

СЕКЦІЯ 11 МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

УДК 330.55.051

Балюнов О.О.*кандидат фізико-математичних наук, доцент,
доцент кафедри інформаційних систем в економіці
Чернігівського національного технологічного університету*

РЕГРЕСІЙНИЙ АНАЛІЗ У МОДЕЛЮВАННІ РЕГІОНАЛЬНОГО ВАЛОВОГО ПРОДУКТУ УКРАЇНИ

REGRESSION ANALYSIS IN THE MODELLING OF GROSS REGIONAL PRODUCT OF UKRAINE

АНОТАЦІЯ

У статті досліджено вплив основних соціально-економічних показників на регіональний валовий продукт України. Використано методи кореляційного та регресійного аналізу. Побудована лінійна модель множинної регресії відповідає теоретичним гіпотезам із великим значенням коефіцієнта детермінації. Це дає змогу використовувати регресійну модель для експертної оцінки передбачуваних значень регресанта. Виконано порівняльний аналіз впливу значущих показників для статистичних даних за 2013–2014 рр.

Ключові слова: кореляційна матриця, лінійний множинний регресійний аналіз, регіональний валовий продукт, капітальні інвестиції, прибуток підприємств, наявний дохід населення, прогнозування регіонального валового продукту.

АННОТАЦИЯ

В статье исследовано влияние основных социально-экономических показателей на региональный валовый продукт Украины. Используются методы корреляционного и регрессионного анализа. Построенная линейная модель множественной регрессии отвечает теоретическим гипотезам с высоким значением коэффициента детерминации. Это позволяет использовать регрессионную модель для экспертной оценки предсказанных значений регрессанта. Выполнен сравнительный анализ влияния значимых показателей для статистических данных за 2013–2014 гг.

Ключевые слова: корреляционная матрица, линейный множественный регрессионный анализ, региональный валовый продукт, капитальные инвестиции, прибыль предприятий, реальный доход населения, прогнозирование регионального валового продукта.

ANNOTATION

The impact of basic social and economic indicators of the regional gross domestic product of Ukraine is being researched in the article. The methods of correlative and regressive analysis are being used. The linear model of multiple regression satisfies theoretical hypotheses with a large value of the factor of determination. It allows to use a regression model for an expert estimation of foreseen values of regressant. The comparative analysis of the impact of significant indicators is conducted for 2013, 2014 statistical data.

Keywords: correlation matrix, linear multiple regression analysis, the regional GDP, capital investments, income tax, available income of the population, regional GDP forecasting.

Постановка проблеми. В економічній теорії та в її прикладних аспектах велика увага приділяється аналізу такої категорії, як добробут населення, та його складникам як то: купівельна

спроможність грошових доходів, індекс споживчого ринку, рівень бідності, інші пов'язані соціальні та екологічні фактори. Рівень доходів населення відповідно до основних макроекономічних показників та соціально-демографічної структури суспільства досліджував С. Кузнець [1]. Один з основоположників поняття внутрішнього валового продукту, він зробив великий внесок у розроблення методів підрахунку національного доходу, що базується на єдиній теоретичній концепції взаємозв'язку між обчисленням об'ємом виробленого національного продукту з певним рівнем добробуту населення. Виходячи з того, що ВВП як макроекономічний індикатор є найбільш точною сукупною мірою загального економічного виробництва, великого значення набувають математичне моделювання та аналіз впливових факторів на ВВП, а отже, і на добробут населення.

Аналіз останніх досліджень і публікацій. Методи оцінки матеріальної нерівності та її вплив на суспільний добробут викладені в роботі А. Сена [2]. У цій та інших працях А. Сен відкрив нові можливості для дослідження широкого кола економічних та соціальних проблем, у тому числі для кількісної оцінки економічної нерівності, виміру ступеня бідності, аналізу безробіття, а також економічного аналізу таких категорій, як базові права і свободи особистості, нерівність і дискримінація жінок і т. ін.

