

МОДЕЛЮВАННЯ ЕКОНОМІЧНИХ ЗАЛЕЖНОСТЕЙ В MS EXCEL ЗА ЕМПІРИЧНИМИ ДАНИМИ З ВИКОРИСТАННЯМ ПОКРОКОВИХ АЛГОРИТМІВ

Наукові праці МАУП, 2014, вип. 43(4), с. 158–162

Проблема визначення ефективного набору предикторів у задачах регресійного аналізу має прикладне значення. Використання покрокових процедур, у тому числі з можливістю примусової заміни змінної, є одним із способів розв'язання задачі. Розробка відповідних надбудов MS Excel дає змогу полегшити економіко-математичне моделювання для користувачів MS Office.

Аналіз даних з використанням ЕОМ давно став буденною справою. Застосування методів багатовимірного статистичного аналізу просто неможливе без використання потужної комп'ютерної техніки та спеціалізованих програм (пакетів). На сьогодні існує багато статистичних інструментів, як платних, так і безплатних, що можуть прийти на допомогу прикладному статистику чи просто рядовому менеджеру торгового підприємства або банку.

Пакет офісних програм від Microsoft поступово (разом з Windows) втрачає свою гегемонію, але ще й досі встановлений на більшості комп'ютерів у світі. Секрет такої “живучості”, зокрема, криється у популярності програми MS Excel, яка має певні засоби (статистичні) для роботи з даними. Серед цих інструментів можна виділити надбудову “Пакет аналізу”, яка містить засоби для побудови лінійної множинної регресії. Робота присвячена проектуванню та розробці надбудови для розширення функціональних можливостей MS Excel, які не представлено у “Пакеті аналізу”, а саме, стандартної покрокової регресії з можливістю заміни змінних на кожному кроці.

Проблема знаходження (відновлення) залежностей між різного роду показниками є однією з основних у прикладному аналізі. До-

слідники завжди намагалися відшукати існуючу залежність, опираючись тільки на окремі розрізнені факти. Наприклад, в [3] виділено три напрями: задача навчання розпізнавання образів, задача відновлення регресії, задача інтерпретації результатів непрямих експериментів. До першого типу можна, наприклад, віднести побудову оцінки для баєсового порогу, побудованої за методом мінімізації емпіричного ризику для вибірки із суміші зі змінними концентраціями [4], і у [5] – для випадку двох порогів. Задачу відновлення регресії також можна звести, як показано в [3], до схеми мінімізації середнього ризику. На практиці ж користуються методом найменших квадратів для оцінки коефіцієнтів регресії, наприклад, [6] або [7]. З розвитком ЕОМ особливої популярності серед дослідників набувають покрокові алгоритми для конструювання ефективної множини предикторів.

Уперше покрокова процедура послідовного включення-видалення змінних була запропонована А. Ефроїмсоном в [1]. Алгоритм, описаний в [2], має незначні відмінності від алгоритму Ефроїмсона і називається стандартною покроковою процедурою або ж F-методом. Окремо слід відзначити, що можливість заміни змінної на кроці не входить до стандартної покрокової процедури.

Відзначимо, що опція заміни змінної відсутня у більшості прикладних статистичних пакетів, що обмежує при побудові моделі. Справді, спеціалісту-практику часто видніше, які ж саме предиктори є найважливішими

