МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ АВТОМОБІЛЬНО-ДОРОЖНІЙ УНІВЕРСИТЕТ

Коваль О. А.

МОДЕЛЮВАННЯ ЗАСОБІВ ВИМІРЮВАЛЬНОЇ ТЕХНІКИ

Методичні рекомендації до проведення лабораторних робіт

Затверджено методичною радою факультету, протокол № 1 від 7 «вересня» 2018 р.

Харків
2018

Зміст

Вступ

В багатьох випадках наукові дослідження спрямовані на знаходження функціональних або стохастичних залежностей об'єктів аналізу від одного або декількох чинників (факторів), які, як передбачається, впливають на цей об'єкт .

Якщо зв'язок між узагальненим показником, що його аналізують, і факторними характеристиками є не функціональним, а має ознаки стохастичної залежності, доцільним вважається застосування статистичних методів, а також теорії ймовірностей. У числі статистичних застосовуються класичні методи одновимірних і багатовимірних сукупностей, варіаційні ряди, закони розподілу, вибір даних, кореляційно-регресійний та дисперсійний аналіз.

Найбільш широко в економічному аналізі застосовуються методи парної і множинної кореляції. За допомогою цих методів є можливим визначення не функціональної, а стохастичної причинно-наслідкової залежності між економічними явищами, тобто вивчення дії факторів, що мають тенденційний вплив на об'єкт дослідження.

Результатом цих видів аналізу є побудова адекватної моделі зв'язку досліджуваного показника та фактора (факторів), яка дозволяє спрогнозувати якісне і кількісне значення показника при тих чи інших умовах. Статистичний прогноз – це імовірнісна оцінка можливості розвитку того або іншого об'єкта (процесу) і величини його ознак у майбутньому, отримана на основі статистичної закономірності, виявленої по даним минулого періоду. Об'єктом статистичного прогнозування можуть бути ті явища й процеси, керування якими, а тим більше планування їхнього розвитку утруднено через дію багатьох факторів, вплив яких не може бути однозначно й повністю визначено. Статистичний прогноз припускає не тільки вірне якісне пророкування, але й досить точний кількісний вимір імовірних можливостей очікуваних значень ознак.

1. План статистичного дослідження

1.1. Мета і задачі дослідження

Мета і задачі дослідження витікають з господарської ситуації. Мета цього завдання — проаналізувати залежності досліджуємого показника від 5-ти факторів $Y = \varphi(X_1), Y = \varphi(X_2), Y = \varphi(X_3)$.

Необхідно дослідити залежність від показників і розробити його статистичну модель, що дозволяє швидко і якісно передбачати значення Y на підставі чинників, що визначають його величину. Передбачається, що на (Y) можуть мати істотний вплив такі чинники: (X_1), (X_2), (X_3).

Для аналізу залежності і побудови моделі необхідно:

1) отримати випадкову 10-відсоткову вибірку з генеральної сукупності, представленої 360 значеннями кожного показника;

2) провести розрахунок значень показників (X_3);

3) розрахувати для всіх показників описову статистику і перевірити вибірку на присутність аномальних спостережень (при необхідності виключити відповідні підприємства з подальшого дослідження);

4) перевірити достатність обсягу вибірки для отримання достовірних результатів;

5) визначити закон розподілу результативної змінної;

6) провести парний кореляційний аналіз залежності і відібрати чинник, що має найвищу кореляцію з результативним показником Y;

7) побудувати регресійну модель залежності від вибраного чинника і оцінити її статистичну значущість;

1.2. Методи дослідження

Методами дослідження є групування даних. описова статистика, парний кореляційно-регресійний аналіз, дисперсійний аналіз і т.і.

2. Збір і систематизація первинних данних

Для формування репрезентативної вибірки необхідний механізм випадкового добору об'єктів з генеральної сукупності. У курсовій роботі для цього механізму служать передостання й остання цифри номера залікової книжки студента.

Відбір вихідних даних проводиться таким чином: за передостанньою цифрою залікової книжки встановлюється номер вибірки Y, за останньою цифрою залікової книжки встановлюється номер вибірки X_1 та X_2.

Данні вибірок знаходяться в додатку А, значення слід виписати в таблицю вихідних даних (табл. 3). Потім на підставі цих первинних показників варто розрахувати X_{i3}.

$$X_{i3} = \frac{X_{i1}}{X_{i2}}.$$

Результати розрахунків наводяться в табл. 3

Таблиця 3 - Вихідні дані до статистичного дослідження

№ спостереження	Значення змінних			
i	Y_1	X_1	X_2	X_3
1	798	706	303	2,3300
2	1500	1026	470	2,1830
3	1778	1544	670	2,3045
4	2146	1605	724	2,2169
5	2441	1997	760	2,6276
6	2678	1992	826	2,4116
7	3013	2184	842	2,5938
8	3729	2352	913	2,5761
9	3859	2311	944	2,4481
10	3980	2632	992	2,6532
11	4078	2912	1086	2,6814
12	4321	2744	1058	2,5936
13	4813	3053	1127	2,7090
14	4879	3425	1247	2,7466
15	4894	3148	1201	2,6211
16	5037	3227	1181	2,7324
17	5231	3452	1257	2,7462
18	5350	3699	1340	2,7604
19	5723	3479	1352	2,5732
20	5811	3717	1351	2,7513
21	5945	4043	1444	2,7999
22	6080	3994	1432	2,7891
23	6210	4081	1484	2,7500
24	6302	4157	1508	2,7566
25	6709	4412	1564	2,8210
26	6775	4438	1584	2,8018
27	6914	4590	1589	2,8886
28	7134	4722	1666	2,8343
29	7301	4799	1713	2,8015
30	7755	5028	1561	3,2210
31	7960	5210	1847	2,8208
32	8053	5333	1681	3,1725
33	8292	5429	1807	3,0044
34	8774	5684	1945	2,9224
35	9118	5773	2088	2,7648
36	9819	7269	2512	2,8937
Сума значень	195200	130167	47069	97,3026
Середнє значення	5422,222222	3615,75	1307,472222	2,7029

2.2. Групування даних. Розрахунок описової статистики і перевірка однорідності вибіркової сукупності

Групування вихідних даних проводиться з метою аналізу структури і закономірностей розподілу досліджуваних показників. У відповідальних дослідженнях групування виконують для кожного досліджуваного показника. У курсовій роботі його виконуємо тільки для результативного показника Y, але різними способами.

У статистичних дослідженнях використовують групування з використанням рівних і нерівних за величиною інтервалів. У курсовій роботі необхідно вибрати найкращий спосіб групування за величиною результативного показника -Y.

2.2.1 Групування з використанням рівних інтервалів

Групування з рівними інтервалами доцільні в тих випадках, коли варіація виявляється в порівняно вузьких інтервалах і розподіл одиниць сукупності за даною ознакою є практично рівномірним. Оптимальну кількість груп АГ з рівними інтервалами визначають за формулою Стерджесса

$$K = 1 + 3{,}322 \cdot \lg n.$$

де n - кількість спостережень (обсяг вибірки);

lg n - десятковий логарифм числа n

Отримане значення K звичайно округлюють до цілого у більший бік. Потім розраховується ширина групувального інтервалу h:

$$h = \frac{Y_{\max} - Y_{\min}}{K}.$$

де Y_{max} – максимальне значення показника, що вивчається, у вибірці;

Y_{min} – мінімальне значення показника, що вивчається, у вибірці.

Значення h також округлюють до цілого у більший бік. Після цього встановлюють межі групувальних інтервалів:

- нижня межа першого групувального інтервалу

$$a_1 = Y_{\min};$$

- верхня межа першого групувального інтервалу

$$b_1 = a_1 + h$$

Межі наступних інтервалів встановлюють за правилом: нижня межа чергового інтервалу береться за рівну верхній межі попереднього інтервалу, а верхня межа дорівнює нижній плюс ширина групувального інтервалу. В результаті весь діапазон зміни значень змінної розбивається на К рівних за величиною інтервалів.

Одночасно зі встановленням меж групувальних інтервалів задають умови віднесення спостережень на інтервал, їх задають у вигляді подвійної нерівності

$$a_k \le Y < b_k, K = 1,2,3..., K$$

Відповідно до цієї умови на інтервал з номером К відносять ті значення досліджуваних ознак, які більше або рівні нижній межі і менше верхньої межі.

