

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ВНУТРІШНІХ СПРАВ  
КРЕМЕНЧУЦЬКИЙ ЛЬОТНИЙ КОЛЕДЖ**

**Циклова комісія економіки та управління**

**ТЕКСТ ЛЕКЦІЇ**

**з навчальної дисципліни «Статистика»  
обов'язкових компонент  
освітньо-професійної програми першого (бакалаврського) рівня вищої освіти**

**Логістика**

**за темою № 6 - Статистичні методи вимірювання взаємозв'язків**

**Харків 2021**

**ЗАТВЕРДЖЕНО**

Науково-методичною радою  
Харківського національного  
університету внутрішніх справ  
Протокол від 23.09.2021 № 8

**СХВАЛЕНО**

Методичною радою  
Кременчуцького льотного  
коледжу  
Протокол від 22.09.2021 № 2

**ПОГОДЖЕНО**

Секцією науково-методичної ради  
ХНУВС з гуманітарних та соціально-  
економічних дисциплін  
Протокол від 22.09.2021 № 8

Розглянуто на засіданні циклової комісії економіки та управління, протокол від 31.08.2021 № 1

**Розробники:** викладач циклової комісії економіки та управління, спеціаліст вищої категорії, викладач-методист Бондарець О.М.

**Рецензенти:**

1. Доктор економічних наук, професор кафедри бізнес адміністрування, маркетингу і туризму Кременчуцького національного університету ім. М. Остроградського – Дружиніна В.В.
2. Кандидат економічних наук, спеціаліст вищої категорії, викладач-методист, викладач циклової комісії управління та адміністрування КЛК ХНУВС - Пушкар О.І.

### План лекції:

1. Зв'язки суспільних явищ як критеріїв статистичного вивчення
2. Загальні методи вивчення зв'язків
3. Основи кореляційно – регресійного аналізу
4. Непараметричні методи виявлення та вимірювання зв'язків

### Рекомендована література:

#### Основна

1. Горкавий В.К. Статистика: підручник. – К.: Алерта, 2020 – 644 с.
2. Карпенко Л. М. Статистика: навчальний посібник. – Одеса: ОРІДУ НАДУ, 2019. – 184 с.
3. Логунова Н. А. Статистика II : підручник. К. : Кондор-Видавництво, 2015. 340 с.
4. Мармоза А. Т. Практикум з теорії статистики : навч. посіб. К. : ЦУЛ, 2013. 484 с.
5. Мармоза А. Т. Теорія статистики : підручник. К. : ЦУЛ, 2013. 592 с.
6. Опря А. Т. Статистика (модульний варіант з програмованою формою контролю знань). К. : ЦУЛ, 2014. 536 с.
7. Теорія статистики : навч. посіб. / М. К. Шапочка, О. М. Маценко. Суми : Університетська книга, 2014. 312 с.

#### Додаткова

1. Бізнес-статистика : навч. посіб. / С. О. Матковський, О.С. Гринькевич, М. Л. Вдовин, О.М. Вільчинська, О. Р. Марець, О. З. Сорочак. Київ : Алерта, 2016. 281 с.
2. Економічна статистика : навч. посіб. / В. М. Соболев, Т. Г. Чала, О. С. Корепанов та ін. ; за ред. В. М. Соболева. Х. : ХНУ імені В. Н. Каразіна, 2017. 388 с.
3. Ковтун Н. В. Теорія статистики : підручник. К. : Знання, 2012. 399 с.
4. Крамченко Л. І. Статистика ринку товарів та послуг : навч. посіб. Вид. 2-ге, переробл. і допов. Львів : Новий світ-2000, 2016. 296 с.
5. Кулинич О. І., Кулинич Р. О. Теорія статистики : підручник. К. : Знання, 2013. 239 с.
6. Моторин Р. М., Чеботовський Е. В. Статистика для економістів : навч. посіб. К. : Знання, 2013. 381 с.
7. Статистика підприємств / С. О. Матковський та ін. Львів : Алерта, 2013. 560 с.
8. Стегній М. І. Статистика : навч. посіб. К. : Кондор, 2012. 306 с.
9. Штагрет А. М. Статистика : навч. посіб. К. : ЦУЛ, 2012. 232 с.

