Тема 1: ***Основні поняття та статистичний розподіл*.**

Математичною статистикою називається наука, що займається розробкою методів отримання, опису та обробки дослідних даних з метою вивчення закономірностей випадкових масових явищ.

Математична статистика включає в себе два основні розділи:

1) теорію оцінювання;

2) теорію перевірки статистичних гіпотез.

Основні завдання, які вирішує математична статистика:

1) вказівки способів збору і угруповання статистичних даних, отриманих в результаті спостережень або в результаті спеціально поставлених експериментів;

2) визначення закону розподілу випадкової величини або системи випадкових величин за статистичними даними;

3) визначення невідомих параметрів розподілу;

4) перевірка правдоподібності припущень про закон розподілу випадкової величини, про форму зв'язку між випадковими величинами або про значення оцінюваного параметра.

Основне завдання математичної статистики - це розробка методів аналізу статистичних даних в залежності від мети дослідження.

Методи математичної статистики ефективно використовуються при вирішенні багатьох завдань науки, організації технологічного процесу, планування, управління та ціноутворення.

Виникала математична статистика в 17 столітті і розвивалася паралельно з теорією ймовірності. Подальший розвиток (кінець 19 - початок 20 століть) математична статистика отримала в роботах: П.Л. Чебишева, А.А. Маркова, О.М. Ляпунова, К. Гаусса, Ф. Гамільтона та інших вчених. У XX столітті найбільший внесок у математичну статистику внесли В.І. Романовський, Е.Е Слуцький, А.Н Колмогоров, Стьюдент, Є. Пірсон, Ю. Нейман, А. Вальд, А.В. Скороход, В.С. Королюк та інші вчені.

Фундаментальними поняттями математичної статистики є поняття генеральної сукупності і вибірки обсягу n.

Генеральною сукупністю будемо називати сукупність всіх мислимих спостережень результатів над випадковою величиною, які в принципі могли б бути проведені в заданих умовах досвіду (позначають N).

Вибіркою обсягу n називається кінцевий набір значень випадкової величини, взятий в заданих умовах досвіду.

Вибірки бувають повторні і безповторні.

Повторної називають вибірку, при якій відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність. Безповторною називають вибірку, при якій відібраний об'єкт в генеральну сукупність не повертається.

Зазвичай використовуються безповторні вибірки.

Вибірка повинна бути репрезентативною або представницької, що означає, що дані вибірки повинні правильно відбивати ознака генеральної сукупності.

Способи відбору:

1. Відбір, що не вимагає поділу генеральної сукупності на частини.

Сюди відносяться: 1) простий випадковий безповторний відбір; 2) простий випадковий повторний відбір.

Простим випадковим називають такий відбір, при якому об'єкти витягують по одному з усієї генеральної сукупності.

2. Відбір, при якому генеральна сукупність ділиться на частини. Сюди відносять: типический відбір, механічний відбір, серійний відбір.

Типовим називається відбір, при якому об'єкти відбираються не з усієї генеральної сукупності, а з кожної її «типової» частини.

Механічним називають відбір, при якому генеральну сукупність «механічно» ділять на стільки груп, скільки об'єктів повинно увійти у вибірку, а з кожної групи випадковим чином вибирається один об'єкт. Щоб механічний відбір був репрезентативним, необхідно враховувати специфіку технологічного процесу.

Серійним називається відбір, при якому об'єкти відбирають з генеральної сукупності не по одному, а «серіями», які піддаються суцільного обстеження. Використовується тоді, коли досліджуваний ознака коливається в різних серіях незначно.

В економічних дослідженнях іноді використовують комбінований відбір.

Нехай з генеральної сукупності N взята вибірка обсягу n для вивчення ознаки X, який прийняв значення x_1 у випадках n_1, $x_2 - n_2$ раз,… ..$x_m - n_m$ раз. Значення $x_1, x_2, ..., x_m$ називаються варіантами ознаки.

Варіанти, записані в зростаючому порядку із зазначенням частоти, з якою вони зустрілися в даній сукупності, утворюють статистичний розподіл вибірки або варіаційний ряд**.**

Варіація ознаки може бути дискретною і безперервною. Дискретної називається варіація, при якій варіанти відрізняються один від одного на деяку кінцеву величину. Безперервної називається варіація, при якій значення ознаки можуть відрізнятися одне від іншого на як завгодно малу величину. При безперервній варіації статистичний розподіл вибірки називається інтервальним і може бути задане у вигляді послідовних інтервалів і відповідних їм частот (в якості частоти інтервалу приймаємо суму частот варіант, що потрапили в цей інтервал). При переході від безперервного статистичного розподілу вибірки до дискретного як значень інтервалу приймають його середину, тобто центральне значення**.**

Кількість спостережуваних варіант $n_1, n_2, ..., n_m$ називаються частотами, для яких виконується наступне $\sum_{k=1}^{m} n_k = n$. Відношення частоти варіанти до обсягу вибірки називається відносною частотою варіанти (позначається ω_k). Сума всіх відносних частот дорівнює одиниці: $\sum_{k=1}^{m} \omega_k = 1$.

x_i	x_1	x_2	x_3	...	x_m
n_i	n_1	n_2	n_3	...	n_m
ω_i	ω_1	ω_2	ω_3	...	ω_m

$x_1 < x_2 < x_3 < ... < x_m$

Емпірична функція розподілу і її властивості.

