**Тема 5. СТАТИСТИЧНА ОБРОБКА Й ОЦІНКА ДАНИХ**

**1. Узагальнення даних**

У процесі дослідження початкові відомості піддаються статистичній обробці за трьома основними напрямами:

1. узагальнення даних;
2. співставлення даних;
3. визначення важливості результатів.

Сукупності вихідних відомостей (змінні) бувають двох різновидів: дискретні і безперервні (континуальні). Різниця між ними проста. Наприклад, у туристичній групі може бути 5, 6, 7 20 осіб, але не може бути 15,5 осіб. Дані, що складаються з таких змінних, і самі змінні нази­ваються *дискретними*. За своєю природою дискретні змінні відрізняються від змінних іншого типу, які теоретично можуть приймати будь-які числові значення на заданому інтервалі. Наприклад, температура води в морі може дорівнювати + 27,0 °С, + 27,1 °С, + 27,2 °С і т. д. У певному сенсі такі змінні неперервні, і тому їх називають *безперервними*. Будь-яке завдання, пов’язана з точним виміром, викликає появу безперервних змінних, а перерахування явищ породжує дискретні дані. Методи обробки безперервних змінних відрізняються від методів обробки дискретних змінних. Тому підміна одних методів іншими неприпустима. Обробку статистичних даних необхідно починати з їх узагальнення для того, щоб характерні особливості стали легко помітними.

Для згортання вихідних даних широко застосовуються чотири основні методи:

1. угрупування відносно порогових числових значень;
2. використання типових числових значень для характеристики деяких «середніх» умов;
3. вимір відхилення нетипових числових значень з метою показу діапазону розкидування розподілу;
4. використання типових числових значень і відхилень для зобра­ження загальної картини за допомогою однієї характеристики, яка вбирала б у себе як «середні» умови, так і розмір «відхилення» від них.

**Статистичне угруповання**

Метод статистичного угруповання використовується для узагаль­нення будь-якої статистичної сукупності даних. З цією метою сукупність розбивається на статистичні групи (наприклад, 0-9 %, 10-19 %, 20-29 %, 30-39 % і т. д.). Потім в кожній групі підраховується кількість спосте­режень (об’єктів). Зі згрупованими даними подібним чином потрібно поводитися як з дискретними.

Кожна група обмежена зверху і знизу. Найменше значення в групі називається нижньою межею групи, найбільше - верхньою межею. Так, для групи, позначеної цифрами 10-19 %, нижньою межею буде 10 %, а верхньою – 19 %. Іноді зручно виділяти групу без очевидного позначення однієї з меж. Так, наприклад, можуть бути корисні групи «менше 10 %» і «більше 30 %». Групи, які не мають меж з обох сторін, називаються відкритими. Головне полягає в тому, щоб виділені групи відповідали завданню дослідження і мали змістовний сенс. Якщо почати з розгляду загальних розподілів спостережень, то це дозволить виявити якісь «природні» групи. Таку процедуру найзручніше виконати, накресливши діаграму розподілу.

Але ідеальні групи треба якимось чином пов’язати з вимогою про чітко обумовлений розмір інтервалу у групі. Межі кожної групи повинні виникати закономірно зростаючим числовим значенням ознаки (основи групування), так щоб ці числа нагадували арифметичну прогресію (наприклад, 1-2; 3-4; 5-6; 7-8), або геометричну прогресію (наприклад, 2-4; 5-8; 9-16). Зрозуміло, що «природно» виникають групи, які навряд чи будуть дотримуватися таких правил, і тому доведеться шукати компромісне рішення. Найбільш вдалу основу для нього надають групи з однаковим розміром інтервалу.

Обираючи відповідні групи, не слід прагнути до багаточисельності груп. При їх надлишку ми ризикуємо втратити цілісне уявлення у більшості випадків. У результаті не вийде узагальненої картини, заради якої вжиті всі дії з даними.

Кількість груп залежить від кількості спостережень, а також від різниці між крайніми числовими значеннями ознаки, що лежить в основі групування. П. Тойн і П. Ньюбі вважають, що кількість груп не повинна перевищувати більш ніж у 5 разів десятковий логарифм числа спосте­режень. Таким чином, можна уникнути виділення надмірної кількості груп. Наприклад, якщо зроблено 100 первинних спостережень, то максимальне число груп має бути:

5 х lg100 = 5 х 2 = 10.

Інтервальний варіаційний ряд слід відобразити у вигляді таблиці (табл. 1) або графіка – гістограми, полігону розподілу, кумуляти, огіви.

Таблиця 1

Угруповання регіонів України за кількістю туристичних підприємств

у 2007 р.

|  |  |  |  |
| --- | --- | --- | --- |
| Група | Абсолютна частота (кількість регіонів) | Відносна частота (% регіонів від загальної кількості) | Кумулятивна частота (частота з нароста­ючим підсумком) |
| Менше 50 | 6 | 24,0 | 6 |
| 51-150 | 12 | 48,0 | 18 (6+12) |
| 151-250 | 3 | 12,0 | 21 (6+12+3) |
| 251-500 | 3 | 12,0 | 24 (6+12+3+3) |
| Більше ніж 500 | 1 | 4,0 | 25(6+12+3+3+1) |

**Гістограма та полігони розподілу**

Більш наочне уявлення щодо розподілу частот одержимо, якщо замість таблиці побудувати графік, де чисельність груп вказується над інтервалом кожної групи. Горизонтальну вісь графіка прийнято називати абсцисой (або віссю х); ця вісь відображає інтервали груп. Вертикальну вісь називають ординатою (або віссю у), яка відображає чисельність груп (рис. 1).



*Рис. 1. Гістограма розподілу регіонів України за кількістю туристичних підприємств*

Ордината відповідає залежній змінній, абсциса - незалежній. Змінна є «залежною» тоді, коли її значення обумовлене іншою змінною. Під час побудови гістограми залежною змінною є чисельність груп, і її відкладають за ординатою, тобто на осі у. Масштаб за ординатою обирається з урахуванням останніх значень чисельності груп. У нашому прикладі максимальна частота спостережень (12) припадає на інтервал 51-150 підприємств.

Отже, вісь ординат слід розбити на діапазоні від 0 до 12 (слід зазначити, що шкала завжди повинна починатися з 0). Шкала на осі абсцис залежить від числа виділених груп. Виразність графіка повинна стати критерієм для вибору його розміру і масштабу шкали.

Після визначення розміру шкали, зібрані дані компонують у вигляді прямокутників, висоти яких показують частоту спостережень у кожній групі (тобто чисельності груп).

Гістограма використовується у роботі з дискретними змінними, тобто згруповані дані. Під час обробки непереривних величин чи при дробному розбитті на групи з дуже малим інтервалом на абсцисі гістограма поступається місцем полігону розподілу.

Спосіб креслення полігонів розподілу схожий зі способом, який використовують для побудови гістограм. На осі відзначаються кількість груп, а на осі х – інтервали, числові значення, ознаки, що вивчається, причому кожен інтервал представлено своєю середньою точкою чи міткою, наприклад, для інтервалу 16-18 середньою є 17. Отже, графік зображує кількість груп, які відмічені над середніми точками інтервалу. Під середніми точками кожного інтервалу на графіку позначаються точки, що розташовані так, наче це їх проекція на шкалу ординати зображала чисельність групи, згідно з інтервалом. Після чого всі відмічені точки з’єднуються, у результаті чого отримуємо полігон розподілу. Якщо округ­лити стики, що містять його відрізки, тобто «згладити» цей полігон, то ми матимемо лінію графіка - розподіл частот. Однак згладження ламаної лінії може бути виконано лише завдяки точним статистичним методам. В іншому випадку бажано обмежитися побудовою полігону розподілу.

На основі даних таблиці 2 побудуємо полігон розподілу кількості туристів у групах (рис. 2).