Широкий огляд статистичних підходів до аналізу соціально-економічної нерівності представлений у роботі В. Крамера [3]. Тут докладно розглядається математичне моделювання проблеми в її історичному розвитку. Економічна нерівність розглядається у вигляді багатовимірної моделі, проводиться аналіз математичних проблем а також методів та теоретичних підходів до їх вирішення

Важливий розвиток економетричної теорії відбувся завдяки працям А. Аткинсона [4] і його

послідовників. Основним припущенням є те, що дохід є випадковою безперервною змінною, що розподіляється в загальному випадку серед скінченного населення.

Впливу коефіцієнта Джині на валовий регіональний продукт (ВРП) на одну особу присвячена робота М.Ю. Малкіної [5]. Виконане тестування гіпотези Кузнеця на масиві даних по країнах та їх регіональних групах, яке не підтвердило справедливості цієї гіпотези. Також встановлено, що в деяких країнах в економіці наявний прямий зв'язок між рівнем розвитку регіону і ступенем диференціації доходів населення. У статті робиться висновок про існування різних типів зв'язків диференціації доходів населення з рівнем економічного розвитку. Це пояснюється впливом на розподіл доходів у різних країнах і регіонах факторів структурно-економічного характеру, а також соціальних і

політичних факторів. Особливу роль у розподілі доходів також відіграє регіональна приналежність країни, що зумовлено спільністю звичаїв та інститутів країн одного регіону.

Відбір пояснювальних змінних та побудова економічної моделі добробуту населення є темою праці [6]. Зазначимо, що значущими пояснювальними змінними, що впливають на добробут, були обрані такі фактори, як інтегральний індикатор з активності малого бізнесу, купівельна спроможність грошових доходів населення, індекс споживчих цін, ціни на послуги готелів, зростання частки витрат на соціальну політику у витратах регіонального бюджету зростання частки витрат на національну економіку, тобто інвестиції.

Зазначимо, що багато дослідників у даній сфері використовують як метод множинний регресійний аналіз, теоретичні основи якого

Таблиця 1

Соціально-економічні показники за 2014 р.

Область	1	2	3	4	5	6	7	8
Region (Reg)	Average Population (AP)	Nativity (N)	Employment Rate (ER)	Salary (S)	Income per Capita (IC)	Capital Investments (CI)	Enterprise Profit (EP)	Per Capita Gross Regional Product (Gr_Reg_Prod)
Вінницька	1607	-8020	56	2810	23422	5491	-1594	27249
Волинська	1039	920	55	2721	20137	3242	-7006	23218
Дніпропетровська	3281	-16225	60	3641	32036	19396	-61954	53749
Донецька	4308	-36204	54	3858	26234	12876	-42131	27771
Житомирська	1260	-6070	56	2763	22102	2667	-2755	23678
Закарпатська	1255	3569	56	2744	17358	2524	-2295	19170
Запорізька	1770	-9060	58	3432	30182	6658	-10046	37251
Івано-Франківська	1380	-784	54	2875	20357	6693	-9278	27232
Київська	4559	-2532	60	4433	45579	78951	-314273	85111
Кіровоградська	978	-6140	54	2789	21954	2990	-3371	29223
Луганська	2225	-11313	52	3377	19788	5116	-45846	14079
Львівська	2520	-2180	55	2961	23595	8880	-9285	28731
Миколаївська	1166	-4674	57	3344	23459	3583	-8480	30357
Одеська	2385	-4690	57	3129	24242	8451	-29898	31268
Полтавська	1446	-10280	56	3179	26196	8561	-3789	48040
Рівненська	1159	2455	57	3033	21781	2561	-7691	24762
Сумська	1126	-9108	57	2877	23938	2705	-1105	26943
Тернопільська	1068	-3463	53	2527	18401	2433	-6170	20228
Харківська	2719	-14201	59	3143	26274	7568	-7592	35328
Херсонська	1069	-3833	56	2617	20728	1918	-2626	21725
Хмельницька	1301	-5777	55	2878	22686	3811	-3563	24662
Черкаська	1252	-8449	56	2829	21761	3005	-6178	30628
Чернівецька	906	60	56	2578	18476	1550	-996	16552
Чернігівська	1053	-10772	57	2690	23093	2431	-2149	26530

викладені в працях Н. Дрейпера і Г. Сміта [7], К. Доугерті [8], Е. Ферстера і Б. Ренца [9].

