

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»

Інститут прикладного системного аналізу

Кафедра системного проектування

ЗВІТ

з виконання лабораторної роботи

з дисципліни «Еколого-економічна організація виробництва»

на тему: «Кореляційно-регресійний аналіз впливу соціально-економічних чинників на рівень доходів населення за регіонами України»

Виконав:

студент групи ДА-82

ННК «ІПСА»

Муравльов Андрій

Викладач: Караєва Н.В.

ЗМІСТ

ЗМІСТ	2
ТЕОРИТИЧНІ ДАНІ	3
ТАБЛИЦІ ДАНИХ	4
АЛГОРИТМ РОЗРАХУНКУ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ	9
ВИСНОВКИ	10
МУЛЬТИКОЛІНЕАРНИЙ АНАЛІЗ	10
РЕГРЕСІЙНИЙ АНАЛІЗ	17
ПЕРШИЙ СПОСІБ	17
ДРУГИЙ СПОСІБ	19
ПЕРЕВІРКА ОТИМАНОЇ МОДЕЛІ	21

ТЕОРИТИЧНІ ДАНІ

Кореляційно-регресійний аналіз — це побудова та аналіз економіко-математичної моделі у вигляді рівняння регресії (рівняння кореляційного зв'язку), що виражає залежність результативної ознаки від однієї або кількох ознак-факторів і дає оцінку міри щільності зв'язку.

Кореляційно-регресійний аналіз складається з таких етапів:

- ✓ попередній (априорний) аналіз,
- ✓ збирання інформації та її первинна обробка,
- ✓ побудова моделі (рівняння регресії),
- ✓ оцінка й аналіз моделі.

Мультиколінеарність моделі означає існування лінійної залежності або сильної кореляції між двома чи більше факторами. Мультиколінеарність між факторами X_i та X_j ($i = j$) називається строгою, якщо існує лінійна залежність $X_j = cX_i$.

На практиці економічні фактори часто пов'язані між собою і це істотно впливає на якість економетричного моделювання.

Задачі роботи

- Обґрунтувати вибір вхідних параметрів і сформулювати таблицю з вхідними даними.
- Побудувати кореляційну матрицю засобами Excel.
- Розробити алгоритм розрахунку коефіцієнту кореляції.
- Проаналізувати результати розрахунку.

Ми вважаємо, що рівень доходів (РД) населення залежить від **заробітної плати (ЗП), прибутку та змішаного доходу (ПЗД), доходів від власності (ДВ), соціальних допомог (СД) та соціальних трансфертів в натурі (СТН).**

Тепер з допомогою даних Держкомстату України створимо таблиці 1, 2 та 3 з вхідними індикаторами, які характеризують рівень доходів населення у різних регіонах України. Будемо досліджувати 2013, 2016 і 2018 роки.

ТАБЛИЦІ ДАНИХ

Таблиця 1. Значення вхідних індикаторів за 2013 рік (у млн грн).

Область	РД	ЗП	ПЗД	ДВ	СД	СТН
Автономна Республіка Крим	57 324	21 010	8 914	2 389	12 573	8 442
Вінницька	46 157	15 135	10 869	2 119	10 640	6 349
Волинська	26 907	8 607	5 446	998	6 726	4 701
Дніпропетровська	124 594	57 783	16 924	7 707	25 394	13 947
Донецька	166 366	74 002	25 361	8 594	38 234	16 750
Житомирська	34 947	11 935	6 670	1 272	9 012	5 340
Закарпатська	29 102	9 192	5 850	790	6 782	5 347
Запорізька	62 671	25 941	11 360	3 370	12 906	7 681
Івано-Франківська	37 310	10 873	8 654	1 221	8 579	6 319
Київська	58 894	25 054	9 910	2 701	13 540	6 938
Кіровоградська	27 695	9 331	5 220	1 578	6 678	4 307
Луганська	71 485	29 941	9 070	3 073	18 978	8 759
Львівська	75 762	28 509	13 123	3 496	16 405	11 868
Миколаївська	35 125	13 704	5 704	1 845	7 732	4 897
Одеська	78 285	28 916	11 828	3 956	14 918	10 636
Полтавська	46 984	19 313	7 427	2 934	10 339	6 074
Рівненська	31 811	10 692	6 296	1 094	7 723	5 457
Сумська	33 469	12 284	6 486	1 618	7 672	4 692
Тернопільська	26 345	7 982	5 267	1 122	6 273	4 604
Харківська	91 333	35 883	16 218	4 744	17 932	13 409
Херсонська	29 489	9 061	6 312	1 290	6 583	4 280
Хмельницька	36 770	11 780	8 148	1 773	8 725	5 407
Черкаська	35 024	12 632	5 296	2 179	8 897	5 231
Чернівецька	22 408	6 431	4 794	795	5 210	3 926
Чернігівська	30 393	11 004	5 382	1 642	7 633	4 317
м. Київ	218 747	118 529	16 202	22 922	24 176	23 218
м. Севастополь	13 336	5 210	937	730	2 863	1 944

Таблиця 2. Значення вхідних індикаторів за 2016 рік (у млн грн).

