https://orcid.org/0000-0002-9399-191X

УДК 519.237.5

С.М. ЛАПАЧ*

ТРИ СПОСОБИ КОДУВАННЯ НОМІНАЛЬНИХ ЗМІННИХ У РЕГРЕСІЙНОМУ АНАЛІЗІ

*Національний технічний університет України «КПІ імені Ігоря Сікорського», м. Київ, Україна

Анотація. *У роботі виконується порівняння трьох способів кодування номінальних змінних у регресійному аналізі: кодування кожного рівня як окремої змінної, кодування двійковим кодом, нумерація рівнів фактора. Хоча способи існують досить давно і навіть мають теоретичне обґрунтування (крім кодування двійковим кодом), для практичного застосування були відсутні їх порівняння і рекомендації про застосування. Аналізуються особливості застосування кожного способу, обмеження. На двох прикладах виконується детальне порівняння цих трьох способів. Порівняльний аналіз виконується за такими напрямами: наявність обмежень у застосуванні; статистичні властивості плану; трудомісткість і складність отримання моделей і кінцевий результат побудови; зручність смислового аналізу і користування. Додатково виконується порівняння з моделями, побудованими на основі поліномів Чебишова. Встановлено, що різні способи кодування номінальних змінних приводять при їх правильному застосуванні до приблизно однакових за статистичними властивостями регресійних моделей. При цьому спосіб кодування кожного рівня як окремої змінної можливий тільки в тому випадку, коли ми маємо експерименти, в яких номінальна змінна (як ефект впливу) відсутня. Спосіб кодування двійковими числами незручно використовувати при великій кількості рівнів варіювання номінальної змінної і важко аналізувати. При кодуванні нумерацією рівнів необхідно, щоб середні значення відгуку за діаграмою розсіяння цього фактора були впорядковані за величиною відповідно до присвоєних номерів. При цьому способі кодування зберігається природна кількість факторів. Відмінно кращі результати досягаються способом нумерації рівнів при застосуванні ортогональних поліномів Чебишова. При цьому забезпечуються найвища точність і найбільша рівномірність апроксимації.*
Ключові слова: *регресійний аналіз,номінальні змінні, чисельне кодування номінальних змінних, поліноми Чебишова, вироджена матриця нормальних рівнянь, інформативність моделі, діаграми розсіяння, точність апроксимації, рівномірність апроксимації.*

Abstract. *The paper compares three methods of coding nominal variables in regression analysis: coding of each level as a separate variable, coding with binary code, numbering of factor levels. Although these methods have existed for a long time and even have a theoretical justification (except for encoding with binary code), there were no recommendations and comparisons for their practical application. The features of the application of each method and the existing limitations are analyzed. In the article, there are considered two examples that provide a detailed comparison of these three methods. Comparative analysis has been carried out in the following areas: the presence of restrictions in use; statistical properties of plans; labour intensity and difficulty of obtaining mathematical models and the final result of their building; convenience of semantic analysis and use. Additionally, there have been made comparisons with models based on Chebyshev orthogonal polynomials. It has been established that different methods of coding nominal variables, when used correctly, lead to regression models that are approximately identical in their properties. Moreover, the method of encoding each level as a separate variable is possible only if there are experiments in which there is no nominal variable as an influence effect. The binary coding method is inconvenient to use with a large number of levels of variation of the nominal variable and inconvenient to analyze. When coding by level numbering, it is necessary that the average response values, according to the dispersion diagram of this factor, are sorted by value in accordance with the assigned numbers. With this encoding method, a natural number of factors is preserved. Sharply*

distinguishable best results are achieved with this coding method using Chebyshev orthogonal polynomials. The highest accuracy and uniformity of approximation are ensured.
Keywords: *regression analysis, nominal variables, numerical coding of nominal variables, Chebyshev polynomials, degenerate matrix of normal equations, model informativeness, scatter diagram, approximation accuracy, approximation uniformity.*

DOI: 10.34121/1028-9763-2021-4-35-45

1. Вступ

У задачах регресійного аналізу частим випадком є наявність факторів різної природи, в тому числі і нечислових. Для використання таких змінних у регресійній моделі необхідне їх чисельне кодування. На даний момент не існує єдиного підходу до виконання цієї операції. Не існує також і порівняння різних способів кодування для обґрунтування вибору, придатного для розв'язання конкретної задачі. В деяких роботах зустрічаються також твердження про правильність тільки одного (в різних дослідників, звісно, іншого) способу. Крім того, для дослідника, який не є спеціалістом із математичної статистики, а тільки користувачем нею як деяким інструментом, спроби використання «за аналогією» деяких способів без розуміння їх обмежень можуть призвести до поганих результатів.