Виділення не вирішених частин загальної проблеми. Нині відсутні емпіричні статистичні дослідження якісної залежності внутрішнього регіонального продукту від основних соціально-економічних показників за окремі роки, зокрема яких змін зазнає регресійна модель ВРП в умовах, коли на економіку України впливають воєнні дії та окупація частини території.

Зростання ВРП супроводжується ростом чисельності зайнятих у виробництві та покращенням рівня життя, що призводить до зростання споживання товарів та послуг. Зростання ВРП визначається інвестиціями, їх часткою у ВРП та перевищенням загального об'єму інвестицій над величиною капіталу, витраченого у процесі виробництва.

Мета статті полягає в аналізі цих індикаторів щодо їх впливу на ВРП та порівнянні результатів математичного моделювання за 2013–2014 рр.

Виклад основного матеріалу дослідження. Як дані для множинного регресійного аналізу розглянемо офіційну статистику за 2014 р. [10] (табл. 1).

Одразу зазначимо, що порівняно з даними за 2013 р. по всіх регіонах України зменшилися капітальні інвестиції, підприємства зазнали значних збитків, майже в половині регіонів зменшився наявний дохід на одну особу, і тільки заробітна плата зросла. Залежною змінною, або регресантом, оберемо змінну Gr_Reg_

Prod – валовий регіональний продукт. Інші сім показників оберемо як регресори, регіони розглядаються як відповідні спостереження.

Отже, специфікація моделі лінійної регресії має вигляд

$$\text{Gr_Reg_Prod} = \beta_1 + \beta_2 AP + \beta_3 N + \beta_4 ER + \beta_5 S + \beta_6 IC + \beta_7 CI + \beta_8 EP + \varepsilon.$$

Множину значущих змінних будемо визначати безпосередньо під час аналізу. Спочатку наведемо матрицю кореляційних коефіцієнтів для перевірки припущення стосовно лінійної залежності та врахування можливих сильних кореляцій між змінними під час побудови регресійної моделі (табл. 2).

Із кореляційної матриці, зокрема, впливає, що деякі змінні сильно корелюють між собою. Зазначимо, що достовірність великих значень кореляції можлива тільки за відсутності викидів у вихідній таблиці, тому діаграма розсіювання для залежної змінної і решти змінних обов'язково повинна враховуватися під час кореляційного аналізу. Великий коефіцієнт кореляції між змінними (мультиколінійність) необхідно враховувати під час побудови регресійної моделі. Тут можливі великі похибки, що виникають під час обчислення коефіцієнтів регресії (погано зумовлена матриця під час оцінювання за методом МНК).

Відзначимо, що змінна Gr_Reg_Prod для статистичних даних за 2014 р. статистично не залежить від приросту населення, слабо корелює з чисельністю населення та рівнем зайнятості, має високий ступінь кореляції із заробітною платою,

Таблиця 2

Матриця кореляційних коефіцієнтів

Correlations (data 2014)								
	AP	N	ER	S	IC	CI	EP	Gr_Reg_Prod
AP	1,000000	-0,562001	0,347781	0,873566	0,752104	0,737712	-0,716444	0,651358
N	-0,562001	1,000000	-0,011894	-0,455650	-0,258358	-0,045914	0,029254	-0,125176
ER	0,347781	-0,011894	1,000000	0,434359	0,669171	0,459994	-0,393961	0,660715
S	0,873566	-0,455650	0,434359	1,000000	0,853078	0,784689	-0,778911	0,764566
IC	0,752104	-0,258358	0,669171	0,853078	1,000000	0,891893	-0,843792	0,949124
CI	0,737712	-0,045914	0,459994	0,784689	0,891893	1,000000	-0,985781	0,884103
EP	-0,716444	0,029254	-0,393961	-0,778911	-0,843792	-0,985781	1,000000	-0,820112
Gr_Reg_Prod	0,651358	-0,125176	0,660715	0,764566	0,949124	0,884103	-0,820112	1,000000

