

ЗЛОЧИННІСТЬ У ЗАХІДНИХ ОБЛАСТЯХ УКРАЇНИ: ОДНОФАКТОРНА МОДЕЛЬ НА ОСНОВІ РІВНЯНЬ РЕГРЕСІЇ

Хаврук В. О.,
*асистент кафедри технічної експлуатації
автомобілів та автосервісу
Національний транспортний університет*

У статті аналізується стан злочинності в областях України та в м. Києві за 2015–2017 рр. з використанням статистичних даних, розрахованих коефіцієнтів злочинності та злочинної активності населення.

Представлена методика отримання однофакторної моделі, яка описує залежність злочинності від чисельності населення в західних областях України, в середовищі Microsoft Excel.

З'ясовано, що найбільш «якісною» однофакторною моделлю є поліноміальне рівняння регресії 4-го та (або) 5-го степенів. Застосування моделі є необхідним у практичній діяльності правоохоронних органів, оскільки вона виявляє залежність кількості зареєстрованих злочинів від чисельності населення.

В статье анализируется состояние преступности в областях Украины и в г. Киеве за 2015–2017 гг. с использованием статистических данных, рассчитанных коэффициентов преступности и преступной активности населения.

Представлена методика получения однофакторной модели, которая описывает зависимость преступности от численности населения в западных областях Украины, в среде Microsoft Excel.

Выяснено, что наиболее «качественной» однофакторной моделью является полиномиальное уравнение регрессии 4-й и (или) 5-й степени. Применение модели является необходимым в практической деятельности правоохранительных органов, поскольку она обнаруживает зависимость количества зарегистрированных преступлений от численности населения.

In the article the state of criminality is analysed in the oblasts of Ukraine and in Kyiv for 2015–2017 with the use of statistical data, expected coefficients of criminality and criminal activity of population.

Presented methodology of receipt of onefactor model that describes dependence of criminality on the quantity of population in the western areas of Ukraine, in the environment of Microsoft Excel.

It was found that the most «qualitative» single-factor model is the polynomial regression equation of the 4th and (or) 5th degree. Application of model is a necessity in practical activity of law enforcement authorities, as she finds out dependence of amount of the registered crimes on the quantity of population.

Ключові слова: зв'язок, злочинність, індекс кореляції, коефіцієнт детермінації, коефіцієнт злочинності, критерій Фішера, населення, похибка, однофакторна модель, регресія, рівняння, тенденція.

Постановка проблеми. Дослідження злочинності ґрунтується, головним чином, на статистичних даних, отриманих з офіційних джерел від правоохоронних органів, і переслідує дві мети: розвиває кримінологічну теорію та вирішує практичні завдання боротьби зі злочинністю. Таким чином, статистичний метод кримінологічних досліджень набуває особливо важливого значення у вивченні злочинності; у з'ясуванні причин, факторів, що впливають на злочинність і т.д. Кількісні результати застосування статистичного методу в кримінології дають математичні методи, зокрема, вивчення злочинності можливо здійснювати шляхом моделювання кримінологічних процесів.

Проблематика пошуку «якісних» моделей злочинності є актуальною. З розвитком комп'ютеризованих програмних продуктів, з однієї сторони, такий пошук прискорюється, а з іншої – потребує від дослідника досить ґрунтовних знань комп'ютерних програм, розділів математики та проведення аналізу причин та умов, що спричиняють злочини. Однофакторна модель злочинності є найпростішим прикладом застосування математичних методів у кримінології, методика її побудови і буде розглянута нижче.

Математичне моделювання в кримінології знайшло втілення в дисертаційних роботах С.М. Корецького [1], К.А. Утарова [2], питання математичного моделювання злочинності досліджує М.О. Ларченко, а також воно представлено в окремих наукових статтях [3]. Хоча В.В. Лунєєв і вказує

на перспективність створення задовільної моделі злочинності [4, с. 232], але на даний час моделювання кримінологічних процесів не знаходить належної уваги серед кримінологів.

Виходячи з вищесказаного, проведення досліджень і з'ясування методики математичного моделювання злочинності є доцільними для розвитку кримінологічної теорії.

Метою статті є аналіз впливу чисельності населення на злочинність та побудова однофакторної моделі, яка описує злочинність у західних областях України з використанням статистичних даних за 2015–2017 рр. на основі рівнянь регресії.

Виклад основного матеріалу. Для вибору початкових даних для побудови однофакторної моделі злочинності проведемо аналіз статистичних даних про чисельність наявного населення і кількість зареєстрованої злочинності в областях України і в м. Києві (табл. 1; рис. 1–3).

Лише зміна населення однозначно не тягне до зміни числа засуджених. Окрім чисельності населення, на чисельність засуджених впливають такі чинники, як склад населення, його статеві і вікова структура, інтенсивність боротьби зі злочинними проявами тощо, чим визначається неоднозначний характер злочинності. З іншої сторони, злочинність є суспільним явищем – злочини вчиняються конкретними індивідуумами суспільства, тобто джерелом злочинності є населення, яке об'єктивно існує в межах правового поля певної держави. На перший погляд, може здатися, що у зв'язку з негативною динамікою – зменшенням чисельності населення України – дослідження залежності злочинності від зміни населення є недоцільним: населення зменшується, значить, зменшується злочинність, і, як наслідок, має покращуватися криміногенна ситуація, але подібні висновки є помилковими, і в межах побудови однофакторної моделі буде проведено порівняння коефіцієнтів злочинності протягом 2015–2017 рр.

На перший погляд, у наведених табл. 1 і рис. 1, 2 немає ніякої узгодженості. Але, як відомо, стан злочинності характеризують, зокрема, кількісні показники, які визначаються кількістю злочинів, зареєстрованих на певній території за певний час, і кількістю виявлених осіб, які вчинили злочини на певній території за певний час. Ці показники характеризують два основних аспекти злочинності – правопорушення їхніх «авторів», тобто осіб, які їх вчинили. Кількісне вираження цих показників не збігається, оскільки не всі злочини розкриваються і, відповідно, не всі особи, які їх вчинили, виявляються. Ці кількісні показники є абсолютними й іноді їх ще називають показниками рівня злочинності. Проте останні доцільно використовувати лише в аналізі стану злочинності та її динаміки в межах одного регіону, оскільки їх неможливо порівняти з відповідними показниками в інших регіонах через те, що вони однобічні, тобто враховують географічні й часові межі, проте не враховують чисельності населення регіону. З метою усунення цього недоліку і уможливлення співвіднесення абсолютних показників із чисельністю населення регіону для кількісної характеристики стану злочинності застосовують відносні показники – коефіцієнти (індекси) інтенсивності злочинності (коефіцієнт злочинності) й злочинної активності населення [6, с. 52].