Метою даної статті є порівняльний аналіз існуючих способів кодування номінальних (за іншими найменуваннями – якісних, нечислових) факторів у регресійному аналізі і розробка рекомендацій щодо їх вибору у прикладних задачах.

2. Способи кодування номінальних змінних

У літературі можемо знайти три способи кодування номінальних змінних, описаних нижче.

1. Кожному рівню варіювання (значенню фактора) номінальної змінної відповідає один новий фактор у робочій матриці, значення якого дорівнює одиниці, якщо в даному експерименті значення номінальної змінної приймає даний рівень, і нуль – в іншому разі [1]. По суті, кожен новий фактор несе інформацію про наявність в експерименті даного рівня варіювання. Кількість факторів у робочій матриці дорівнює кількості значень даної номінальної змінної. В роботі скорочено позначається як ОРОФ.

2. Кожен рівень кодується у двійковій формі таким чином, щоб у кожному експерименті не було більше однієї одиниці (один рівень кодується, як набір нулів) [2]. Тобто, фактично фактор кодується як двійкове число (номер). Кількість факторів дорівнює числу рівнів варіювання даної номінальної змінної мінус одиниця. Надалі позначається КДЧ.

3. Кожному рівню номінального фактора відповідає один числовий рівень [3–5]. Кількість факторів у робочій матриці дорівнює кількості номінальних змінних. Такий спосіб теоретично обґрунтовано в [6, 7]. Надалі скорочено позначається НРВ.

Для наочності порівняння виконується на реальних прикладах.

3. Приклад 1

Розглянемо задачу з [2]. В роботі досліджується вплив на параметр шорсткості крихких матеріалів при обробці алмазними кругами таких факторів, як середня зернистість і матеріал, який обробляється. X_1 – середній розмір зерна, а X_2 – матеріал, який підлягає обробці (M_1, M_2, M_3). В табл. 1 представлене оригінальне кодування (за роботою) двійковими числами.

Таблиця 1 – Кодування двійковими числами (КДЧ)

№ експерименту	X_1	X_{20}	X_{21}	$У$
1	90	1	0	7,50
2	71,5	1	0	5,5
3	50	1	0	3,7
4	34	1	0	3,4
5	24	1	0	2,7
6	17	1	0	2,3
7	90	0	1	5
8	71,5	0	1	3
9	50	0	1	2,2
10	34	0	1	1,5
11	24	0	1	1,3
12	17	0	1	1
13	112,5	0	0	7,25
14	71,5	0	0	6,8
15	50	0	0	2,65
16	34	0	0	2
17	17	0	0	1,46
18	12	0	0	1,35

Розглянемо кодування цієї задачі різними способами. Звернемо увагу на принципові обмеження двох перших способів. Перший спосіб (ОРОФ) можливий тільки в тому випадку, коли ми маємо експерименти, в яких номінальна змінна (як ефект впливу) відсутня. Наприклад, вивчення дії трьох препаратів на певну хворобу. Ми маємо три групи пацієнтів, в яких лікували кожну різним препаратом, і контрольну групу, яка не лікувалася. Приклад подібного експерименту приведено в [1]. Якщо ж у нас такі експерименти відсутні, то при запропонованому способі кодування ми отримаємо погано обумовлену матрицю. Це пов'язано з тим, що в такому випадку один із факторів завжди буде лінійною комбінацією всіх інших. Детальніше пояснення приведено далі на прикладі.

Візьмемо цю ж задачу з [2] і виконаємо кодування за першим способом (ОРОФ) (табл. 2).

Таблиця 2 – Кодування один рівень – один фактор (ОРОФ)

№ експерименту	X_1	X_{200}	X_{201}	X_{202}	$У$
1	90	1	0	0	7,50
2	71,5	1	0	0	5,5
3	50	1	0	0	3,7
4	34	1	0	0	3,4
5	24	1	0	0	2,7
6	17	1	0	0	2,3
7	90	0	1	0	5
8	71,5	0	1	0	3
9	50	0	1	0	2,2
10	34	0	1	0	1,5
11	24	0	1	0	1,3
12	17	0	1	0	1
13	112,5	0	0	1	7,25

Продовження табл. 2

14	71,5	0	0	1	6,8
15	50	0	0	1	2,65
16	34	0	0	1	2
17	17	0	0	1	1,46
18	12	0	0	1	1,35

Звертаємо увагу на те, що при побудові кореляційної матриці в Excel здається, що все в порядку (табл. 3).

