

ЛАБОРАТОРНА РОБОТА №1. ОПИСОВА СТАТИСТИКА

Мета

Навчитись розраховувати числові характеристики вибірки програмними засобами.

Теоретичні відомості

Множину однорідних об'єктів називають *статистичною сукупністю*. *Вибірковою сукупністю (вибіркою)* називають сукупність випадково взятих об'єктів із статистичної сукупності.

Генеральною називають сукупність об'єктів, з яких зроблено вибірку. *Об'ємом* сукупності (вибіркової або генеральної) називають кількість об'єктів цієї сукупності.

Полігоном частот вибірки називають ламану з вершинами в точках (x_i, n_i) . *Полігоном відносних частот* вибірки називають ламану з вершинами в точках $(x_i, n_i/n)$. Полігони частот є аналогами щільності ймовірностей.

Гістограмою частот називають ступінчасту фігуру, яка складається з прямокутників, основами яких є інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють n_i/h .

Нехай x_1, x_2, \dots, x_n – спостереження (значення величини X) у вибірці з об'ємом n . Тоді *вибіркове середнє* можна знайти за формулою:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Медіана вибірки, що має n відсортованих значень, визначається як центральне значення, якщо n непарне число, або як середнє значення двох центральних значень, якщо n парне число.

Мода вибірки визначається як значення, що зустрічається найчастіше у вибірці. Вибірка даних може мати більше однієї моди, і в цьому випадку вона називається *мультимодальною вибіркою*.

Якщо $x_1, x_2, \dots, x_n \in$ вибіркою з об'ємом n , тоді *вибіркова дисперсія* розраховується за формулою:

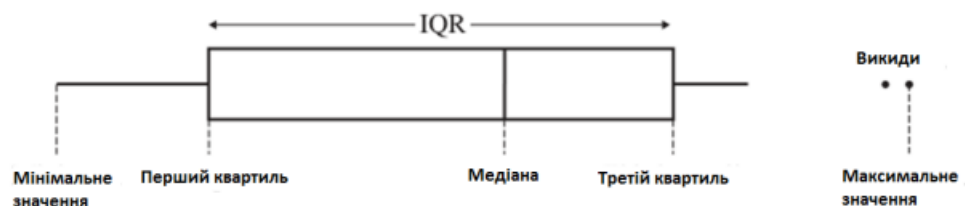
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

або

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

Значення $s = \sqrt{s^2}$ називається *вибірковим середнім квадратичним відхиленням*.

Діаграма розмаху або *коробкова діаграма* – це схематичне представлення положення даних, включаючи найменші та найбільші значення, нижню та верхню чверть вибірки (нижній та верхній квартилі), медіану та статистичні викиди. Виглядає діаграма наступним чином:



Діаграма Парето – це діаграма, де категорії розташовані в порядку зменшення частоти. Діаграма Парето використовується для контролю якості та вдосконалення процесів з метою визначення декількох основних причин більшості проблем та першочергового їх вирішення.

Кругова діаграма або секторна діаграма створюється діленням кола на сектори, де розмір сектора відображає відносну частоту категорії у відсотках.

Завдання

Формула розрахунку варіанта:

$$\text{№ варіанта}^1 = 120 \% \text{ № групи}^2 + \text{№ групи} \times \text{№ команди}^3$$

1. Згенерувати вибірку об'ємом n (див. варіанти завдань нижче) з нормальної популяції. Можна використати онлайн генератор (наприклад, <https://pinetools.com/gaussian-random-number-generator>) або стандартні функції з додаткових бібліотек (наприклад, `numpy.random.normal` для Python). Значення математичних сподівань обрати самостійно.

2. Написати програму, що:

- а) будує полігон та гістограму частот;
- б) розраховує вибіркове середнє, медіану, моду, вибіркві дисперсію та середньоквадратичне відхилення заданої вибірки (написати власні реалізації розрахунків відповідних характеристик);
- в) будує діаграми розмаху, Парето та кругову;
- г) виводить результати пунктів а)-в).

3. Скласти звіт до виконаної роботи, в якому навести значення математичних сподівань, згенерованої вибірки, скріншоти результатів відповідно до п. 2 та посилання на репозиторій з кодом (лінк з qr-кодом на останній сторінці звіту).

Для виконання завдання використовувати мову програмування за власним вибором. На захисті потрібно продемонструвати роботу програми та відповісти на контрольні запитання.

Контрольні питання:

1. Що таке вибіркве середнє, медіана та мода вибірки, як їх знайти?

¹ Орієнтовно 8 команд на академгрупу

² № групи – від 11 до 14

³ № команди – від 1 до 8

2. Що таке вибіркова дисперсія та середньоквадратичне відхилення, як вони позначаються, як їх знайти?

3. Що таке полігон частот, що позначає?

4. Що таке гістограма частот, які її особливості?

Таблиця 1 – Варіанти завдань

№	n	σ	№	n	σ	№	n	σ	№	n	σ
1.	100	1,5	31.	150	2	61.	123	2	91.	120	1,5
2.	99	1,6	32.	134	1,7	62.	146	1,7	92.	142	1,6
3.	103	1,3	33.	145	1,8	63.	143	1,8	93.	124	1,3
4.	102	1,2	34.	125	1,2	64.	122	1,2	94.	117	1,2
5.	124	1,7	35.	128	1,3	65.	121	1,3	95.	110	1,7
6.	146	1,9	36.	147	1,4	66.	147	1,4	96.	139	1,9
7.	109	2,1	37.	101	1,5	67.	116	1,5	97.	101	2,1
8.	135	1,1	38.	106	1,9	68.	143	1,9	98.	132	1,1
9.	140	2	39.	117	2,1	69.	147	2,1	99.	106	2
10.	140	1,8	40.	124	2,2	70.	145	2,2	100.	129	1,8
11.	114	1,4	41.	120	1,1	71.	148	1,1	101.	135	1,4
12.	104	2,2	42.	132	2	72.	115	2	102.	105	2,2
13.	104	1,5	43.	122	1,7	73.	123	1,7	103.	132	1,5
14.	139	1,6	44.	134	1,8	74.	118	1,8	104.	107	1,6
15.	150	1,3	45.	137	1,2	75.	146	1,2	105.	131	1,3
16.	117	1,2	46.	114	1,3	76.	148	1,3	106.	124	1,2
17.	150	1,7	47.	99	1,4	77.	128	1,4	107.	116	1,7
18.	149	1,9	48.	150	1,5	78.	100	1,5	108.	128	1,9
19.	138	2,1	49.	115	1,9	79.	115	1,9	109.	108	2,1
20.	141	1,1	50.	136	2,1	80.	98	2,1	110.	124	1,1
21.	122	2	51.	124	2,2	81.	121	2,2	111.	146	2
22.	138	1,8	52.	109	1,1	82.	110	1,1	112.	148	1,8
23.	128	1,4	53.	101	2	83.	114	2	113.	120	1,4
24.	145	2,2	54.	148	1,7	84.	143	1,7	114.	146	2,2
25.	116	1,5	55.	144	1,8	85.	129	1,8	115.	115	1,5
26.	138	1,6	56.	98	1,2	86.	147	1,2	116.	103	1,6
27.	109	1,3	57.	131	1,3	87.	99	1,3	117.	124	1,3
28.	104	1,2	58.	103	1,4	88.	105	1,4	118.	107	1,2
29.	149	1,7	59.	98	1,5	89.	149	1,5	119.	98	1,7
30.	123	1,9	60.	127	2,1	90.	115	2,1	120.	125	1,9