

Міністерство освіти і науки України
Львівський національний університет імені Івана Франка

Звіт
з індивідуального завдання з курсу
«Математична Статистика»
студента групи ПМІ-31
Васька Юрія

Львів 2018

Постановка задачі

Згенерувати вибірку заданого об'єму з певного проміжку для неперервної та дискретної статистичної змінної (та зчитати з файлу у вигляді частотної таблиці) .

На підставі отриманих вибірових даних:

1. побудувати варіаційний ряд;
2. побудувати статистичний розподіл варіанти;
3. представити графічно статистичний матеріал,
4. побудувати емпіричну функцію розподілу;
5. обчислити всі числові характеристики.

Користуючись критерієм Пірсона, на підставі наведених статистичних даних при заданому рівні значущості перевірити правильність висунутої гіпотези щодо закону розподілу генеральної сукупності.

Короткі теоретичні відомості

Варіаційний ряд – це зростаючий числовий ряд.

Припустимо, що вивчають деяку генеральну випадкову величину X . Для цього проводять низку незалежних дослідів або спостережень, у кожному з яких величина X набуває того чи іншого

значення. Сукупність отриманих значень x_1, x_2, \dots, x_n величини X (де n - кількість дослідів) і є утворена нами вибірка. Цю сукупність часто називають **статистичним рядом**. Статистичний ряд відіграє роль вихідного числового матеріалу, що підлягає подальшій обробці та аналізу.

Наступний етап обробки вихідного статистичного ряду – побудова **статистичного (емпіричного) закону розподілу**. Форма його запису залежить від характеру досліджуваної випадкової величини X .

Для неперервної випадкової величини:

Для вибірки з неперервного розподілу або для вибірки великого об'єму використовують представлення її у вигляді інтервального статистичного ряду, який називають ще статистичним рядом за згрупованими параметрами.

Інтервали	(x_1, x_2)	(x_2, x_3)	...	(x_{k-1}, x_k)
Частоти	n_1	n_1	...	(n_k)

Для дискретної випадкової величини:

Нехай у варіаційному ряді варіанта x_1 повторюється n_i разів, тоді число n_i називають його частотою, а його відношення до об'єму вибірки - відносною частотою.

За отриманими значеннями варіант, частот і відносних частот можна побудувати таблиці, які називають відповідно розподілами частот і відносних частот. Розподіли частот і відносних частот називають також статистичним розподілом вибірки.

Розподіл частот:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

Розподіл відносних частот:

x_i	x_1	x_2	...	x_k
v_i	v_1	v_2	...	v_k

$$\frac{n_i}{n} = v_i$$

де, n — загальна кількість спостережень.

Графічне представлення:

Статистичний розподіл вибірки можна задати графічно полігоном або гістограмою частот (відносних частот).

Полігон розподілу вибірки використовується для зображення як дискретних, так і інтервальних варіаційних рядів, а гістограма — лише для інтервальних рядів.

Полігоном частот називають ламану, відрізки якої послідовно з'єднують точки $(x_1; n_1), (x_2; n_2), \dots, (x_k; n_k)$ координатної площини.

Емпірична функція розподілу:

Емпіричною функцією розподілу випадкової величини X (функцією розподілу вибірки) називають функцію $F_n(x)$, що визначає для будь-якого дійсного числа x відносну частоту події $\{X < x\}$, тобто

$$F_n(x) = \frac{n_x}{n}$$
$$F_n(x) = \begin{cases} 0, & x \leq x_1; \\ \frac{n_1}{n}, & x_1 \leq x \leq x_2; \\ \frac{n_1 + n_2}{n}, & x_2 \leq x \leq x_3; \\ \dots\dots\dots \\ \frac{n_1 + n_2 + \dots + n_{k-1}}{n} & x_{k-1} \leq x \leq x_k; \\ 1 & x > x_k. \end{cases}$$

Числові характеристики розподілу:

Медіана - це варіанта, що ділить варіаційний ряд на дві частини з рівною кількістю варіант. Якщо кількість варіант непарна ($n=2k+1$), то медіана дорівнює x_{k+1}

$$\underbrace{x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)}}_{k\text{-елементів}} \leq x_{(k+1)} \leq \underbrace{x_{(k+2)} \leq \dots \leq x_{(n)}}_{k\text{-елементів}}.$$

У випадку парної кількості варіант ($n=2k$) медіана дорівнює $(x_k + x_{k+1})/2$.

$$\underbrace{x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)}}_{k\text{-елементів}} \leq \underbrace{x_{(k+1)} \leq \dots \leq x_{(n)}}_{k\text{-елементів}}.$$

Мода статистичного матеріалу - це варіанта, яке найчастіше трапляється у вибірці.

Середнє вибіркове або **середнє арифметичне** - це величина, яка дорівнює середньому арифметичному елементів даної вибірки.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Девіація – це величина, яка дорівнює сумі квадратів відхилень елементів вибірки від її середнього вибіркового.

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

Розмах вибірки – це величина, яка дорівнює різниці між максимальним та мінімальним елементами цієї вибірки і позначається p

$$p = \max(x_1, \dots, x_n) - \min(x_1, \dots, x_n).$$

Варіанса - це величина, яка дорівнює сумі квадратів відхилень елементів вибірки від її середнього вибіркового поділену на обсяг вибірки без одного елемента і позначається s^2 .

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

Стандарт – це величина, яка дорівнює значенню кореня квадратного з варіанси і позначається s .

$$s = +\sqrt{s^2}.$$

Варіація – це величина яка дорівнює відношенню стандарту вибірки до її середнього вибіркового.

$$v = \frac{s}{x},$$

Квантиль вибірки порядку α – це елемент цієї вибірки (якщо він існує), до якого, включно з ним, є α процентів елементів статистичного матеріалу.

$$\underbrace{x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(r)}}_{\alpha \% \text{ від } n} x_{(r+1)} \leq \dots \leq x_{(n)};$$

Момент порядку k відносно константи c вибірки x_1, x_2, \dots, x_n – це величина, яка дорівнює середньому арифметичному k -их степенів усіх відхилень елементів цієї вибірки від константи c .

$$M_k(c) = \frac{1}{n} \sum_{i=1}^n (x_i - c)^k, \quad (k = 0, 1, \dots).$$

Якщо c дорівнює нулю, то такий момент називається початковим, а якщо c дорівнює середньому вибіркового цієї вибірки – центральним.

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k; \quad \mu_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k,$$

Асиметрія – це величина, яка дорівнює відношенню третього центрального моменту до другого в степені півтора.

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}.$$

Ексцес – це величина, яка дорівнює відношенню четвертого центрального моменту до другого в другому степені мінус 3.

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3.$$

Критерій узгодженості Пірсона – це один з найвідоміших критеріїв, тому його часто і називають просто «критерій хі-квадрат». Використовується для перевірки гіпотези про закон розподілу.

Отримані результати та їх аналіз

Згенерувати вибірку заданого об'єму (50 елементів) з певного проміжку (від 10 до 25) для неперервної статистичної змінної.

ВВЕДІТЬ ПРОМІЖОК

ВІД

ДО

ВВЕДІТЬ ПОТУЖНІСТЬ
ВИБІРКИ

Неперервна статистична змінна

Дискретна статистична змінна

Зчитати дані про неперервну статистичну змінну з файлу

Зчитати дані про дискретну статистичну змінну з файлу

Частотна таблиця

Елемент	Частота	Ймовірність
11.25	7	0.14
13.75	8	0.16
16.25	12	0.24
18.75	10	0.2
21.25	6	0.12
23.75	7	0.14
*		