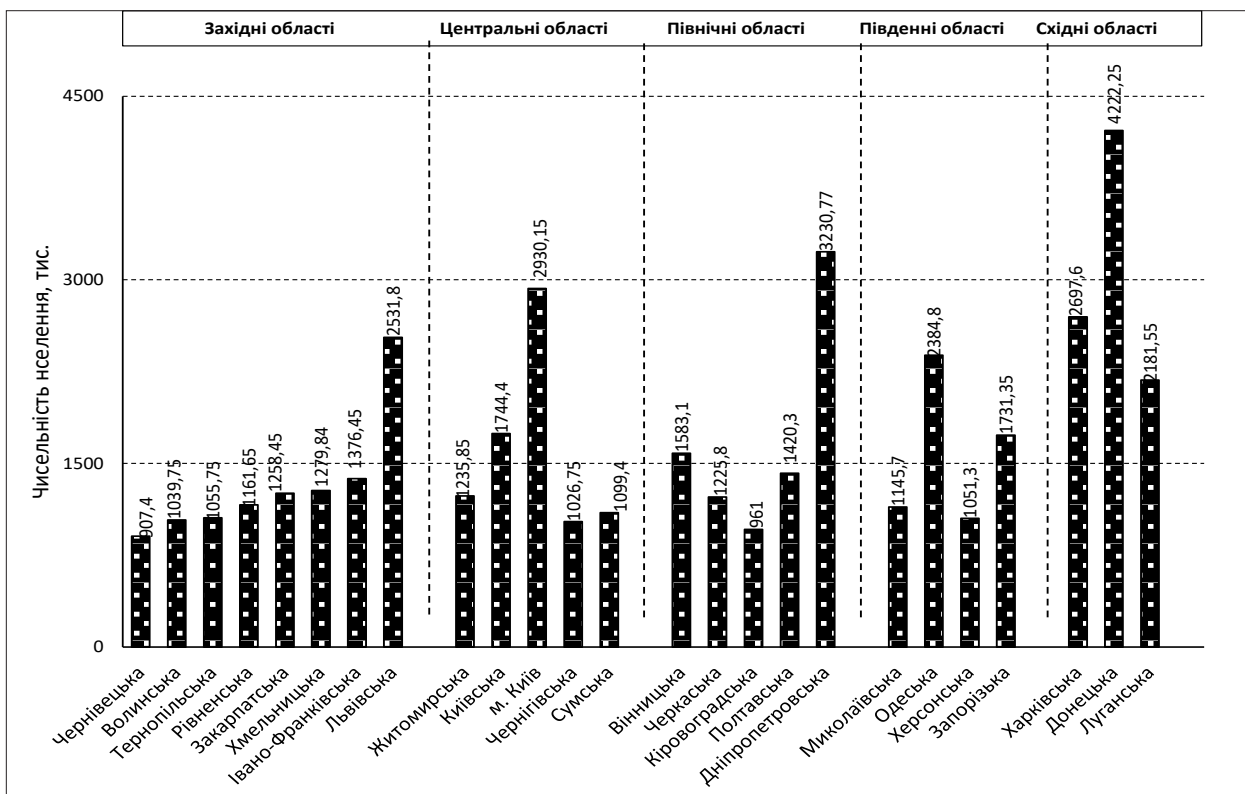


Рис. 1. Чисельність наявного населення в областях України і в м. Києві (2017 р.)

Таблиця 1

Чисельність наявного населення, кількості облікованих кримінальних правопорушень (кількість виявлених злочинів) та виявлених осіб, які вчинили злочини в областях України та в м. Києві за 2017 р. [5]

	Чисельність наявного населення ¹	Кримінальні правопорушення ²		Кількість осіб, які вчинили злочини ²	
		всього виявлено	виявлених на 10 тис. населення	всього виявлено	виявлених на 10 тис. населення
Вінницька	1583100	14325	90,48701914	4865	30,73084455
Волинська	1039750	10834	104,1981245	2465	23,70762202
Дніпропетровська	3230770	44343	137,2521102	11455	35,45594394
Донецька	4222250	260913	61,79406714	7960 ³	18,85250755
Житомирська	1235850	14836	120,0469313	4201	33,99279848
Закарпатська	1258450	11183	88,8632842	3164	25,14203981
Запорізька	1731350	31944	184,5034222	6394	36,9307188
Івано-Франківська	1376450	8244	59,89320353	2475	17,98103818
Київська	1744400	25720	147,443247	5915	33,90850722
Кіровоградська	961003	16711	173,8912366	3558	37,02381782
Луганська	2181550	113633	52,086819	38703	17,7396805
Львівська	2531800	27526	108,721068	5743	22,68346631
Миколаївська	1145700	16798	146,6177883	4133	36,07401589
Одеська	2384800	32576	136,5984569	5529	23,18433412
Полтавська	1420300	22676	159,6564106	5290	37,24565233
Рівненська	1161650	11413	98,24818147	2991	25,74785865
Сумська	1099400	12450	113,2435874	4140	37,65690377
Тернопільська	1055750	7627	72,24248165	2424	22,95998106
Харківська	2697600	34986	129,6930605	8194	30,37514828
Херсонська	1051300	16760	159,4216684	4278	40,69247598
Хмельницька	1279839	11644	90,98022915	3050	23,83113182
Черкаська	1225800	17756	144,8523413	3282	26,77435144
Чернівецька	907400	9213	101,5318492	2109	23,24223055
Чернігівська	1026750	14932	145,4297541	2872	27,97175554
м. Київ	2930150	70662	241,1548897	7357	25,10792963
Всього по Україні	42483162	522613	123,0165024	117714	27,70838983

Примітка: ¹Середня чисельність наявного населення отримана шляхом ділення суми чисельності на 1 січня 2017 р. і на 1 січня 2018 р. на 2; ²За даними обласних прокуратур та прокуратури м. Києва; ³Ураховуючи всі повідомлення про злочини, які вчинені на тимчасово окупованій території та зареєстровані іншими територіальними органами досудового розслідування

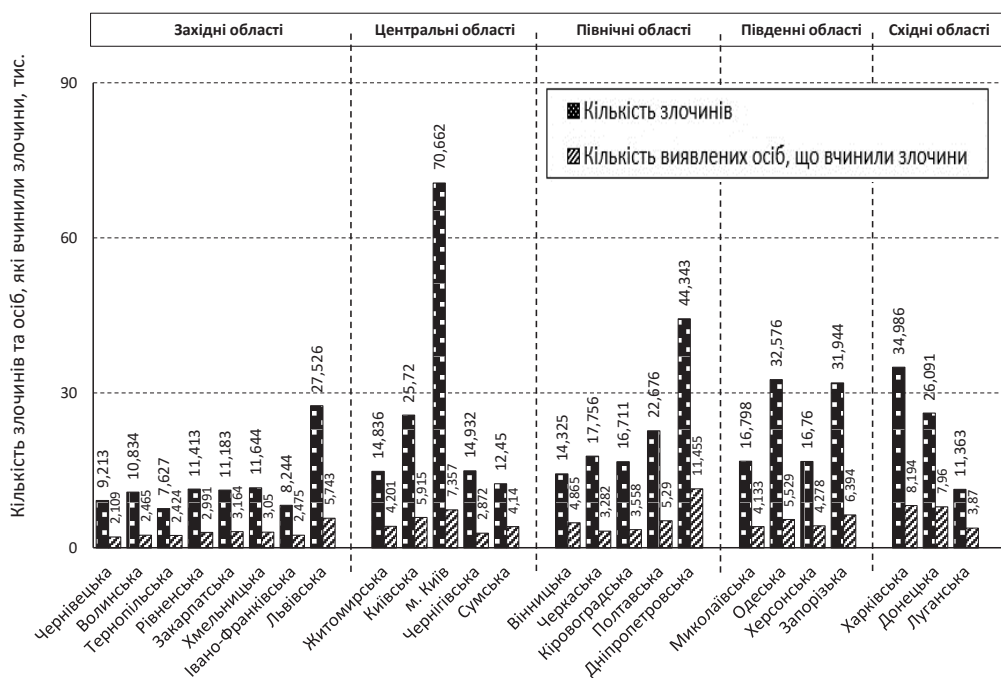


Рис. 2. Кількості облікованих кримінальних правопорушень (кількість виявлених злочинів) та виявлених осіб, які вчинили злочини в областях України і в м. Києві (2017 р.)