Далі слід розподілити одиниці вибіркової сукупності за інтервалами у залежності від величини результативної ознаки. Для цього рекомендується, скласти таблицю такого вигляду:

Таблиця 4 - Групування результативного показника Y

№ інтервалу	Межі інтервалів		Частота, f_k	Кумулятивна частота, S_{fk}	Частка, $W_k = f_k/n$	Кумулятивна частка
	X_{min}	X_{max}				
1	798	2086,714286	3	3	0,083	0,083
2	2086,714286	3375,428571	4	7	0,111	0,194
3	3375,428571	4664,142857	5	12	0,139	0,333
4	4664,142857	5952,857143	9	21	0,250	0,583
5	5952,857143	7241,571429	8	29	0,222	0,806
6	7241,571429	8530,285714	4	33	0,111	0,917
7	8530,285714	9819	3	36	0,083	1
Разом			36		1	

Після групування даних за інтервалами побудуємо гістограму розподілу.

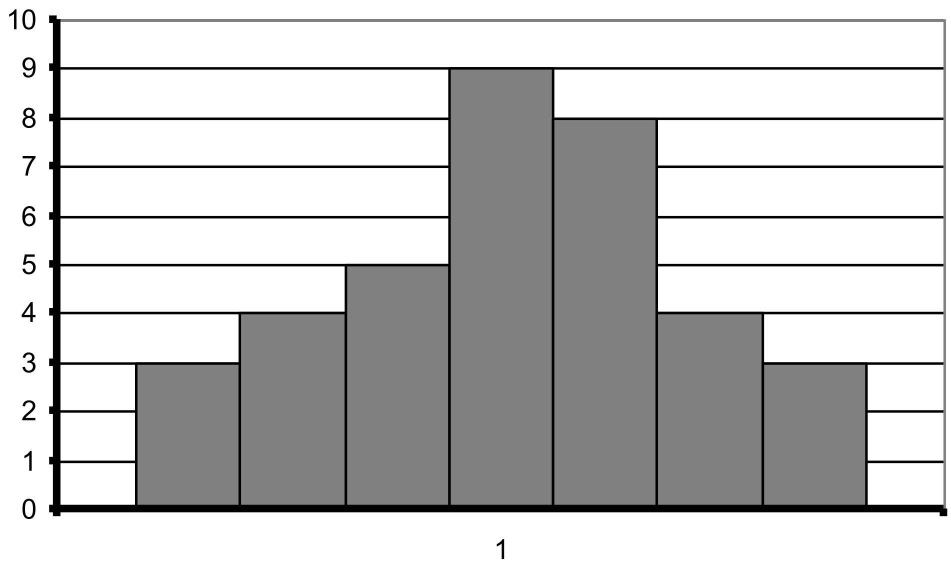

Рисунок 2.1 Гістограма розподілу

2.2.2. Групування з використанням нерівновеликих інтервалів

Групування з нерівновеликими інтервалами застосовуються для опису статистичних даних розподілу, що мають явну асиметрію частот і часток. Ширину і межі цих інтервалів встановлюють на основі логічного аналізу попередніх відомостей про якісні і кількісні характеристики досліджуваного явища.

У курсовій роботі в якості одного з можливих рішень задачі групування по Y рекомендується використовувати досить просту формалізовану процедуру поділу на групи.

Процедура виділення груп об'єктів з нерівними інтервалами досліджуваної ознаки така. Необхідно ранжирувати значення ознаки. Потім весь інтервал її можливих значень $[F_{\min}; F_{\max}]$ розділити на два інтервали, відокремлюваних одне від одного середнім значенням ознаки Y.

На першому інтервалі $[Y_{\min}; Y]$ будуть розташовані варіанти досліджуваної ознаки менше середнього значення Y, на другому $[Y; Y_{\max}]$- більше, ніж середнє значення Y.

У випадку симетричного розподілу точка, що відповідає середньому значенню ознаки Y, не поділятиме інтервал $[Y_{min};Y_{max}]$ на рівні частини, а буде зміщена до якого-небудь з кінців інтервалу.

Вибираємо з двох інтервалів, розділених значенням середньої величини, інтервал найменшої довжини, для чого порівнюємо за модулем величини $[Y_{\min} - Y]$ і $[Y - Y_{\max}]$. Довжину найменшого з двох порівнюваних інтервалів поділяємо навпіл і отримане значення ΔY додаємо до середнього Y і віднімаємо від нього. Одержуємо координати двох точок $(Y - \Delta Y)$ і $(Y + \Delta Y)$, які відзначаємо на числовій осі варіаційного ряду вліво і вправо від середнього значення.

В результаті числова вісь, що відповідає ранжируваному варіаційному ряду досліджуваної ознаки, розділяється на три інтервали $[Y_{\min}; Y - \Delta Y); [Y - \Delta Y; Y + \Delta Y]; [Y + \Delta Y; Y_{\max}]$, довжини яких можуть бути інтерпретовані як величини, що відмежовують дрібні, середні і великі одиниці сукупності. Після встановлення меж інтервалів занесемо у таблицю частот і часток, побудуємо гістограму розподілу Y

№ інтервалу	Межі інтервалів		Частота, f_k	Кумулятивна частота, S_{fk}	Частка, $W_k = f_k/n$	Кумулятивна частка
	X_{min}	X_{max}				
1	798	3223,833	7	7	0,194	0,194
2	3223,833	7620,611	22	29	0,611	0,806
3	7620,611	9819	7	36	0,194	1
Разом			36		1	

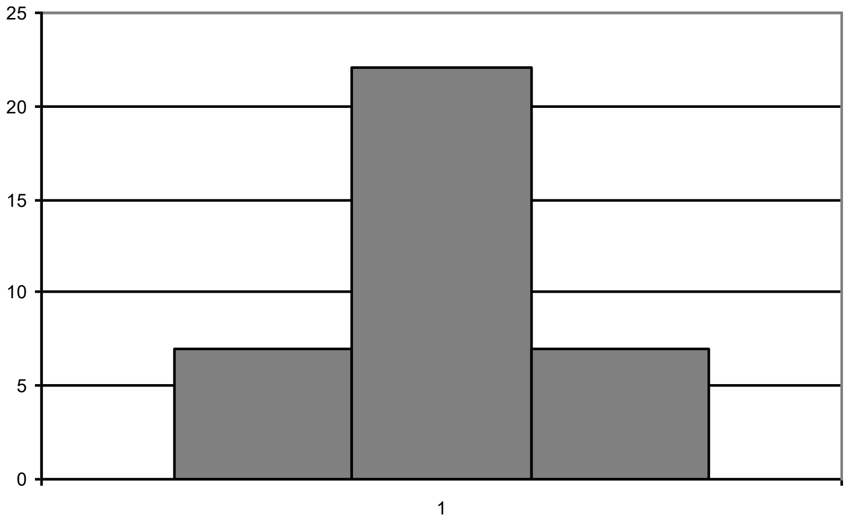

Рисунок 2.1 Гістограма розподілу з нерівновеликими інтервалами

Результати групувань з використанням рівновеликих і нерівновеликих інтервалів проаналізовано і зроблено попередній висновок щодо закономірності розподілу за величиною досліджуваної ознаки Y. Закон розподіленім дуже схожий на нормальний. Необхідно вказати що перший спосіб групування дав більш чітке уявлення про закономірності розподілу показника, що досліджується.

2.2.3. Розрахунок узагальнюючих характеристик і перевірка однорідності вибіркової сукупності

Наступнім етапом аналізу сукупності спостережень є розрахунок узагальнюючих характеристик (описової статистики) досліджуваної статистичної сукупності. Цей розрахунок можна виконати за не згрупованими або згрупованими даними (на підставі частотної таблиці). Більш точними є результати, отримані з використанням не згрупованих даних.

У курсовій роботі розрахунок зазначених показників описової статистики варто виконати для кожної змінної за не згрупованими даними. Для розрахунку рекомендуються формули, наведені в табл. 5.

Таблиця 5 Формули для розрахунку узагальнюючих показників вибіркових сукупностей

Узагальнюючі показники	Результативна змінна Y	Факторні зміні X
Середнє	$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$	$\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_j$
Середнє квадратичне відхилення	$\sigma_Y = \sqrt{\overline{Y^2} - \left(\bar{Y}\right)^2}$	$\sigma_X = \sqrt{\overline{X_j^2} - \left(\bar{X}_j\right)^2}$
Коефіцієнт варіації	$V = \frac{\sigma_Y}{\bar{Y}} \cdot 100\%$	$V = \frac{\sigma_X}{\bar{X}_j} \cdot 100\%$

Примітка.