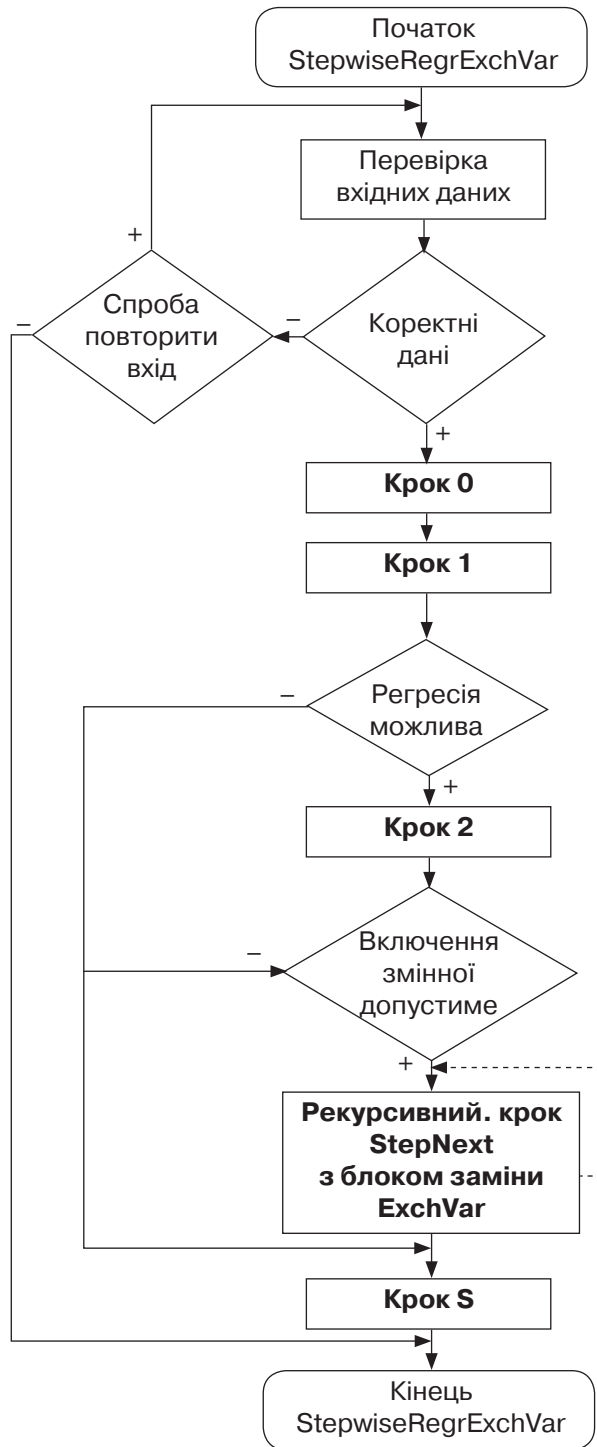


Рис. 1. Загальна схема системи

ми у його задачі. Тобто при побудові моделі працюють не тільки статистичні критерії, а й досвід та інтуїція дослідника.

Здійснимо дослідження існуючих покрокових процедур та правил зупинки для побудови регресії з можливістю заміни змінної на кроці та розробку відповідних надбудов MS Excel.

Розглядається зв'язок між однією змінною, яку називають залежною змінною та однією чи більше іншими змінними, які називаються незалежними змінними. *Математичною моделлю* є рівняння, що пов'язує залежну змінну з незалежними з урахуванням певних припущень. *Функція регресії* зв'язує незалежні змінні з залежною. Причому, функція регресії також залежить від невідомих параметрів. Якщо функція лінійна відносно цих параметрів (але необов'язково лінійна відносно незалежних змінних), то кажуть про *лінійну модель регресії*. В іншому випадку модель називають *нелінійною*. Але в будь-якому разі говорять про *регресію* залежної змінної по незалежним змінним. У роботі розглядаються виключно лінійні моделі. Таким чином, *регресійний аналіз* — метод дослідження взаємозв'язку між двома або більше змінними.

Покрокові процедури

Статистикою, що вимірює ефективність набору незалежних змінних як предикторів залежної змінної Y , є множинний коефіцієнт