Таблиця 3 – Матриця кореляцій до табл. 2

	X_1	X_{200}	X_{201}	X_{202}	$У$
X_1	1				
X_{200}	-0,01416	1			
X_{201}	-0,01416	-0,5	1		
X_{202}	0,02832	-0,5	-0,5	1	
У	0,880929	0,277572	-0,35164	0,07407	1

Але спроба побудувати регресійну модель буде невдалою (табл. 4).

Таблиця 4 – Фрагмент побудови регресійної моделі за табл. 2 з використанням Excel

	Коефіцієнти	Стандартна похибка	t -статистика	P -значення
Y -пересічення	0,477739	0,40359825	1,18369989	0,25624154
X_1	0,062773	0,00571124	10,991123	2,8629E-08
X_{200}	0,708186	0,40747659	1,73797958	0,10414965
X_{201}	-1,141814	0,40747659	-2,8021586	0,0141202
X_{202}	0	0	65535	#ЧИСЛО!

Це пов'язано з лінійною залежністю між стовпчиками матриці: $X_{202} = I - X_{200} - X_{201}$. Оскільки оцінки регресійних коефіцієнтів знаходяться як корені системи так званих нормальних лінійних рівнянь B = $(X^TX)^{-1}X^TY$ (при цьому $(X^T X)^{-1}(X^T X) = I$), то в результаті такої залежності матриця $X^T X$ буде виродженою. Більшість сучасних програм масового застосування про це не попереджають у процесі роботи [8], хоча деякі мають засоби для перевірки поганої обумовленості. Таке кодування у наведених умовах призводить до поганої ситуації: результат ніби є, але він непридатний для практичного застосування і навіть може бути небезпечний за своїми наслідками при використанні на практиці. Задачу, подібну до цієї, неможливо привести до способу ОРОФ навіть переформулюванням: обробляти матеріал, якого немає, – нонсенс.

Розглянемо цю ж задачу щодо моделювання шорсткості з кодуванням іншими допустимими способами і побудуємо регресійні моделі. Варіанти кодування приведені в табл. 5.

Для наочності висновків за подальшими діями на рис. 1 і 2 приведені маргінальні залежності відгуку від кожної з незалежних змінних. Ці залежності допоможуть зрозуміти неоднозначність у деяких ситуаціях способу кодування нумерацією рівнів варіювання.

Таблиця 5 – Варіанти кодування при дослідженні шорсткості

№ експерименту	Відгук і числова змінна		Спосіб кодування двійковими числами		Два варіанти кодування нумерацією рівнів	
	$У$	X_1	X_{20}	X_{21}	$X_{2в1}$	$X_{2в2}$
1	7,50	90	1	0	3	3
2	5,5	71,5	1	0	3	3
3	3,7	50	1	0	3	3
4	3,4	34	1	0	3	3
5	2,7	24	1	0	3	3
6	2,3	17	1	0	3	3
7	5	90	0	1	2	1
8	3	71,5	0	1	2	1
9	2,2	50	0	1	2	1
10	1,5	34	0	1	2	1
11	1,3	24	0	1	2	1
12	1	17	0	1	2	1
13	7,25	112,5	0	0	1	2
14	6,8	71,5	0	0	1	2
15	2,65	50	0	0	1	2
16	2	34	0	0	1	2
17	1,46	17	0	0	1	2
18	1,35	12	0	0	1	2

За приведеними таблицями були побудовані регресійні моделі. Оскільки якість побудови регресійних моделей залежить не тільки від плану експерименту, а й від перетворень вихідної матриці і структури моделі [3, 4, 8, 10], додатково побудовані моделі з використанням поліномів Чебишова, базуючись на кодуванні нумерацією рівнів.