Таблиця 2

Розподіл туристів за групами

|  |  |  |
| --- | --- | --- |
| Кількість туристів | Середня кількість | Число туристичних груп |
| Абсолютна частота | Кумулятивна частота |
| 13-15 | 14 | 5 | 5 |
| 16-18 | 17 | 7 | 12 |
| 19-21 | 20 | 10 | 22 |
| 22-24 | 23 | 8 | 30 |
| 25-27 | 26 | 6 | 36 |

Як ми бачимо на рис. 2, на відміну від ступеневого відображення згрупованих даних на гістограмі, лінія графіка неперервна.



Фактично, це кумулятивні цифри, тобто цифри, які можливо вичислити лише за умови наростаючого підсумку. Тому, табульований розподіл можна назвати *розподілом кумулятивних частот* (табл. 2). Зрозуміло, що у цьому випадку графік є полігоном кумулятивних частот, у цьому разі графік називають огівою. Для огіви абсолютні частоти можна замінювати на відносні.



*Рис. 3. Огіва. Розподіл туристів за групами (кумулятивна частота)*

Гістограми та полігони розподілу – це найпростіший спосіб відображення перемінного характеру вихідних даних.

Використовуючи згрупування даних при побудові розподілу і графі­ків, отримаємо у доступній формі загальне уявлення про мінливість даних у вихідній сукупності.

**2. Змінення типового**

При розгляді будь-якої сукупності даних можна:

1) обрати числові значення, які частіше зустрічаються;

2) підрахувати певного виду середню для того, щоб зобразити «типову» цифру даної сукупності.

**1. Мода**

Термін «мода» походить від французького виразу «а 1а modе», що значить модний. Саме «найбільш модне» інколи перетворюється на те, що «найбільш часто зустрічається», тож не важко зрозуміти, чому саме використовують це слово. У теорії статистики модою є таке значення, що варіюється та зустрічається найчастіше. Мода зображає найбільш типовий рівень, який характерний для даної сукупності.

Залежно від того, чи зображена ознака, що варіюється у вигляді дискретного ряду чи інтервального, поняття моди виконується по різному. Для дискретного ряду встановлення моди виконується без підрахування шляхів легкого перегляду стовпця частот. У цьому стовпчику, треба знайти найбільше число зі значенням у ньому, яке буде зображати найбільшу частоту порівняно з іншими частотами у стовпчику. Певне значення ознаки, що відповідає їй є мода. Так, наприклад, розглянемо розподіл сімей у місті N (табл. 3).

Таблиця 3

Розміщення сімей у місті N, %

|  |  |
| --- | --- |
| Величина сім’ї, осіб | % |
| 2 | 34,2 |
| 3 | 28,0 |
| 4 | 25,2 |
| 5 | 8,3 |
| 6 | 2,8 |
| 7 | 0,9 |
| 8 | 0,3 |
| 9 | 0,1 |
| 10 і більше | 0,1 |
| Середня величина сім’ї | 3,23 |

З таблиці видно, що найбільша частота – 34,2, якій відповідає певне значення варіюючої ознаки, а саме 2. Воно і є модою для відповідного ряду. Це означає, що в місті N найчастіше зустрічаються родини, щоскладаються з двох чоловік. Таким чином, у дискретному ряду мода визначається як варіант, що має максимальну частоту.

У разі інтервального ряду перегляд стовпця частот виявляє не моду, а лише інтервал, в межах якого знаходиться мода. Як приклад розглянемо розподіл кількості народжених дітей за віком матері на Україні (табл. 4).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Вік | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| Всього | 408 589 | 427 259 | 426 086 | 460 368 | 472 657 | 510 589 |
| до 15 | 138 | 136 | 99 | 149 | 143 | 172 |
| 15-19 | 54 290 | 53 874 | 50 424 | 49 672 | 48 425 | 48 392 |
| 20-24 | 163 197 | 169 505 | 163 792 | 172 622 | 174 577 | 183 069 |
| 25-29 | 113 417 | 119 641 | 122 045 | 135 711 | 139 976 | 154 457 |
| 30-34 | 54 269 | 58 746 | 63 015 | 71 739 | 76 252 | 85 774 |
| 35-39 | 18 051 | 19 696 | 21 215 | 24 774 | 27 174 | 32 086 |
| 40-44 | 3 924 | 4 148 | 4 260 | 4 433 | 4 871 | 5 394 |
| 45-49 | 162 | 194 | 196 | 210 | 213 | 283 |
| 50-54 | 3 | 3 | 4 | 7 | 14 | 8 |
| старші за 55 | 2 | - | - | - | 3 | 4 |
| невідомо | 1 136 | 1 316 | 1 036 | 1 051 | 1 009 | 950 |

Таблиця 4.

Динаміка кількості народжених дітей за віком матері на Україні (ос.)

УРозглядаючи стовпчик за кожен рік, ми побачимо, що кількість народжених дітей залежно від віку матері на Україні змінюється від групи до групи і у певному інтервалі досягають максимального. Цей інтервал є модальним (числа частот модального інтервалу виділені). Варто виділити, що модальний інтервал за цей період не змінився. Це свідчить про те, що середній вік матері в 2003-2008 рр. був 20-24 роки, хоча спостерігаються зміни в темпах зростання частот.

У межах модального інтервалу необхідно визначити значення варійованої ознаки, яка має максимальну щільність розподілу. З геометричної точки зору, моду також можна розглядати як точку, на яку опускається перпендикуляр з вершини кривої розподілі на ось абсциси. Ця точка максимальної концентрації, максимального накопичення випадків.

Найбільш спрощений спосіб переходу від модального інтервалу до моди, тобто від лінії на осі абсциси до точки на цій же осі, полягає у тому, щоб зобразити модою середину модального інтервалу. Але подібне рішен­ня буде правильним у випадку:

1) загальної симетричності розподілу;

2) якщо інтервали, суміжні з модальним, майже не відрізняються один від одного за кількістю випадків чи за питомою вагою.

Коли є розподіл іншого типу, то середина модального інтервалу вже не може розглядатися як мода. Справжня мода буде тією, що розташована близько до середини інтервалу, праворуч чи ліворуч від неї, залежно від характеру розподілу. Цей характер розподілу виявляють частоти, суміжні з частотою модального інтервалу. Частота, яка передує модальній, назива­ється *частотою передмодального інтервалу* (d1), а частота, наступна за модальною, – *частотою післямодального інтервалу* (d2). Місцезнаходження моди всередині модального інтервалу цілком визначається тим, як саме частоти передмодального інтервалу та після модального інтервалу відрізняються від частоти модального інтервалу.

Відношення однієї до будь-якої з двох зазначених різниць до їх суми показує те місце, яке всередині модального інтервалу має займати мода. Так, розрахуємо відношення першої різниці до суми двох інших:

$$\frac{d\_{1}}{d\_{1}+d\_{2}}$$

За симетрією d1 = d2 зазначений вираз дорівнює $\frac{1}{2}$ Це свідчить про те, що для отримання моди потрібно взяти половину модального інтервалу. Якщо асиметрія вже існує, то відношення $\frac{d\_{1}}{d\_{1}+d\_{2}}$ вловлює той нахил, який він викликає.