«...» — відсутність даних з області.

Область	РД	ЗП	ПЗД	ДВ	СД	СТН
Автономна Республіка Крим
Вінницька	69 654	23 458	19 043	2 447	12 522	10 746
Волинська	39 359	13 537	9 255	1 071	8 064	6 733
Дніпропетровська	184 138	86 057	33 836	7 462	29 886	21 892
Донецька	111 547	55 007	11 260	3 370	27 512	11 588
Житомирська	51 920	18 436	11 822	1 424	10 509	8 634
Закарпатська	42 235	14 501	10 474	804	8 034	7 243
Запорізька	94 160	37 880	22 191	3 282	16 472	11 402
Івано-Франківська	54 492	16 483	15 608	1 219	10 155	8 950
Київська	87 937	39 426	17 543	2 252	16 030	11 529
Кіровоградська	40 427	14 247	9 073	1 999	7 899	6 418
Луганська	38 022	17 685	3 094	1 193	10 434	5 014
Львівська	112 697	44 323	24 725	3 707	19 520	17 117
Миколаївська	50 728	20 881	9 976	1 718	8 977	7 191
Одеська	115 025	44 524	21 667	3 675	17 654	13 606
Полтавська	69 789	28 707	12 483	4 335	12 319	10 502
Рівненська	45 716	16 201	11 027	1 133	8 966	7 609
Сумська	50 951	18 803	11 858	1 943	9 030	8 256
Тернопільська	38 727	12 275	8 922	1 072	7 330	7 506
Харківська	131 681	52 212	28 455	4 281	23 011	18 727
Херсонська	42 707	13 768	10 773	1 391	7 756	5 694
Хмельницька	55 542	18 123	14 690	2 180	10 433	8 786
Черкаська	51 710	18 901	8 840	3 004	10 792	8 929
Чернівецька	32 397	9 664	8 471	825	6 215	5 544
Чернігівська	44 283	16 288	8 934	1 797	8 985	7 566
м. Київ	333 927	185 379	34 193	17 868	29 268	33 378
м. Севастополь

Таблиця 3. Значення вхідних індикаторів за 2018 рік.

Область	РД	ЗП	ПЗД	ДВ	СД	СТН
Автономна Республіка Крим
Вінницька	114 480	45 927	28 588	3 593	16 185	17 361
Волинська	63 810	26 013	13 537	1 572	10 347	10 967
Дніпропетровська	307 844	157 772	53 211	7 725	43 590	34 490
Донецька	174 771	92 064	13 630	4 433	39 928	19 934
Житомирська	84 830	34 727	18 881	2 130	13 869	13 276
Закарпатська	69 194	28 437	16 030	846	10 306	11 825
Запорізька	147 627	67 656	30 873	3 932	22 430	18 620
Івано-Франківська	87 479	31 557	23 376	1 316	13 509	14 292
Київська	150 606	74 615	26 930	3 402	22 387	20 797
Кіровоградська	63 999	26 431	13 222	2 382	10 347	10 157
Луганська	58 880	29 470	3 837	1 193	14 971	8 638
Львівська	189 077	83 401	39 619	4 144	26 831	28 501
Миколаївська	81 581	37 991	14 839	2 031	12 226	11 009
Одеська	193 923	80 374	35 080	4 254	23 826	23 741
Полтавська	114 656	53 539	19 263	5 502	16 977	16 738
Рівненська	73 661	30 444	16 485	1 426	11 610	12 206
Сумська	79 848	33 411	17 306	2 663	12 234	12 521
Тернопільська	61 731	23 505	13 107	1 434	9 363	11 384
Харківська	216 227	96 555	41 142	6 859	31 726	32 606
Херсонська	68 064	25 098	15 128	1 582	10 318	10 128
Хмельницька	86 821	34 060	19 781	3 025	13 701	13 976
Черкаська	82 600	34 194	13 771	4 077	14 357	13 801
Чернівецька	52 108	18 522	12 561	830	7 957	9 043
Чернігівська	69 247	29 818	12 779	2 082	11 732	11 528
м. Київ	555 666	333 786	59 089	18 731	45 049	58 654
м. Севастополь

Маючи дані, можемо побудувати матриці кореляції в Excel. На рис. 1 наведені результати побудови матриці кореляції за даними 2013 року.

	РД	ЗП	ПЗД	ДВ	СД	СТН
РД	1					
ЗП	0,99091	1				
ПЗД	0,86543	0,79539	1			
ДВ	0,93304	0,96834	0,64618	1		
СД	0,89717	0,84241	0,954	0,68618	1	
СТН	0,98146	0,96141	0,8742	0,90531	0,87525	1

Рис. 1. Результати побудови матриці кореляції за даними 2013 року

Коефіцієнти кореляції виділені різними кольорами. Покажемо межі значень:

- 0 ... 0.3 — зв'язку немає (червоний);
- 0.3 ... 0.5 — слабкий зв'язок (жовтий);
- 0.5 ... 0.7 — середній зв'язок (світло-синій);
- 0.7 ... 0.9 — сильний зв'язок (блакитний);
- 0.9 ... 1 — дуже сильний зв'язок (синій).