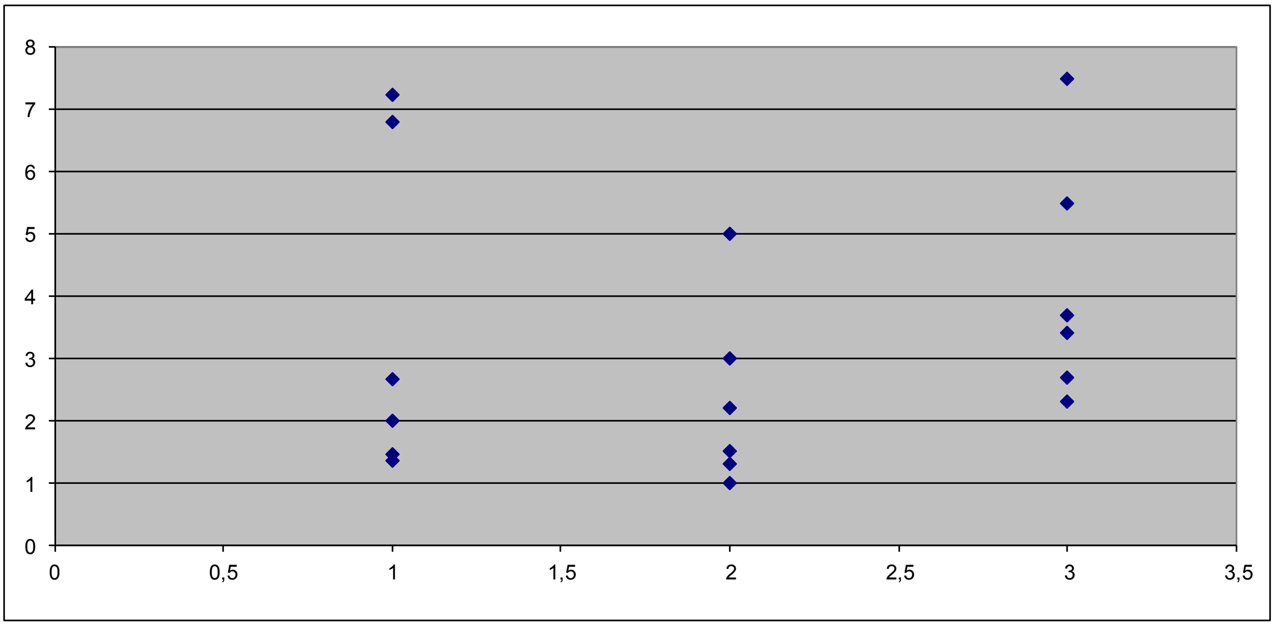

Рисунок 1 – Діаграма розсіяння для матеріалу обробки в задачі дослідження шорсткості

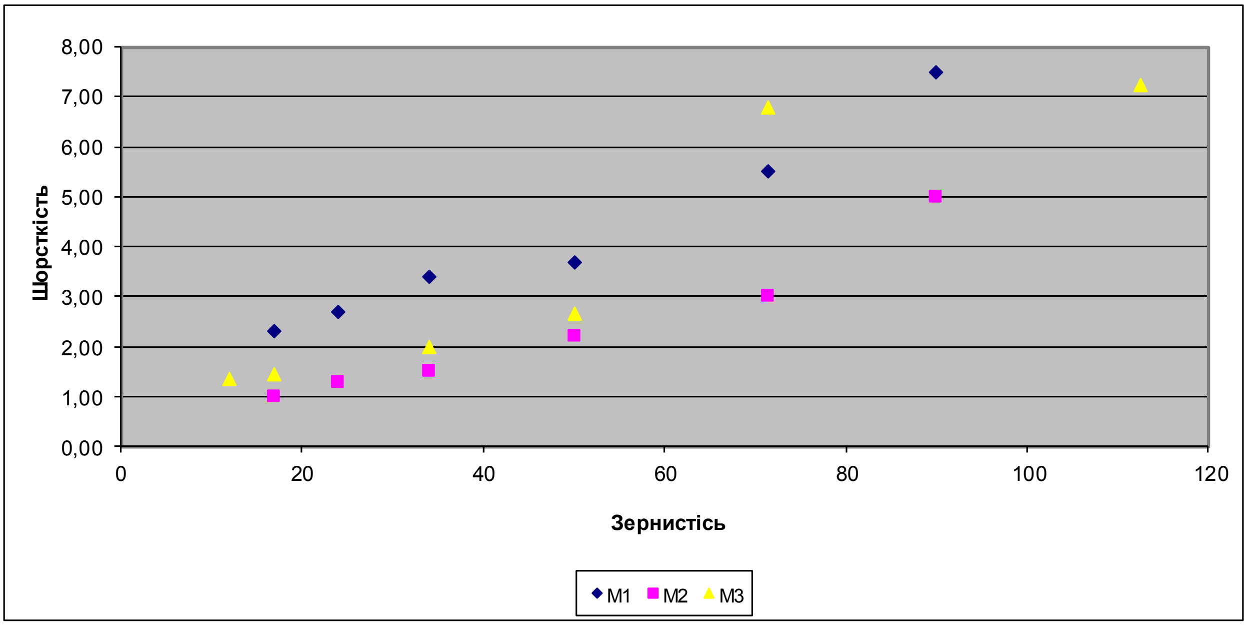

Рисунок 2 – Залежність шорсткості від зернистості для різних оброблюваних матеріалів