Припустимо, що d1 і d2 співвідносяться між собою як 1:2. Тоді дріб $\frac{d\_{1}}{d\_{1}+d\_{2}}$ буде дорівнювати вже не 1/2, а 1/3, тобто модальна точка буде розташована на відстані однієї третьої від початку інтервалу, а не всередині інтервалу. Проте визначення долі інтервалу ще не визначає моду, бо необхідно:

1. Від оперування долею інтервалу перейти до оперування в тих одиницях, в яких подано всі розподіли, тобто це відношення потрібно помножити на величину інтервалу *(і):*

$\frac{d\_{1}}{d\_{1}+d\_{2}}∙$*i*

2. Отриманий вираз визначає тільки відстань від початку модаль­ного інтервалу. Тому, щоб знайти моду, потрібно до цього виразу додати ще значення нижньої межі модального інтервалу *(l),* чи верхньої межі передмодального, якщо ознака є перервною. У результаті формула визна­чення моди буде виглядати так:

Мо = *l* +$\frac{d\_{1}}{d\_{1}+d\_{2}}∙$*i*

Мода використовується в різних галузях економіки. Обчислення модальної собівартості, модального навантаження на готелі, підприємства громадського харчування і т. д. дозволяють економісту робити висновки щодо переважаючого рівня у даний момент. Ця характеристика повинна бути використана для того, щоб виявити резерви економіки.

Обчислення моди можливе лише тоді, коли за гістограмою ховається крива розподілу з очевидно вираженою вершиною. Якщо такої вершини немає, то і обчислення моди неможливе. Крім того, моду можна обчис­лювати тільки стосовно рівних інтервалів.

**2. Середні величини**

Будь-яке явище, що вивчається, має як спільні для всієї сукупності, так і особливі, індивідуальні властивості. Обчислення середньої є одним з найбільш поширених заходів узагальнення. *Середня –* зведена, що уза­гальнює величину, віддзеркалюючи певний розмір варіювальної ознаки, характерної для деяких якісних однорідних сукупностей у цілому чи для її окремої частини. Статистичні середні розраховуються за допомогою бази масових даних статистично організованого спостереження (суцільне або вибіркове).

Середню величину обчислюють у випадку необхідності виключення впливу випадковості окремих факторів, коли є коливання у значенні ознаки від одного випадку до іншого . Говорять про середній прибуток, середній виробіток, середнє відвідування музеїв екскурсантами і т. п. У таких випад­ках мається на увазі встановлення якого-небудь значення ознаки, в якій буде відображатися певне відношення ознаки сукупності як цілого.

Для отримання повного і всебічного уявлення про сукупність рядів значущих ознак у цілому, варто привернути систему середніх величин, які можуть описати явища з різного боку.

Так, зміни обсягу платежів туристичними підприємствами до бюджету характеризують показники середнього обороту на одне підприємство, середнього розміру доходу на одне підприємство, середньої кількості туристів, яким надало послуги підприємство, та ін. Вивчення цих показників виявить більш чітку загальну тенденцію . Середні величини між собою розрізняються за видом. Виділяють просту середню арифметичну, середню арифметичну дискретного ряду розподілу (зважену) і середню арифметичну інтервального ряду розподілу.

*Проста середня арифметична* обчислюється шляхом ділення суми окремих величин, що характеризують значення конкретної ознаки, на число випадків. Так, наприклад, якщо є дані про кількість екскурсантів, які відвідували музей у робочі дні тижня, то для того, щоб найти середню кількість відвідувань на день, потрібно скласти кількість відвідувань за днями тижня і отриману суму розділити на кількість робочих днів – 6.

Можна говорити, що проста середня арифметична – це результат ділення обсягу явища на чисельність сукупності. У даному прикладі обсяг явища, тобто сума значень варіювальної ознаки – це загальна кількість відвідувань; чисельність сукупності – число робочих днів у тижні. Розді­ливши перше на друге, отримуємо середню кількість відвідувань екскур­сантами музею на день. Припустимо, кількість відвідувань за днями тижня склали: 1 120, 1 150 1 300, 1 360, 1 400 1 350; загальна кількість відвіду­вань склало 7 680, а середня кількість відвідувань на день дорівнюватиме 1 280 (7 680 відвідувань: 6). Ця величина є простою середньою арифме­тичною для даної сукупності.

З урахуванням обчислення середніх величин стоїть поняття, що визна­чає показник, тобто величина, яка визначає вид середньої. Якщо визна­чальний показник дорівнює сумі значень індивідуальних величин, то його середня величина буде дорівнює середній арифметичній, яка обчислюється за допомогою простого рівняння першого ступеня з одним невідомим.

Для зазначених вище цифр відвідин музею за 6 робочих днів визна­чальним показником буде загальна кількість відвідувань. Відшукування середньої арифметичної, що позначається символом *х*, зводиться до того, щоб знайти таку величину, яка при множенні на число робочих днів дала б загальну кількість відвідувань екскурсантами музею:

1120 + 1150 + 1300 + 1360 + 1400 + 1350 = $\overbar{х}+\overbar{х}$ +$\overbar{х}+\overbar{х}$ +$\overbar{х}+\overbar{х}$

7680 = $\overbar{6X}$

*X* = 1280;

у загальному вигляді:

$$\sum\_{i=1}^{i=n}x\_{i}=\overbar{х} n$$

$$\overbar{x}=\frac{\sum\_{i=1}^{i=n}x\_{i}}{n}$$

Це і є формула простої середньої арифметичної. З точки зору арифметики, в основі її лежать дві операції: додавання окремих величин

$$\left(\sum\_{i=1}^{i=n}x\_{n}\right)$$

і ділення цієї суми на число випадків *(п).*

Варіанти варіювальної ознаки мають різні частоти. Деякі варіанти зустрічаються частіше, тобто їм відповідає більше число випадків. Інші варіанти зустрічаються рідше. Обчислення середньої арифметичної такого ряду розподілу відрізняється від обчислення простої середньої арифметичної.

При обчисленні простої середньої арифметичної мається на увазі, що кожен варіант зустрічається лише один раз. Таким чином, обчислення простої середньої арифметичної є обчисленням середньої з такого ряду розподілу, в якому частоти всіх варіантів рівні одиниці.

Якщо варіант зустрічається з різною частотою, то множник для варіанта буде вже не одиниця, а відповідна частота, і тому чисельник дробу визначиться як сума добутків варіантів на частоти. Ця сума добутків може бути отримана також і шляхом простого підсумовування.

Уявімо собі, що потрібно визначити середню кількість екскурсантів в одній групі, що відвідали музей за 1 день. Припустимо, музей відвідало 10 груп, серед яких було 3 групи по 20 осіб, 4 - по 25 осіб, 3 - по 30 осіб. Щоб визначити середню величину, необхідно визначити обсяг явища. В даному випадку обсягом явища буде загальна кількість екскурсантів, які відвідали музей. Цю величину можна отримати двома способами:

1. шляхом простого підсумовування числа екскурсантів у всіх екскурсіях (осіб):

20+20 + 20 + 25 + 25 + 25 + 25 + 30 + 30 + 30 = 250;

2)або шляхом множення варіантів на частоту:

3 х 20 + 4 х 25 + 3 х 30 = 250.

Середню арифметичну отримаємо після розподілу обсягу явища на число випадків:

250 ос.

10 (3+4+3) груп

= 25 ос.

Це і буде *середньою арифметичною дискретного ряду розподілу*.

У загальному вигляді обчислення середньої арифметичної дискрет­ного ряду розподілу може бути представлене в такий спосіб у таблиці 5.

Як видно, при обчисленні середньої немає принципової різниці, чи мають справу з окремими спостереженнями, чи для ряду розподілу. У першому випадку частота дорівнює одиниці і тому множення на неї не проводиться, у другому випадку частоти не дорівнюють одиниці і тому, для того, щоб отримати обсяг явища необхідно зробити множення кожно­го варіанта на число випадків, в яких цей варіант зустрічається.