Знаючи це, поглянемо на рис.1. Аж 7 пар індикаторів дають результат більше за 0.9, тобто дуже сильний зв'язок між ними. Між ЗП та РД коефіцієнт максимально наближений до 1 (0,990914718), це можна пояснити тим, що зарплата є основним джерелом доходів населення. Дві пари показують середній зв'язок (див. рис. 1). Інші пари мають коефіцієнти 0.7...0.9, що каже нам про сильний зв'язок між ними. Нийнижчий коефіцієнт — 0,646182572 між ДВ і ПЗД.

Отже, існує залежність між індикаторами на високому рівні.

Перейдемо до наступної таблиці даних.

	РД	ЗП	ПЗД	ДВ	СД	СТН
РД	1					
ЗП	0,99175	1				
ПЗД	0,85722	0,79012	1			
ДВ	0,96522	0,97811	0,74999	1		
СД	0,8644	0,82694	0,82523	0,74439	1	
СТН	0,97946	0,9532	0,91357	0,93171	0,8666	1

Рис. 2. Результати побудови матриці кореляції за даними 2016 року

На рис. 2 (дані за 2016 рік), кореляція між ЗП та РД знову найвища (0,991748727). Знову бачимо 7 пар величин, що дуже сильно між собою корелюють (хоч і дещо інших порівняно з попередньою матрицею): показники більші за 0.9. Інші пари показують значення кореляції 0.7–0.9, що показує сильний зв'язок між ними. Нийнижчий коефіцієнт — 0,744390433 між ДВ і СД.

Так, залежність між індикаторами є на дуже високому рівні (особливо порівнюючи з 2013 роком).

Розглянемо останній набір даних — за 2018 рік.

	РД	ЗП	ПЗД	ДВ	СД	СТН
РД	1					
ЗП	0,99452	1				
ПЗД	0,89356	0,84961	1			
ДВ	0,97392	0,97967	0,82253	1		
СД	0,87784	0,85713	0,80784	0,82953	1	
СТН	0,99342	0,98789	0,89994	0,97221	0,83702	1

Рис. 3. Результати побудови матриці кореляції за даними 2018 року

Тут 6 пар мають коефіцієнти більші за 0.9, тобто мають дуже сильний зв'язок між собою. Вже третє пара ЗП та РД має коефіцієнт майже 1 (0,9945195). До речі, тут вже з'явився конкурент у найкращої пари, а саме — пара СТН/РД (0,993422925). Серед інших пар показники у межах 0.8–0.9, що вказує на сильний зв'язок між ними. Нийнижчий коефіцієнт — 0,807839911 має пара СД/ПЗД.

Отже, залежність залишається на високому рівні.

АЛГОРИТМ РОЗРАХУНКУ КОЕФІЦІЄНТА КОРЕЛЯЦІЇ



ВИСНОВКИ

Результати, показані на рис. 1–3 свідчать, що у всіх взятих роках (а саме 2013, 2016 і 2018) найбільший вплив на доходи населення мала заробітна плата (ЗП), а найменший — прибуток та змішаний дохід (ПЗД) у 2012 і 2016 роках та соціальні допомоги (СД) у 2018 році. Провівши аналіз результатів кореляції можна помітити, що вплив майже усіх показників є дуже високим (коефіцієнти більші за 0.85). Вплив ЗП протягом усіх років був найсильнішим: коефіцієнт вище за 0.99. Варто зазначити, що на рівні з ЗП у 2018 році на РД впливали й СДН (0,993422925).

Вплив прибутку та змішаних доходів і соціальних допомог був нижче за вплив інших індикаторів. Це прослідковується постійно, тому можна сказати про це як про закономірність.

Результати розрахунків матриць на рис. 1–3 показують, що між показниками впливу існує сильний кореляційний зв'язок, тому для побудови оптимальної економетричної прогнозної моделі необхідно надалі провести мультиколінеарний аналіз.

МУЛЬТИКОЛІНЕАРНИЙ АНАЛІЗ

Задачі:

1. Побудувати блок-схему алгоритму мультиколінеарного аналізу методом Фаррара-Глобера.
2. Обчислити середні значення за кожним фактором.
3. Обчислити дисперсію за кожним фактором.
4. Сформувати нормалізовану матрицю.
5. Розрахувати кореляційну матрицю на основі нормалізованих даних.
6. Розрахувати детермінант кореляційної матриці.
7. Обчислити критерій «хі-квадрат».
8. Визначити обернену матрицю.
9. Обчислити F-критерій (Фішера).
10. Обчислити t -критерій Стюдента.
11. Вибір показників для регресійної моделі.