Таблиця 3

Результати стандартної множинної регресійної моделі

Regression Summary for Dependent Variable: Gr_Reg_Prod (data 2014) R=,97529634 R ² =,95120294 Adjusted R ² =,92985423 F(7,16)=44,556 p						
	Beta	Std.Err. – of Beta	B	Std.Err. – of B	t(16)	p-level
Intercept			-50087,6	34840,82	-1,43761	0,169814
AP	-0,332059	0,164857	-4,7	2,34	-2,01422	0,061115
N	-0,068637	0,110690	-0,1	0,20	-0,62008	0,543937
ER	0,089157	0,096855	656,7	713,35	0,92053	0,370969
S	0,222767	0,170723	7,2	5,54	1,30484	0,210402
IC	0,357664	0,274546	0,9	0,69	1,30275	0,211100
CI	1,887245	0,568699	1,8	0,53	3,31853	0,004346
EP	1,314838	0,437694	0,3	0,10	3,00402	0,008409

капітальними інвестиціями та прибутком підприємств та має найбільшу залежність від наявного доходу на одну особу. Отже, зробимо таке передбачення: якщо для наявних даних можливо побудувати адекватну лінійну регресійну модель, то серед значущих пояснювальних змінних присутність таких факторів як заробітна плата, капітальні інвестиції, наявний дохід, при-

буток підприємств (сильна від'ємна кореляція). Чисельність населення, рівень зайнятості будуть малоймовірними. Тут ми виходили з тих міркувань, що наявність високого ступеня кореляції будь-якого знаку може призвести до причинно-наслідкової залежності. Однак за наявності багатьох факторів можливий опосередкований вплив на пояснювальну змінну через різні приховані

Таблиця 4

Остаточні результати множинної регресійної моделі, 2014 р.

Regression Summary for Dependent Variable: Gr_Reg_Prod (data_2014) R=,94912423 R²=,90083680 Adjusted R²=,89632938 F(1,22)=199,86 p						
	Beta	Std.Err. - of Beta	B	Std.Err. - of B	t(22)	p-level
Intercept			-26811,4	4170,274	-6,42917	0,000002
IC	0,949124	0,067137	2,4	0,170	14,13706	0,000000

Таблиця 5

Остаточні результати множинної регресійної моделі, 2013 р.

Regression Summary for Dependent Variable: Gr_Reg_Prod (data_2013) R = 0,97091993 R² = 0,94268551 Adjusted R² = 0,93408834 F(3,20)=109,65 p						
	Beta	Std.Err. - of Beta	B	Std.Err. - of B	t(20)	p-level
Intercept			-8593,79	6757,661	-1,27171	0,218066
IC	0,531408	0,123938	1,34	0,313	4,28770	0,000359
CI	0,596564	0,141810	0,45	0,106	4,20678	0,000434
EP	0,235445	0,073881	0,88	0,275	3,18683	0,004633