Коефіцієнт (індекс) інтенсивності злочинності або коефіцієнт злочинності – це кількість зареєстрованих злочинів на 10 000 або 100 000 населення регіону. Коефіцієнт злочинності розраховують за формулою [6, с. 52; 7, с. 65]:

$$K_i = \frac{3}{H} \cdot 10000 \text{ (або } 100000), \quad (1)$$

де 3 – кількість злочинів; H – чисельність населення регіону. Коефіцієнти злочинності, розраховані за формулою (1), наведені в 4-му стовпці табл. 1 та представлені на рис. 3.

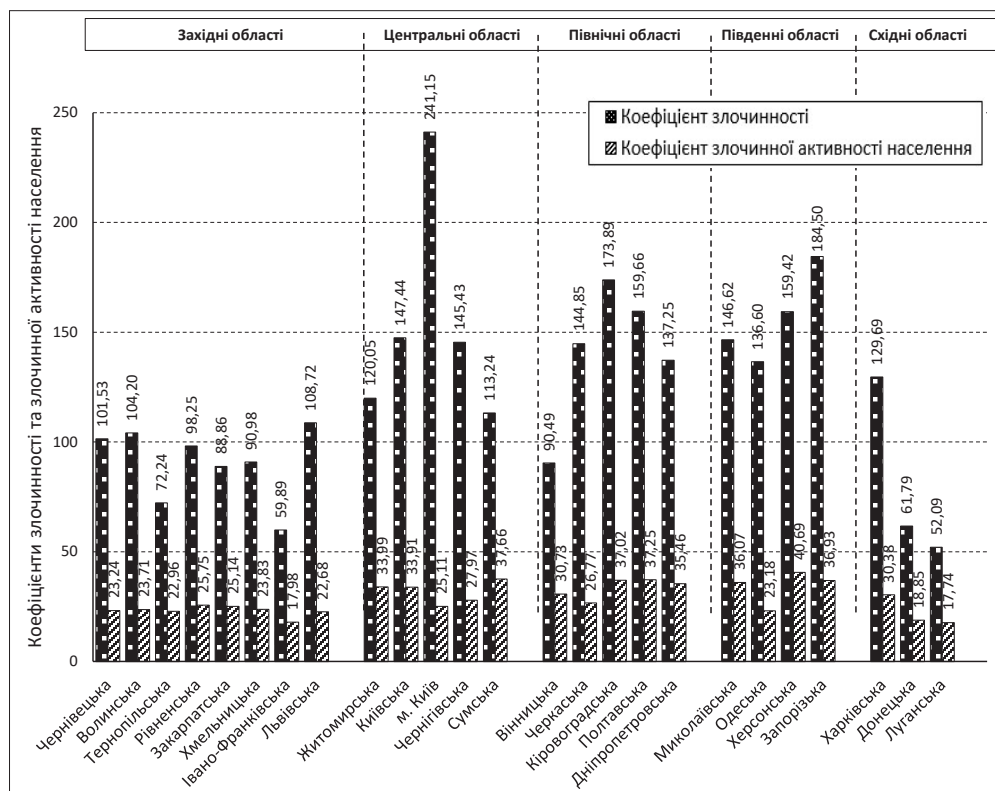


Рис. 3. Коефіцієнти злочинності та злочинної активності населення (на 10 тис.) в областях України і в м. Києві (2017 р.)

Аналізуючи коефіцієнт злочинності (табл. 1; рис. 3), ми бачимо, що він високий в м. Києві (241,1548897), в Запорізькій (184,5034222) та Кіровоградській (173,8912366) областях. Високий коефіцієнт злочинності в м. Києві можна пояснити тим, що місто є столицею України з найбільшою чисельністю населення, об'єктами посягань стають матеріально забезпечені жителі міста, внутрішня міграція в столицю також підсилює рівень злочинності і т.д. Досить високий коефіцієнти злочинності в Запорізькій області (184,5034222), пояснюється тим, що промислові об'єкти та інші сфери діяльності не здатні забезпечити наявне населення роботою і належним доходом, як наслідок – безробіття і втрата відчуття на покращення умов життя, що і спричинює злочинність. Аналогічні висновки можна зробити щодо Кіровоградської області, лише зазначивши, що область не є такою промислово розвинутою, як Запорізька, а тому і джерела доходів для наявного населення є меншими, крім того, область є однією з двох областей (разом із Чернівецькою) з найменшою чисельністю населення, як результат – досить високий коефіцієнт злочинності.

Низькі коефіцієнти злочинності – 72,24 в Тернопільській області, 59,89 – в Івано-Франківській області. Пояснення тому – зайнятість більшої частини населення сільським господарством. Одним зі стримуючих чинників злочинності в цих областях можна назвати сімейні стосунки. Питання безробіття досить актуальне, і частково населення змушене виїжджати на заробітки в країни Європейського Союзу, тому в цих областях спостерігається відтік працездатного населення, і тим самим знижуються криміногенні явища. Досить занижені коефіцієнти злочинності в Луганській та Донецьких областях – 52,09 та 61,79 відповідно. Це можна пояснити тим, що облік злочинності в даних областях ведеться неналежним чином у зв'язку з проведенням операції об'єднаних сил на сході України, також наявна чисельність населення потребує уточнення.

Коефіцієнт злочинної активності населення відображує частоту вчинення кримінальних правопорушень мешканцями регіону, іншими словами, це кількість виявлених осіб, що вчинили злочини, з розрахунку на 10 000 або 100 000 населення регіону, тобто [6, с. 52]:

$$K_a = \frac{O_z}{H} \cdot 10000 \text{ (або } 100000 \text{)}, \quad (2)$$

де O_z – кількість виявлених осіб, що вчинили злочини. Коефіцієнти злочинної активності населення, розраховані за формулою (1), наведені в 6-му стовпці табл. 1 та представлені на рис. 3.

Коефіцієнт злочинної активності населення (табл. 1; рис. 3) високий в Херсонській (40,69247598), Сумській (37,65690377), Кіровоградській (37,02381782), Запорізькій (36,9307188), Миколаївській (36,07401589), Дніпропетровській (35,45594394) областях, в інших областях – не перевищує 30. Але зіставляючи кількість злочинів і кількість осіб, яких виявлено, можна стверджувати про низький рівень розкриття злочинів, а отже, в дійсності (при 100% розкриття злочинів) коефіцієнти злочинної активності населення були вищими – в областях України в межах 120–180, а в м. Києві – в межах 220–235. Отже, розраховані коефіцієнти злочинної активності населення не відображають реальний стан злочинності, а тому для побудови однофакторної моделі доцільно брати до уваги кількість виявлених кримінальних правопорушень.

Чисельність населення є одним із важливих чинників, що впливає на рівень злочинності. Проте прямої і жорсткої залежності рівня злочинності від чисельності населення немає, оскільки окрім чисельності населення на рівень злочинності впливає його соціальна якість: вік, стать, освіта, зайнятість, забезпеченість тощо. Але, незважаючи на це, між чисельністю населення і рівнем злочинності існує значима пряма кореляційна залежність.