1) i – номер спостереження, i = 1,2..., n;
2) j – номер чинника, j = 1,2,3;
3) $\overline{Y^2}, \overline{X^2}$ – середня квадратів показників;
4) $(\overline{Y})^2, (\overline{X}_j)^2$ – квадрат середніх значень.

Розрахунок виконується на підставі вихідних даних, наведених у табл. 3. Результати розрахунків представимо в окремій таблиці.

№ спостереження	Значення змінних			
i	$\overline{Y^2}$	$\overline{X_1^2}$	$\overline{X_2^2}$	$\overline{X_3^2}$
1	636804	498436	91809	5,429
2	2250000	1052676	220900	4,765
3	3161284	2383936	448900	5,311
4	4605316	2576025	524176	4,914
5	5958481	3988009	577600	6,904
6	7171684	3968064	682276	5,816
7	9078169	4769856	708964	6,728
8	13905441	5531904	833569	6,636
9	14891881	5340721	891136	5,993
10	15840400	6927424	984064	7,040
11	16630084	8479744	1179396	7,190
12	18671041	7529536	1119364	6,727
13	23164969	9320809	1270129	7,338
14	23804641	11730625	1555009	7,544
15	23951236	9909904	1442401	6,870
16	25371369	10413529	1394761	7,466
17	27363361	11916304	1580049	7,542
18	28622500	13682601	1795600	7,620
19	32752729	12103441	1827904	6,621
20	33767721	13816089	1825201	7,570
21	35343025	16345849	2085136	7,839
22	36966400	15952036	2050624	7,779
23	38564100	16654561	2202256	7,563
24	39715204	17280649	2274064	7,599
25	45010681	19465744	2446096	7,958
26	45900625	19695844	2509056	7,850
27	47803396	21068100	2524921	8,344
28	50893956	22297284	2775556	8,033
29	53304601	23030401	2934369	7,849
30	60140025	25280784	2436721	10,375
31	63361600	27144100	3411409	7,957

32	64850809	28440889	2825761	10,065
33	68757264	29474041	3265249	9,027
34	76983076	32307856	3783025	8,540
35	83137924	33327529	4359744	7,644
36	96412761	52838361	6310144	8,374
Середнє значення квадратів	34409571,06	15181768,36	1920759,417	7,356118495
Середнє квадратичне відхилення	2238,096787	1451,936741	459,6474788	0,225204593
Варіація	41,27637517	40,15589409	35,15542977	8,332113966

Аналізуючи отримані результати можна зробити висновки що перші три показника мають значну варіацію, а останній відрізняється достатньою однорідністю досліджуваної сукупності об'єктів. Чим менше значення коефіцієнта варіації, тим однорідніше об'єкти досліджуваної сукупності і надійніше рішення, прийняті з використанням описової статистики. Сукупність вважається однорідною, якщо $V_0 < 33\%$. Неоднорідні сукупності характеризуються великими значеннями коефіцієнтів варіації, що може бути наслідком присутності у вибірках аномальних спостережень. Значення змінних, що різко виділяються, прийнято вважати аномальними. Об'єкти з такими значеннями змінних необхідно виключити з вибірки, оскільки, як правило, вони мають іншу структуру або/і зовнішні умови існування (відмінні від інших об'єктів, що потрапили у вибірку) і спотворюють висновки про досліджуване явище. Далі, при необхідності, ці об'єкти вивчають окремо.

У курсовій роботі перевірка значень змінних на присутність аномальних спостережень може бути виконана з використанням правила «трьох сигм», відповідно до якого значення змінної вважається аномальним, якщо воно виходить за межі припустимого інтервалу:

$$\bar{Y} - 3\cdot\sigma_Y \le Y \le \bar{Y} + 3\cdot\sigma_Y \text{ та } \bar{X}_j - 3\cdot\sigma_X \le X \le \bar{X}_j + 3\cdot\sigma_X$$

Для досягнення однорідності треба поступово одне за одним виключати підприємства з сукупності до тих пір, поки коефіцієнти варіації за змінними Y та X_j не будуть задовольняти умові однорідності сукупності ($V_a < 33\%$) при виконанні правила "трьох сигм".

Розрахуємо за правилом трьох сигм і занесемо в таблицю

	Значення змінних			
	Y	X_1	X_2	X_3
$\bar{Y} - 3\cdot\sigma_Y$ $\bar{X}_j - 3\cdot\sigma_X$	-1292,07	-740,06	-71,47	2,03
$\bar{Y} + 3\cdot\sigma_Y$ $\bar{X}_j + 3\cdot\sigma_X$	12136,51	7971,56	2686,41	3,38

Отже по цім даним усі значення попадають в заданий інтервал тому ніякі дані виключати ми не будемо.

2.3. Поширення вибіркових результатів на генеральну сукупність. Оцінка достатності обсягу вибірки

Кінцевою метою вибіркового спостереження є характеристика генеральної сукупності. Враховуючи, що на основі вибіркового обстеження не можна точно оцінити досліджуваний параметр генеральної сукупності, потрібно знайти межі, у яких він знаходиться. Для цього необхідно з імовірністю 0,95 визначити граничну помилку Δy , вибіркової середньої результативного показника (середнього доходу підприємств, що входять у вибірку) і довірчі межі середнього доходу Y всіх підприємств генеральної сукупності:

$$\Delta Y = t \cdot \sqrt{\frac{\sigma_Y^2}{n}(1-\frac{n}{N})}$$

$$\bar{Y} - \Delta Y \leq \bar{\bar{Y}} \leq \bar{Y} + \Delta Y,$$

Де n- обсяг вибірки;

N- обсяг генеральної сукупності;

Y - генеральна середня.

Знайдемо граничну помилку ΔY

$\Delta Y = 693{,}59$

Отже маємо результат довірчі межі Y генеральної сукупності

$$\bar{Y} - \Delta Y \leq \bar{\bar{Y}} \leq \bar{Y} + \Delta Y,$$

$$4728{,}63 \leq \bar{\bar{Y}} \leq 6115{,}82$$

Обсяг вибірки повинен бути достатнім для одержання достовірних висновків про досліджуване явище. У зв'язку з цим у курсовій роботі варто визначити, яким повинен бути обсяг вибірки для проведення дослідження. З цією метою варто розрахувати мінімальний обсяг вибірки, необхідний для оцінки генеральної середньої результативного показника з відносною помилкою Π_d= 5% на рівні довірчої імовірності 0,95. Розрахунок мінімального обсягу вибірки виконують за формулою без повторної власне випадкової вибірки:

$$n = \frac{t^2 \sigma_y^2 N}{\Delta_{nd}^2 \cdot N + t^2 \sigma_y^2}$$

де t — коефіцієнт довіри, що відповідає рівню довірчої імовірності 0,95 (t = 1,96);

Δ_{nd} - припустима гранична абсолютна величина помилки оцінки генеральної середньої:

$$\Delta_{nd} = \frac{\bar{Y}\Pi_d}{100};$$

$$\Delta_{nd} = 271{,}111$$

Отже $n = 151{,}57$

Повинно бути 152 спостереження

В результаті розрахунків виявляється, що фактичний обсяг вибірки більше або менше мінімального. У цьому випадку слід знайти фактичну величину помилки оцінки генеральної середньої $\Delta_{факт}$.Для її визначення формулу, за якою проводився розрахунок мінімального обсягу вибірки, необхідно перетворити.

$$\Delta_{факт} = \sqrt{\frac{\frac{t^2\sigma_y^2 N}{n} - t^2\sigma_y^2}{N}}$$

Потім підставити в неї фактичні значення обсягу вибірки n, дисперсії ознаки σ_y, коефіцієнта довіри t і знайти відповідне їм значення помилки оцінки генеральної середньої $\Delta_{факт}$

$$\Delta_{факт} = 693{,}59$$

2.4. Аналіз закономірностей розподілу досліджуваних показників

З метою найбільш повного опису поводження досліджуваної ознаки в статистичних дослідженнях часто потрібно визначити закон її розподілу. У курсовій роботі це необхідно зробити тільки для результативної змінної. Якщо вона виявиться розподіленою за нормальним законом, то надалі можна буде використовувати метод кореляційно-регресійного аналізу її залежності від факторної змінної. У протилежному випадку доведеться використовувати інші методи і прийоми статистичного аналізу.