## Інформаційні ресурси в Інтернеті

1. Офіційний портал Верховної Ради України: Законодавство України [Електронний ресурс]. – Режим доступу: <https://zakon.rada.gov.ua>
2. Офіційний сайт Державної служби статистики України [Електронний ресурс]. – Режим доступу: [www.ukrstat.gov.ua](http://www.ukrstat.gov.ua)

## Текст лекції:

**1. Зв'язки суспільних явищ як критеріїв статистичного вивчення**

Одним із найзагальніших законів об'єктивного світу є закон загального зв'язку і залежності між явищами суспільного життя. Усі явища суспільного життя існують не ізольовано, а у нерозривному взаємозв'язку, тобто залежать одне від одного, тому вивчення будь-якого явища буде неповним, якщо не досліджені його зв'язки з іншими явищами і процесами. Статистичне дослідження взаємозв'язків дає можливість виявити не тільки наявність і напрямок зв'язку, але дозволяє кількісно оцінити і виразити його аналітично.

Визначення зв'язків між явищами дає змогу перейти від констатації фактів до пояснення і використання їх на практиці. Так, при вивченні урожайності сільськогосподарських культур можна визначити кількісні характеристики впливу багатьох факторів на урожайність. Це дозволяє виявити резерви зростання урожайності, встановити ступінь залежності їх як від об'єктивних причин, так і від умов діяльності сільськогосподарських підприємств. Визначення взаємозв'язків дозволяє проводити науково обґрунтовані прогнози.

Зв'язки між явищами, окремими їх ознаками досить різноманітні, однак у будь-якому випадку одні ознаки виступають як фактори, що впливають на інші і зумовлюють їх зміну, інші — ж результати дії цих факторів. Одні із них є причиною, інші наслідком.

**Якщо перші прийнято називати** ознаками-факторами, або факторними (причинними) **ознаками, то другі —** результативними (наслідковими) **ознаками.**

Різнманітність зв'язків, в яких перебувають явища, зумовлює необхідність їх класифікації, зведення зв'язків до певних типів, форм за їх істотними рисами і властивостями.

В основу класифікації зв'язків у статистиці покладено відмінність і подібність зв'язків за такими їх особливостями, як ступінь тісноти, спрямованість, аналітичне вираження, одиничність або множинність факторів. Відповідно до цього розрізняють зв'язки *функціональні і кореляційні, прямі і обернені, прямолінійні, криволінійні, однофакторні і багатфакторні.*

За статистичною природою зв'язки поділяють на *функціональні і стохастичні.*

При **функціональному** зв'язку кожному можливому значенню факторної

ознаки  $x$  відповідає чітко визначене значення результативної ознаки —  $y$ , тобто функціональні зв'язки характеризуються повною відповідністю між причиною і наслідком, факторною і результативною ознаками. Така залежність притаманна фізичним, хімічним явищам тощо. У суспільних процесах це найчастіше зв'язок складових елементів розрахункових формул відповідних показників, наприклад, залежність валового збору від урожайності сільськогосподарської культури і розміру посівної площі.

На відміну від функціональних, **стохастичні** зв'язки неоднозначні. Стохастичні зв'язки проявляються як узгодженість варіації двох чи більше ознак. У ланці зв'язку " $x - y$ " кожному значенню ознаки  $x$  відповідає певна множина значень ознаки  $y$ , які утворюють так званий *умовний розподіл*. Стохастичний зв'язок, відбиваючи множинність причин і наслідків, виявляється в зміні умовних розподілів.

Якщо умовні розподіли замінюються одним параметром — середньою  $y_i$  то такий зв'язок називають **кореляційним**. Отже, кореляційний зв'язок є різновидом стохастичного і виявляється в зміні середніх умовних розподілів.

За напрямком дії (спрямованістю) розрізняють зв'язок *прямий і обернений*.

*Прямий* — це такий зв'язок, при якому зі збільшенням або зменшенням значень факторної ознаки відповідно збільшується або зменшується значення результативної ознаки, тобто факторна і результативна ознаки змінюються в одному напрямку.

Прикладом прямого зв'язку може бути зв'язок між фондоозброєністю і продуктивністю праці, між собівартістю продукції і рівнем рентабельності.