Тепер детальніше з використанням математичних методів розглянемо вплив чисельності населення на злочинність. На основі наведеної статистики (табл. 1; рис. 1) можливо стверджувати, що дані по Донецькій, Луганській щодо чисельності населення не відповідають дійсності, як результат – занижені коефіцієнти злочинності 61,79 і 52,09 у Донецькій, Луганській областях відповідно. Досить високі коефіцієнти злочинності в Запорізькій, Кіровоградській і Полтавській областях – 184,5; 173,89 і 159,656 відповідно (табл. 1; рис. 3), що також потребує з'ясування й уточнення. З огляду на дані аспекти, доцільно розглянути певний регіон України, в якому сформувався усталений процес соціального життя. Обираємо західні області (Чернівецька, Закарпатська, Волинська, Тернопільська, Рівненська, Івано-Франківська, Львівська, Хмельницька).

На основі статистичних даних (табл. 2) можна перекоонатися, що протягом 2015–2017 рр. негативна демографічна ситуація (зменшення наявного населення) і надалі спостерігається в усіх західних областях України, причому чітких тенденцій щодо зменшення злочинності не спостерігається. Так, наприклад, в Закарпатській, Тернопільській та Чернівецькій областях злочинність у порівнянні з 2015 р. спочатку зменшилась у 2016 р., а потім знову зросла; у Волинській, Івано-Франківській, Львівській та Рівненській областях злочинність, навпаки, спочатку зросла у 2016 р., а у 2017 р. знизилась, і лише в Хмельницькій області спостерігається тенденція зниження злочинності (12620, 12213, 11644). Зважаючи на відсутність однорідних тенденцій зниження злочинності у зв'язку зі зниженням чисельності

Таблиця 2

Кількість облікованих кримінальних правопорушень і чисельність наявного населення в західних областях України за 2015–2017 рр. [5]

Області	Чисельність населення ¹			Всього виявлено злочинів ²			Виявлених злочинів на 10 тис.		
	2015	2016	2017	2015	2016	2017	2015	2016	2017
Волинська	1042800	1041850	1039750	10722	11947	10834	102,819	114,671	104,198
Закарпатська	1259400	1259000	1258450	11378	9477	11183	90,345	7,5274	88,863
Івано-Франківська	1382500	1381150	1376450	8512	9893	8244	61,570	71,629	59,893
Львівська	2536000	2534100	2531800	33204	36887	27526	130,93	145,563	108,721
Рівненська	1161500	1162250	1161650	11093	12016	11413	95,506	103,386	98,248
Тернопільська	1067750	1062400	1055750	7195	7042	7627	67,385	66,284	72,242
Хмельницька	1297800	1289850	1279839	12620	12213	11644	97,241	94,685	90,980
Чернівецька	909950	909000	907400	9186	8424	9213	100,951	92,673	101,532

Примітка: ¹Середня чисельність наявного населення отримана шляхом ділення суми чисельності на 1 січня поточного року і наступного року на 2; ²За даними обласних прокуратур.

Таблиця 3

Усереднені показники чисельності наявного населення та злочинності в західних областях України за 2015–2017 рр.¹

Області	Середня чисельність населення	Середня кількість злочинів	Виявлених злочинів на 10 тис.
Волинська	1041467	11168	107,2301882
Закарпатська	1258950	10679	84,82730318
Івано-Франківська	1380033	8883	64,36801043
Львівська	2533967	32539	128,4113182
Рівненська	1161800	11507	99,0474551
Тернопільська	1061967	7288	68,62738943
Хмельницька	1289163	12159	94,31701034
Чернівецька	908783	8941	98,38428668

Примітка: ¹Усереднені показники отримані шляхом ділення суми стовпців табл. 2 для чисельності населення та виявлених злочинів на 3, коефіцієнт злочинності (злочинів на 1 тис.) отриманий шляхом ділення 3-го стовпця на 2-й стовпець даної таблиці з наступним множенням на 1000.

населення в західних областях України, для побудови однофакторної моделі доцільно використати усереднені показники чисельності наявного населення та злочинності за період 2015–2017 рр. (табл. 3).

Для отримання моделі парної кореляції необхідно визначити:

1. Характеристики залежності чисельності зареєстрованої злочинності від чисельності населення.
2. Розрахувати коефіцієнт парної кореляції.

Для цього, згідно з усередненими показниками (табл. 3), спочатку побудуємо для західних областей України точкову діаграму залежності кількості злочинів від чисельності населення (рис. 4).

Для вирішення даних завдань застосовується розрахунковий метод, відомий під назвою «метод найменших квадратів» (далі – МНК). Суть цього методу полягає в тому, що до статистичного матеріалу на основі математичних методів необхідно підібрати теоретичну криву розподілу. Ця крива знаходить своє математичне вираження у вигляді деякого рівняння (формули). На математичній мові це рівняння буде функцією, а в математичній статистиці застосовується інший термін – регресія. Необхідно зауважити, що МНК має недолік – надмірну чутливість до окремих «викидів» (outliers) спостережуваної величини, тобто МНК-пряма «сильно» змінює кут нахилу зі збільшенням віддаленості «викиду» від області розміщення основної сукупності величин [8, с. 358; 9, с. 34].

Побудова рівняння регресії зводиться до оцінки її параметрів. Для оцінки параметрів регресії, лінійних за параметрами, використовують метод найменших квадратів (МНК). МНК дозволяє отримати такі оцінки параметрів, за яких сума квадратів відхилень фактичних значень результуючої ознаки у від теоретичних \hat{y}_i мінімальна, тобто [10, с. 42; 11, с. 52]:

$$\sum (y - \hat{y}_i)^2 \rightarrow \min, \quad (3)$$

Для лінійних ($y = a + bx$) і нелінійних рівнянь, які зводяться до лінійних, знаходять коефіцієнти a і b .

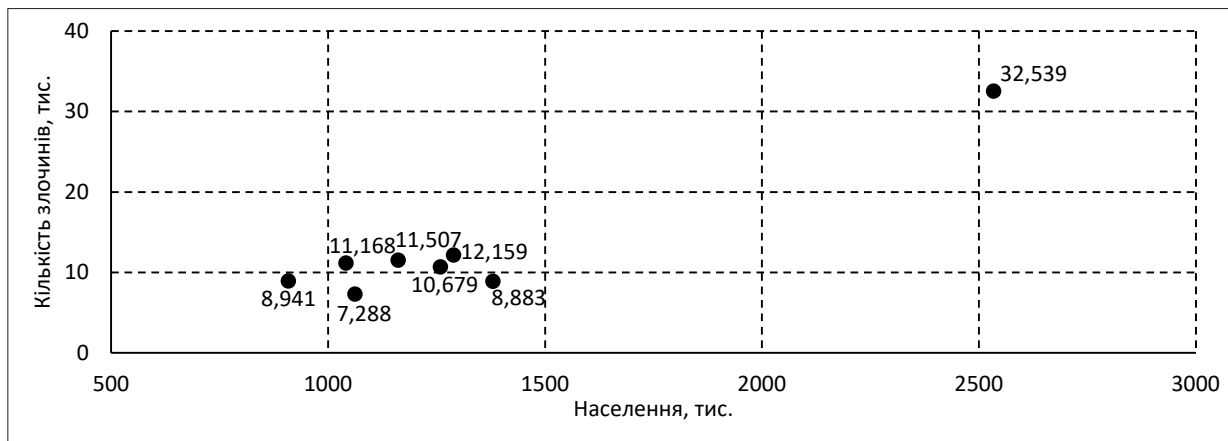


Рис. 4. Залежність злочинності від чисельності населення в західних областях України (2015–2017 рр.)