Таблиця 5

Обчислення середньої арифметичної
дискретного ряду розподілу

|  |  |  |
| --- | --- | --- |
| Варіанти *х* | Частоти */* | Добуток варіантів на частоти *х/* |
| *Х1* | $$f\_{1}$$ | *Х1*$f\_{1}$ |
| *Х2* |  | *Х2*$f2$ |
| *Хз* | $$f\_{2}$$$$f\_{3}$$ | *Хз/*$f3$ |
| $$x\_{n}$$ | $$f\_{n}$$ | $$X\_{n}f\_{n}$$ |
|  | $$\sum\_{}^{}f\_{i}$$ | $$\sum\_{}^{}x\_{i}f\_{i}$$ |

Середня арифметична розраховується як частка від ділення обсягу явища $\sum\_{}^{}x\_{i}f\_{i} $на чисельність сукупності$ \sum\_{}^{}f\_{i} $ Отже, формула середньої арифметичної дискретного ряду розподілу виразиться в такому вигляді:

$$\overbar{x}=\frac{\sum\_{i=1}^{i=n}x\_{i}f\_{i}}{\sum\_{i=1}^{i=n}f\_{i}}$$

і в більш простому записі:

$$\overbar{x}=\frac{\sum\_{}^{}xf}{\sum\_{}^{}f}$$

Отже, щоб обчислити середню арифметичну для дискретного ряду розподілу, треба виконати такі дії:

1. знайти добуток варіантів на частоти в кожному рядку (знайти *x1; f1; x2; f2* і т. д.);
2. підрахувати суму цих творів ($\sum\_{}^{}xf$*);*
3. визначити загальне число випадків ($\sum\_{}^{}f$*);*);
4. обчислити частку від ділення суми добутків варіантів і частот на суму частот $\left(\frac{\sum\_{}^{}x\_{f}}{\sum\_{}^{}f}\right)$.

Отримана величина буде середньою арифметичною.

Середня з дискретного ряду розподілу так само точна, як середня з окремих спостережень. Це правильно лише для дискретних рядів, які не мають невизначених груп, тобто груп з невизначеною межею (наприклад, 100 і більше). Якщо дискретний ряд має хоча б одну таку групу, то для правильного обчислення середньої потрібен підрахунок обсягу явища для даної групи за окремими одиницями, що входять до неї.

У статистичних публікаціях щодо відкритої групи повинно вказу­ватися не тільки число випадків, але і підсумкова величина ознаки - обсяг явища в даній групі за безпосередніми даними про окремі одиниці сукупності.

Так, наприклад, у статистичних публікаціях про розподіл кількості осіб у родинах використовуються відкриті групи, наприклад, «10 і більше». Ця група об’єднує всі багатодітні родини. Але в тих публікаціях наво­дяться дані про загальну кількість осіб, які потрапили в цю відкриту групу. Завдяки додатковій інформації звичайний ряд розподілу надає можливість визначити правильну середню арифметичну для дискретного ряду, навіть якщо в ньому є відкрита група. Додаючи вказану додаткову кількість членів сімей в останній групі до суми добутків варіантів на частоти в інших групах, отримаємо загальну кількість членів сімей, тобто обсяг явища, необхідний для обчислення середнього розміру родини.

Якщо варіанти ряду розподілу виражені в дискретній формі, то обчислення середньої цього ряду не має будь-яких особливих труднощів. При цьому ранжована ознака приймає певні значення і тому добуток *х* на відповідну частоту */* дає обсяг явища для даної групи, а підсумовування творів приводить до обсягу явища для всієї сукупності.

У деяких випадках ряд розподілу виражається в інтервальній формі, тобто у стовпці варіантів є не одне значення *х*, а два значення, що показують нижню і верхню межі інтервалу. Тоді, щоб отримати обсяг явища в даній групі як множник зазвичай беруть середину між верхньою і нижньою межею кожного інтервалу, тобто центр інтервалу. При цьому вважають, що випадки, що потрапили в кожну групу, більш-менш рівномірно розподіляються в межах кожного інтервалу і тому середина буде найкраще відображати цю групу. Центр інтервалу знаходиться шляхом розрахунку простої середньої арифметичної з нижніми межами даного інтервалу і суміжними з ним. Число випадків множиться на цю величину і знаходять обсяг явища за даною групою. Розглянемо розрахунок середньої арифметичної інтервального ряду на наступному прикладі (табл. 6).

Таблиця 6

Обчислення середнього віку туристів, які відпочивають у пансіонаті N

|  |  |  |  |
| --- | --- | --- | --- |
| Інтервали за віком | Вік у вигляді середини інтервалу | Число туристів у зазначеному віці | Об’єм явищ |
| 20-24 | 22,5 | 10 | 225 |
| 25-29 | 27,5 | 15 | 412,5 |
| 30-34 | 32,5 | 18 | 585,0 |
| 35-39 | 37,5 | 20 | 750,0 |
| 40-44 | 42,5 | 22 | 850,0 |
| 45-49 | 47,5 | 25 | 1187,5 |
| 50-54 | 52,5 | 30 | 1575,0 |
| 55-59 | 57,5 | 28 | 1610,0 |
| 60-64 | 62,5 | 26 | 1625,0 |
| 65-69 | 67,5 | 20 | 1350,0 |
| 70-74 | 72,5 | 10 | 725,0 |
| 75-79 | 77,5 | 5 | 387,5 |
| Разом |  | 229 | 11282,5 |

 (у серпні 2016 р.)

$\overbar{x}=\frac{\sum\_{}^{}xf}{\sum\_{}^{}f}$=$\frac{11282,5}{229}$=49,3 року

Таким чином, середній вік туристів, які відпочивали в пансіонаті N в серпні 2016 р. склав 49,3 року. Але слід мати на увазі, що обчислена величина не є абсолютно правильною середньою. Якби обчислювали середню на основі даних про кожного з 229 туристів, то отримана проста середня арифметична була б точніше, ніж середня ряду розподілу. Однак практична розбіжність буває майже незначною для висновків. Ступінь роз­біжності між зазначеними середніми залежить від низки причин, а саме:

* *число випадків*. Чим більше число випадків, тим більша ймовірність того, що середина інтервалу буде мало відрізнятися від групової середньої. І навпаки, якщо на кожну групу припадає кілька одиниць, то можна бути впевненим, що вони будуть групуватися навколо середини інтервалу. Можливо, що всі вони виявляться поблизу верхньої межі або, навпаки, поблизу нижньої межі. У таких випадках середина інтервалу буде «поганою» середньою для групи. Якщо ж спостережень багато, то немає підстав вважати, що більшість їх будуть групуватися тільки навколо верхньої або тільки навколо нижньої межі. Швидше за все, вони будуть більш-менш рівномірно розподілятися в межах інтервалу;
* *величина інтервалу*. Якщо інтервал невеликий, тобто якщо верхня межа не дуже далеко відстає від нижньої, то і помилка, пов’язана з прий­няттям середини інтервалу в основу середньої величини, буде незначна. Фактична групова середня буде мало відставати від середини інтервалу через вузькі межі коливань окремих величин у межах даного інтервалу. У результаті, чим більше число груп, тим більш детально описано розподіл, тим менша буде розбіжність між серединою інтервалу і серед­ньою групою;
* *характер розподілу.* Чим більш симетричний розподіл, тим менша помилка, що виникає при користуванні центрами інтервалів;
* *принцип побудови інтервального ряду*. При рівних інтервалах центр його буде ближче знаходитися до середньої арифметичної за даною групою. Якщо ж інтервали нерівні і за розміром різко відрізняються один від одного, то розбіжність між центром інтервалу і відповідної середньої арифметичної можливо буде значним.

Застосування зазначених способів знаходження середньої арифме­тичної інтервального ряду доречно застосовувати тоді, коли відсутні дані про підсумковий підрахунок щодо окремих випадків.

Правильною середньою арифметичною буде середня, обчислена за даними окремих одиниць сукупності. Це пов’язано з тим, що при цьому немає умовностей, пов’язаних зі встановленням середини інтервалу і, особливо з встановленням середини для відкритих інтервалів. Якщо ж відсутні окремі дані за одиницями спостереження, підсумкові дані за цими одиницями («обсяг явища»), то тоді єдиним методом віднайдення середньої арифметичної є метод, заснований на застосуванні середини інтервалів.