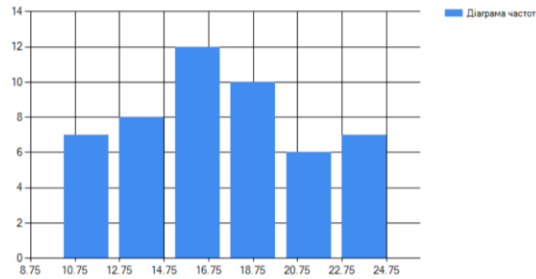
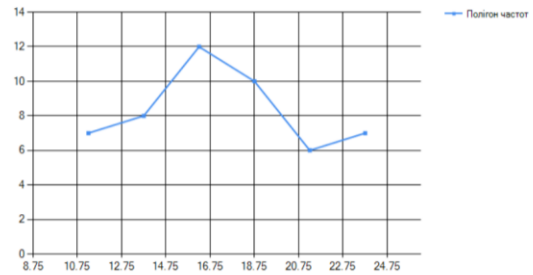
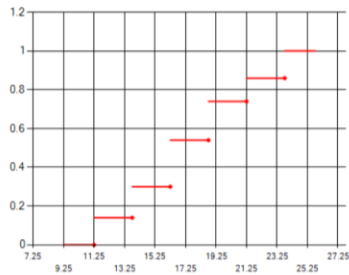
Переглянути графічне представлення статистичного матеріалу

Переглянути числові характеристики статистичного матеріалу

Перевірити правильність висунутої гіпотези, щодо нормального розподілу генеральної сукупності

Побудував статистичний розподіл варіанти, записав варіаційний ряд, побудував відповідні графіки (діаграми та полігон частот) та емпіричну функцію розподілу, обчислив усі числові характеристики та перевірів справедливість гіпотези про нормальний розподіл за критерієм Пірсона.

Графічне представлення статистичного матеріалу



Числові характеристики статистичного матеріалу

Величина	Значення
Варіаційний ряд	10.395; 10.432; 10.55; 10.587; 10.616; 12.323; 12.433; 12.573; 12.954; 13.71; 13.746; 14.068; 14.157; 14.651; 14.794; 15.132; 15.203; 15.234; 15.412; 15.465; 15.52; 15.696; 15.837; 15.978; 16.218; 16.514; 16.795; 17.825; 18.058; 18.151; 18.596; 18.785; 18.829; 18.886; 19.115; 19.156; 19.418; 20.356; 20.401; 20.965;
Об'єм вибірки	50
Медіана	16.25
Мода	16.25
Середнє вибіркове	17
Девіація	776.125
Розмах	12.5
Варіанса	15.839
Стандарт	3.98
Вибіркова дисперсія	15.522
Варіація	0.23
Центральний момент першого порядку	-0.0824790964287145
Початковий момент першого порядку	17.2175209035713
Асиметрія	0.169
Екセス	-0.86

Хі емпіричне

0.310181260

Виберіть рівень значущості

0.3

Перевірити гіпотезу

Хі критичне

0.1485

Оскільки Хі емпіричне більше за Хі критичне
Гіпотезу відхилено

Критерій пірсона

m(i)	np(i)
7	5.655
8	8.395
12	11.945
10	11.59
6	7.66
7	4.755

Кінцева таблиця (np(i)>=10 та m(i)>=5)

m(i)	np(i)
15	14.05
12	11.945
10	11.59
13	12.415

Висновок

На цьому індивідуальному завданні я навчився створювати стільникову аплікацію, яка дає змогу:

- Згенерувати вибірку заданого об'єму з певного проміжку для неперервної та дискретної
- статистичної змінної(також можна зчитати з файлу)
- Будувати варіаційний ряд з отриманих даних
- Будувати статистичний розподіл варіанти
- Графічно представити дані
- Обчислювати числові характеристики
- На підстав критерію Пірсона, перевірити правильність висунутої гіпотези щодо закону генеральної сукупності.

Для перевірки правильності роботи програми звірів результати з обчисленнями, отриманими в процесі вивчення матеріалу для виконання індивідуального завдання.