Нехай маємо статистичний розподіл частот деякої ознаки Х. Позначимо через n- обсяг вибірки, n_x- кількість спостережень, при яких спостережуваний ознака $X < x$. Тоді відносна частота події $X < x$ дорівнює n_x/n. При зміні x змінюється відносна частота n_x/n, тобто величина n_x/n є функція від x. Так як ця функція знаходиться емпіричним шляхом, то її називають емпіричною.

Емпіричної функцією розподілу (і функцією розподілу вибірки) називають функцію $F^*(x)$, що визначає для кожного значення x відносну частоту події $X < x$. За визначенням, $F^*(x) = n_x/n$, де n_x- число варіант, менших $x; n$ - обсяг вибірки.

Властивості емпіричної функції:

.

1. $0 \le F^*(x) \le 1$.
2. $F^*(x)$ - неспадаюча функція.
3. $F^*(x) = 0, x \le x_1 \quad F^*(x) = 1, x > x_m$,

Графічне зображення статистичних розподілів – полігон частот, полігон відносних частот, гістограма частот, гістограма відносних частот.

Для того, щоб наочно мати уявлення про характер розподілу, застосовують графічне зображення рядів розподілу. Основними способами їх зображення є полігон частот, гістограма, кумулята.

Гістограму застосовують для зображення інтервальних варіаційних рядів. При її побудові на осі абсцис відкладають відрізки, які зображують інтервал. На кожному з відрізків будують прямокутник з висотою, яка дорівнює частоті, що відповідає інтервалу, поділеній на довжину інтервалу. В результаті площа фігури, що складається з прямокутників, дорівнює одиниці.

Полігон розподілу – це ламана лінія з вершинами в точках, координатами яких є варіанти (в інтервальних рядах середини інтервалів) і частоти, що відповідають цим варіантам. Полігон розподілу застосовують для зображення дискретних і інтервальних варіаційних рядів.

Кумулята – графічне зображення варіаційного ряду з накопиченими частотами (відносними частотами). Для її побудови на осі абсцис відкладають варіанти, а на осі ординат – накопичені частоти (відносні частоти). Кумулята має вигляд східчастої функції.

Кумулята или кумулятивная кривая в отличие от полигона строится по накопленным частотам или частостям. При этом на оси абсцисс помещают значения признака, а на оси ординат — накопленные частоты или частости (рис. 3)

При построении кумуляты накопленная частота (частость) соответствующего интервала присваивается его верхней границе:

Числові характеристики статистичного розподілу вибірки.

1. Показники центру розподілу.

1. Модою називається значення варіанти, що має найбільшу частоту. Для дискретного ряду мода визначається по частотах варіант і відповідає варіанті з найбільшою частотою. У разі інтервального розподілу з рівними інтервалами, інтервал, що містить моду, визначається за найбільшою частоті, а при нерівних інтервалах - по найбільшої щільності. Мода розраховується за формулою

$$M_o = x_{M_{o(\min)}} + h\frac{n_{M_o} - n_{M_o - 1}}{\left(n_{M_o} - n_{M_o - 1}\right) + \left(n_{M_o} - n_{M_o + 1}\right)}.$$

2.Медіана, це таке значення варьирующего ознаки, яке припадає на середину впорядкованого варіаційного ряду. Для дискретного ряду: $M_e = \begin{cases} x_{\frac{n+1}{2}}, & n - непарне \\ \frac{1}{2}\left[x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right], & n \text{ - } парне \end{cases}$.

При розрахунку медіани інтервального варіаційного ряду спочатку знаходять інтервал, що містить медіану, шляхом використання накопичених частот. Медианному інтервалу відповідає перша з накопичених частот, що перевищує половину всього обсягу сукупності. Для знаходження медіани при сталості щільності всередині інтервалу, що містить медіану, використовують формулу:

$$M_e = x_{M_{e(\min)}} + h\frac{\frac{1}{2}\sum_{i=1}^{m} n_i - V_{M_e - 1}}{n_{M_e}}.$$

3.Вібіркове середнє:

$$x_B = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$$

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

2. Показники міри розсіяння.

Вибіркова дисперсія

$$D_B = \frac{1}{n}\sum_{i=1}^{m} n_i \left(x_i - \overline{x_B}\right)^2 = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\overline{x_B}\right)^2$$

Вибіркове середньоквадратичне відхилення

$$\sigma_B = \sqrt{D_B} \; .$$

3. Відносні показники варіації.

1. **Коефіцієнт варіації**

$$V = \frac{\sigma_B}{x_B} \cdot 100\%$$

2. Вибіркова асиметрія:

$$A = \frac{\mu_3}{\sigma_B^3}.$$

3. Вибірковий ексцес

$$E = \frac{\mu_4}{\sigma_B^4} - 3.$$