Тіснота зв'язку досліджуваних явищ оцінюється за допомогою лінійного коефіцієнта парної кореляції r_{xy} , який показує, на яку частину величини свого середнього квадратичного відхилення зміниться в середньому значення результативної ознаки в разі зміни фактора на величину його середньоквадратичного відхилення за наявності фіксованого на постійному рівні значення інших незалежних змінних. Для лінійної регресії ($-1 \leq r_{xy} \leq 1$) [12, с. 6]:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y} \quad (4)$$

та індекс кореляції для p_{xy} для нелінійної регресії ($0 \leq p_{xy} \leq 1$) [11, с. 262; 12, с. 6]:

$$p_{xy} = \sqrt{1 - \frac{\sigma_{\text{зал}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}} \quad (5)$$

Зв'язки між ознаками можуть бути слабкими і сильними (тісними). Оцінка зв'язку між ознаками за лінійним коефіцієнтом кореляції здійснюється за шкалою Чеддока: $0,1 < r_{xy} < 0,3$: слабкий; $0,3 < r_{xy} < 0,5$: помірний; $0,5 < r_{xy} < 0,7$: помітний; $0,7 < r_{xy} < 0,9$: високий; $0,9 < r_{xy} < 1,0$: дуже високий.

Оскільки залежності між ознаками соціальних явищ мають статистичний характер, значенню однієї ознаки відповідає безліч значень іншої ознаки, які варіюють біля середньої цієї ознаки і виражаються його середньою величиною. Це є наслідком множинності причин, що впливають на соціальні явища одночасно і у взаємному зв'язку. Такий неповний зв'язок між двома причинами, що виявляється не в кожному окремому випадку, а лише в разі масового зіставлення, під час порівняння середніх значень ознак, називається кореляційним зв'язком або просто кореляцією.

Оцінку «якості» побудованої моделі дає коефіцієнт (індекс) детермінації, а також середня похибка апроксимації. Середня помилка (похибка) апроксимації – середнє відхилення розрахункових значень від фактичних [11, с. 260; 12, с. 6]:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% \quad (6)$$

Прийнятна межа значень $\bar{A} = 8 \dots 10\%$.

Долю дисперсії, визначену регресією, в загальній дисперсії результативної ознаки у характеризує коефіцієнт (індекс) детермінації [12, с. 6] (R^2):

$$R^2 = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \bar{y})^2} \quad (7)$$

Коефіцієнт детермінації – квадрат коефіцієнта або індексу кореляції. F -тест – оцінювання якості рівняння регресії, полягає в перевірці гіпотези H_0 про статистичну незначущість рівняння регресії і показника тісноти зв'язку. Для цього виконується порівняння фактичного $F_{\text{факт}}$ і критичного (табличного) $F_{\text{табл}}$ значень F -критерію Фішера. $F_{\text{факт}}$ визначається зі співвідношення значень факторної і залишкової дисперсій, розрахованих на один ступінь свободи [12, с. 9; 13, с. 80]:

$$F_{\text{факт}} = \frac{\sum (y_i - \bar{y}_i)^2 / m}{\sum (y_i - \bar{y}_i)^2 / (n - m - 1)} = \frac{R^2 \cdot (n - m - 1)}{(1 - R^2) \cdot m} \quad (8)$$

де n – число одиниць сукупності (обсяг вибірки), m – число параметрів при змінних x .

$F_{\text{табл}}$ – це максимально можливе значення критерію під впливом випадкових чинників за даних ступенях свободи і рівня значущості α . Рівень значущості α – ймовірність відкинути правильну гіпотезу за умови, що вона вірна. Зазвичай α приймається рівним 0,05 або 0,01. Якщо $F_{\text{табл}} < F_{\text{факт}}$, тоді H_0 – гіпотеза про випадкову природу оцінюваних характеристик відхиляється і признається їх статистична значущість і надійність. Якщо $F_{\text{табл}} > F_{\text{факт}}$, тоді гіпотеза H_0 не відхиляється і признається статистична незначущість, ненадійність рівняння регресії.

З метою математичного вираження взаємозв'язку між чисельністю населення і кількістю зареєстрованих злочинів розглянемо три найбільш поширених моделі, які використовуються в соціальних і економічних дослідженнях: $y = a + b \cdot x$ – лінійна функція, $y = a + b^x$ – показникова функція, $y = ax^b$ – степенева функція, де y – кількість зареєстрованих злочинів, x – чисельність населення в західних областях України, b – показник впливу чисельності населення на чисельність злочинних діянь, a – показник сукупного впливу інших факторів.

Для розрахунків рівнянь регресії однофакторної моделі злочинності можна використати відповідні вбудовані статистичні функції в Excel (ЛИНЕЙН, ЛГРФПРИБЛ), скориставшись рекомендаціями [12, с. 23–24]:

1) для лінійної функції $y = a + b \cdot x$, знаходимо: значення коефіцієнта $a = -7,225665128$; значення коефіцієнта $b = 0,0151342$; коефіцієнт детермінації $R^2 = 0,90714937$; F -статистика дорівнює 58,6199171;

2) для показникової (експоненціальної) функції $y = a \cdot b^x$, отримаємо: значення коефіцієнта $a = 3,916089004$; значення коефіцієнта $b = 1,000813132$; коефіцієнт детермінації $R^2 = 0,841080598$; F -статистика дорівнює 31,75498737;

3) для степеневих функції $y = ax^b$ маємо: значення коефіцієнта $a' = -6,755680936$; значення коефіцієнта $b = 1,287962084$; коефіцієнт детермінації $R^2 = 0,784150283$; F -статистика дорівнює 21,79711775; коефіцієнт $a = e^{a'} = e^{-6,755680936} = 0,001164247$.

Таким чином, отримуємо три рівняння, що пов'язують кількість зареєстрованих злочинів із чисельністю населення в західних областях України: лінійна $y = -7,225665128 + 0,0151342 \cdot x$; показникова $y = 3,916089004 \cdot 1,000813132^x$; степенева $y = 0,001164247 \cdot x^{1,287962084}$.

Для опису взаємозв'язку між злочинністю і населенням у даному випадку приймаємо лінійну модель. Коефіцієнт детермінації для лінійної моделі дорівнює $R^2 = 0,90714937$, який вищий за коефіцієнти детермінації показникової моделі ($R^2 = 0,841080598$) та степеневих ($R^2 = 0,784150283$). Оцінимо якість рівняння регресії $y = -7,225665128 + 0,0151342 \cdot x$. Середня відносна похибка апроксимації – середнє відхилення розрахункових значень від фактичних становить:

$$\bar{A}_{\text{розн.}} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{8} \cdot 1,5184 \cdot 100\% = 18,98\%$$

Середня помилка (похибка) апроксимації не повинна бути вищою за прийнятну межу значень $\bar{A} = 8 \dots 10\%$. Оскільки $\bar{A}_{\text{розн.}} = 18,98\% > 8 \dots 10\%$, тоді в нашому випадку рівняння не має достатньої точності, та його використання не є доцільним.

2) перевірка статистичної значимості рівняння регресії в цілому за допомогою F -критерія Фішера. Для $k_1 = 1$ і $k_2 = n - 2 = 6$ критерій Фішера табличний становить $F_{\text{табл}} = 5,99$ [12, с. 187; 13, с. 149; 14, с. 76], за розрахунками $F_{\text{факт}} = 58,6199171$. Через те, що $F_{\text{факт}} = 58,6199171 > F_{\text{табл}} = 5,99$, гіпотеза про випадковість розбіжності факторної і залишкової дисперсій відкидається, тобто a , b і R^2 не випадково відрізняються від нуля і сформувалися під впливом систематично діючого фактора населення. Тобто рівняння регресії статистично значиме взагалі.

