

Лекція 13.
Вибірковий метод аналізу властивостей генеральної сукупності.

План лекції

1. Предмет та основні поняття математичної статистики.....	2
2. Вибірковий метод і його основні поняття. Випадкова вибірка, об'єм вибірки.....	2
3. Способи відбору	3
4. Варіаційний ряд для дискретних і неперервних випадкових величин. ...	5
5. Полігон і гістограма	6

Питання, що розглядаються:

Предмет математичної статистики, статистична сукупність, статистичні дані, дві основні задачі математичної статистики, випадкова вибірка, генеральна сукупність, об'єм вибірки, повторні і безповторні вибірки, простий випадковий відбір, типовий відбір, механічний відбір, серійний відбір, комбінований відбір, варіанти, варіаційний ряд, частоти, відносні частоти, полігон, гістограма.

1. Предмет та основні поняття математичної статистики.

Предметом математичної статистики є вивчення випадкових подій і випадкових величин за результатами спостережень. Сукупність предметів або явищ, об'єднаних якою-небудь загальною ознакою, називається **статистичною сукупністю**. Результатом спостережень над статистичною сукупністю є **статистичні дані** - відомості про те, яких значень набула у результаті спостережень ознака (випадкова величина X), що нас цікавить.

Обробка статистичних даних методами математичної статистики призводить до встановлення певних закономірностей, властивих масовим явищам. При цьому **точність** статистичних висновків підвищується із зростанням числа спостережень.

Статистичні дані, як правило, є рядом значень x_1, x_2, \dots, x_n деякої випадкової величини. Обробка цього ряду значень є першим етапом дослідження випадкової величини.

Перша задача математичної статистики - **вказати способи збору і групування статистичних даних**, отриманих в результаті спостережень або в результаті спеціально поставлених експериментів.

Другою задачею математичної статистики є розробка **методів аналізу** статистичних даних залежно від цілей дослідження. До цієї задачі відносяться:

- Оцінка невідомої **ймовірності події**; оцінка невідомої **функції розподілу**; оцінка **параметрів розподілу**, вид якого відомий; оцінка залежності випадкової величини від однієї або декількох випадкових величин і тому подібне
- Перевірка статистичних гіпотез про вид **невідомого** розподілу або про **величину параметрів розподілу**, вид якого відомий.

У сучасній математичній статистиці є багато спільного з **наукою про ухвалення рішень в умовах невизначеності**, оскільки вона розробляє способи визначення числа необхідних випробувань до початку дослідження (планування експерименту), в процесі дослідження (послідовний аналіз) і вирішує багато інших аналогічних завдань.

2. Вибірковий метод і його основні поняття. Випадкова вибірка, об'єм вибірки.

Нехай вимагається вивчити сукупність **однорідних** об'єктів відносно деякої якісної або кількісної ознаки, що характеризує ці об'єкти. Наприклад, для партії деталей якісною ознакою може служити стандартність деталі, а кількісною - контрольований розмір деталі.

В принципі, можливе проведення суцільного обстеження, тобто обстеження усіх об'єктів. На практиці таке обстеження застосовується рідко, наприклад,

- ✓ через велике числа об'єктів
- ✓ через дорожнечу проведення операції контролю,
- ✓ через те, що контроль часто пов'язаний з руйнуванням об'єкту (перевірка електролампи на довговічність її роботи), і так далі

У таких випадках випадково відбирається і вивчається **обмежене** число об'єктів з сукупності.

Вибірковою сукупністю або **випадковою вибіркою** називають сукупність випадково відібраних об'єктів.

Генеральною сукупністю називають сукупність об'єктів, з яких робиться вибірка.

Об'ємом сукупності (вибіркової або генеральної) називають число об'єктів цієї сукупності. Наприклад, якщо з 1000 деталей відбирається для обстеження 100, то об'єм генеральної сукупності $N=1000$, а об'єм вибірки $n = 100$.

Приклад: Число одиниць товару N , виробленого деяким підприємством впродовж року, є генеральна сукупність. Для дослідження якості продукції на практиці розглядається вибірка, що складається з n одиниць товару. Ознакою, або випадковою величиною, може бути число одиниць товару, що задовольняють сертифікаційним вимогам.

При складанні вибірки можна поступати двома способами: після того, як об'єкт відібраний і досліджений, його можна повернути або не повертати в генеральну сукупність. У зв'язку з цим вибірки підрозділяються на **повторні** і **безповторні**.

Повторною називають вибірку, при якій відібраний об'єкт (перед відбором наступного) повертається в генеральну сукупність. При **безповторній** вибірці відібраний об'єкт в генеральну сукупність не повертається.

Для того, щоб за даними вибірки можна було досить упевнено судити про ознаку генеральної сукупності, що нас цікавить, необхідно, щоб об'єкти вибірки правильно його представляли. Вибірка повинна правильно представляти пропорції генеральної сукупності, тобто вибірка має бути репрезентативною (представницькою). Приклад - вивчення громадської думки.