*Оберненим* зв'язком називають такий, при якому значення результативної ознаки змінюється в протилежному напрямку відносно зміни значення факторної ознаки. Прикладом такого зв'язку може бути продуктивність праці і собівартість продукції. За формою аналітичного вираження в загальній класифікації виділяють зв'язки *прямолінійні та криволінійні*. Якщо певний зв'язок явищ можна точно або наближено зобразити рівнянням будь-якої прямої лінії, то його називають *лінійним (прямолінійним)* зв'язком, а якщо рівнянням будь-якої кривої лінії (параболи, гіперболи і т.п.) — *нелінійним (криволінійним)*.

Аналітичним рівнянням можна описувати лише функціональні зв'язки. Кореляційні зв'язки описуються рівнянням лише наближено. Однак, навіть наближене описування за допомогою аналітичних рівнянь кореляційних зв'язків суспільних явищ дає можливість отримати результати, цілком придатні для наукових і практичних потреб.

## 2. Загальні методи вивчення зв'язків

Для відповіді на питання про наявність або відсутність кореляційного зв'язку використовують ряд специфічних методів:

1) елементарні прийоми (паралельне порівняння рядів значень факторної і результативної ознак, балансовий метод, графічне зображення,

метод аналітичного групування);

- 2) дисперсійний аналіз;
- 3) кореляційно-регресійний аналіз.

Елементарним способом виявлення зв'язку є паралельне порівняння ряду значень факторної ознаки і відповідних значень результативної ознаки. Значення факторної ознаки розташовують у порядку зростання і потім відслідковують напрям зміни величини результативної ознаки. В тих випадках, коли збільшення величини факторної ознаки призводить до збільшення результативної ознаки, можна стверджувати про можливу наявність прямого кореляційного зв'язку. Якщо ж із збільшенням факторної ознаки, величина результативної ознаки має тенденцію до зменшення, то можна передбачати обернений зв'язок між ознаками.

Найбільш поширеним і простим методом дослідження взаємозв'язків у статистиці є *метод аналітичного групування* і побудова кореляційних таблиць.

**Метод аналітичного групування** полягає в тому, що всі елементи сукупності групують, як правило, за факторною ознакою і в кожній групі обчислюють середні значення результативної ознаки тобто лінія регресії оцінюється лише в окремих точках, які відповідають певному значенню  $x$ .

На **першому етапі аналізу** кореляційного зв'язку при обґрунтуванні моделі постають два питання: вибір факторних ознак і визначення числа груп і меж інтервалів. При визначенні числа груп і меж інтервалів слід зважати на той факт, що типовість та сталість групових середніх залежить від чисельності груп. На практиці аналітичне групування часто виконується за принципом рівних інтервалів, що значно спрощує подальший аналіз зв'язку.

На **другому етапі** проводиться оцінка лінії регресії — у кожній групі за факторною ознакою обчислюють середні значення результативної та факторної ознак. Групові середні обчислюємо за вихідними незгрупованими даними.

У нашому прикладі:

**Третій етап аналітичного групування** вимірювання тісноти зв'язку за допомогою дисперсійного аналізу.

Основною метою дисперсійного аналізу є виявлення впливу окремих факторів чи умов, які визначають варіацію ознаки. В основі дисперсійного аналізу лежить закон розкладання загальної дисперсії на складові, згідно з яким загальна дисперсія

### 3. Основи кореляційно – регресійного аналізу

Важливою характеристикою кореляційного зв'язку є *лінія регресії* — емпірична в моделі аналітичного групування і теоретична в моделі регресійного аналізу.

**Емпірична лінія регресії** представлена груповими середніми результативної ознаки  $\bar{y}_j$ , кожна з яких належить до відповідного інтервалу значень груповального фактора  $x_j$ .

**Теоретична лінія регресії** описується певною функцією  $Y = f(x)$ , яку називають **рівнянням регресії**, а  $Y$  — **теоретичним рівнем результативної ознаки**.

На відміну від емпіричної, теоретична лінія регресії неперервна. Так, вважають, що маса дорослої людини в кілограмах має бути на 100 одиниць менша за її зріст у сантиметрах. Співвідношення між масою і зростом можна записати у вигляді рівняння:  $Y = -100 + x$ , де  $y$  — маса;  $x$  — зріст.

Безперечно, така форма зв'язку між масою та зростом людини надто спрощена. Насправді збільшення маси не жорстко пропорційне до збільшення зросту. Люди одного зросту мають різну масу, проте в середньому зі збільшенням зросту маса зростає. Для точнішого відображення зв'язку між цими ознаками в рівняння слід увести другий параметр, який був би коефіцієнтом пропорційності при  $x$ , тобто  $Y = -100 + bx$ .