Розглянемо мультиколінеарний аналіз для даних за 2016 рік, де значення коефіцієнтів кореляції між факторами найбільші.

Блок-схему алгоритму мультиколеніарного аналізу методом Фаррара-Глобера наведено на рис. 1.

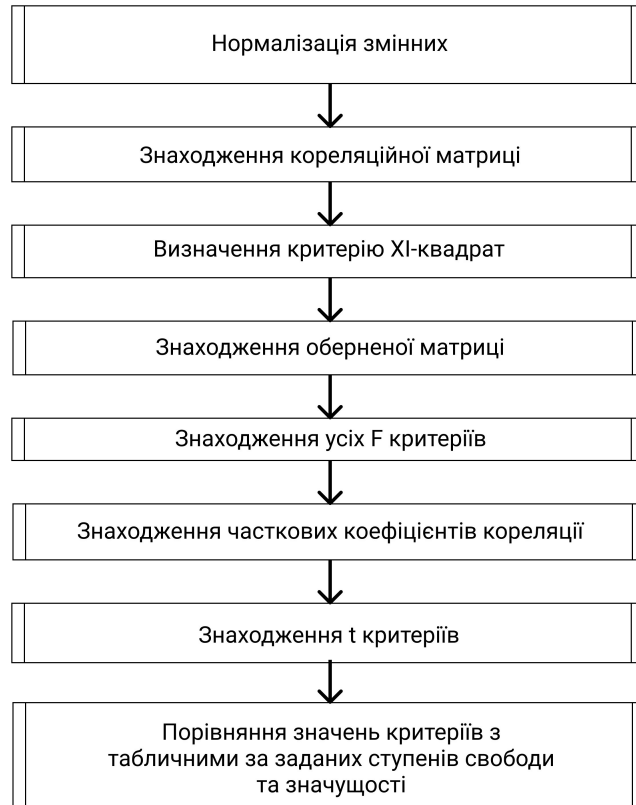


Рис. 1. Блок-схема алгоритму мультиколеніарного аналізу методом Фаррара-Глобера

Згідно з рис. 2, для обчислення *середнього значення відповідного фактору* необхідно за допомогою функції «СРЗНАЧ» розрахувати середнє значення відповідного діапазону. Для першого стовпчика необхідно виділити комірку В31 та ввести формулу «=СРЗНАЧ(В3:В27)», де В3:В27 — діапазон значень першого стовпчика. Для п'яти інших колонок операція аналогічна.

Під кожною коміркою із середнім значенням необхідно розрахувати дисперсію. Для першого стовпчика (рис. 2) вибираємо клітинку В32 і вводимо формулу «=ДИСПР(В3:В27)». Теж саме в наступні п'ять комірок.

Регіон	РД	ЗП	ПЗД	ДВ	СД	СТН
Вінницька	69,654	23,458	19,043	2,447	12,522	10,746
Волинська	39,359	13,537	9,255	1,071	8,064	6,733
Дніпропетровська	184,138	86,057	33,836	7,462	29,886	21,892
Донецька	111,547	55,007	11,260	3,370	27,512	11,588
Житомирська	51,920	18,436	11,822	1,424	10,509	8,634
Закарпатська	42,235	14,501	10,474	804	8,034	7,243
Запорізька	94,160	37,880	22,191	3,282	16,472	11,402
Івано-Франківська	54,492	16,483	15,608	1,219	10,155	8,950
Київська	87,937	39,426	17,543	2,252	16,030	11,529
Кіровоградська	40,427	14,247	9,073	1,999	7,899	6,418
Луганська	38,022	17,685	3,094	1,193	10,434	5,014
Львівська	112,697	44,323	24,725	3,707	19,520	17,117
Миколаївська	50,728	20,881	9,976	1,718	8,977	7,191
Одеська	115,025	44,524	21,667	3,675	17,654	13,606
Полтавська	69,789	28,707	12,483	4,335	12,319	10,502
Рівненська	45,716	16,201	11,027	1,133	8,966	7,609
Сумська	50,951	18,803	11,858	1,943	9,030	8,256
Тернопільська	38,727	12,275	8,922	1,072	7,330	7,506
Харківська	131,681	52,212	28,455	4,281	23,011	18,727
Херсонська	42,707	13,768	10,773	1,391	7,756	5,694
Хмельницька	55,542	18,123	14,690	2,180	10,433	8,786
Черкаська	51,710	18,901	8,840	3,004	10,792	8,929
Чернівецька	32,397	9,664	8,471	825	6,215	5,544
Чернігівська	44,283	16,288	8,934	1,797	8,985	7,566
м. Київ	333,927	185,379	34,193	17,868	29,268	33,378
Сер. Значення	79590.84	33470.64	15128.52	3018.08	13510.92	10822.4
Дисперсія	4020481466	1271361492	64067000.4	11371532.6	48703691.5	37889463

Рис. 2. Скріншот даних, середніх значень та дисперсії

Нормалізована матриця має розмір вихідної (25 × 6). Для розрахунку першого нормалізованого значення виберемо клітинку Т3 та задамо формулу «=(В3-В\$31)/КОРЕНЬ(25*В\$32)», де В\$31 — зафіксоване середнє значення, а В\$32 — зафіксована дисперсія, 25 — кількість регіонів.