У табл. 6 узагальнено результати побудови регресійних моделей при різних способах кодування і додаткових перетвореннях вихідної матриці. В таблиці прийняті такі позначення способів кодування:

1 – модель за [2] у відповідності з кодуванням автора (спосіб кодування двійковими числами);

2 – перший варіант способу кодування нумерацією рівнів варіювання (НРВ);

3 – другий варіант способу кодування нумерацією рівнів варіювання (НРВ);

4 – спосіб кодування нумерацією рівнів (НРВ) з побудовою ортогональних контрастів (поліномів Чебишова);

5 – спосіб кодування нумерацією рівнів (НРВ) з побудовою ортогональних контрастів (поліномів Чебишова) і формуванням самої інформативної моделі.

Таблиця 6 – Характеристики регресійних моделей шорсткості

Характеристики моделі	Модель				
	1	2	3	4	5
R (коефіцієнт множинної кореляції)	0,95416	0,89185	0,952894	0,959635	0,95685
F_R (розрахункове Фішера для значимості R)	47,43	29,16	74,03	27,94	50,60
F_{kp} (критичне Фішера)	3,34	3,68	3,68	3,11	3,34
S^2 залишкова дисперсія	0,497812	1,061235	0,47715	0,51285	0,469268
% середнє відхилення апроксимації	15,74	27,96	15,37	13,94	14,25

Продовження табл. 6

Max % максимальне відхилення апроксимації		59,69	60,98	52,78	36,54	29,01
t_{kp} критичне Стьюдента для значимості коефіцієнтів регресії		2,14	2,13	2,13	2,18	2,14
Розрахункові t для коефіцієнтів регресії	t_{b1}	10,99	7,569706	11,24886	10,27451	11,3683
	t_{b2}	1,74	1,19135	4,638793	4,497318	4,702811
	t_{b3}	2,80	–	–	0,59769	1,119485
	t_{b4}	–	–	–	1,070568	–
	t_{b5}	–	–	–	0,66578	–

Звертаємо увагу на варіант 2, який виділяється найгіршими показниками (за множинним коефіцієнтом кореляції, залишковою дисперсією, точністю апроксимації і формально незначущим номінальним фактором). Цей варіант показує залежність результатів моделювання від порядку нумерації рівнів факторів при кодуванні способом нумерації рівнів варіювання при нелінійній залежності відгуку від номінального фактора (рис. 1). Всі інші варіанти за статистичними характеристиками приблизно однакові, за винятком значущості факторів. І це теж може мати суттєві наслідки, адже в багатьох підручниках і монографіях рекомендується залишати в моделі тільки ті регресори, коефіцієнти регресії, при яких статистично значимі. Автор особисто (і не тільки він [9]) не підтримує даний підхід щодо формування структури моделі. Звернемо увагу на цю сторону досліджуваних моделей.

При способі КДЧ за другий фактор відповідають дві змінні. Одна з них визначається за моделлю незначимою, а друга значимою. При даному кодуванні це фактично означає статистично значиму різницю при переході від матеріалу М2 до М3 і незначиму при переході від М1 до М3. При способі кодування нумерацією рівнів (стовпчики 3–5 табл. 6) другий фактор визначається як значущий (тобто вибір матеріалу впливає статистично значущо), при цьому складова другого порядку формально незначима, але підвищує точність опису.

Звернемо увагу також на описові характеристики, на які мало звертають теоретики, але які суттєво важливі для практиків. Якщо середня точність апроксимації приблизно однакова у всіх моделей, то максимальне відхилення дуже відрізняється. Моделі з використанням поліномів Чебишова забезпечують значно менше значення цієї характеристики, а у випадку вибору найбільш інформативної моделі майже удвічі менше. Тобто, поліноми Чебишова забезпечують більш рівномірну апроксимацію у всьому діапазоні, що взагалі теоретично відомо [11, 12].

3. Приклад 2

Розглянемо тепер задачу з [1]. У табл. 7 представлені різні види кодування для даної задачі. Вони позначені таким чином: 1 – спосіб кодування один рівень – один фактор; 2 – спосіб кодування двійковими числами; 3 і 4 спосіб кодування нумерацією рівнів із різним порядком нумерації рівнів фактора.