Рівняння регресії в такому вигляді описує числове співвідношення варіації ознак  $x$  і  $y$  в середньому. Коефіцієнт пропорційності при цьому відіграє визначальну роль. Він показує, на скільки одиниць у середньому змінюється  $y$  зі зміною  $x$  на одиницю. У разі прямого зв'язку  $b$  — величина додатна, у разі оберненого — від'ємна.

Подаючи  $y$  як функцію  $x$ , тим самим абстрагуються від множинності причин, штучно спрощуючи механізм формування варіації  $y$ . Аналіз причинних комплексів здійснюється за допомогою множинної регресії.

Різні явища по-різному реагують на зміну факторів. Для того щоб відобразити характерні особливості зв'язку конкретних явищ, статистика використовує різні за функціональним видом регресійні рівняння. Якщо зі зміною фактора  $x$  результат  $y$  змінюється більш-менш рівномірно, такий зв'язок описується лінійною функцією  $Y = a + bx$ . Коли йдеться про нерівномірне співвідношення варіацій взаємозв'язаних ознак (наприклад, коли прирости значень  $y$  зі зміною  $x$  прискорені чи сповільнені або напрям зв'язку змінюється), застосовують нелінійні регресії, зокрема:

степеневу  $Y = ax^b$ ;

гіперболічну  $Y = a + \frac{b}{x}$ ;

параболічну  $Y = a + bx + cx^2$  тощо.

Вибір та обґрунтування функціонального виду регресії ґрунтується на теоретичному аналізі суті зв'язку. Нехай вивчається зв'язок між урожайністю та кількістю опадів. Надто мала і надто велика кількість опадів спричиняють зниження врожайності, максимальний її рівень можливий за умови оптимальної кількості опадів, тобто зі збільшенням факторної ознаки (опадів) урожайність спершу зростає, а потім зменшується. Залежність такого роду описується параболою  $Y = a + bx + cx^2$ .

Вивчаючи зв'язок між собівартістю  $y$  та обсягом продукції  $x$ , використовують рівняння гіперболи  $Y = a + \frac{b}{x}$ , де  $a$  — пропорційні витрати на одиницю продукції,  $b$  — постійні витрати на весь випуск.

Зауважимо, що теоретичний аналіз суті зв'язку, хоча й дуже важливий, лише окреслює особливості форми регресії і не може точно визначити її функціонального виду. До того ж у конкретних умовах простору і часу межі варіації взаємозв'язаних ознак  $x$  і  $y$  значно вужчі за теоретично можливі. І якщо кривина регресії невелика, то в межах фактичної варіації ознак зв'язок між ними досить точно описується лінійною функцією. Цим значною мірою пояснюється широке застосування лінійних рівнянь регресії:

$$Y = a + bx.$$

Параметр  $b$  (*коефіцієнт регресії*) — величина іменована, має розмірність результативної ознаки і розглядається як *ефект впливу*  $x$  на  $y$ . Параметр  $a$  — вільний член рівняння регресії, це значення  $y$  при  $x = 0$ . Якщо межі варіації  $x$  не містять нуля, то цей параметр має лише розрахункове значення.

Параметри рівняння регресії визначаються методом найменших квадратів, основна умова якого — мінімізація суми квадратів відхилень емпіричних значень  $y$  від *теоретичних*  $Y$ :

$$\sum (y - Y)^2 = \min.$$

Математично доведено, що значення параметрів  $a$  та  $b$ , при яких мінімізується сума квадратів відхилень, визначаються із системи нормальних рівнянь:

$$\sum y = na + b \sum x,$$

$$\sum xy = a \sum x + b \sum x^2.$$

Розв'язавши цю систему, знаходимо такі значення параметрів:

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \sum x \sum x},$$

$$a = \bar{y} - b\bar{x}.$$

Рівняння регресії відбиває закон зв'язку між  $x$  і  $y$  не для окремих елементів сукупності, а для сукупності в цілому; закон, який абстрагує вплив інших факторів, виходить з принципу «за інших однакових умов».

Вплив інших окрім  $x$  факторів зумовлює відхилення емпіричних значень  $y$  від теоретичних  $Y$  у той чи інший бік. Відхилення  $(y - Y)$  називають *залишками* і позначають символом  $e$ . Залишки, як правило, менші за відхилення від середньої, тобто  $(y - Y) \leq (y - \bar{y})$ .