Отримавши результат у комірці Т3, тягнемо формулу вправо та вниз. Маємо таблицю, що є нормалізованою матрицею (рис. 3).

Нормалізована матриця:					
-0.031343	-0.05616	0.097811	-0.03387	-0.028341	-0.00248
-0.1269	-0.11181	-0.146761	-0.115479	-0.156099	-0.13287
0.329764	0.294964	0.467442	0.263564	0.46928	0.359669
0.1007965	0.1208	-0.096662	0.020872	0.401246	0.024876
-0.08728	-0.08433	-0.08262	-0.094543	-0.08603	-0.0711
-0.117828	-0.1064	-0.116302	-0.131315	-0.156959	-0.1163
0.0459542	0.024733	0.17647	0.015653	0.084859	0.018832
-0.079167	-0.09529	0.011981	-0.106702	-0.096175	-0.06084
0.0263256	0.033404	0.06033	-0.045435	0.072192	0.022959
-0.123531	-0.10783	-0.151309	-0.060441	-0.160828	-0.14311
-0.131117	-0.08854	-0.300706	-0.108244	-0.088179	-0.18872
0.1044239	0.060872	0.239787	0.040859	0.172209	0.204521
-0.09104	-0.07062	-0.128746	-0.077106	-0.129934	-0.11799
0.1117669	0.062	0.163377	0.038961	0.118733	0.090444
-0.030917	-0.02672	-0.066103	0.078105	-0.034158	-0.01041
-0.106848	-0.09687	-0.102484	-0.111802	-0.130249	-0.10441
-0.090336	-0.08227	-0.08172	-0.063762	-0.128415	-0.08339
-0.128893	-0.11889	-0.155082	-0.11542	-0.177134	-0.10776
0.1643034	0.105123	0.332988	0.074903	0.272255	0.256833
-0.116339	-0.11051	-0.108831	-0.0965	-0.164926	-0.16663
-0.075855	-0.08609	-0.010957	-0.049706	-0.088208	-0.06617
-0.087942	-0.08172	-0.157131	-0.000835	-0.077919	-0.06152
-0.148859	-0.13353	-0.166351	-0.130069	-0.209088	-0.1715
-0.111368	-0.09638	-0.154782	-0.072421	-0.129705	-0.10581
0.8022303	0.852074	0.476363	0.880734	0.45157	0.732867

Рис. 3. Нормалізована матриця

Для розрахунку кореляційної матриці на основі нормалізованих даних необхідно визначити квадратний діапазон комірок розміром 6×6. Нехай кореляційна матриця буде зберігатися в комірках K16:P21. В рядку формул задаємо вираз «=МУМНОЖ(ТРАНСП(Т3:Y27); Т3:Y27)» (рис. 4).

Кореляційна матриця:						
	1	0.991445	0.849304723	0.965253	0.864522	0.979032
	0.991445	1	0.779155253	0.977904	0.82689	0.951778
	0.849305	0.7791553	1	0.74235	0.817881	0.908749
	0.965253	0.977904	0.742349528	1	0.744634	0.931238
	0.864522	0.8268902	0.81788136	0.744634	1	0.866312
	0.979032	0.9517777	0.908748634	0.931238	0.866312	1

Рис. 4. Коефіцієнти кореляційної матриці

Детермінант квадратної матриці в Ексел визначимо за допомогою функції «МОПРЕД». Виділимо комірку M23 і введемо формулу «=МОПРЕД(K16:P21)», де K16:P21 — діапазон комірок кореляційної матриці (рис. 5).

Детермінант	6.7746E-08
-------------	------------

Рис. 5. Детермінант кореляційної матриці

Для розрахунку фактичного значення χ^2 -критерію у комірку M25 вводять формулу « $=(25-1-(2*6+5)/6)*LN(M23)$ », де 25 — кількість об'єктів, 2*6 — подвоєна кількість факторів, M23 — детермінант кореляційної матриці (рис. 6).

Хі2-критерій	349.4086077
--------------	-------------

Рис. 6. Фактичне значення χ^2 -критерію

Для визначення мультиколінеарності розраховане значення необхідно порівняти з табличним при визначеному ступені вільності та заданому рівні значимості. Табличне значення розрахуємо до комірки M27 за допомогою формули «=ХИ2.ОБР(0,05;6*(6-1)/2)» (рис. 7).

Хі2-критерій	349.4086077
Хі2 табл	7.260943928

Рис. 7. Табличне та фактичне значення χ^2 -критерію

Порівнюючи фактичне значення із табличним, приходимо до висновку, що між показниками існує мультиколінеарність.

Для визначення оберненої матриці до матриці кореляції нормалізованого набору даних необхідно виділити квадратний діапазон комірок K31:P36. В рядку формул введемо вираз «=МОБР(K16:P21)» (рис. 8).