Статистичні характеристики побудованих регресійних моделей приведено в табл. 8.

Таблиця 7 – Матриця кодування до прикладу за Карлсбергом [1]

№ експерименту	Відгук	Спосіб кодування						
		1			2		3	4
1	164	0	0	0	0	0	0	0
2	141	0	0	0	0	0	0	0
3	144	0	0	0	0	0	0	0
4	138	0	0	0	0	0	0	0
5	153	0	0	0	0	0	0	0
6	153	1	0	0	0	1	1	2
7	191	1	0	0	0	1	1	2
8	192	1	0	0	0	1	1	2
9	126	1	0	0	0	1	1	2
10	162	1	0	0	0	1	1	2
11	165	0	1	0	1	0	2	3
12	168	0	1	0	1	0	2	3
13	175	0	1	0	1	0	2	3
14	189	0	1	0	1	0	2	3
15	182	0	1	0	1	0	2	3
16	150	0	0	1	1	1	3	1
17	132	0	0	1	1	1	3	1
18	123	0	0	1	1	1	3	1
19	155	0	0	1	1	1	3	1
20	159	0	0	1	1	1	3	1

На рис. 3 представлена діаграма розсіяння для пояснення особливостей кодування способом нумерації рівнів варіювання.

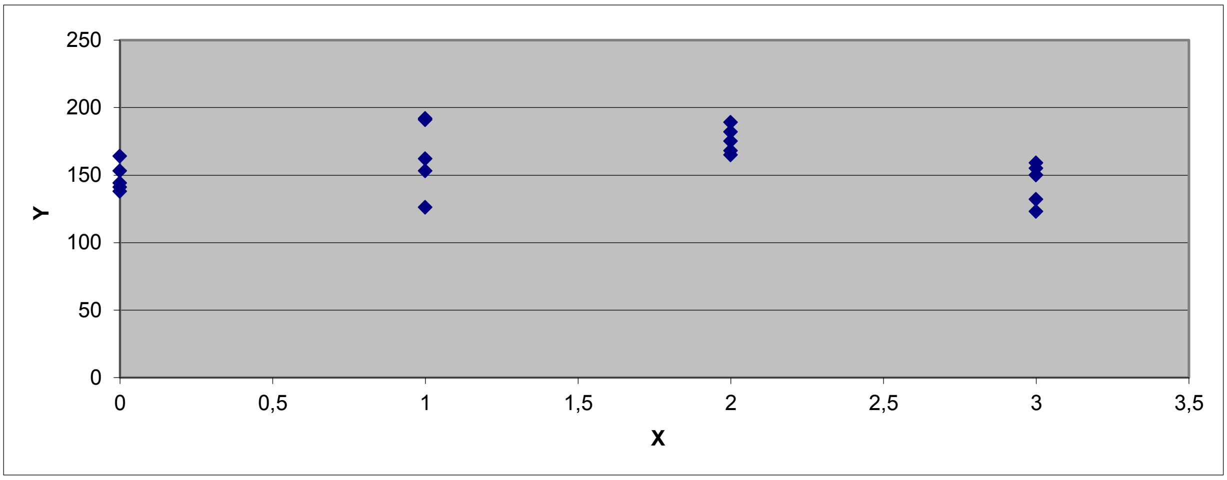

Рисунок 3 – Діаграма розсіяння до задачі із Карлсберга [1]

У табл. 8 прийняті такі позначення для моделей за прикладом із Карлсберга [1]:
1 – кодування способом один рівень – один фактор;
2 – невдале кодування способом нумерації рівнів варіювання;
3 – правильне кодування способом нумерації рівнів варіювання;
4 – кодування двійковим кодом;
5 – ортогональний поліном при способі кодування нумерацією рівнів варіювання.

Таблиця 8 – Статистичні характеристики моделей за прикладом із Карлсберга [1]

Характеристика	Модель				
	1	2	3	4	5
R	0,636401	0,008831	0,576249	0,608121	0,606018
F_R	3,63	0,001404	8,95	4,99	4,93
F_{kp}	3,24	3,16	3,16	3,20	3,59
S^2	305,15	455,8422	304,4978	304,1882	305,4212
%	8,25	10,76	8,06	8,24	8,02
Max %	-38,8	28,34	24,29	33,81	16,92
t_{kp}	2,12	2,10	2,10	2,11	2,11
t_{b1}	1,520627	-0,03747	2,99	3,128266	2,986902
t_{b2}	2,516275			0,435906	0,972409
t_{b3}	-0,38016				

Звернемо увагу на колонку 2, де приведено результати помилкової нумерації за способом нумерацією рівнів. На відміну від попереднього випадку модель абсолютно невдала і зовсім непридатна. Правильне кодування виконується таким чином. Будується діаграма розсіяння для даного номінального фактора (рис. 1, 3). Потім виконується нумерація рівнів фактора таким чином, щоб вона відповідала впорядкуванню середнього значення рівнів діаграми розсіяння за величиною.