У статистиці для опису поведінки випадкових дискретних і безперервних величин використовуються різні закони розподілу. Нормальний закон використовується для опису розподілу випадкових безперервних величин У курсовій роботі слід дати відповідь на питання, чому досліджуваний результативний показник є випадковою безперервною величиною, для чого необхідно знати закон його розподілу і чому бажано, щоб це був нормальний закон

Для перевірки гіпотези про нормальність розподілу результативного показника необхідно:

1)за даними вибірки побудувати гістограму, полігон і кумуляту розподілу емпіричних значень результативного показника Y;

2)визначити графічним і аналітичним способами моду (Mo) і медіану (Me), використовуючи для цього дані табл. 4 і побудовані гістограму і кумуляту. За співвідношенням Mo, Me і Y, а також за допомогою коефіцієнта асиметрії Пірсона (FQ) й ексцесу (Ex) описати форму розподілу емпіричних значень;

3)розрахувати теоретичні частоти, припускаючи, що має місце нормальним закон розподілу, і побудувати теоретичну криву розподілу;

4)в одних координатних осях побуду чат й емпіричну і теоретичну криві розподілу досліджуваного показникаY ;

5)за критерієм x· на рівні довірчої імовірності 0,95 перевірити гіпотезу про близькість емпіричного і теоретичного розподілів.

Для розрахунку моди спочатку за найбільшою частотою (див. табл. 4) визначають модальний інтервал, тобто інтервал, що містить моду, а потім приблизно розраховують її за формулою:

$$Mo = a_{мо} + h_{мо}\frac{f_{мо} - f_{мо-1}}{\left(f_{мо} - f_{мо-1}\right)+\left(f_{мо} - f_{мо+1}\right)}$$

де $a_{мо}$ - нижня межа модального інтервалу;

$h_{мо}$ - величина модального інтервалу;

$f_{мо-1}$, $f_{мо+1}$ - частоти відповідно в попередньому і наступним за модальним інтервалах.

Отже інтервал, що містить моду

мін	макс	Частота, $f_к$
4664,142857	5952,857143	9

Mo = 5695,114286

Графічно моду визначають за гістограмою. Для цього обирають найвищий прямокутник, який і є модальним. Далі праву верхню вершину прямокутника. Ідо передує модальному (частота $f_{мо-1}$), з'єднують із правою верхньою вершиною модального прямокутника (частота $f_{мо}$), а ліву верхню вершину цього прямокутника - з лівою верхньою вершиною прямокутника, що йде за модальним (частота $f_{мо+1}$). З точки перетину опускають перпендикуляр на горизонтальну вісь. Основа перпендикуляра покаже значення моди Mo. Точність визначення залежить ВІД масштабу графіка.

При обчисленні медіани спочатку знаходять інтервал, що містить медіану, для чого використовують накопичені частоти або частки. Медіанним є інтервал, накопичена частота якого дорівнює чи перевищує половину всього обсягу сукупності. Потім значення медіани розраховується за формулою:

$$Me = a_{me} + h_{me} \frac{\sum f - S_{f_{mo-1}}}{2 \cdot f_{Me}}$$

Де a_{Me} - нижня межа медіанного Інтервалу;

h_{Me} - ширина медіанного інтервалу;

$S_{f_{ME-1}}$ - накопичена частота інтервалу, передуючого медіанному;

f_{ME} - частота медіанного інтервалу.

Отже Me = 5523,285714

Графічно медіана визначається за кумулятою. Для цього останню ординату кумуляти, що дорівнює сумі всіх частот або часток, ділять навпіл. З отриманої точки відновлюють перпендикуляр до перетину з кумулятою. Абсциса точки перетину і дає значення медіани.

Співвідношення моди, медіани і середньої арифметичної використовують для розпізнавання симетричності варіації. Необхідною, але недостатньою умовою симетричності є рівність трьох характеристик: Y= Me = Mo. У рядах із правосторонньою асиметрією Y > Me > Mo, з лівосторонньою асиметрією Y<Me<Mo.

Як показники формоутворення застосовуються:

- коефіцієнт асиметрії Персона:
-

$$K_a = \frac{\overline{Y} - Mo}{\sigma_y}$$

(якщо K_a > О, то скошеність правостороння, якщо K_a < 0 - лівостороння; якщо K_a= О, то розподіл симетричний);

- ексцес

(якшо Ex = 0, то розподіл близький до нормального, якщо Ex > 0, розподіл гостровершинний, Ex<0 - розподіл низьковершинний).

Отже $K_a = \frac{\overline{Y} - Mo}{\sigma_y} =$ -0,12;

$$Ex = \frac{\left\lfloor \sum \left(Y_i - \bar{Y}\right)^4 \right\rfloor}{n \cdot \sigma_y^4} - 3 = \text{-0,611}.$$

За цими даними варіація з лівосторонньою асиметрією, скошенність лівосторонння, розподіл низько вершинний

Розрахунок теоретичних частот виконується за методом оцінки імовірності того, що нормоване відхилення від середньої не вийде за межі ± z з використанням функції щільності нормального розподілу φ_z. Розрахунки теоретичних частот і перевірку гіпотези про близькість емпіричного і теоретичного розподілів рекомендується виконувати за допомогою таблиці.

Порядок розрахунку теоретичних частот кривої нормального розподілу:

1) визначають середини інтервалів $\frac{Y_{\max} - Y_{\min}}{2}$;

2) визначають $z_{K\min} = \frac{Y_{K\min} - \bar{Y}}{\sigma_y}$ та $z_{K\max} = \frac{Y_{K\max} - \bar{Y}}{\sigma_y}$

3) за таблицею розподілу функції $\Phi(z)$ (додаток Б) визначають її значення $\Phi(z_{\kappa\min})$ та $\Phi(z_{\kappa\max})$;

4) знаходять нормовану ймовірність $P_K = \Phi(z_{k\min}) - \Phi(z_{k\max})$

5) обчислюють теоретичні частки за формулою: nP_K.

Сума теоретичних і емпіричних частот повинна бути рівною, але може не збігатися через округлення в розрахунках.

Для перевірки гіпотези про близькість емпіричного і теоретичного розподілів розраховують критерій згоди Персона:

$$\chi^2_{розр} = \sum_{i=1}^{k} \frac{(nP_r - f_r)^2}{nP_k}.$$

і порівнюють його з табличним значенням $\chi^2_{табл}$ визначають для рівня значущості а = 0,05 і числа ступенів свободи $df = k - 3$ за додаток В.

Таблиця 6 - Розрахунок теоретичних частот кривої нормального розподілу

$Y_{\min}$	798	2086,714	3375,428	4664,142	5952,857	7241,571	8530,286
$Y_{\max}$	2086,714	3375,428	4664,142	5952,857	7241,571	8530,285	9819
$\frac{Y_{k\max} - Y_{k\min}}{2}$	644,4	644,4	644,4	644,4	644,4	644,4	644,4
f_k	3	4	5	9	8	4	3
f_k/n	0,083	0,111	0,139	0,250	0,222	0,111	0,083
$z_{K\min} = \frac{Y_{K\min} - \bar{Y}}{\sigma_y}$	-2,07	-1,49	-0,91	-0,34	0,24	0,81	1,39
$z_{K\max} = \frac{Y_{K\max} - \bar{Y}}{\sigma_y}$	-1,49	-0,91	-0,34	0,24	0,81	1,39	1,96
$\Phi(z_{\kappa\min})$	-0,48	-0,43	-0,32	-0,13	0,10	0,30	0,42
$\Phi(z_{\kappa\max})$	-0,43	-0,32	-0,13	0,10	0,30	0,42	0,48
$P_K = \Phi(z_{k\min}) - \Phi(z_{k\max})$	0,05	0,11	0,19	0,23	0,20	0,12	0,05
nP_k	1,80	3,96	6,86	8,12	7,11	4,42	1,98
$nP_r - f_r$	-1,20	-0,04	1,86	-0,88	-0,89	0,42	-1,02
$\frac{(nP_r - f_r)^2}{nP_k}$	0,80	0,0004	0,50	0,10	0,11	0,04	0,53

У нашому випадку $\chi^2_{табл} = 9{,}49$ для k = 7 – 3 і $\alpha = 0{,}05$.
Отже $\chi^2_{розр} \le \chi^2_{табл}$ тобто 1,73 < 9,49

На підставі отриманих даних можна сказати, що гіпотеза близькості емпіричного і теоретичного розподілу приймається.

