що зустрічається найчастіше.

```
def moda(xs: List[N]) -> N:
    return max(xs, key=xs.count)
```

Таке значення – 98.