Всі інші моделі мають приблизно однакові статистичні характеристики. Найвищу інформативність має модель про кодуванні способом нумерацією рівнів без використання ортогональних контрастів. І знову ж найбільш рівномірна апроксимація забезпечується при використанні поліномів Чебишова. Слід зауважити, що спосіб нумерацією рівнів забезпечує в цілому більш високі показники апроксимації, ніж інші способи. Це видно і на попередньому прикладі, хоч розрив не надто визначений.

Для даного прикладу на рис. 4 і табл. 9 подається порівняльна діаграма розподілу коефіцієнтів кореляції між регресорами за функцією «Аналіз плану» ПРІАМ [13].

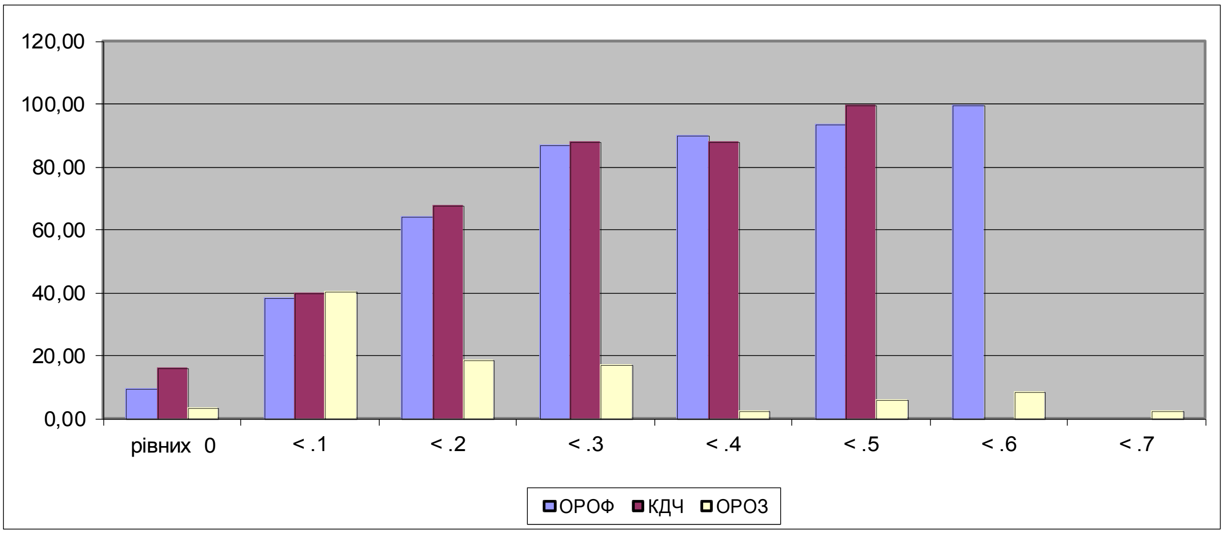

Рисунок 4 – Розподіл значень коефіцієнтів кореляції для різних видів кодування номінальної змінної

Таблиця 9 – Розподіл значень коефіцієнтів кореляції для різних видів кодування номінальної змінної

Спосіб коду-вання	Інтервал								Середнє коеф. кор.	Сер. кв. відхил. коеф. кор.
	$=0$	$\le 0,1$	$\le 0,2$	$\le 0,3$	$\le 0,4$	$\le 0,5$	$\le 0,6$	$\le 0,7$		
КДЧ	9,68	29,03	25,81	22,58	3,23	3,23	6,45	0,00	0,090682	0,153953
ОРОФ	3,70	40,74	18,52	17,28	2,47	6,17	8,64	2,47	0,154242	0,194528
НРВ	16	24	28	20	0	12	0	0	0,055572	0,081791

За узагальненими характеристиками найкращі показники має спосіб нумерацією рівнів, за ним спосіб кодування двійковими числами і найгірший спосіб один рівень – один фактор.