Оскільки $\chi^2_{розр} \le \chi^2_{табл}$, то з імовірністю 95% можна стверджувати, що в основі емпіричного розподілу підприємств за величиною валового доходу лежить закон нормального розподілу, а розбіжності між теоретичними й емпіричними частотами пояснюються випадковими факторами.

3. Парний кореляційно-регресійний аналіз залежностей

3.1. Кореляційний аналіз парних зв'язків $Y = \varphi(X_j)$

Кореляційний аналіз проводиться з метою виявлення наявності зв'язку між результативною і факторною змінними й оцінки його сили й істотності. Якщо факторних змінних декілька (у загальному випадку m), то проводять аналіз залежності результативної змінної Y від кожної факторної змінної

$$Y = \varphi(X_j),\ j = 1,2,3\ldots,m$$

Таким чином, у курсовій роботі необхідно провести кореляційний аналіз залежності валового доходу підприємства від середньорічної вартості основних фондів, $Y=\varphi(X_1)$ середньоспискової чисельності працюючих $Y=\varphi(X_2)$, фондовіддачі $Y=\varphi(X_3)$, фондоозброєності $Y=\varphi(X_4)$, і від продуктивності праці $Y=\varphi(X_5)$.

Для виявлення наявності залежності однієї змінної від іншої необхідно побудувати кореляційне поле (за допомогою пакета "Статистика" або вручну) і розрахувати коефіцієнт лінійної кореляції r_{YX_j}. У курсовій роботі рекомендується використовувати таку розрахункову формулу:

$$r_{YX_j} = \frac{\overline{YX_j} - \overline{Y}\cdot\overline{X}_j}{\sqrt{\left(\overline{Y^2} - \overline{Y}^2\right)\left(\overline{X_j^2} - \overline{X}_j^2\right)}};$$

Розрахунок середніх для добутків і середніх для квадратів значень досліджуваних змінних наводиться в табл. 7.

Таблиця 7 - Розрахунок середніх значень квадратів і добутків змінних

№ п/п	Y^2	X_1^2	X_2^2	X_3^2	$Y\cdot X_1$	$Y\cdot X_2$	$Y\cdot X_3$
1	636804	498436	91809	5,429	563388	241794	1859,366337
2	2250000	1052676	220900	4,765	1539000	705000	3274,468085
3	3161284	2383936	448900	5,311	2745232	1191260	4097,361194
4	4605316	2576025	524176	4,914	3444330	1553704	4757,361878
5	5958481	3988009	577600	6,904	4874677	1855160	6414,048684
6	7171684	3968064	682276	5,816	5334576	2212028	6458,324455
7	9078169	4769856	708964	6,728	6580392	2536946	7815,192399
8	13905441	5531904	833569	6,636	8770608	3404577	9606,361446
9	14891881	5340721	891136	5,993	8918149	3642896	9447,191737
10	15840400	6927424	984064	7,040	10475360	3948160	10559,83871
11	16630084	8479744	1179396	7,190	11875136	4428708	10934,7477
12	18671041	7529536	1119364	6,727	11856824	4571618	11206,82798
13	23164969	9320809	1270129	7,338	14694089	5424251	13038,23336
14	23804641	11730625	1555009	7,544	16710575	6084113	13400,62149
15	23951236	9909904	1442401	6,870	15406312	5877694	12827,90341
16	25371369	10413529	1394761	7,466	16254399	5948697	13763,25064
17	27363361	11916304	1580049	7,542	18057412	6575367	14365,4829
18	28622500	13682601	1795600	7,620	19789650	7169000	14768,39552
19	32752729	12103441	1827904	6,621	19910317	7737496	14726,56583
20	33767721	13816089	1825201	7,570	21599487	7850661	15987,7772
21	35343025	16345849	2085136	7,839	24035635	8584580	16645,17659
22	36966400	15952036	2050624	7,779	24283520	8706560	16957,76536
23	38564100	16654561	2202256	7,563	25343010	9215640	17077,5

24	39715204	17280649	2274064	7,599	26197414	9503416	17372,29045
25	45010681	19465744	2446096	7,958	29600108	10492876	18925,90026
26	45900625	19695844	2509056	7,850	30067450	10731600	18981,97601
27	47803396	21068100	2524921	8,344	31735260	10986346	19971,84393
28	50893956	22297284	2775556	8,033	33686748	11885244	20220,13685
29	53304601	23030401	2934369	7,849	35037499	12506613	20453,88149
30	60140025	25280784	2436721	10,375	38992140	12105555	24978,94939
31	63361600	27144100	3411409	7,957	41471600	14702120	22453,49215
32	64850809	28440889	2825761	10,065	42946649	13537093	25548,27424
33	68757264	29474041	3265249	9,027	45017268	14983644	24912,71057
34	76983076	32307856	3783025	8,540	49871416	17065430	25640,83085
35	83137924	33327529	4359744	7,644	52638214	19038384	25209,87261
36	96412761	52838361	6310144	8,374	71374311	24665328	28413,34037
Сума	1238744558	546543661	69147339	264,820	821698155	291669559	543073,262
Серед	34409571,06	15181768,36	1920759,417	7,356	22824948,7	8101932,194	15085,368