Через закон великих чисел можна стверджувати, що вибірка буде репрезентативною, якщо її здійснити випадково: кожен об'єкт вибірки відібраний випадково з генеральної сукупності, якщо усі об'єкти мають однакову ймовірність потрапити у вибірку.

Якщо об'єм вибірки досить великий, а вибірка складає лише незначну частину сукупності, то відмінність між повторною і безповторною вибіркою стирається.

3. Способи відбору

На практиці застосовуються різні способи відбору, які можна підрозділити на два види:

- Відбір, що не вимагає розчленовування генеральної сукупності на частини. Сюди відносяться а) простий випадковий неповторний відбір і б) простий випадковий повторний відбір.
- Відбір, при якому генеральна сукупність розбивається на частини. Сюди відносяться а) типовий відбір, б) механічний відбір і в) серійний відбір.

Простим випадковим називають відбір, при якому об'єкти витягаються по одному з генеральної сукупності. Здійснити такий відбір для генеральної сукупності з N об'єктів можна, наприклад, за допомогою запису на картках номерів від 1 до N , наступному перемішуванню карток і виїманню їх навмання. При цьому обстеженню підлягають об'єкти, що мають номери, співпадаючі з номерами карток. Якщо картки повертаються в пачку, то маємо просту випадкову повторну вибірку, інакше - просту неповторну. При великому об'ємі генеральної сукупності раціональнішим є використання таблиць випадкових чисел. Наприклад, щоб вибрати 50 об'єктів з пронумерованої генеральної сукупності, відкривають будь-яку сторінку таблиці випадкових чисел і виписують 50 чисел підряд; у вибірку потрапляють ті об'єкти, номери яких співпадають з виписаними випадковими числами. Якщо випадкове число таблиці перевершує число N , таке число пропускають. При проведенні неповторної вибірки пропускають також випадкові числа, що вже зустрічалися раніше.

Типовим називають відбір, при якому об'єкти відбираються не з усієї генеральної сукупності, а з кожної її "типової" частини. Наприклад, якщо деталі виготовлені на декількох верстатах, то відбір виробляють з продукції кожного верстата окремо.

Механічним називають відбір, при якому генеральна сукупність механічно ділиться на стільки груп, скільки об'єктів повинно увійти до вибірки, а з кожної групи вибирається один об'єкт. Наприклад, якщо треба відібрати 20% виготовлених верстатом деталей, то відбирають кожну п'яту деталь.

Серійним називають відбір, при якому об'єкти відбирають з генеральної сукупності не по одному, а "серіями", які піддаються суцільному обстеженню. Наприклад, якщо вироби виробляються великою групою верстатів-автоматів, то піддають суцільному обстеженню продукцію тільки декількох верстатів. Цим видом відбору користуються тоді, коли обстежувана ознака коливається в різних серіях незначно.

На практиці часто застосовують **комбінований** відбір, при якому поєднуються вказані вище способи. Наприклад, розбивають генеральну сукупність на серії однакового об'єму, потім простим випадковим відбором вибирають декілька серій і, нарешті, з кожної серії простим випадковим відбором витягають окремі об'єкти.

4. Варіаційний ряд для дискретних і неперервних випадкових величин.

Нехай з генеральної сукупності витягнута вибірка, причому значення досліджуваного параметра x_1 спостерігалось n_1 раз, x_2 - n_2 раз і так далі. При цьому $\sum_i n_i = n$ об'єм вибірки. Спостережувані значення x_i називають **варіантами**, а послідовність варіант, записаних в зростаючому порядку - **варіаційним рядом**. Числа спостережень називають **частотами**, а їх відношення до об'єму вибірки n_i/n - **відносними частотами**. Варіаційний ряд можна представити таблицею:

X	x_1	x_2	..	x_m
n	n_1	n_2	.	n_m

Статистичним розподілом вибірки називають перелік варіант і відповідних їм відносних частот. Статистичний розподіл можна представити як

X	x_1	x_2	..	x_m
w	w_1	w_2	.	w_m

де відносні частоти $w_k = \frac{n_k}{n}$.

Зауважимо, що в теорії ймовірності під розподілом розуміють відповідність між можливими значеннями випадкової величини і їх **ймовірностями**, а в математичній статистиці - відповідність між спостережуваними **варіантами** і їх **частотами** або **відносними частотами**.