**3. Варіаційний розмах**

Бувають випадки, що необхідно вивчити ряд розподілу не окремо від середньої арифметичної, а разом з нею, виходячи з неї. Окремі значення досліджуваної ознаки можуть бути близькі до середньої. У цьому випадку середня буде добре відображати всю сукупність. І, навпаки, в інших випадках окремі значення далеко віддалені від середньої. Тоді середня не буде становити всю сукупність.

Візьмемо, наприклад, середній рівень споживання непродовольчих товарів. Він може бути обчислений як середня арифметична з рівня споживання громадян будь-якої країни. Але значення середньої величини для країн, в яких немає різких відмінностей у рівні споживання непродовольчих товарів, буде набагато більшим, ніж для країн, в яких спостерігаються різкі відмінності.

Тому обмежуватися обчисленням тільки однієї середньої величини не можливо. Необхідні й інші показники, які характеризували б відхилення окремих значень від загальної середньої.

Не можна задовольнятися однією середньою арифметичною. Треба вимірювати характер всього ряду розподілу. Бувають навіть випадки, коли в центрі уваги знаходиться лише характер розподілу навколо середньої, а не середня. Тому необхідний показник, що визначає ступінь зосередженості окремих значень близько центру всіх значень. Для цього використовуються *показники варіації*, серед яких найпростішим є варіаційний розмах.

*Варіаційний розмах* (або амплітуда коливань) – це різниця між максимальним і мінімальним значеннями певної кількісної ознаки в деякій сукупності. Наприклад, розглянемо таку ознаку, як вік туристів у групі. Наймолодшому туристу 15 років, найстаршому – 55 років. Різниця становить 40 років.

За величиною варіаційного розмаху можна визначити, наприклад, відмінності в роботі підприємств. Так, наприклад, якщо заповнюваність одного готелю складає 90 %, а іншого 40 %, то, варіаційний розмах, тобто різниця, дорівнює 50 %.

Якщо в одному кафе міста можуть обслужити 300 осіб за зміну, а в іншому – 1 500, то варіаційний розмах складе 1 200 осіб за зміну.

Однак цей показник має недоліки:

1. він повністю залежить від окремих випадків, які опинилися на двох полюсах рангового ряду. І в цьому його істотний недолік;
2. орієнтація на окремі випадки може дати хибне уявлення про характер їх відмінностей.

Тому краще використовувати інший показник, який спирався б не тільки на крайні значення, а на все значення розглянутої ознаки в даній сукупності. Таким показником є середньоквадратичне відхилення.

**4. Відхилення від середнього арифметичного**

Для того щоб отримати уявлення про всі відхилення в сукупності даних, можна розрахувати кілька показників:

1. повне відхилення сукупності даних;
2. дисперсія;
3. середнє квадратичне відхилення.

З них найпростішим є перший. Розглянемо розрахунки відхилення сукупності даних ще на прикладі розподілу туристичних фірм за регіо­нами України.

Спочатку обчислюється середнє арифметичне: $\overbar{x}$=$\frac{\sum\_{}^{}x}{n}=153,32$; потім – відхилення кожного спостереження від середнього арифметичного. Всі індивідуальні відхилення додаються (табл. 7, стовпець 3). Таким чином, повне відхилення - це сума різниць кожного індивідуального спостереження та середнього арифметичного сукупності даних, причому підсумовування виконується без урахування знака різниць: $\sum\_{}^{}\left|x-\overbar{x}\right|$

Таблиця 7

Обчислення дисперсії і середнього квадратичного відхилення для розподілу туристичних фірм за регіонами України

|  |  |  |  |
| --- | --- | --- | --- |
| Регіони | х | *(х — х)* | *(х — х)2* |
| АР Крим | 465 | 311,68 | 97 144,42 |
| Вінницька | 50 | -103,32 | 10 675,02 |
| Волинська | 48 | -105,32 | 11 092,30 |
| Дніпропетровська | 231 | 77,68 | 6 034,18 |
| Донецька | 257 | 103,68 | 10 749,54 |
| Житомирська | 37 | -116,32 | 13 530,34 |
| Закарпатська | 49 | -104,32 | 10 882,66 |
| Запорізька | 125 | -28,32 | 802,02 |
| Івано-Франківська | 63 | -90,32 | 8 157,70 |
| Київська | 1102 | 948,68 | 899 993,74 |
| Кіровоградська | 28 | -125,32 | 15 705,10 |
| Луганська | 96 | -57,32 | 3 285,58 |
| Львівська | 156 | 2,68 | 7,18 |
| Миколаївська | 59 | -94,32 | 8 896,26 |
| Одеська | 191 | 37,68 | 1 419,78 |
| Полтавська | 95 | -58,32 | 3 401,22 |
| Рівненська | 60 | -93,32 | 8 708,62 |
| Сумська | 59 | -94,32 | 8 896,26 |
| Тернопільська | 53 | -100,32 | 10 064,10 |
| Харківська | 302 | 148,68 | 22 105,74 |
| Херсонська | 55 | -98,32 | 9 666,82 |
| Хмельницька | 64 | -89,32 | 7 978,06 |
| Черкаська | 70 | -83,32 | 6 942,22 |
| Чернігівська | 46 | -107,32 | 11 517,58 |
| Чернівецька | 72 | -81,32 | 6 612,94 |
| Всього | 3 833 |  | 11 94 269,44 |

Ігнорування знака зручно, але математично некоректно. Математично коректним способом позбавлення від різноманітності в знаках індивідуальних відхилень є спосіб стандартизації відхилень від середньої, який полягає в тому, щоб звести в квадрат індивідуальні відхилення.

Сума квадратів різниць буде повним відхиленням, або в математич­ному записі: $\sum\_{}^{}\left(x-\overbar{x}\right)^{2}$. Якщо розділити цю суму на число спостережень у сукупності (тобто на п), ми обчислимо «середнє повне відхилення від середнього арифметичного» або, іншими словами, – дисперсію ($σ^{2}$):

$$σ^{2}=\frac{\sum\_{}^{}\left(x-\overbar{x}\right)^{2}}{n}$$

Для обчислення дисперсії необхідні індивідуальні відхилення (табл. 5.7, стовпець 3), звести їх у квадрат (табл. 5.7, стовпець 4), підсуму­вати квадрати і використати отриману суму в наведеній вище формулі. Для даних про туристичні фірми дисперсія (табл. 5.7) дорівнює:

$$\frac{\sum\_{\left(x-\overbar{x}\right)}^{}2}{n}=\frac{1194269,44}{25}=47770,78$$

Отримуючи з дисперсії квадратний корінь, ми маємо «стандартне відхилення», тобто середньоквадратичне відхилення від арифметичного середнього, або середньоквадратичне відхилення ($σ$2):

$$σ^{2}=\frac{\sum\_{\left(x-\overbar{x}\right)}^{}2}{n}$$

Для нашого прикладу

$σ=\sqrt{\frac{\sum\_{}^{}\left(x-\overbar{x}\right)^{2}}{n}}$=$\sqrt{47770,78=218,57}$

Так як дисперсія є квадрат середнього квадратичного відхилення, що позначається як о, то для дисперсії використовують символ (о2). Для знаходження о і при більш широкій сукупності даних, ніж у нашому прикладі обсяг обчислень набагато зріс би. Розрахунки за визначенням індивідуальних відхилень, зведення їх у квадрат і потім підсумовування квадратів дуже трудомісткі. Шляхом тотожних перетворень можна спростити розрахункові формули і привести їх до виду:

$$σ=\sqrt{\frac{\sum\_{}^{}x^{2}}{n}}-\overbar{x}^{2}; $$

$σ^{2}$=$\frac{\sum\_{}^{}x^{2}}{n}-\overbar{x}^{2}$

що зменшує обсяг обчислень.