4. Порівняння способів кодування

Порівнювати вище описані способи кодування будемо за такими характеристиками:

• наявність обмежень у застосуванні;
• статистичні властивості отриманого плану;
• якість отримуваних моделей;
• зручність смислового аналізу і користування.

Порівняльні характеристики зведені в табл. 10.

Таблиця 10 – Порівняння характеристик різних видів кодування

Вид кодування	Переваги	Недоліки
Один рівень – один фактор (ОРОФ)	• Простота аналізу з відповідністю природному процесу відносно окремих рівнів фактора	• Неможливість використання при відсутності експериментів без ефектів. • Велика кількість факторів замість одного. • Найгірша рівномірність апроксимації. • Найгірші статистичні характерристики плану
Кодування двійковими числами	• Менша кількість факторів порівняно зі способом один рівень – один фактор	• Складності в інтерпретації при кількості рівнів більше 2. • При кількості рівнів більше 2 порушується природність або співпадає зі способом ОРОФ
Нумерація рівнів фактора	• Природна кількість факторів. • Спрощений аналіз відносно фактора в цілому. • Сама висока рівномірність апроксимації. • Найвища точність апроксимації. • Найкращі статистичні характерристики плану	• Більш складний аналіз відносно окремих рівнів фактора. • При нелінійній залежності результат залежить від порядку нумерації. • Необхідність враховувати природу фактора при обчислювальному експерименті

5. Висновки

1. Різні методи кодування номінальних змінних при їх правильному застосуванні приводять до приблизно однакових за статистичними властивостями регресійних моделей.
2. Вибір методу кодування залежить від особливостей задачі і мети дослідження.
3. Спосіб кодування один рівень – один фактор можливий тільки в тому випадку, коли ми маємо експерименти, в яких номінальна змінна (як ефект впливу) відсутня.
4. Спосіб кодування двійковими числами незручно використовувати при великій кількості рівнів варіювання номінальної змінної.
5. Найкращі результати досягаються способом нумерації рівнів варіювання. Нумерація рівнів варіювання номінальної змінної повинна бути виконана так, щоб середні значення відгуку за діаграмою розсіяння цього фактора були впорядковані за величиною відповідно до присвоєних номерів. При цьому способі кодування зберігається природна кількість факторів.
6. Суттєві кращі по точності і рівномірності апроксимації досягаються способом кодування рівнів варіювання при застосуванні ортогональних поліномів Чебишова.

СПИСОК ДЖЕРЕЛ

1. Карлсберг К. Регрессионный анализ в Microsoft Excel. СПб.: Альфа книга, 2017. 400 с.
2. Рыжов Э.В., Горленко О.А. Математические методы в технологических исследованиях. К.: Наукова думка, 1980. 184 с.
3. Лапач С.Н., Пасечник М.Ф., Чубенко А.В. Статистические методы в фармакологии и маркетинге фармацевтического рынка. К.: ЗАТ “Укрспецмонтаж”, 1999. 312 с.
4. Радченко С.Г. Математическое моделирование технологических процессов в машиностроении К.: ЗАО «Укрспецмонтажпроект», 1998. 274 с.
5. Хикс Ч. Основные принципы планирования эксперимента. М.: Мир, 1967. 406 с.
6. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: исследование зависимостей. М.: Финансы и статистика, 1986. 487 с.
7. Розин Б.Б., Ягольницер М.А. Конструирование экономико-статистических моделей с заданными свойствами. Новосибирск: Наука, 1981. 175 с.
8. Огороднік С.В., Лапач С.М. Поліноми Чебишова і звичайні поліноми в регресії при використанні масових програмних засобів. *Загально університетська науково-технічна конференція молодих вчених та студентів*. Київ, 2012. С. 105–107.
9. Pardoux C. Sur la selection de variables en regression multiple. *Cah. Bur. Univ. rech. oper*. 1982. N 39–40. P. 101–133.
10. Лапач С.Н. Регрессионный анализ. Процессный подход. *Математичні машини і системи*. 2016. № 1. С. 129–138.
11. Суетин П.К. Классические ортогональные многочлены. М.: ГРФМЛ, 1976. 328 с.
12. Пашковский С. Вычислительные применения многочленов и рядов Чебышева. М.: ГРФМЛ, 1983. 384 с.
13. Лапач С.Н., Радченко С.Г., Бабич П.Н. Планирование, регрессия и анализ моделей PRIAM (ПРИАМ). *Каталог программные продукты Украины*. К., 1993. С. 24–27.

Стаття надійшла до редакції 25.10.2021