3) перевірка статистичної значимості параметрів рівняння регресії за допомогою t -критерія Стьюдента. Середньоквадратичні похибки параметрів становлять:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{42,65667664 / (8 - 2)}{1819561,711}} = 0,001976671;$$

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{(n - m - 1)} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{\frac{42,65667664}{(8 - 2)} \cdot \frac{15960469,38}{8 \cdot 1819561,711}} = 2,79198014$$

Розрахункові значення критерію дорівнюють:

$$t_b = \frac{b}{m_b} = \frac{0,0151342}{0,001976671} = 7,65640645, \quad t_a = \frac{a}{m_a} = \frac{-7,225665128}{2,79198014} = -2,588007352.$$

При $df = n - m - 1 = 8 - 1 - 1 = 6$ і рівні значимості $\alpha = 0,05$ критичне значення t -критерія Стьюдента становить [9, с. 557; 13, с. 144; 14, с. 200; 15, с. 391;]: $t_{\text{табл}} = 1,943$, оскільки $t_b > 1,943$ і $t_a > 1,943$, тоді параметри b і a статистично значимі.

Таблиця 4

Основні показники оцінки «якості» рівнянь регресії

Регресія	Рівняння	Коефіцієнт детермінації, R^2	Середня відносна похибка апроксимації, $\bar{A}_{розр}$	Розрахунковий критерій Фішера, $F_{факт}$	Розрахункові значення критерію Стьюдента	
					t_b	t_a
лінійна	$y = -7,225665128 + 0,0151342 \cdot x$	0,907149370	18,980%	58,61991710	7,6564	2,5881
показникова	$y = 3,916089004 \cdot 1,000813132^x$	0,841080598	14,697%	31,75498737	10,008	3,3828
степенева	$y = 0,001164247 \cdot x^{1,287962084}$	0,784150283	17,573%	21,79711775	7,2487	2,4502

Таблиця 5

Основні показники оцінки «якості» поліноміальних рівнянь регресії 2–4 степенів, отримані в Excel за допомогою оператора ЛИНЕЙН

Регресія поліноміальна	Рівняння	Коефіцієнт детермінації, R^2	Середня відносна похибка апроксимації, $\bar{A}_{розр}$	Розрахунковий критерій Фішера, $F_{факт}$	Розрахункові значення критерію Стьюдента	
					p_{xy}	Оцінка зв'язку за шкалою Чеддока
2-го степеня	$y = 17,7505 - 0,01739 \cdot x + 9,15767 \cdot 10^{-6} \cdot x^2$	0,95899723	13,589%	58,47149209	0,9793	дуже високий
3-го степеня	$y = -60,3836 + 0,1539 \cdot x - 0,0001 \cdot x^2 + 2,4486 \cdot 10^{-8} \cdot x^3$	0,96691987	11,470%	38,97283605	0,9833	
4-го степеня	$y = 685,428 - 2,161898 \cdot x + 0,002495 \cdot x^2 - 1,21975 \cdot 10^{-6} \cdot x^3 + 2,098 \cdot 10^{-10} \cdot x^4$	0,97707552	8,73%	31,96611887	0,9885	

Отже, за F -критерієм Фішера та t -критерієм Стьюдента параметри рівняння лінійної регресії є статистично значимими, але середня похибка апроксимації (\bar{A}) вказує на те, що розрахункові значення кількості злочинів за лінійним рівнянням $y = -7,225665128 + 0,0151342 \cdot x$ будуть відрізнятися від усереднених дійсних у середньому на 19% (або занижені, або завищені на 19%). Аналогічні розрахунки \bar{A} та перевірку проведемо для рівнянь показникової та степеневої функцій (табл. 4). Можна зробити висновок, що ні одна із функцій неприйнятна для використання як модель злочинності, отже, необхідно провести додатково пошук найбільш прийнятних рівнянь регресії, наприклад, рівнянь функцій 2–4 степенів (поліноми).

Для опису взаємозв'язку між злочинністю і населенням у західних областях України приймаємо поліноміальну регресію 4-го степеня. Коефіцієнт детермінації для даної моделі дорівнює $R^2 = 0,97707552$. У подібному випадку зазвичай говорять, що зміна числа зареєстрованих злочинів для першої моделі на 97,707552% визначається зміною населення і тільки на 2,292448% – іншими чинниками (тут йде мова про зіставлення впливу на злочинність чисельності населення і всіх інших чинників. Якщо абстрагуватися від впливу чисельності населення, тоді дія основних чинників має бути прийнята за 100%).

Оцінимо якість рівняння регресії $y = 685,428 - 2,161898 \cdot x + 0,002495 \cdot x^2 - 1,21975 \cdot 10^{-6} \cdot x^3 + 2,098 \cdot 10^{-10} \cdot x^4$:

1) середня відносна похибка апроксимації – середнє відхилення розрахункових значень від фактичних становить:

$$\bar{A}_{розр} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{8} \cdot 0,69839959 \cdot 100\% = 8,73\% .$$

Середня помилка (похибка) апроксимації не повинна бути вищою за прийнятну межу значень $\bar{A} = 8 \dots 10\%$. У нашому випадку рівняння має достатню точність, і його використання є можливим.

2) перевірка статистичної значимості рівняння регресії в цілому за допомогою F -критерія Фішера. Для $k_1 = 1$ і $k_2 = n - m - 1 = 3$ критерій Фішера табличний становить $F_{табл} = 10,13$ [12, с. 187; 13, с. 149; 14, с. 76], за розрахунками $F_{факт} = 31,96611887$. Через те, що $F_{факт} = 31,96611887 > F_{табл} = 10,13$, гіпотеза про випадковість розбіжності факторної і залишкової дисперсій відкидається, тобто a , b , c , d , e і R^2 не випадково відрізняються від нуля і сформувалися під впливом систематично діючого фактора населення. Отже, рівняння регресії статистично значиме в цілому.

Індекс кореляції ρ_{xy} для поліноміальної регресії 4-го порядку становить $\rho_{xy} = 0,9885$, що вказує на дуже високий зв'язок між фактором чисельності населення та кількістю злочинів.

Зважаючи на те, що в Україні загалом і в областях зокрема спостерігається негативна демографічна ситуація (зменшення населення) (табл. 2), проведемо довільний розрахунок. Згідно зі статистичною динамікою населення в Тернопільській області за рік зменшується на 5000–6000 осіб, припустимо, що у 2018 р. середня чисельність населення зменшиться на 5 тис. і становитиме 1 050 750 осіб, тоді злочинність у нас буде такою: підставимо в параболічну регресію 4-го степеня $y = 685,428 - 2,161898 \cdot x + 0,002495 \cdot x^2 - 1,21975 \cdot 10^{-6} \cdot x^3 + 2,098 \cdot 10^{-10} \cdot x^4$ чисельність населення $x = 1055750$ та $x = 1050750$ і знайдемо різницю: $\Delta y = -685,428 - 2,161898 \cdot 1055750 + 0,002495 \cdot 1055750^2 - 1,21975 \cdot 10^{-6} \cdot 1055750^3 + 2,098 \cdot 10^{-10} \cdot 1055750^4 - 685,428 + 2,161898 \cdot 1050750 - 0,002495 \cdot 1050750^2 + 1,21975 \cdot 10^{-6} \cdot 1050750^3 - 2,098 \cdot 10^{-10} \cdot 1050750^4 = 76,42017 \approx 76$. Тобто зі зменшенням населення в Тернопільській області на 5 тис. чоловік кількість злочинів зменшиться на 76 і становитиме у 2018 р. $7627 - 76 = 7551$.