Обернена матриця:						
	810.5706	-583.48876	-109.9012788	-29.92474	-41.65373	-74.3983
	-583.489	468.7509	84.67381564	-12.16629	20.45137	41.77289
	-109.901	84.673816	26.47675558	9.188742	5.451805	-10.3343
	-29.9247	-12.166287	9.188742104	40.31333	10.98613	-14.5321
	-41.6537	20.45137	5.45180493	10.98613	8.614397	-1.3326
	-74.3983	41.772888	-10.33433932	-14.53211	-1.332598	58.15849

Рис. 8. Обернена матриця до матриці кореляції

Розрахуємо фактичні значення F-критерію. Результати зберігаємо в масив комірок K40:P40. При цьому всі значення розраховуються окремо, оскільки в розрахунках необхідно використовувати діагональні елементи оберненої матриці. Перше значення для комірки K40 обчислюється за допомогою виразу «=(K31-1)*(25-6)/(6-1)» (рис. 9).

Фактичні значення F-критерію:						
	3076.368	1777.4534	96.81167121	149.3907	28.93471	217.2023

Рис. 9. Фактичні значення F-критеріїв

Табличне значення F-критерію розраховується за допомогою статистичної функції «=FРАСПОБР(0,05;(6-1);(25-6))» (рис. 10).

Фактичні значення F-критерію:						
	3076.368	1777.4534	96.81167121	149.3907	28.93471	217.2023
	2.740058					

Рис. 10. Фактичні й табличне значення F-критеріїв

Всі фактичні критерії більші за критичне значення, отже всі вони є значимими.

Частинні коефіцієнти кореляції розраховуються на основі значень елементів оберненої матриці за формулою: $F_{ij} = -\frac{A_{ij}}{\sqrt{A_{ii}A_{jj}}}$,

де A_{ij} — алгебраїчне доповнення елемента k_{ij} кореляційної матриці K .

Значення t-критеріїв Стюдента розраховується на основі частинних коефіцієнтів кореляції за формулою $t_{ij} = F_{ij} \frac{\sqrt{22}}{\sqrt{1-F_{ij}^2}}$.

На цьому етапі тестування мультиколінеарності має сенс лише для пар різних показників, тому пропускають розрахунки діагональних елементів. Табличне значення t-критерію Стюдента в Excel обчислюється за допомогою статистичної формули «СТЮДЕНТ.ОБР» (рис. 11).

Частинні коеф. кореляції						Факт знач t-крит							
	-1	0.94659894	0.750196345	0.1655429	0.498478	0.3426583	РД	ЗП	ПЗД	ДВ	СД	СТН	
	-1	-0.760056171	0.0885039	-0.3218392	-0.252998			13.771	5.32161563	0.7873281	2.6970385	1.7107805	РД
		-1	-0.2812544	-0.3609903	0.2633563				-5.48578741	0.4167557	-1.5943901	-1.2265694	ЗП
			-1	-0.5895328	0.3001219					-1.374692	-1.8156226	1.2804524	ПЗД
				-1	0.0595361						-3.4233034	1.4757261	ДВ
					-1							0.2797451	СД
						-1							СТН
							крит t						
							-2.07387						

Рис. 11. Матриця частинних коефіцієнтів та t-критеріїв

Порівнявши значення t-критеріїв з критичним значенням робимо висновок, що між всіма парами факторів, крім ПЗД-ЗП і СД-ДВ, є мультиколінеарність, оскільки відповідні фактичні значення більші за критичну величину (рис. 11).

Для виключення із складу регресійної моделі обираємо:

— ЗП, оскільки значення фактичного критерію Стюдента для пари РД-ЗП приймає найбільше значення, а саме 13,77;

— ПЗД та СТН з найнижчими показниками порівняння з критичним значенням.

Взагалі тут можна було виключати чи не всі параметри, але ми користуємося принципом вибору найменш придатного кандидата.

Так, згідно з мультиколінеарним аналізом, основними параметрами економетричної прогновної моделі є: РД, ДВ, СД.

РЕГРЕСІЙНИЙ АНАЛІЗ

Побудувати множинну лінійну регресійну модель в Excel можна:

1. Засобами матричні функції МУМНОЖ та МОБР.
2. Засобом "Регрессия" із надбудови "Пакет анализа".

ПЕРШИЙ СПОСІБ

Використовуючи таблицю з вихідною інформацією (рис. 12) підготувати матрицю значень, у якій в першому стовпчику містяться одиниці (тобто назви регіонів), в наступних – значення стовпчиків факторів впливу, останній стовпчик містить значення залежного (прогнозованого) фактору.