Якщо вихідна інформація може надаватися тільки у вигляді згрупо­ваних (дискретних) даних, наведені вище методи обчислення $σ і σ^{2}$ засто­совувати не можна і формули потрібно піддати додатковій обробці. Повний виклад цих модифікацій можна знайти в інших посібниках.

Отже, загальна картина розподілу сукупності даних складається з характеристики центру розподілу і характеристики відхилень. Об’єднаний захід отримаємо, скориставшись ставленням середнього квадратичного відхилення до середньої арифметичної. У цьому випадку говорять про коефіцієнт варіації (*V). У* нашому прикладі він дорівнює 142,56 %.

$$V=\frac{σ}{x}∙100=\frac{218, 57}{153,32}∙100=142.56\%$$

Зрозуміло, що при зменшенні середнього квадратичного відхилення зменшується і значення *V.* І навпаки, зростання середнього квадра­тичного відхилення приводить до зростання *V.* Так як середнє відхилення є мірою середньої величини відхилення від середнього арифметичного, отже, якщо *V* велике, то це означає, що значне число спостережень істотно відхиляється від середньої (тобто діапазон даних високий). І, навпаки, якщо середнє відхилення мале за відношення до середнього арифметичного, то розкид величин не значний. Таким чином, чим менше *V,* тим більше цифри в сукупності прагнуть до середньої величини. Сукупність даних вважається однорідною, якщо коефіцієнт варіації не перевищує 33 %.

**5. Функціональна і кореляційна залежності**

Між явищами може існувати залежність, яка буває двох видів:

1) функціональна і 2) кореляційна.

При *функціональній залежності* кожному значенню однієї величини відповідає одне або кілька певних значень іншої величини. У такому вигляді такі залежності поширені в математиці і фізиці.

Деякі відкриття функціональних залежностей використовується мандрівниками. Наприклад, для тих, хто піднімається в гору, необхідно знати, що температура повітря зменшується на кожні 100 м на 0,6°, а тиск залежить від температури повітря і висоти місцевості над рівнем моря. У нижніх шарах атмосфери воно зменшується приблизно на 1 мм на кожні 10,5 м висоти.

Завдяки відкриттю функціональних залежностей почали широко використовуватися вимірювальні прилади. Так, для вимірювання зміни тиску в просторі використовують баричний градієнт, який вказує різницю в тиску на одиницю відстані. Одиниця відстані дорівнює довжині 1° мери­діана, тобто 111 км. Атмосферний тиск вимірюють за допомогою баромет­рів, широке поширення з яких отримав барометр-анероїд. Він складається з металевої коробочки, звідки викачане повітря. При збільшенні тиску повітря дно коробочки вдавлюється, а при зменшенні - згинається. Зміни, що відбуваються, передаються на стрілку, яка переміщається по шкалі, градуйованою в мб або мм (1 мм тиску становить 1,33 мб).

Вимірювання температури повітря за допомогою термометрів також здійснюється за допомогою функціональної залежності, тобто залежності між температурою повітря і висотою ртутного стовпа. Фізики, обчис­ливши коефіцієнт лінійного розширення ртуті, почали застосовувати термометри: при кожному підвищенні температури на 1 градус ртутний стовпчик підвищується, наприклад, на 2 мм за умови однакової висоти над рівнем моря. Залежно від пристрою термометра це співвідношення може бути й іншим.

Отже, функціональна залежність є суворою, повною залежністю, діючою в кожному окремому випадку.

Термін кореляція (від пізньолатинської correlation – співвідношення) застосовується в різних галузях науки і техніки для позначення взаємо­залежності, взаємної відповідності, співвідношення понять, предметів, функцій підприємств.

На відміну від функціональної, *кореляційна залежність* виникає у випадку, коли одна з ознак залежить не тільки від другої, а й від низки випадкових факторів або коли серед умов, від яких залежить та чи інша ознака, є загальні для них обох умови. Основою теорії кореляції є припу­щення про те, що досліджувані явища підпорядковані певним імовірніс­ним закономірностям. Кореляційна залежність – це неповна і неточна залежність, яка відображає закон множинності причин і наслідків. Йдеться про зв’язок явищ, кожне з яких відчуває вплив різноманітних факторів, що діють з різною силою. Із загальної кількості чинників необхідно виділяти головні з метою їх обліку, вимірювання, порівняння та контролю, якщо це знадобиться.

У цілому кореляційна залежність відображає числове співвідношення між величинами, що виражаються у вигляді тенденцій до збільшення або зменшення однієї змінної величини за умови зростання іншої змінної величини.

Прикладом застосування кореляційної залежності є виділення одного фактора, однієї причини; вплив цього фактора стає в центрі нашої уваги. Але слід враховувати, що цей фактор ще не є єдиним, який керує якимось явищем. Поряд з дією цього фактора є низка інших, у результаті чого вдається кореляційна залежність. Наприклад, доходи від туризму залежать від обсягу туристичного потоку, але не визначають їх, бо на доходи впливають і інші фактори: рівень розвитку туристичної інфра­структури, обсяг наданих послуг, розвиток екскурсійної діяльності та ін.

У реальному житті безліч різноманітних факторів перехрещу­ються, в результаті чого більш поширені кореляційні залежності, а не функціональні.

Коефіцієнт кореляції демонструє як напрям зв’язку, так і тісноту зв’язку двох змінних. Один із способів відображення кореляційної за­лежності - розрахунок коефіцієнта рангової кореляції Спірмена (rс) за формулою:

$$r\_{c}=1-\frac{6\sum\_{}^{}d^{2}}{N^{3}-N}=1-\frac{6\sum\_{}^{}d^{2}}{N\left(N^{2}-1\right)}$$

де $d^{2}$ *-* квадрат різниці рангу;

*N -* кількість елементів сукупності або «пар» порівнюваних рангів.

Числове значення змінюється в межах від +1 до -1, тобто -1< *r* <1. Якщо зв’язок прямий, то має позитивні значення, якщо зворотний, то має негативні значення. При ослабленні кореляційної залежності абсолютна величина г зменшується; при відсутності залежності *r = 0* або близькій до неї. Кореляційний зв’язок сильніший, коли ближче *r* до +1 або -1. Про силу кореляції дає уявлення градація коефіцієнтів кореляції:

*0,7 <\r | <1* - кореляція висока;

*0,4 <\r \* $\leq $ *0,7*- кореляція середньої сили;

*0,2 <\ г\* $\leq $ *0,4* - кореляція низька;

*\r \ < 0,2* - кореляція мізерно мала.

Для розрахунку коефіцієнта кореляції статистичні дані проран- жуємо, тобто присвоїмо ранг у порядку зменшення ознаки. Далі необхідно переглянути всі ранги. Справа в тому, що бувають випадки, коли кілька об’єктів мають один і той же ранг. Наприклад, якісь два регіони ділять друге і третє місця за площею, а три регіони за кількістю внутрішніх туристів займають 4, 5 і 6 місця в ранжированому ряді. Ці випадки відомі під назвою «об’єднаних рангів». У такій ситуації перед розрахунком коефіцієнта кореляції необхідно присвоїти кожному регіону (об’єкту) середнє значення об’єднаних рангів. Так, наприклад, для регіонів значення рангів за площею дорівнюватиме 2,5, а за кількістю внутрішніх туристів – 5. Потім обчислюється коефіцієнт кореляції, використовуючи значення, надані об’єктам, які мали «об’єднані ранги».