Прогнозне значення y_p визначається шляхом підстановки в рівняння регресії відповідного (прогнозного) значення x_p . Обчислюється стандартна середня похибка прогнозу m_{y_x} [12, с. 9]:

$$m_{y_x} = \sigma_{\text{зал}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}}, \quad (9)$$

де $\sigma_{\text{зал}} = \sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{n - m - 1}}$ – середньоквадратична похибка рівняння регресії [11, с. 264]. І будуватиметься довірчий інтервал прогнозу:

$$\gamma_{\hat{y}_p} = \hat{y}_p \pm \Delta_{\hat{y}_p}, \quad \gamma_{\hat{y}_{p\min}} = \hat{y}_p - \Delta_{\hat{y}_p}, \quad \gamma_{\hat{y}_{p\max}} = \hat{y}_p + \Delta_{\hat{y}_p}, \quad (10)$$

$$\text{де } \Delta_{\hat{y}_p} = t_{\text{табл}} \cdot m_{\hat{y}_p}$$

$$\sigma_{\text{зал}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - m - 1}} = \sqrt{\frac{10,53189262}{8 - 4 - 1}} = 1,87366776.$$

$$m_{\hat{y}_x} = 1,87366776 \cdot \sqrt{1 + \frac{1}{8} + \frac{(5 - 1329,51625)^2}{1819561,711}} \approx 2,708184$$

Для розрахунку злочинності формуємо розрахункову табл. 6.

Для прогнозного значення \hat{y}_x 95%-ні довірчі інтервали становлять $\hat{y}_p \pm t_{\text{табл}} \cdot m_{\hat{y}_p}$; при $df = n - m - 1 = 8 - 4 - 1 = 3$ і рівні значимості $\alpha = 0,05$ критичне значення t -критерія Стюдента становить [9, с. 557; 13, с. 144; 14, с. 200; 15, с. 391]: $t_{(0,05,3)} = 2,353363$, $\hat{y}_p \pm 2,708184 \cdot 2,353363$, тобто $\hat{y}_p = \pm 6,37334$.

Наприклад, для Тернопільської області в разі зменшення населення на 5 тис. осіб зменшення кількості злочинів з урахуванням довірчих інтервалів становитиме: $76,42017 - 6,37334 \approx 70 \leq \hat{y}_p \leq 76,42017 + 6,37334 \approx 83$; $70 \leq \hat{y}_p \leq 83$. Отже, функція $y = 685,428 - 2,161898 \cdot x + 0,002495 \cdot x^2 - 1,21975 \cdot 10^{-6} \cdot x^3 + 2,098 \cdot 10^{-10} \cdot x^4$ показує, що зі зменшенням населення на кожні 5 тис. осіб у західних областях України число зареєстрованих злочинів зменшилось би в середньому на 80 злочинів.

Таблиця 6

Розрахунок параметрів для середньої похибки полінома 4-го степеня
 $y = 685,428 - 2,161898 \cdot x + 0,002495 \cdot x^2 - 1,21975 \cdot 10^{-6} \cdot x^3 + 2,098 \cdot 10^{-10} \cdot x^4$

Область України	Населення (тис.) x	Злочинів (тис.) y	\hat{y}	$(y - \hat{y})^2$	$(x - \bar{x})^2$
Чернівецька	908,783	8,941	9,034527917	3,944146477	82972,37043
Волинська	1041,467	11,168	9,182012468	0,741300395	4979,595639
Тернопільська	1061,967	7,288	9,492246828	0,021390123	2551,942031
Рівненська	1161,8	11,507	11,05708606	1,06364 · 10 ⁻⁸	1450701,609
Закарпатська	1258,95	10,679	11,53998803	0,20242255	28128,74051
Хмельницька	1289,163	12,159	11,2899882	4,85870408	71582,60118
Івано-Франківська	1380,033	8,883	9,029253625	0,755181517	1628,384786
Львівська	2533,967	32,539	32,53889687	0,008747471	177016,4677
Σ	10636,13	103,164	x	10,53189262	1819561,711
Середнє значення	1329,516	12,8955	x	x	x

На практиці моделі злочинності (зокрема, й однофакторні) можуть бути найрізноманітнішими. Їх можна отримати в різних програмних середовищах (Excel, STATISTICA, SPSS Statistics, Wolfram Mathematica, MATLAB). Так, наприклад, в Excel є вбудована функція поліноміальних трендів, за допомогою якої можна отримати модель злочинності, що описується поліноміальною регресією 5-го степеня (рис. 5), для якого $\bar{A}_{\text{розрах.}} = 8.75\%$ (рівняння має достатню точність і його використання є можливим) та $F_{\text{факт}} = 20,99$. $F_{\text{табл}} = F_{(0,05, 1, 2)} = 18,5128$. В силу того, що $F_{\text{факт}} = 20,99 > F_{\text{табл}} = 18,5128$, рівняння поліноміальної регресії 5-го степеня статистично значиме в цілому.

Індекс кореляції r_{xy} для поліноміальної регресії 5-го степеня становить $r_{xy} \approx 0,9856$, що вказує на дуже високий зв'язок між фактором чисельності населення та кількістю злочинів.

Використання функції $y = 0,0923x^5 - 1,9159x^4 + 14,689x^3 - 50,909x^2 + 77,726x - 30,713$ не дуже зручне з огляду на те, що замість x у нього підставляються порядкові точки ($x = 1, 2, 3, 4, 5, 6, 7, 8$), а не кількість злочинів. Для Тернопільської області, припустимо, відбудеться зменшення населення на 5 тис. осіб у 2018 р. в порівнянні з 2017 р. і становитиме 1050750 осіб. Тоді $x = (3 \cdot 1050,750) / 1061,967 = 2,968$, а кількість злочинів зміниться на $\Delta y = (0,0923 \cdot 3^5 - 1,9159 \cdot 3^4 + 14,689 \cdot 3^3 - 50,909 \cdot 3^2 + 77,726 \cdot 3 - 30,713) - (0,0923 \cdot 2,968^5 - 1,9159 \cdot 2,968^4 + 14,689 \cdot 2,968^3 - 50,909 \cdot 2,968^2 + 77,726 \cdot 2,968 - 30,713) \approx -24$, тобто може збільшитись на 24 злочини. Тим самим можна дійти до суперечності: в разі зменшення населення Тернопільської області за рівнянням полінома 4-го степеня отримуємо зменшення злочинів на 76, а за поліноміальним рівнянням 5-го степеня, навпаки, – збільшення на 24.