Регіон	ДВ	СД	РД
1	2,447	12,522	69,654
1	1,071	8,064	39,359
1	7,462	29,886	184,138
1	3,370	27,512	111,547
1	1,424	10,509	51,920
1	804	8,034	42,235
1	3,282	16,472	94,160
1	1,219	10,155	54,492
1	2,252	16,030	87,937
1	1,999	7,899	40,427
1	1,193	10,434	38,022
1	3,707	19,520	112,697
1	1,718	8,977	50,728
1	3,675	17,654	115,025
1	4,335	12,319	69,789
1	1,133	8,966	45,716
1	1,943	9,030	50,951
1	1,072	7,330	38,727
1	4,281	23,011	131,681
1	1,391	7,756	42,707
1	2,180	10,433	55,542
1	3,004	10,792	51,710
1	825	6,215	32,397
1	1,797	8,985	44,283
1	17,868	29,268	333,927

Рис. 12. Вихідні дані

Матриця коефіцієнтів рівняння регресії в Excel визначається за допомогою множення попередньо обчисленої матриці та оберненої матриці, використовуючи функцію МУМНОЖ. Для цього необхідно обчислити транспоновану матрицю до створеної. Спершу виділяємо квадратний діапазон комірок розміром 4×4 (відповідно до кількості

стовпців нової матриці). Для першої комірки G3 вводимо формулу «=МУМНОЖ(ТРАНСП(A2:D28);A2:D28)», де A2:D28 – діапазон значень матриці. Приклад розрахованої транспонованої матриці наведено на рис. 13.

Транспонована матриця:			
25	75452	337773	1989771
75452	512008486	1457526082	11165047752
337773	1457526082	5781216269	36447560645
1989771	11165047752	36447560645	2.5888E+11

Рис. 13. Скріншот розрахунку транспонованої матриці

Обернена матриця (рис. 14) має той самий розмір як і вихідна (4 × 4). Обернену матрицю бажано розмістити під вихідною, наприклад, в діапазоні комірок G10:J13.

Для розрахунку першого значення необхідно вибрати клітинку G10 та задати формулу «=МОБР(G3:J6)», де G3:J6 — діапазон значень транспонованої матриці.

Обернена матриця:			
2.18E-01	4.57E-06	-1.85E-05	7.36E-07
4.57E-06	9.68E-08	1.66E-08	-6.55E-09
-1.85E-05	1.66E-08	6.11E-09	-1.44E-09
7.36E-07	-6.55E-09	-1.44E-09	4.83E-10

Рис. 14. Скріншот розрахунку оберненої матриці

Наступним кроком є побудова матриці, яка є результатом множення транспонованої матриці на матрицю залежної змінної.

Результат виводиться в комірках G15:G18 (рис. 15). Виділяємо першу комірку і вводимо значення «=МУМНОЖ(ТРАНСП(A2:D28);D2:D28)».

1989771
11165047752
36447560645
2.5888E+11

Рис. 15. Результат множення транспонованої матриці на матрицю залежної змінної

Матриця коефіцієнтів рівняння регресії в Excel визначається за допомогою множення попередньо обчисленої матриці та оберненої матриці, використовуючи функцію МУМНОЖ. Виділимо діапазон комірок G22:J25 і введемо формулу «=МУМНОЖ(G10:J13; G15:G18)», де G10:J13 — діапазон комірок оберненої матриці, G15:G18 — діапазон комірок попередньо обчисленої матриці.

Результат побудови матриці коефіцієнтів рівняння регресії на рис. 16.

Матриця коеф. регресії:			
1.34E-09	1.34E-09	1.34E-09	1.34E-09
0.00	0.00	0.00	0.00
-5.68E-14	-5.68E-14	-5.68E-14	-5.68E-14
1	1	1	1

Рис. 16. Матриця коефіцієнтів рівняння регресії

Згідно з рис. 16 рівняння регресії матиме вигляд:

$$РД = 1,34 * 10^{-9} - 5,68 * 10^{-14} * СД.$$

ДРУГИЙ СПОСІБ

Для побудови лінійної регресійної моделі необхідно:

1) Викликати засіб "Регрессия" із надбудови "Пакет анализа", натиснувши на кнопку "Анализ данных" на закладці "Данные" панелі інструментів Excel.

2) Викликати Сервис – Анализ данных – Регрессия – ОК. З'явиться вікно для надання вхідних даних. У вікні "Регрессия" вибрати вхідний інтервал для прогнозованого фактора у вікні вибору "Входной интервал Y" (комірки D2:D26). Задати інтервал для пояснювальних факторів у вікні вибору "X" (комірки B2:C26). У полі "Новый рабочий лист" ввести коротку назву листа. Натиснути кнопку "ОК".