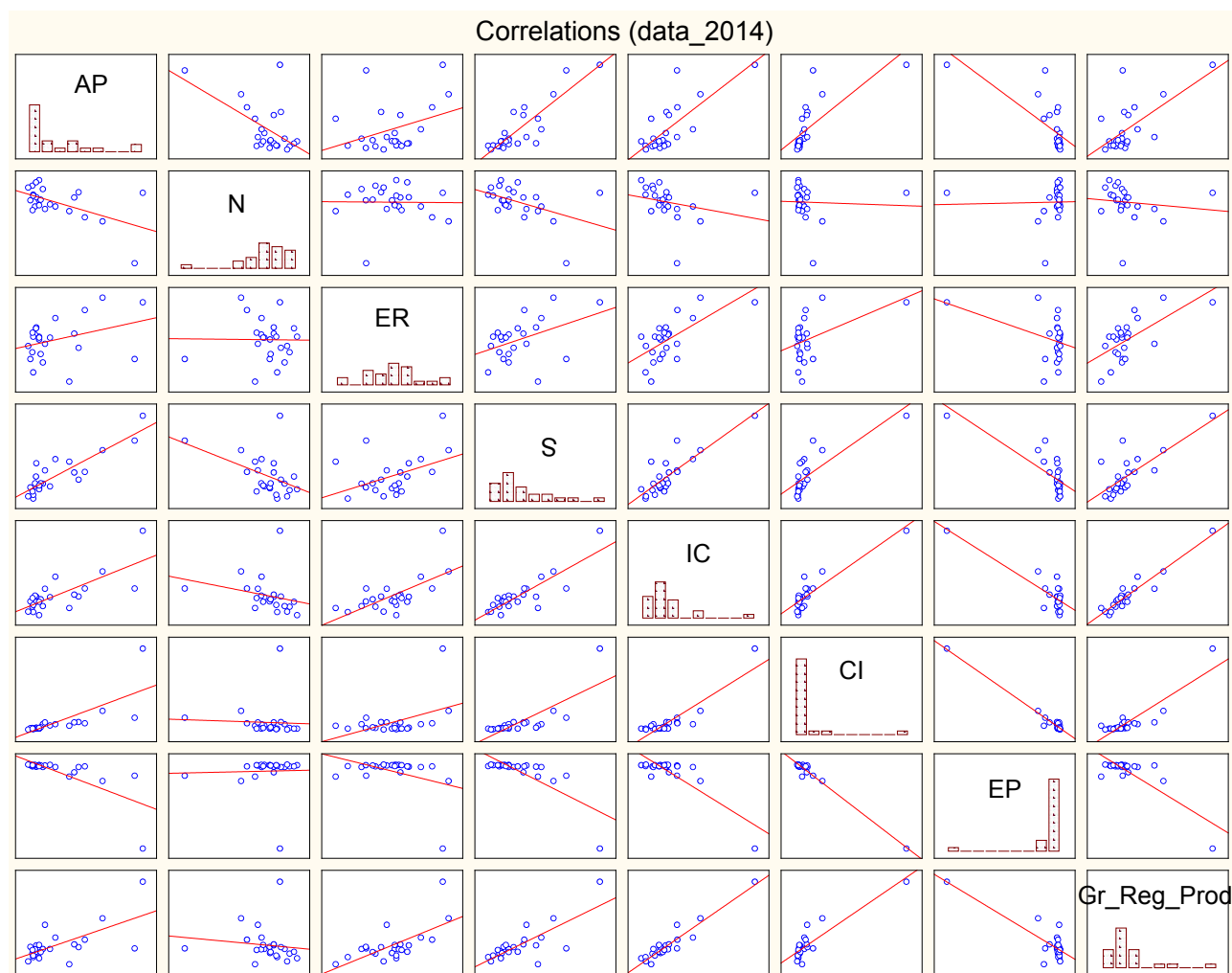


Рис. 1. Діаграми розсіювання

кореляції, тому далі звернемося до більш потужного метода дослідження – множинної лінійної регресії, яка дає змогу отримати обґрунтовані та значущі оцінки. Зазначимо лише, що пояснювальна змінна *Gr_Reg_Prod* має попарну лінійну кореляцію з рештою змінних, що дає змогу використовувати коефіцієнт кореляції Пірсона.

Одним із найбільш відомих способів усунення мультиколінійності є використання покрокових процедур із включенням/виключенням предикторів. Реалізація цього процесу відбувається на основі метода виключення незначущих змінних, порівнюючи значення F-критерія для кожної змінної. Статистичний аналіз даних виконуємо за допомогою пакета STATISTICA. До даних табл. 1 застосуємо модуль стандартної множинної регресії, отримаємо результати, наведені в табл. 3.

Як бачимо, можливими незначущими змінними є змінні AP, N, ER, S, IC. Коефіцієнт

детермінації $R^2=0,9512$ наближається до 1, що вказує на високий ступінь наближення регресійної моделі до варіації змінної *Gr_Reg_Prod*. Можливими пояснювальними змінними є CI, EP. Подальший аналіз змінних проведемо за допомогою опцій «Регресія вперед» та «Регресія назад», включаючи або виключаючи пояснювальні змінні. Остаточо отримаємо:

Отже, отриманий результат

$$Gr_Reg_Prod = -26811,4 + 2,4IC + \varepsilon$$

дає змогу зробити висновок: валовий регіональний продукт тим вищий, чим більший наявний дохід населення, що не суперечить нашим апріорним уявленням для статистичних даних за 2014 р. Зазначимо, що розрахунки на базі аналогічної групи статистичних показників за 2013 р. дали такий результат (табл. 5).