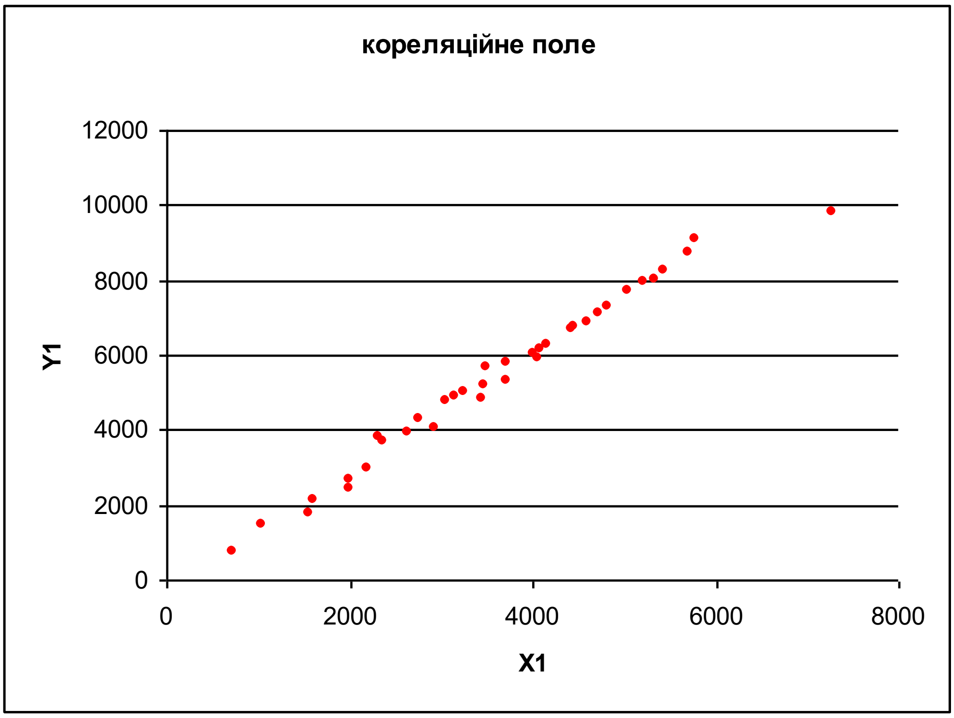

Рисунок 3.1 Графік кореляційного поля між Y та X_1

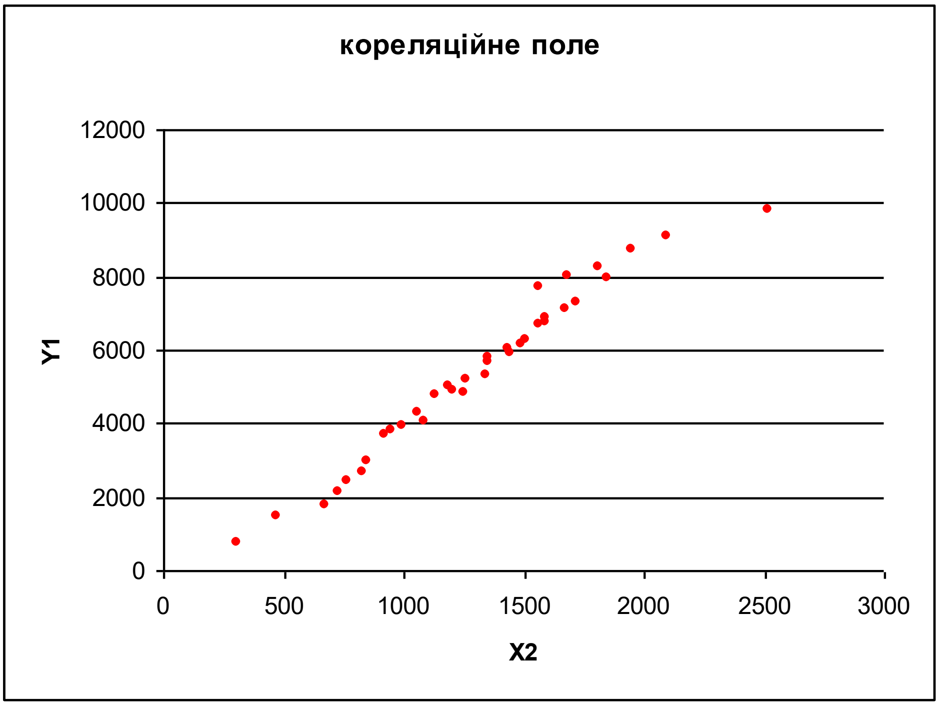

Рисунок 3.2 Графік кореляційного поля між Y та X_2

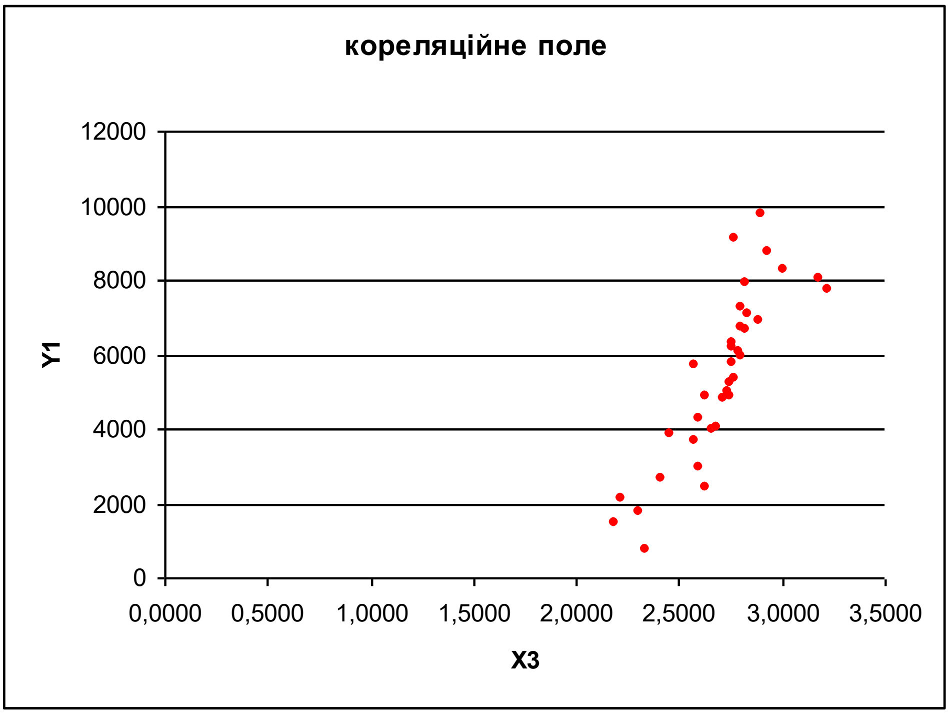

Рисунок 3.3 Графік кореляційного поля між Y та X_3

Лінійний коефіцієнт кореляції може приймати значення від -1 до +1. Якщо r_{YX_j} дорівнює нулю, це означає відсутність лінійної залежності між X_j і Y.

Рівність коефіцієнту лінійної кореляції одиниці ($r_{YX_j} = \pm1$) свідчить про наявність прямої чи зворотної функціональної залежності між X_j і Y. Всяке проміжне значення від 0 до 1 характеризує ступінь наближення кореляційного зв'язку між X_1 і Y до функціонального. Для якісної оцінки тісноти зв'язку між ознаками використовують співвідношення Чеддока

п	0	0 – 0,2	0,2 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9–0,99	1
Сила зв‘язку	Відсутня	Дуже слабка	Слабка	Помірна	Помітна	Тісна	Дуже тісна	Функціональна

Використовуючи дані таблиці 7 розрахуємо коефіцієнти кореляції для відповідних залежностей.

Залежність	$Y=\varphi(X_1)$	$Y=\varphi(X_2)$	$Y=\varphi(X_3)$
r_{YX_j}	0,9908	0,9842	0,8529

Використовуючи співвідношення Чеддока можна сказати, що сила зв'язку між Y та X_1 та X_2 дуже тісна, сила зв'язку між Y та X_3 тісна.

Для того, щоб підтвердити або відкинути реальність вимірюваного за допомогою коефіцієнта кореляції зв'язку між змінними Y і X_j необхідно, використовуючи t-критерій Стьюдента, перевірити значущість самого r_{YX_j}. Для цього визначається розрахункове значення критерію і зіставляється з $t_{табл}$, що визначається за додатком Г для рівня значущості a = 0,05 і числа ступенів свободи v= n – 2.

$$t_{розр}=|r|\cdot\sqrt{\frac{n-2}{1-r^2}};$$

Розрахунки зведемо в таблицю

Залежність	$Y=\varphi(X_1)$	$Y=\varphi(X_2)$	$Y=\varphi(X_3)$
r_{YX_j}	0,9908	0,9842	0,8529
$t_{розр}$	42,595	32,458	9,528

$t_{табл}=2{,}0317;$

Оскільки $t_{розр}>t_{табл}$, то лінійний коефіцієнт кореляції вважається значущим, а зв'язок між X і Y - істотним.

3.2. Регресійнñ аналіз парного зв‘язку Y = ф(Xy)

У даному розділі курсової роботи за результатами кореляційного аналізу слід відібрати фактор X_j що має найвищу кореляцію з результативним показником Y (тобто пару змінних X_j і Y, що мають максимальне значення лінійного коефіцієнта кореляції). Для цієї пари залежних змінних повинні бути представлені найважливіші результати регресійного аналізу. Зокрема повинні бути отримані відповіді на такі питання:

1) яка форма зв'язку (лінійна чи нелінійна) має місце між Y і досліджуваною факторною змінною X_j

2) яке рівняння регресії щонайкраще описує залежність між Y від X_j

3) чи є це рівняння статистично значущим?

3.2.1. Вибір рівняння регресії між двома ознаками

Оскільки найвищу кореляцію з результативним показником Y має фактор X_1, то будемо робити парний аналіз між валовим доходом та середньорічної вартості основних фондів.

Для вибору форми зв'язку застосуємо раніше побудований графік із зображенням кореляційного поля (графік залежності змінних Y, обраної X_1).