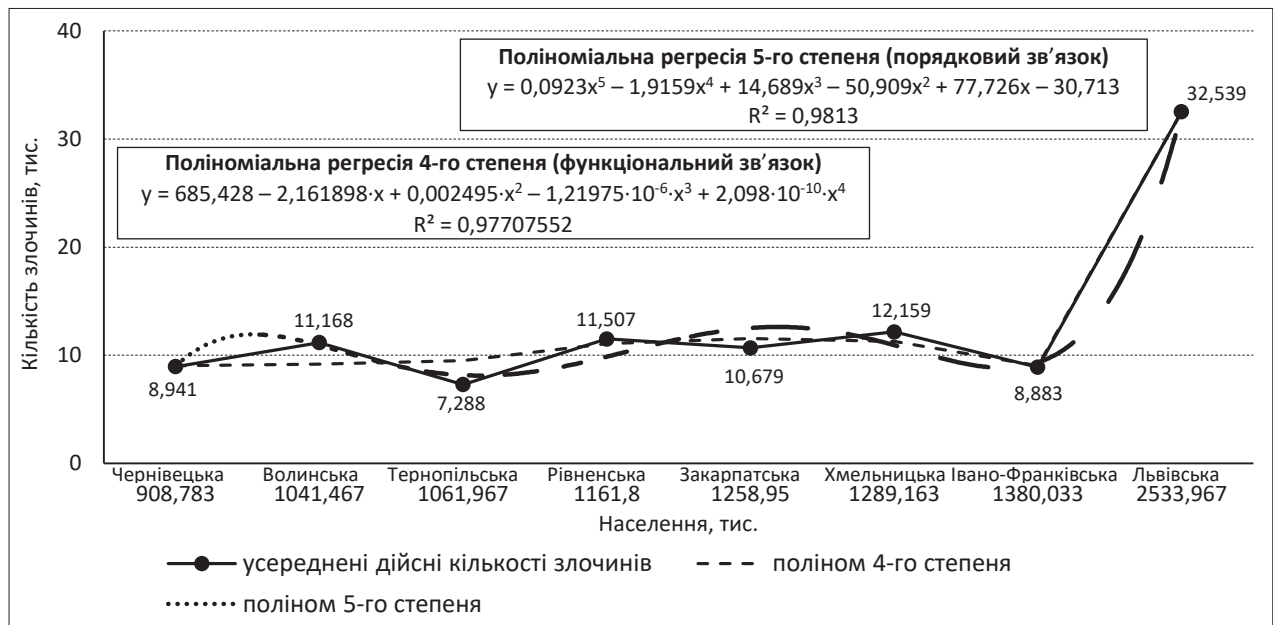


Рис. 5. Усереднені дійсні значення кількості злочинів та поліноміальні регресії 4-го, 5-го степенів

Аналізуючи статистичні дані (табл. 2), можна стверджувати, що, незважаючи на зменшення чисельності населення протягом 2015–2017 рр., наприклад в Тернопільській області, кількість злочинів не має тенденції до зменшення (7195, 7042, 7627), тенденція зменшення злочинності спостерігається лише в Хмельницькій області. Тим самим, як статичні дані, так і формалізований опис злочинності (однофакторна модель на основі поліноміального рівняння регресії 5-го степеня) дають переконливі докази того, що зменшення чисельності населення однозначно не спричинює зменшення злочинності.

Таким чином, провівши аналіз статистичних даних щодо чисельності населення і кількості злочинів за 2015–2017 рр. в західних областях України, отримана однофакторна модель на основі полінома 4-го степеня ($y = 685,428 - 2,161898 \cdot x + 0,002495 \cdot x^2 - 1,21975 \cdot 10^{-6} \cdot x^3 + 2,098 \cdot 10^{-10} \cdot x^4$) та на основі полінома 5-го степеня ($y = 0,0923x^5 - 1,9159x^4 + 14,689x^3 - 50,909x^2 + 77,726x - 30,713$). В окремих випадках модель із використанням даних рівнянь регресії (рис. 5) є необхідною для виявлення залежності кількості зареєстрованих злочинів від чисельності населення.

Висновки. Вплив населення зовсім не зводиться до прямих залежностей типу «більше населення – більше злочинів» або, навпаки, «менше населення – менше злочинів». Ці чинники входять у набагато складніші комплекси взаємодії, які посилюють або послаблюють елементарні прямі залежності і включають такі форми, як характер, рівень загальної і правової культури (в тому числі нефор-

мального соціального контролю), добробут, здоров'я населення і багато інших, що інтенсивно (безпосередньо або побічно) впливають на типові ціннісні орієнтації і поведінкові стереотипи вікових і соціально-професійних груп населення, розподілених у свою чергу за певними місцями роботи (навчання) і проживання [2, с. 129].

Розглянута однофакторна модель на основі рівнянь регресії (зокрема, поліноміальні регресії 4-го та 5-го степенів) може надати кримінологічно-теоретичну важливу практичну допомогу. Модель дає змогу порівнювати отримані результати кількості злочинів у західних областях України.

Подальші дослідження питання використання методів моделювання злочинності необхідно проводити в напрямі побудови математичних моделей злочинності на основі рівнянь множинної регресії.

Література:

1. Корецький С.М. Кримінологічна характеристика девіантної поведінки неповнолітніх: дис. ... канд. юрид. наук: 12.00.08. Київ, 2003. 189 с.
2. Утаров К.А. Математические методы в криминологии: дис ... канд. юрид. наук: 12.00.08. Москва, 2004. 165 с.
3. Потапов Д.К., Евстафьева В.В. Математическое моделирование преступности в Санкт-Петербурге эконометрическими методами. Россия в ВТО: проблемы, задачи, перспективы: сборник науч. статей, вып. 13; под общей редакцией проф. В.В. Тумалева. Санкт-Петербург: НОУ ВПО «Институт бизнеса и права», 2012. С. 110–113.
4. Криминология: учеб. / А.И. Гуров и др.; под ред. Н.Ф. Кузнецовой, В.В. Лунеева. 2-е изд., перераб. и доп. Москва: Волтерс Клувер, 2004. 640 с.
5. Сайти головних управлінь Державної служби статистики України (дата звернення: 22.10.2018).
6. Александров Ю.В., Гель А.П., Семаков Г.С. Криминология: Курс лекций. Київ: МАУП, 2002. 295 с.
7. Криминология: учеб. / Ю.М. Антонян и др.; под ред. В.Н. Кудрявцева и В.Е. Эминова. 4-е изд., перераб. и доп. Москва: Норма, 2009. 800 с.
8. Лагутин М.Б. Наглядная математическая статистика: учеб. пособ. 2-е изд., испр. Москва: БИНОМ. Лаборатория знаний, 2009. 472 с.
9. Магнус Я.Р., Катышев П.К., Пересецкий А.Л. Эконометрика. Начальный курс: учеб. 6-е изд., перераб. и доп. Москва: Дело, 2004. 576 с.
10. Эконометрика: учеб. / И.И. Елисеева и др.; под ред. И.И. Елисеевой. 2-е изд., испр. и доп. Москва: Финансы и статистика, 2007. 576 с.
11. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: учеб. для студ. вузов / под ред. М.Р. Ефимовой. 2-е изд., испр. и доп. («Высшее образование»). Москва: Инфра-М, 2008. 412 с.
12. Практикум по эконометрике: учеб. пособ. / И.И. Елисеева и др.; под ред. И.И. Елисеевой. Москва: Финансы и статистика, 2002. 192 с.
13. Осипов А.Л., Рапоцевич Е.А. Основы математического моделирования социально-экономических процессов: учеб. пособие. Новосибирск: Изд-во СибАГС, 2014. 154 с.
14. Мардиа К., Земрох К. Таблицы F-распределений и распределений, связанных с ними; пер. с англ. М.С. Никулина. Москва: Наука, 1984. 255 с.
15. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. Москва: ФИЗМАТЛИТ, 2006. 816 с.