Отже, пояснювальна змінна має вигляд:

$$Gr_Reg_Prod = -8593,79 + 1,34IC + 0,45CI + 0,88EP + \varepsilon,$$

Таблиця 6

Дані розсіювання

Analysis of Variance; DV: Gr_Reg_Prod (data_2014)					
	Sums of – Squares	df	Mean – Squares	F	p-level
Regress.	4,417179E+09	1	4,417179E+09	199,8565	0,000000
Residual	4,862386E+08	22	2,210176E+07		
Total	4,903418E+09				

Таблиця 7

Таблиця залишків

Predicted & Residual Values (data_2014) Dependent variable: Gr_Reg_Prod			
	Observed – Value	Predicted – Value	Residual
Вінницька	27249,00	29396,20	-2147,20
Волинська	23218,00	21514,02	1703,98
Дніпропетровська	53749,00	50069,35	3679,65
Донецька	27771,00	36146,14	-8375,14
Житомирська	23678,00	26229,41	-2551,41
Закарпатська	19170,00	14844,71	4325,28
Запорізька	37251,00	45619,14	-8368,15
Івано-Франківська	27232,00	22040,78	5191,22
Київська	85110,50	82569,97	2540,53
Кіровоградська	29223,00	25874,24	3348,76
Луганська	14079,00	20676,73	-6597,73
Львівська	28731,00	29812,56	-1081,56
Миколаївська	30357,00	29484,51	872,49
Одеська	31268,00	31364,76	-96,76
Полтавська	48040,00	36053,27	11986,73
Рівненська	24762,00	25458,83	-696,83
Сумська	26943,00	30635,46	-3692,46
Тернопільська	20228,00	17346,28	2881,72
Харківська	35328,00	36241,17	-913,17
Херсонська	21725,00	22931,59	-1206,59
Хмельницька	24662,00	27630,90	-2968,90
Черкаська	30628,00	25409,63	5218,36
Чернівецька	16552,00	17526,50	-974,50
Чернігівська	26530,00	28608,34	-2078,34
Minimum	14079,00	14844,71	-8375,14
Maximum	85110,50	82569,97	11986,73
Mean	30561,85	30561,85	-0,00
Median	27240,50	28119,62	-805,00