У невеликих за обсягом сукупностях коефіцієнт регресії схильний до випадкових коливань. Тому слід перевірити його істотність. Коли зв'язок лінійний, істотність коефіцієнта регресії перевіряють за допомогою  $t$ -критерію (Стюдента), статистична характеристика якого для гіпотези  $H_0: b = 0$  визначається відношенням коефіцієнта регресії  $b$  до власної стандартної похибки  $\mu_b$ , тобто  $t = b / \mu_b$ .

Стандартна похибка коефіцієнта регресії залежить від варіації факторної ознаки  $\sigma_x^2$ , залишкової дисперсії  $\sigma_e^2$  і числа ступенів свободи  $k = n - m$ , де  $m$  — кількість параметрів рівняння регресії:

$$\mu_b = \sqrt{\frac{\sigma_e^2}{\sigma_x^2(n - m)}}.$$

Для коефіцієнта регресії, як і для будь-якої іншої випадкової величини, визначаються довірчі межі  $b \pm t_{\alpha} b$ .

Важливою характеристикою регресійної моделі є відносний ефект впливу фактора  $x$  на результат  $y$  — **коефіцієнт еластичності**:

$$\gamma = b \frac{\bar{x}}{\bar{y}}.$$

Він показує, на скільки процентів у середньому змінюється результат  $y$  зі зміною фактора  $x$  на 1%.

#### 4. Непараметричні методи виявлення та вимірювання зв'язків

Поряд із вивченням кореляційної залежності між кількісними показниками статистика встановлює також зв'язки і між якісними ознаками. Для виявлення та вимірювання зв'язків між якісними ознаками використовують непараметричні методи, які називають *ранговими методами кореляції*. Вони простіші в обчисленнях. Найголовніша їх відмінність від параметричних методів полягає у тому, що непараметричні методи не вимагають таких попередніх уявлень про характер розміщення вихідних статистичних даних. Для обчислень використовують не самі значення ознак, а їх *знаки, ранги, частоти*.

Розглянемо чотири основних показники вимірювання зв'язків між ознаками, які є основою рангових методів: коефіцієнт Фехнера; коефіцієнт асоціації; коефіцієнт контингенції; коефіцієнт Спірмена.

**Коефіцієнт Фехнера** (коефіцієнт збігу знаків) визначають як відношення різниці числа знаків лінійних відхилень факторної та результативної ознак, що збігаються, та числа знаків, що не збігаються, до загального числа відхилень ознак від середніх.

Наприклад, якщо лінійні відхилення ознак  $x$  та  $y$  мають такі знаки для різних значень:

$X_i$	$Y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$
$x^1$	$y^1$	+	-
$x^2$	$y^2$	+	+
$x^3$	$y^3$	+	+
$x^4$	$y^4$	-	-
$x^5$	$y^5$	+	-

Тоді коефіцієнт Фехнера дорівнює:

$$f = \frac{3-2}{5} = \frac{1}{5} = 0.2$$

Значення коефіцієнта  $i$  змінюється від -1 до +1, і чим воно ближче до нуля, тим зв'язок між ознаками тісніший. Цей коефіцієнт може бути застосований як до якісних, так і до кількісних ознак.

**Коефіцієнти асоціації та контингенції** застосовують для вимірювання тісноти зв'язків якісних альтернативних ознак. При дослідженні щільності зв'язку між якісними альтернативними ознаками (протилежними за змістом) використовують розрахункову таблицю, яка складається із чотирьох комірок, кожна з яких відповідає відомій альтернативі того чи іншого показника.

	ТАК	НІ	РАЗОМ
Так	a	b	a + b
Ні	c	d	c + d
Разом	a + c	b + d	a + b + c + d

За даними такого макету таблиці можна обчислити коефіцієнт асоціації Д. Юла і коефіцієнт контингенції К. Пірсона.

**Коефіцієнт асоціації (A)** обчислюють за формулою:

$$A = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c}$$

**Коефіцієнт контингенції (K)** обчислюють за формулою:

$$K = \frac{a \cdot d - b \cdot c}{\sqrt{(a+b) \cdot (b+d) \cdot (a+c) \cdot (c+d)}}$$

Зв'язок між ознаками підтверджується, якщо:

$$A \geq 0,5 \text{ або } K \geq 0,3$$