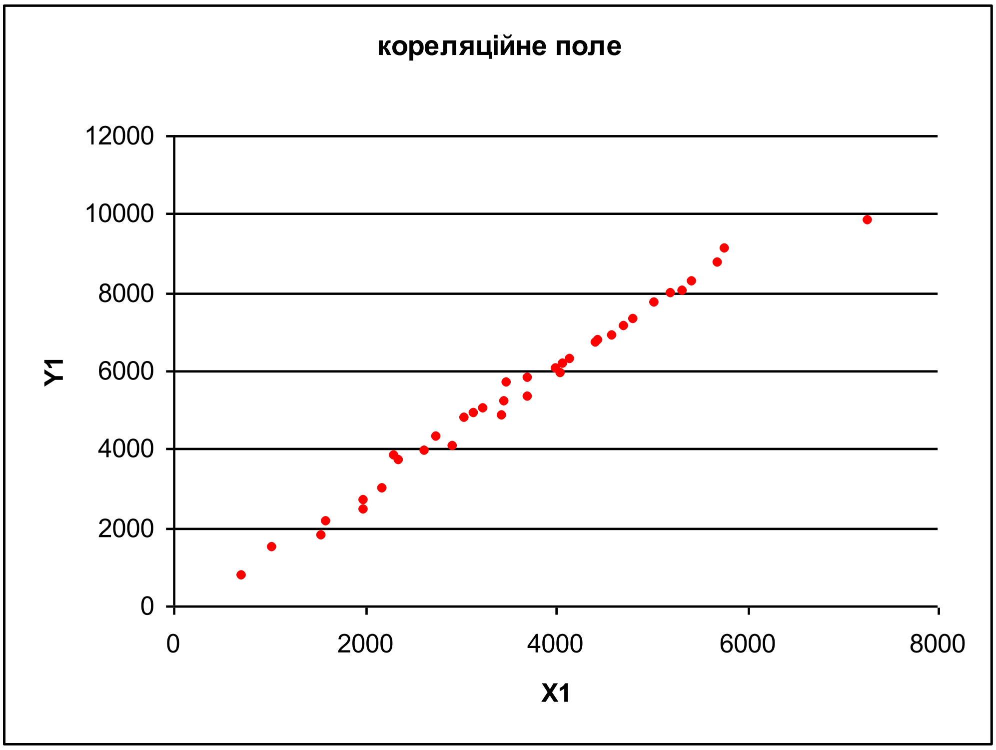

Рисунок 3.6 графік кореляційного поля між Y та X_1

За його виглядом визначим, для апроксимації емпіричної лінії регресії доцільно використовувати рівняння парної регресії у вигляді лінії.

У випадку, коли між змінними X і Y передбачається пряма чи зворотна лінійна залежність, обирають, рівняння вигляду

$$\hat{Y}_i = a + bX_{ij};$$

де $\hat{Y}_i$ – теоретичне значення результативної змінної, обчислене за рівнянням регресії за умови, що і-тий об'єкт має значення факторної змінної, рівне $X_{j.}$

a, b – параметри рівняння;

X_j – значення j-n факторної змінної і-тому спостереженні.

Далі розрахуємо невідомі значення параметрів a і b за даними вибірки. Значення параметра b можна розрахувати за кожною з нижче приведених формул, використовуючи дані табл. 7

$$b = \frac{\overline{Y_i X_{ij}} - \overline{Y}_i \cdot \overline{X}_{ij}}{\overline{X_{ij}^2} - \left(\overline{X_{ij}}\right)^2};$$

Для розрахунку параметра а краще використовувати формулу

$$a = \overline{Y} - b \cdot \overline{X};$$

Використовуючи дані, розраховані раніше маємо

$b = 1{,}527.$

Після знаходження рівняння регресії необхідно розрахувати теоретичне кореляційне відношення η_{TCOP}: порівняти його з розрахованим раніше лінійним коефіцієнтом кореляції r і підтвердити або спростувати гіпотезу про обрану ферму зв'язку. Кореляційне відношення - це універсальний вимірник тісноти зв'язку, застосований до усіх випадків кореляційної залежності незалежно від форми цього зв'язку. Факт збігів або розбіжностей значень теоретичного кореляційного відношення і лінійного коефіцієнта кореляції використовують для підтвердження обраної форми зв'язку. Оскільки за допомогою теоретичного кореляційного відношення виміряється тіснота зв'язку будь-якої форми, а за допомогою лінійного коефіцієнта кореляції - тільки прямолінійної, то збіг η_{TCOP} і r свідчить про наявність прямолінійного зв'язку, розбіжність — про криволінійну. Якщо різниця квадратів η_{TCOP} і r не перевищує 0,1 (тобто $\left|\eta_{TCOP}^2 - r^2\right| \le 0{,}1$ то гіпотезу про прямолінійну форму зв'язку можна вважати підтвердженою.

Визначити теоретичне кореляційне відношення рекомендується за допомогою табл. 8.

Таблиця 8. Розрахункова таблиця для визначення теоретичного кореляційного відношення і перевірки адекватності рівняння регресії і його параметрів

	X_{i1}	Y_i	$\hat{Y}_i$	$Y_i-\bar{Y}$	$\left(Y_i-\bar{Y}\right)^2$	$\hat{Y}_i-\bar{Y}$	$\left(\hat{Y}_i-\bar{Y}\right)^2$	$Y_i-\hat{Y}_i$	$\left(Y_i-\hat{Y}_i\right)^2$
1	706	798	978,41416	-4624,22	21383431,16	-4443,81	19747430	-180,41	32549,27
2	1026	1500	1467,1223	-3922,22	15383827,16	-3955,10	15642815	32,88	1080,94
3	1544	1778	2258,2186	-3644,22	13280355,60	-3164,00	10010918	-480,22	230609,97
4	1605	2146	2351,3786	-3276,22	10733632,05	-3070,84	9430080	-205,38	42180,39
5	1997	2441	2950,0461	-2981,22	8887685,94	-2472,18	6111654,4	-509,05	259127,99
6	1992	2678	2942,41	2678,00	7171684,00	-2479,81	6149468,1	-264,41	69912,70
7	2184	3013	3235,6349	-2409,22	5804351,72	-2186,59	4781163,7	-222,63	49566,34
8	2352	3729	3492,2067	-1693,22	2867001,49	-1930,02	3724959,6	236,79	56071,03
9	2311	3859	3429,5910	-1563,22	2443663,72	-1992,63	3970579	429,41	184392,05
10	2632	3980	3919,8264	-1442,22	2080004,94	-1502,40	2257193,1	60,17	3620,86
11	2912	4078	4347,446	-1344,22	1806933,38	-1074,78	1155143,7	-269,45	72601,18
12	2744	4321	4090,8742	-1101,22	1212690,38	-1331,35	1772487,3	230,13	52957,85
13	3053	4813	4562,7831	4813,00	23164969,00	-859,44	738635,60	250,22	62608,50
14	3425	4879	5130,9063	-543,22	295090,38	-291,32	84864,94	-251,91	63456,80
15	3148	4894	4707,8683	4894,00	23951236,00	-714,35	510301,47	186,13	34645,00
16	3227	5037	4828,5181	-385,22	148396,16	-593,70	352484,51	208,48	43464,68
17	3452	5231	5172,141	-191,22	36565,94	-250,08	62540,57	58,86	3464,37
18	3699	5350	5549,3627	5350,00	28622500,00	127,14	16164,70	-199,36	39745,49
19	3479	5723	5213,3758	300,78	90467,27	-208,85	43616,81	509,62	259716,78
20	3717	5811	5576,8525	388,78	151148,16	154,63	23910,54	234,15	54825,03
21	4043	5945	6074,7239	522,78	273296,60	652,50	425758,55	-129,72	16828,31
22	3994	6080	5999,8905	657,78	432671,60	577,67	333700,69	80,11	6417,52
23	4081	6210	6132,7580	787,78	620593,83	710,54	504861,20	77,24	5966,31
24	4157	6302	6248,8262	879,78	774008,94	826,60	683274,24	53,17	2827,45
25	4412	6709	6638,2655	1286,78	1655797,05	1216,04	1478761,4	70,73	5003,36
26	4438	6775	6677,9731	1352,78	1830007,72	1255,75	1576910,3	97,03	9414,22
27	4590	6914	6910,1094	1491,78	2225400,94	1487,89	2213808,5	3,89	15,14
28	4722	7134	7111,7016	1711,78	2930183,16	1689,48	2854340,6	22,30	497,22
29	4799	7301	7229,297	1878,78	3529805,94	1807,07	3265519	71,70	5141,32
30	5028	7755	7579,0287	2332,78	5441852,16	2156,81	4651814,6	175,97	30965,86
31	5210	7960	7856,9815	2537,78	6440316,05	2434,76	5928053	103,02	10612,80
32	5333	8053	8044,8287	2630,78	6920991,72	2622,61	6878065	8,17	66,77
33	5429	8292	8191,4412	2869,78	8235624,49	2769,22	7668573	100,56	10112,07
34	5684	8774	8580,8805	3351,78	11234414,27	3158,66	9977122,3	193,12	37295,13
35	5773	9118	8716,8024	3695,78	13658773,38	3294,58	10854259	401,20	160959,44
36	7269	9819	11001,513	4396,78	19331654,83	5579,29	31128487	-1182,51	1398337,3
$\sum$	130167	195200	195200	21688,88	255051027,1	0	177009722	0	3317057,4

Використовуючи дані табл. 8, значення теоретичного кореляційного відношення можна розрахувати за однією з нижченаведених формул

$$\eta_{TCOP} = \sqrt{\frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}}$$

$$\eta_{TEOP} = \sqrt{1 - \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}}$$

Отже $\eta_{TCOP} = 0{,}83307676\ 6$

Далі за даними табл. 8 в одних координатних осях побудувати емпіричну та теоретичну лінії регресії, тобто графіки залежності Y_i, від X_{ij} та $\hat{Y}_i$, від X_{ij}.