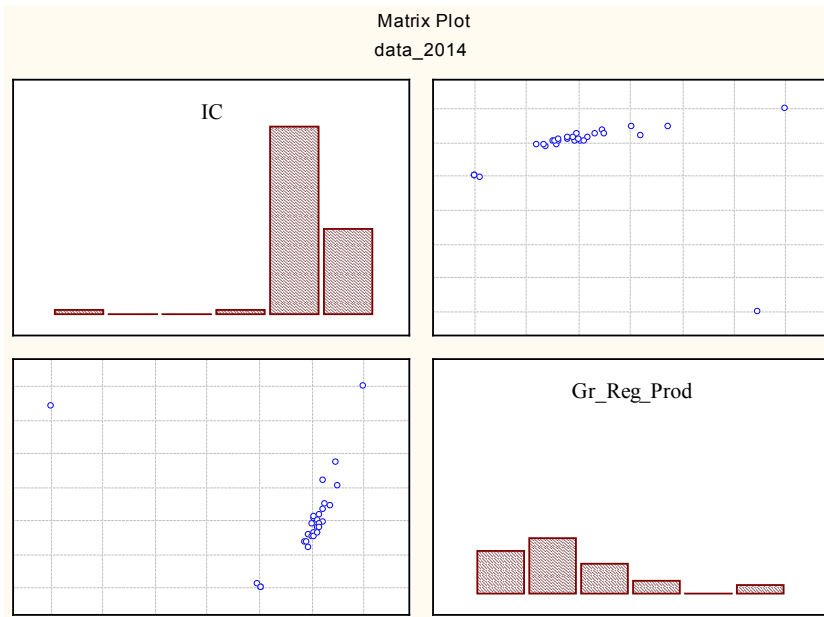


Рис. 2. Діаграма розсіювання, гістограми для змінних IC, Gr_Reg_Prod

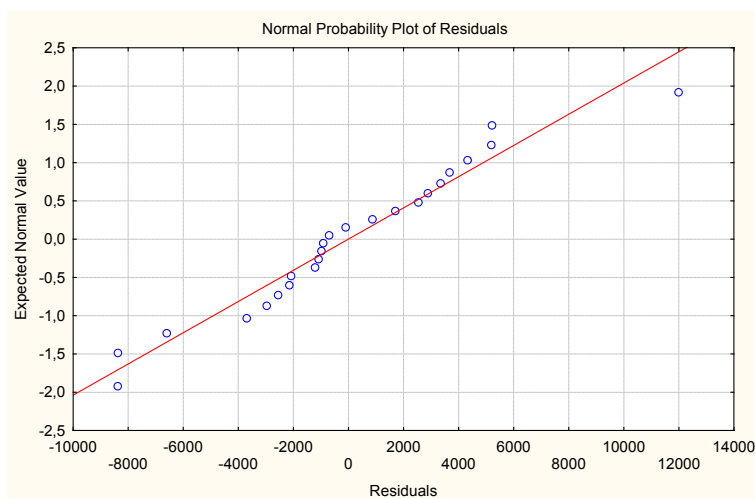


Рис. 3. P-P діаграма

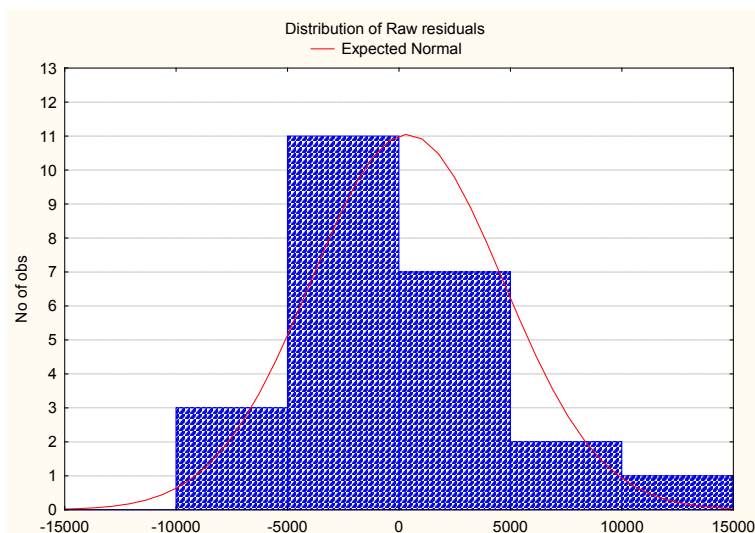


Рис. 4. Гістограма залишків

тобто валовий регіональний продукт також суттєво залежить від капітальних інвестицій та прибутку підприємств.

Для того щоб регресійна модель давала змогу робити певні статистичні висновки, необхідно перевірити основні теоретичні гіпотези [7]. Для оцінки значущості моделі у цілому виконаємо аналіз дисперсії (табл. 6).

Бачимо, що рівень значущості p наближається до нуля, тобто значно менше допустимих 0,05.

Далі проаналізуємо значення коефіцієнта детермінації R^2 . Коефіцієнт детермінації вказує на ступінь зміни регресанта, що відбувається під одночасним впливом усіх значущих змінних побудованої моделі. У нашому випадку коефіцієнт детермінації $R^2=0,9008$ наближається до 1, що вказує на високий ступінь наближення регресійної моделі до варіації змінної Gr_Reg_Prod. Скоригований коефіцієнт детермінації $Adjusted R^2$ також близький до 1, хоча за кількості значущих змінних менше 10 можна обмежитися контролем R^2 .

Подальший аналіз моделі лінійної регресії полягає в побудові матричної діаграми розсіювання даних. Оберемо змінними отримані вище регресори та регресант, для яких будемо попарні діаграми розсіювання, на діагоналі знаходяться гістограми (рис. 2).