Розрахуємо коефіцієнт кореляції між кількістю туристів і обсягом наданих послуг туристам за 27 регіонами України, попередньо проран- жувавши відповідні показники (табл. 8).

|  |  |  |
| --- | --- | --- |
| Регіон | Кількість туристів, тис. | Обсяг наданих послуг, млн грн |
| всього | ранг | всього | ранг |
| АР Крим | 392,7 | 3 | 632,9 | 2 |
| Вінницька | 47,6 | 20 | 11,7 | 27 |
| Волинська | 68, 7 | 15 | 24,2 | 18 |
| Дніпропетровська | 168,4 | 5 | 273,4 | 3 |
| Донецька | 151,9 | 6 | 236,5 | 4 |
| Житомирська | 12,9 | 27 | 12,8 | 25 |
| Закарпатська | 73, 0 | 12 | 28,5 | 17 |
| Запорізька | 114, 7 | 10 | 64,6 | 9 |
| Івано-Франківська | 1 268,9 | 2 | 54,7 | 10 |
| Київська | 20,2 | 26 | 45,9 | 12 |
| Кіровоградська | 37,5 | 22 | 16,1 | 23 |
| Луганська | 59,1 | 16 | 12,8 | 26 |
| Львівська | 117,2 | 9 | 123,7 | 7 |
| Миколаївська | 58,1 | 17 | 36,8 | 13 |
| Одеська | 133,0 | 8 | 124,7 | 6 |
| Полтавська | 72,3 | 13 | 49,4 | 11 |
| Рівненська | 57,80 | 18 | 22,3 | 19 |
| Сумська | 20,6 | 25 | 18,8 | 21 |
| Тернопільська | 33,1 | 23 | 15,2 | 24 |
| Харківська | 135,2 | 7 | 105,4 | 8 |
| Херсонська | 71,1 | 14 | 29,2 | 15 |
| Хмельницька | 56,8 | 19 | 30,4 | 14 |
| Черкаська | 40,9 | 21 | 19,3 | 20 |
| Чернігівська | 32,9 | 24 | 16,5 | 22 |
| Чернівецька | 75, 8 | 11 | 29,0 | 16 |
| м. Київ | 1 814,1 | 1 | 3 209,3 | 1 |
| м. Севастополь | 175,5 | 4 | 212,8 | 5 |

Таблиця 8

PP

$r\_{s}$=1$\frac{6∙562}{27 (27^{2}-1)}=1-\frac{3372}{27∙728}=1-\frac{3372}{19656}=1-0,17=0.83$

Розрахунки показали, що існує сильна позитивна кореляція між кількістю туристів і обсягом наданих послуг туристам за регіонами України.

* 1. **Вивчення рядів динаміки в туризмі**

Важливими показниками туристичного потоку є загальна кількість туристів, кількість туристів, величина сумарних туристичних витрат. Загальна кількість туристів дорівнює сумі кількості туристів за певний період, прийнятих на обслуговування за днями регістрації, тобто в перший день обслуговування. Кількість туроднів вимірюється в людино-днях і визначається за формулою:

$$D=P×t\_{cр}$$

де$ D$ - кількість туроднів, людино-днів;

Р - кількість туристів, осіб;

іср - середня тривалість перебування одного туриста в даній країні (регіоні).

Величина сумарних туристичних витрат (Тр) визначається за формулою:

$$Т\_{р=}D×T\_{ср}$$

де Тср - середні витрати туриста на добу.

У більшості країн світу туристичний потік дуже нерівномірний протягом усього року. Під час аналізу рядів динаміки за рік отримують кількісні характеристики, які відображають характер зміни показників за місяцями року. Для вивчення нерівномірності туристичного потоку можна розрахувати коефіцієнти нерівномірності трьома методами:

$$К\_{н}=\frac{D\_{max}}{D\_{min}}×100\%; $$

$$К\_{н}=\frac{D\_{max}}{D\_{год}}×100\%; К\_{н}=\frac{D\_{max}}{D\_{см}}×100\%; $$

де $D\_{max}, D\_{min} $- число туроднів в місяці максимального і мінімального туристичного потоку (людино-днів);

$D\_{год}$,$ D\_{см}$ - річне і середньомісячне число туроднів (людино-днів);

$D\_{см}$= $D\_{год}$,: 12 [27, с. 177].

Обсяг реалізованих послуг в туризмі має чітко виражений сезонний характер, що пов’язано з часом року, періодом відпусток, канікулами і т. д. Сезонними коливаннями називають постійні повторювані цілорічні зміни досліджуваних явищ.

Сезонні коливання описують *індексами сезонності*, які розрахо­вуються як відношення фактичного значення показника до деякого теоретичного (розрахункованого) рівня.

Сукупність цих показників відображає сезонну хвилю. Для виявлен­ня сезонних коливань зазвичай беруть дані за кілька років (не менше трьох), розподілені за місяцями, щоб виявити стійку сезонну хвилю, яка б не відображала випадкові умови одного року.

Для обчислення індексів сезонності застосовують різні методи: простої середньої, аналітичного вирівнювання, відносних чисел, ковзної середньої та ін. Якщо ряд динаміки не містить яскраво вираженої тенденції в розвитку, то індекси сезонності обчислюють безпосередньо за емпіричними даними методом простої середньої за формулою:

$$i\_{c}=\frac{\overbar{y\_{i}}}{\overbar{y}\_{i}}×100\%$$

де *іс -* індекс сезоності;

$\overbar{y\_{i}}$- середній рівень обсягу реалізованих послуг окремого місяця, грн;

$\overbar{y\_{i}}$- загальний середньомісячний обсяг реалізованих за розрахунковий період, грн.

Приклад розрахунку індексу сезонності методом простої середньої наведено в табл. 5.9.

Таблиця 5.9

Розрахунок сезонності прибутку
від реалізації туристичного продукту в регіоні (дані умовні)

|  |  |  |
| --- | --- | --- |
| Рік | Сума прибутку в місяць, млн грн | Разом |
| I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII |
| 2013 | 15 | 16 | 18 | 28 | 28 | 29 | 38 | 37 | 33 | 29 | 22 | 17 | 310 |
| 2014 | 18 | 19 | 17 | 27 | 30 | 31 | 39 | 38 | 34 | 33 | 21 | 17 | 324 |
| 2015 | 17 | 18 | 19 | 26 | 31 | 36 | 36 | 42 | 36 | 31 | 22 | 18 | 332 |
| Разом | 50 | 53 | 54 | 81 | 89 | 96 | 113 | 117 | 103 | 93 | 65 | 52 | 966 |
| Середній рівень | 16,7 | 17,7 | 18 | 27 | 29,7 | 32 | 37,7 | 39 | 34,3 | 31 | 21,7 | 17,3 | 26,8 |
| Індекс сезон­ності, % | 62,3 | 66,0 | 67,2 | 100,7 | 110,8 | 119,4 | 140,7 | 145,5 | 128,0 | 115,7 | 81,0 | 64,6 |  |

З таблиці видно, що загальний середньомісячний обсяг прибутку від реалізації туристичного продукту за три роки становить:

 $ \overbar{y\_{0}}$=$\frac{\sum\_{}^{}y}{n}=\frac{966}{36}=26,8 млн грн.$

Наприклад, для січня індекс сезонності розраховується так:

$i\_{c}=\frac{\overbar{y}\_{i}}{\overbar{y}\_{0}}×100\%$=$\frac{16,7}{26,8}$×100%=62,3%

Індекси сезонності показують, що найменше відхилення від серед­нього рівня прибутку від реалізації туристичного продукту спостерігається в січні (62,3 %), а найбільше - в серпні (145,5 %). Таким чином, розрахунок індексу сезонності дозволяє виявити наявні сезонні коливання і тенденції до зміни обсягу реалізації послуг туризму.

Аналіз сезонних коливань може сприяти вжиттю заходів щодо усу­нення нерівномірності попиту в туризмі. Ця проблема успішно вирішується в США, Канаді, Франції та ін. розвинених країнах за рахунок диверсифікації туристичного продукту. У даний час з метою цілорічного використання наявних туристичних ресурсів, у Туреччині основний акцент роблять на розвиток ділового, медичного, спортивного та інших видів туризму.