Приведений спосіб представлення статистичних даних застосовують у разі дискретних випадкових величин. Для неперервних випадкових величин зручніше розбити відрізок $[a, b]$ можливих значень випадкової величини на часткові напівінтервали $\Delta_k = [a_{k-1}, a_k)$, $(k = 1, \dots, m)$ (Δ_m замкнутий також і справа) за допомогою деякої системи точок $a = a_0 < a_1 < \dots < a_m = b$. Часто розбиття $[a, b]$ проводять на рівні частини, тоді $\Delta_k = [a + (k-1)h, a + kh)$, $k = 1, \dots, m$, де $h = \frac{b-a}{m}$

В якості частот n_k тепер потрібно брати кількість спостережуваних значень, що потрапили до кожного з часткових інтервалів Δ_k . Варіаційний ряд має у такому разі вигляд

X	Δ_1	Δ_2	..	Δ_m
n	n_1	n_2	.	n_m

а статистичний розподіл -

X	Δ_1	Δ_2	..	Δ_m
n	w_1	w_2	.	w_m

Число інтервалів k часто вибирають на підставі формули Стерджерса

$$k = 1 + 1,4 \ln n .$$

5. Полігон і гістограма

Графічно статистичний розподіл представляється зокрема, за допомогою *полігону і гістограми*.

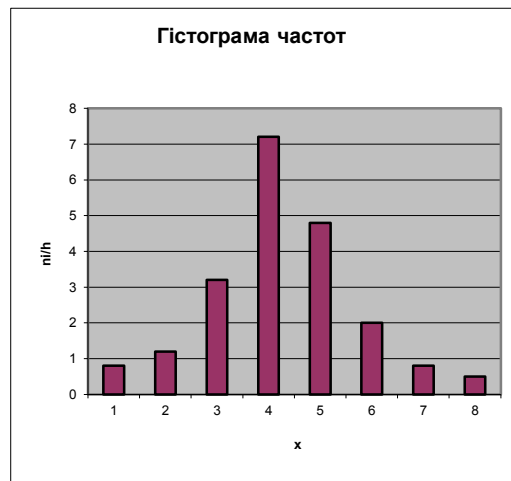
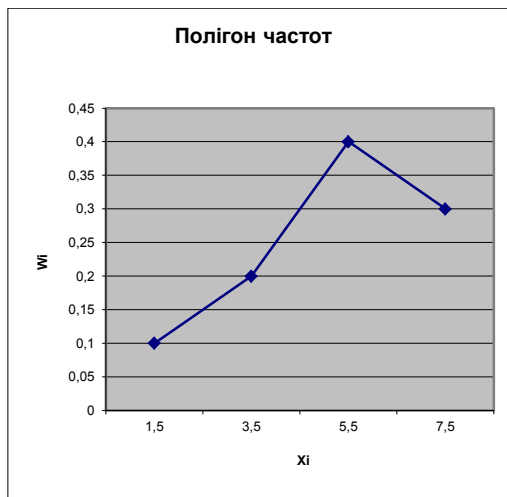
Полігоном частот називають ламану лінію, відрізки якої сполучають точки $(x_1; n_1), (x_2; n_2), \dots, (x_k, n_k)$. Для побудови полігону частот на осі абсцис відкладають варіанти x_i , а на осі ординат - відповідні ним частоти n_i і сполучають точки $(x_i; n_i)$ відрізками прямих.

Полігон відносних частот будується аналогічно, за винятком того, що на осі ординат відкладаються відносні частоти w_i .

У разі неперервної ознаки будується *гістограма*, для чого інтервал, в якому знаходяться усі спостережувані значення ознаки, розбивають на декілька часткових інтервалів завдовжки h і знаходять для кожного часткового інтервалу n_i - суму частот варіант, що потрапили в i -й інтервал.

Гістограмою частот називають ступінчасту фігуру, що складається з прямокутників, основами якої служать часткові інтервали завдовжки h , а висоти дорівнюють відношенню n_i / h . Для побудови гістограми частот на осі абсцис відкладають часткові інтервали, а над ними проводять відрізки, паралельні осі абсцис на відстані (висоті) n_i / h . Площа i -го прямокутника рівна $hn_i / h = n_i$ - сумі частот варіант i -го інтервалу, тому площа гістограми частот дорівнює сумі усіх частот, тобто **об'єму вибірки**.

У разі гістограми відносних частот по осі ординат відкладаються відносні частоти w_i , на осі абсцис - часткові інтервали, над ними проводять відрізки, паралельні осі абсцис на висоті w_i / h . Площа i -го прямокутника дорівнює відносній частоті варіант w_i , що потрапили в i -й інтервал. Тому площа гістограми відносних частот дорівнює сумі усіх відносних частот, тобто одиниці.



Питання для самоперевірки

1. Що вивчає математична статистика?
2. Сформулювати дві основні задачі математичної статистики.
3. Дати означення вибіркової та генеральної сукупності.
4. Як визначається об'єм вибірки?
5. Що таке повторка і безповторна вибірки?
6. Перерахувати види відборів.
7. Дати означення варіанти, частоти, варіаційного ряду.
8. Що таке статистичний розподіл вибірки?
9. Якими способами можна графічно представити вибірку?
10. Як будується полігон частот? Відносних частот?
11. Як будується гістограма частот? Відносних частот?