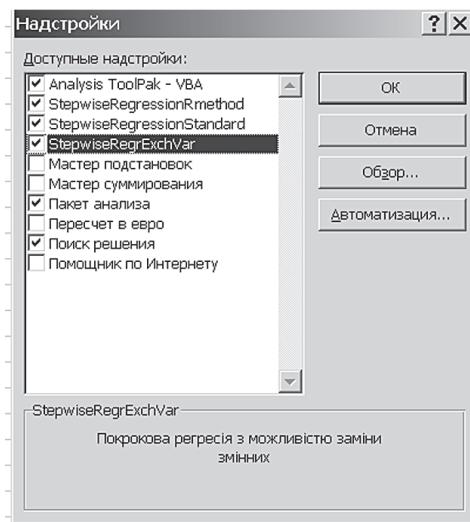


Рис. 2. Надбудова StepwiseRegrExchVar

№	Тип функції	Ім'я	Дія
1	Function	AddSummaryTable	Додає запис з результатами поточного кроку до підсумкової таблиці
2	Function	CalcFenter	Обчислює F -включення для змінних, які не увійшли до моделі
3	Function	CalcFleave	Обчислює F -видалення для тих змінних, які увійшли до моделі
4	Function	CalcMaxFenter	Повертає максимум F -включення та номер змінної, на якій цей максимум досягнуто
5	Function	CalcMinFleave	Повертає мінімум F -виключення і номер змінної, на якій його досягнуто
6	Function	ContinueExchUnsucc	Інформує користувача про те, що остання заміна була небажаною, оскільки зменшився множинний коефіцієнт кореляції (точніше, коефіцієнт детермінації)
7	Function	ContinueExchVar	Запитує користувача, чи бажає він замінити змінну, що входить на даний момент у модель, на змінну, що в нього не входить
8	Function	ContinuePrintSummary	Запитує користувача, чи бажає він побачити підсумкову таблицю
9	Function	ContinueProcedure	Перехоплює ситуацію некоректно заданих вхідних даних користувачем
10	Function	ContinueProcedureNotExistRegr	Перехоплює ситуацію, коли регресія не має смислу для вибраного користувачем максимуму рівня значущості альфа
11	Sub	ExchVar	Модуль заміни змінних. Дає можливість перебору у діалоговому режимі всіх пар, трійок і т. д. змінних для моделі покрокової регресії. Вперше підключається після другого кроку (в моделі УЖЕ має міститися ДВІ змінні)
12	Function	IColon	Повертає i -ту колонку двовимірного масиву
13	Function	InModel	Змінна x В моделі
14	Function	MakeSummaryTable	Повертає заголовок підсумкової таблиці
15	Function	MakeTitleLinest	Створює заголовок з іменами незалежних змінних, що увійшли до моделі на кроці
16	Function	MakeVarOut	Створює масив маркерів для порожнього набору змінних, що увійшли до моделі
17	Function	OutModel	Змінна x НЕ в моделі
18	Function	PartCor	Обчислення часткового коефіцієнта кореляції використовуючи рекурсію
19	Sub	PrintMatrToRange	Переносить двовимірний масив у діапазон
20	Sub	PrintResultsStep	Дає можливість користувачу вибрати комірку на листі для відображення результатів на кроці та відображає їх
21	Sub	PrintSummary	Переносить підсумкову таблицю на лист Excel. Причому дає можливість користувачу вибрати місце під таблицю
22	Sub	ShowVar	Демонструє користувачу, які змінні увійшли, а які не увійшли на даний момент. Тим самим дає можливість вибору для заміни

23	Sub	StepNext	Рекурсивний крок
24	Function	TitleAddToLinest	Додає до таблиці (з результатами функції ЛИНЕЙН) заголовки з іменами змінних
25	Function	Xln	Формує масив з тих змінних, що увійшли до моделі

кореляції. Тому, серед всіх підмножин незалежних змінних X_1, X_2, \dots, X_p розмірності k , $k = 1, \dots, p$ вибирається підмножина S_k , якій відповідає максимальне значення цього коефіцієнта. Але, у випадку великої кількості незалежних змінних такий підхід для визначення оптимальної множини складно реалізувати, оскільки доведеться змоделювати $2^p - 1$ регресійних рівнянь. Одним з виходів є покрокова регресія, коли незалежні змінні включаються, замінюються, видаляються з множини, що будується відповідно до наперед заданого критерію. Сукупність цих критеріїв називається покроковою процедурою. У роботі використано стандартну покрокову процедуру (F-метод), яка передбачає F-видалення і F-включення змінних, причому з можливістю перебору усіх двійок, трійок і

т. д. незалежних змінних на кожному кроці для врахування проблематики прикладної задачі. Окремо розглядається модифікація F-методу – R-метод.