Простим способом виявлення сезонної лінії тренду є механічне вирівнювання динамічного ряду, або метод ковзної середньої. Для цього розраховують середню величину з трьох (або більше) рівнів ряду, утворених послідовним винятком початкового члена ряду і заміщенням його таким за порядком:

$\overbar{y}\_{I}$=$\frac{y\_{1+}y\_{2+}y\_{3}}{3};$

$\overbar{y}\_{II}$=$\frac{y\_{2+}y\_{3+}y\_{4}}{3};$

$\overbar{y}\_{III}$=$\frac{y\_{3+}y\_{4+}y\_{5}}{3};$

де $\overbar{y}\_{I}, \overbar{y}\_{II}, \overbar{y}\_{III} $ - рівні диинамічного ряду, згладжені за трирівневою ковзною середньою;

$y\_{1, }y\_{2, }y\_{3}$ *–* емпіричні рівні динамічного ряду.

Приклад розрахунку ковзної середньої наведено в таблиці 5.10 (дані умовні).

Таблиця 5.10

Розрахунок тримісячної ковзної середньої

|  |  |  |  |
| --- | --- | --- | --- |
| Місяць | Кількість туристів | За три місяці | Ковзна середня |
| 1 | 1 950 |  |  |
| 2 | 1 770 | 5 260 | 1 753,3 |
| 3 | 1 540 | 4 620 | 1 540 |
| 4 | 1 310 | 3 650 | 1 216,7 |
| 5 | 800 | 3 525 | 1 175 |
| 6 | 1 415 | 3 775 | 1 258,3 |
| 7 | 1 560 | 4 905 | 1 635 |
| 8 | 1 930 | 5 640 | 1 880 |
| 9 | 2 150 | 6 180 | 2 060 |
| 10 | 2 100 | 6 300 | 2 100 |
| 11 | 2 050 | 6 150 | 2 050 |
| 12 | 2 000 | - | - |

Розрахунок тримісячної ковзної середньої певною мірою згладжує гострі піки і провали сезонних коливань. Недоліком способу є те, що отримані середні не дають теоретичних закономірностей рядів, в основі

Розрахунок тримісячної ковзної середньої певною мірою згладжує гострі піки і провали сезонних коливань. Недоліком способу є те, що отримані середні не дають теоретичних закономірностей рядів, в основі яких була б математично виражена закономірність і це дозволяло б не лише виконати аналіз, але і прогнозувати динаміку ряду на майбутнє. Більш досконалим прийомом вивчення загальної тенденції в рядах динаміки є *аналітичне вирівнювання.* Його суть полягає в знаходженні рівняння, що виражає закономірність зміни явища як функцію часу $y=f(t)$ Шляхом теоретичного аналізу виявляється характер розвитку явища і на цій основі використовуються такі функції:

* лінійна: $\overbar{y\_{t}}=a\_{0+}a\_{1}t; $
* парабола 2-го порядка: $\overbar{y\_{t}}=a\_{0+}a\_{1}t+a\_{2}t^{2}$
* показова: $y\_{t}$=$a\_{o×}a\_{1t}$
* гіперболічна: $\overbar{y}\_{t}=a\_{0}+\frac{a\_{1}}{t}$

Найчастіше при комплексному дослідженні динамічних рядів постає завдання подальшого прогнозування їх рівнів. Одним з найпростіших способів прогнозу є метод екстраполяції. Він ґрунтується на прогно­зуванні подій з урахуванням аналізу показників за минулі роки (при цьому, не менше ніж за 5-8 років) і застосовується за таких умов:

* період часу, для якого будується функція, повинен бути достатнім для виявлення тенденції розвитку;
* значних змін характеристик досліджуваного процесу не відбудеться;
* не очікується значних зовнішніх впливів на процеси, які можуть істотно вплинути на хід розвитку.

Для прогнозування можна використовувати методи, засновані на використанні: середнього абсолютного приросту, середнього темпу зрос­тання і функції аналітичного вирівнювання.

Прогнозування на основі середнього абсолютного приросту можливо в разі, якщо розвиток процесу можна описати лінійною функцією за формулою:

$y\_{n+1}=y\_{n}+\overbar{Δ }$× t

де *уп* - останній рівень динамічного ряду;

$\overbar{Δ }$- середній абсолютний приріст ряду динаміки;

t - кількість періодів екстраполяції.

$$\overbar{Δ}=\frac{y\_{n-}y\_{1}}{n-1}$$

де У1, У2 ...., Уп - рівні динамічного ряду;

*п* - кількість рівнів ряду.

Середній абсолютний приріст ряду динаміки показує, на скільки одиниць у середньому змінювалося значення показника за аналізований період.

У разі опису загальної тенденції розвитку процесу за допомогою показової функції, прогнозоване значення рівня має такий вигляд:

$$y\_{n+1}=y\_{n}×\left(\overbar{T}\_{p}\right)^{t}$$

де *Тр -* середній темп зростання динамічного ряду, виражений у коефі­цієнтах [27, с. 127].

$\overbar{T}\_{p}$=$\sqrt[n-1]{T\_{1}}×T\_{2}×…. ×T\_{n-1}$

де Т1, Т2, .... Тn-1 - темпи зростання, виражені в коефіцієнтах.

Темп зростання *(*$\overbar{T}\_{p}$) виражений у коефіцієнтах, показує, у скільки разів змінився поточний рівень показника порівняно з попереднім.

$$T\_{p}=\frac{y\_{i}}{y\_{1}}$$

Мета аналітичного вирівнювання - отримання рівняння тренду, який описує зміну рівнів динамічного ряду в часі *t*. Якщо продовжити позначення умовного показника часу *і* до прогнозного періоду, підставити відповідне *t* в рівняння тренду, то отримаємо прогнозну оцінку показника.

Прогнозування за допомогою методу екстраполяції виправдано при недостатньому знанні про природу досліджуваного явища або відсутності даних, необхідних для застосування більш досконалих методів прогнозування*.*

**Питання для самоконтролю**

1. За якими основними напрямами вихідні відомості піддаються статистичній обробці?
2. Які методи використовуються для згортання вихідних даних?
3. У чому суть методу статистичного угруповання?
4. Зробіть угруповання регіонів України за кількістю населення.
5. Намалюйте гістограму розподілу регіонів України за обсягами відра­хування в бюджет туристичними підприємствами. Поясніть.
6. Чим відрізняється побудова полігону розподілу від гістограми?
7. Який графік називається огіви?
8. Які статистичні сукупності називаються бімодальною і мульти- модальні?
9. Яку роль і значення мають середні величини в туризмі? Наведіть приклади.
10. Які властивості має середня арифметична величина?
11. Як розраховується дисперсія і середньоквадратичне відхилення для статистичного ряду розподілу?
12. Розрахуйте коефіцієнт варіації для розподілу екскурсантів за регіонами України. Результат поясніть.
13. Чим функціональна залежність відрізняється від кореляційної?
14. Наведіть приклади використання функціональної залежності на практиці.
15. Розрахуйте рангові коефіцієнти кореляції між розподілом турис­тичних підприємств за регіонами України та обсягом відрахування в бюджет. Результат поясніть.
16. Визначте залежність між забезпеченістю туристично-рекреацій­ним потенціалом регіонів України і кількістю в’їзних туристів. Результат поясніть.
17. Розрахуйте рангові коефіцієнти кореляції між кількістю виїзних туристів з регіонів України та середньомісячною заробітною платою. Результат поясніть.
18. Як розраховується індекс сезонності? Які для цього використо­вуються методи?
19. У чому полягає суть аналітичного вирівнювання? Які для цього використовуються функції?
20. Дайте визначення екстраполяції. За яких умов можна застосо­вувати метод екстраполяції?