Відзначаємо, що Gr_Reg_Prod залежить від IC майже лінійно, отже, очевидно, нема потреби будувати нелінійну регресійну модель, для якої коефіцієнт детермінації буде гіршим.

Далі проаналізуємо нормальність залишків моделі, отримаємо таблицю залишків (табл. 7). Перевіримо, чи виходять залишки за межі інтервалу $(-3s, 3s)$, де s – емпіричне середньоквадратичне відхилення залишків (у таблиці залишок позначений *). Бачимо, що не виходять. Значення залишків знаходяться в третьому стовпчику таблиці. Середньоквадратичне відхилення залишків дорівнює 0.

Порівняти розподіл залишків із нормальним розподілом допомагає p - p -діаграма (рис. 3).

Графік показує незначне відхилення від нормального розподілу, отже, вважаємо, що гіпотеза нормального розподілу залишків справджується. Цей висновок підтверджує також і гістограма розподілу залишків порівняно з нормальним розподілом (рис. 4).

Проаналізувавши викиди, встановлюємо, що вони відсутні. Це означає, що побудована модель є прийнятною для всієї множини вихідних даних, жодний рядок не виключається з розрахунку.

Для повноти побудованої моделі нам залишається прорахувати прогноз для заданих значень регресорів, а також знайти інтервал надійності. На 2015 р., за даними Державної служби статистики України, наявний дохід на одну особу в Чернігівській області становить 27 690,2 грн. [10]. Модельне передбачуване значення залежної змінної $Gr_Reg_Prod = 39639,8$ (табл. 8). З імовірністю 0,95 дійсне значення Gr_Reg_Prod повинно потрапити в інтервал (37245,2; 42034,4).

Таблиця 8
Передбачуване значення залежної змінної,
ймовірний інтервал на 2015 р.

Predicting Values for (data 2014) variable: Gr_Reg_Prod			
	B-Weight	Value	B-Weight – * Value
IC	2,399809	27690,20	66451,2
Intercept			-26811,4
Predicted			39639,8
-95,0% CL			37245,2
+95,0% CL			42034,4

Висновки. Таким чином, першочерговий вплив на ВРП у 2014 р. здійснює лише індикатор наявного доходу населення, який витрача-

ється на покупку товарів і послуг, у результаті чого підтримується сукупний попит на корисності та блага, що формує ВРП. Очевидно, це пояснюється тим, що економіка зазнала впливу війни та анексії територій. Цей висновок підтверджується порівнянням із результатами за 2013 р., коли в період сталого розвитку економіки значущими індикаторами виступали також капітальні інвестиції та прибуток підприємств.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Kuznets S. Economic Growth of Nations: Total Output and Production Structure, Harvard University Press, Cambridge (USA), 1971.
2. Sen A. On Economic Inequality. Expanded edition. Oxford: Clarendon Press, 1997. – 260 p.
3. Krämer W. Measurement of Inequality//Handbook of Applied Economic Statistics, 1998. – P. 39–61.
4. Atkinson, A.B. On the Measurement of Poverty, Econometric. – 1987. – Vol. 55. – № 4. – P. 749–776.
5. Малкина М.Ю. Исследование взаимосвязи неравномерности распределения доходов с уровнем экономического развития / М.Ю. Малкина // Вестник Нижегородского университета. Экономические науки. – 2013. – № 3. – С. 170–176.
6. Бобков В.Н., Степанов В.С. Модель «Благосостояние» для оценки и прогноза качества и уровня жизни населения региона / В.Н. Бобков, В.С. Степанов // Уровень жизни населения регионов России. – 2014. – № 1(191). – С. 104–110.
7. Draper N., Smith H. Applied Regression Analysis. A Wiley-Interscience Publication, 1998. – 736 p.
8. Dougherty C. Introduction to Econometrics. Oxford, 2001. – 608 p.
9. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа / Э. Ферстер, Б. Ренц. – М.: Финансы и статистика, 2008. – 267 с.
10. Державна служба статистики України Електронний ресурс. – Режим доступу: http://www.ukrstat.gov.ua/operativ/menu/menu_u/senstr.htm.