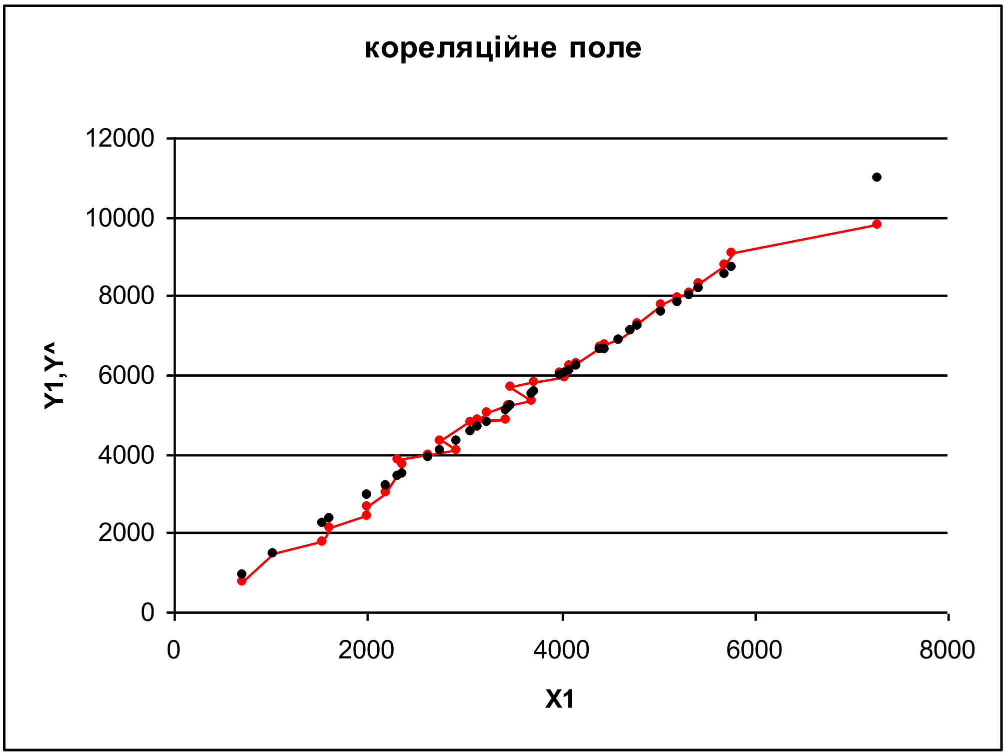

Рисунок 3.2 - Графіки залежності Y_i, від X_{ij} та $\hat{Y}_i$, від X_{ij}.

3.2.2. Оцінка істотності параметрів регресії і рівняння зв'язку

Розраховані для обмеженого числа спостережень параметри a і b рівняння регресії не єдиноможливі, не однозначні, оскільки являють собою лише оцінку реальних параметрів зв'язку в генеральній сукупності. Тому, знайшовши параметри рівняння регресії, здійснюють перевірку їхньої значущості (істотності) і з заданою імовірністю визначають межі, у яких ці параметри можуть знаходитися. Для цього виконують такі дії:

а) використовуючи дані табл. 8, знаходять залишкову дисперсію

$$\sigma^2_{ocm} = \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{n}$$

$$\sigma^2_{ocm} = 92140{,}486$$

б) обчислюють факторну дисперсію, використовуючи розрахунки табл. 5

$$\sigma^2_{Xj} = 2108120{,}29\ 9$$

в) розраховують середні помилки параметрів регресії

$$\mu_a = \sqrt{\frac{\sigma^2_{ocm}}{n-2}}$$

$$\mu_a = 52{,}057$$

$$\mu_b = \sqrt{\frac{\sigma^2_{ocm}}{\sigma^2_{xj}(n-2)}}$$

$$\mu_b = 0{,}03585$$

г) визначають фактичні значення t-критерію Стьюдента для параметрів a і b

$$t_a = \frac{a}{\mu_a}$$

$$t_a = -1{,}92$$

$$t_b = \frac{b}{\mu_b}$$

$$t_b = 42{,}59$$

д) за додатком Г знаходять критичне (табличне) значення t-критерію Стьюдента для числа ступені свободи ν = n-2 і рівня значущості a=0.05 і порівнюють його з фактичними значеннями t-критерію для параметрів a і b.

$$t_{Табл} = 2{,}0317$$

Яісщо $t_{розр} > t_{Табл}$, то параметр вважається значущим, тобто з імовірністю 95% гіпотеза про те, що кожний з цих параметрів насправді дорівнює нулю і лише в силу випадкових обставин виявився рівний величині, що перевіряється, відхиляється;

У нашому випадку 1,92 > $t_{Табл}$, та 42,59 > $t_{Табл}$, то параметри вважаються значущими, тобто з імовірністю 95% гіпотеза про те, що кожний з цих параметрів насправді дорівнює нулю і лише в силу випадкових обставин виявився рівний величині, що перевіряється, відхиляється.

е) будують довірчі інтервали для оцінки істинних значень параметрів a і b , що можуть мати місце в генеральній сукупності

$$a - t_{Табл} \cdot \mu_a \le a \le a + t_{Табл} \cdot \mu_a$$
$$b - t_{Табл} \cdot \mu_b \le b \le b + t_{Табл} \cdot \mu_b$$
$$-205,56 \le a \le 5,97$$
$$1,45 \le b \le 1,60$$

Поряд з перевіркою окремих параметрів необхідно здійснити перевірку значущості рівняння регресії в цілому, тобто перевірку адекватності моделі. Ця задача розв'язується за допомогою F-критерію Фішера

$$F = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2} \cdot \frac{n-m}{m-1};$$
$$F = 23,5965745\ 4$$

де m- число параметрів у рівнянні регресії (для моделі парної регресіїm = 2).

Це значення порівнюють із критичним значенням, яке знаходять за табл. додаток Е для обраного рівня значущості, рівного 0,05, на перетинанні стовпця, що відповідає числу ступенів свободи v_1 = m-1, і рядка, що відповідає числу ступенів свободи v_2 = n-m.

Маємо v_1=1, v_2=34, $F_{Табл}$=4,14

Оскільки розрахункове значення більше критичного $F_{розр} > F_{Табл}$, то модель вважається значущою на обраному рівні довірчої імовірності.

Висновки

В процесі виконання комплексу лабораторних досліджень було проведено групування даних, розрахунок описової статистики і перевірка однорідності вибіркової сукупності.

Групування було проведено з використанням як рівних інтервалів так і з використанням нерівновеликих інтервалів.

Групування з використанням як рівних інтервалів має переваги.

Потім був проведений розрахунок узагальнюючих характеристик і перевірка однорідності вибіркової сукупності, який показав, що вибіркова сукупність однорідна.

Проведена оцінка достатності обсягу вибірки, яка показала, що 10 процентів для заданої значимості недостатньо.

Проведено аналіз закономірностей розподілу досліджуваних показників

Проведено парний кореляційно-регресійний аналіз залежностей парних зв'язків $Y = \varphi(X_j)$.

Проведено регресійний аналіз парного зв'язку $Y = \varphi(X_1)$, як самого щільного зв'язку.

Проведена оцінка істотності параметрів регресії і рівняння зв'язку, яка показала що коефіцієнти регресії значущі, тобто модель вважається значущою на обраному рівні довірчої імовірності.

СПИСОК ЛІТЕРАТУРИ

1. Ірина А.М., Пальян З.О. Теорія статистики: Практикум. Київ: Знання, 1997.-325 с

2. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. М.: ИНФРА-М, 1998. -416с.

3. Статистика: Підручник / А.В. Головач, А.М. Єріна, О .'В. Козирєв та ін. К.: Вища школа, 1993. - 623 с.

4. Статистика: Підручник / С.С. Герасименко, А.В. Головач, А.М. Єріна та ін. -К.:КНЕУ, 1998. -468 с.

5. Теория статистики: Учебник / Под ред. проф. Г.Л. Громыїко. - М.: ИНФРА-М, 2000.-144с.