Результати роботи функції "Регрессия" наведені на рис. 17.:

Регрессионная статистика								
Множественный R	0.989647343							
R-квадрат	0.979401863							
Нормированный R-квадрат	0.977529305							
Стандартная ошибка	9700.891928							
Наблюдения	25							
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	2	98441675957	49220837978	523.0288806	2.83209E-19			
Остаток	22	2070360693	94107304.2					
Итого	24	1.00512E+11						
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	-1523.528351	4516.37195	-0.337334561	0.739061974	-10889.9105	7842.853802	-10889.9105	7842.853802
Переменная X 1	13.56884942	0.861982738	15.74143984	1.85347E-13	11.78120663	15.3564922	11.78120663	15.3564922
Переменная X 2	2.972595153	0.416511626	7.136883983	3.72034E-07	2.108802909	3.836387397	2.108802909	3.836387397

Рис. 17. Результати регресійного аналізу

На рис. 17 у графі «Коеффициенты» вказані значення параметрів моделі. Отже, побудована лінійна регресійна модель має вигляд:

$$РД = - 1523,53 + 13,57 * ДВ + 2,97 * СД.$$

Для перевірки статистичної значущості моделі надається значення F- статистики у графі F : $F = 523,03$.

Коефіцієнт детермінації моделі R^2 надається у графі R -квадрат, $R^2 = 0,98$.

На рис. 18 наведені основні графіки, що надані пакетом.

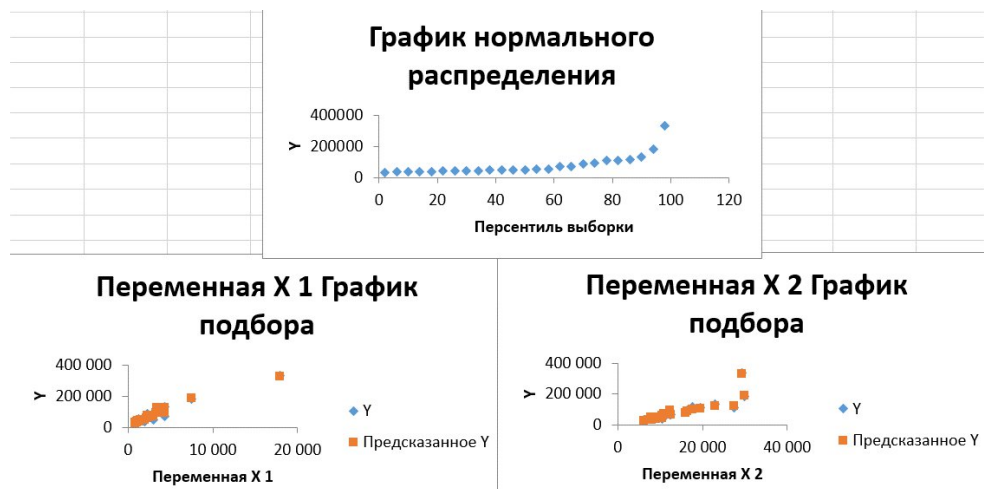


Рис. 18. Графіки для різних параметрів

ПЕРЕВІРКА ОТИМАНОЇ МОДЕЛІ

На рис. 19 наведено значення фактичних і прогнозованих регіональних показників рівня доходів населення за 2018 рік згідно економетричної прогнозної моделі.

	A	B	C	D	E
1	ДВ	СД	Фактичні значення	Прогнозовані	Відсоток відхилення
2	4,407	19,422	129,061	142182.28	10.17%
3	2,234	12,449	72,185	78031.33	8.10%
4	10,748	51,386	360,385	373405.6	3.61%
5	6,260	48,301	199,322	283801.77	42.38%
6	2,676	16,837	97,301	102334.97	5.17%
7	1,127	11,880	78,182	56657.3	27.53%
8	4,708	26,406	169,384	173877	2.65%
9	1,768	15,857	98,587	82857.89	15.95%
10	4,001	27,609	173,511	166162.85	4.23%
11	2,945	12,205	71,713	89449.53	24.73%
12	1,609	17,854	66,287	87666.28	32.25%
13	5,687	31,909	216,876	211728.81	2.37%
14	2,807	14,290	92,529	94956	2.62%
15	5,515	27,912	225,458	193594.26	14.13%
16	6,659	20,375	129,647	184888.87	42.61%
17	1,737	14,106	82,555	75683.44	8.32%
18	3,108	14,822	89,702	102197.64	13.93%
19	1,860	11,235	68,282	66937.63	1.97%
20	8,771	38,359	245,934	289712.55	17.80%
21	1,899	11,966	76,449	70385.16	7.93%
22	3,333	16,634	97,560	112971.12	15.80%
23	4,739	17,504	92,887	140676.58	51.45%
24	1,173	9,307	58,028	47704.19	17.79%
25	2,545	14,140	76,808	89839.18	16.97%
26	24,687	55,772	675,427	632009.58	6.43%
27					
28				Середнє:	15.88%

Рис. 19. Перевірка моделі на адекватність

Згідно з рисунком, відсоток у багатьох випадках дає непогані результати. Середнє відхилення — 15,88%, тобто прогнозна модель працює на достатньому рівні. Має сенс спробувати використати у моделі й інші чинники — це дасть змогу покращити результати.

Усі матеріали до лабораторної роботи викладено на системі контролю версій GitHub за посиланням: <https://github.com/lakub-muravlov/fourth-course-projects/tree/main/EEOV/Lab1>