ВВА-програми

Надбудова – це програма, яка додається в Excel з метою розширення його функціональності [8]. Прикладами надбудов, що входять до стандартної поставки Excel, є *Поиск решения* і *Анализ данных*. Розроблено нові надбудови для статистичного аналізу даних:

1. StepwiseRegressionStandard.
2. StepwiseRegressionRmethod.
3. StepwiseRegrExchVar.

StepwiseRegressionStandard реалізує стандартну покрокову процедуру (F-метод). StepwiseRegressionRmethod реалізує метод множинної кореляції (R-метод). Stepwise-

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	X1	X2	X3	Y		x3	x0						
2	0,5	13	40	50		-0,2797	64,56947						
3	0,5	7	21	52		0,143261	3,75473						
4	1	15	45	51		0,275984	5,131282						
5	1	6,5	19	55		3,811846	10						
6	2	6	18	56		100,3661	263,3006						
7	1,5	10	31	54		x3	x2	x0					
8	2	5,3	16	60		-0,90082	1,737022	64,38268					
9	1	5,8	17	58		0,304298	0,780489	3,179763					
10	0,5	7,3	15	62		0,532996	4,344008	#Н/Д					
11	1	17	34	64		5,135899	9	#Н/Д					
12	1,5	3,8	11	65		193,833	169,8337	#Н/Д					
13	2	7,3	22	67		x3	x2	x1	x0				
14						-0,94647	2,067354	4,383674	57,30495				
15						0,258613	0,678946	2,055855	4,274414				
16						0,702229	3,679154	#Н/Д	#Н/Д				
17						6,288761	8	#Н/Д	#Н/Д				
18						255,3773	108,2894	#Н/Д	#Н/Д				
19						Крок	Число змі	Включена	Виключен	F-включеі	F-видалеі	R ² між	Y та включеними
20						0	0	-	-	-	-	-	-
21						1	1	3	-	3,811846	-	0,275984	-
22						2	2	2	-	4,953093	-	0,532996	-
23						Заміна	-	1	3	-	-	0,203681	-
24						3	3	1	-	4,546653	-	0,702229	-

Рис. 3. Покрокова регресія та підсумкова таблиця, у якій також відображено заміну

RegrExchVar доповнює StepwiseRegression-Standard блоком заміни змінної на кроці.

Продемонструємо можливості надбудов на прикладі StepRegrExchVar. На (рис. 1) зображено загальний вигляд системи у нотаціях [2]. У таблиці наведено функції, підпрограми та їх призначення.

Найкращий спосіб завантажити\вивантажити надбудову — використати вікно Excel Надстройки, доступ до якого отримується за допомогою команди *Сервис\Надстройки*. Припустимо, що необхідний нам код VBA скопільовано, а xls-файл з робочою книгою збережено як xla-файл надбудови MS Excel. Тоді надбудову StepwiseRegrExchVar буде видно у вікні диспетчера надбудов (рис. 2). Для зручності, у підпису вікна диспетчера будемо бачити інформацію про призначення цієї надбудови.

У підсумковій таблиці (рис. 3), яка, до речі, друкується за бажанням, відображено всі етапи побудови регресії, у тому числі інформація про зроблені заміни.

Висновки

У роботі реалізовано стандартну покрокову процедуру (*F*-метод), яка передбачає *F*-видалення і *F*-включення змінних разом із можливістю заміни змінних на кожному кроці. Заміна змінних дає змогу перебрати всі двійки, трійки, четвірки і т. д. незалежних змінних, виходячи з особливостей конкретної прикладної задачі. Тобто за такого підходу, статистична “раціональність” поєднується з досвідом та інтуїцією дослідника.

Програмну VBA-реалізацію (надбудову Excel) можна використовувати на практиці

для статистичного аналізу реальних даних. Відзначимо, що численні статистичні пакети, маючи у своєму арсеналі засоби побудови покрокової регресії, не підтримують функції заміни змінних. При відповідному доопрацюванні розроблена надбудова суттєво розширює статистичний інструментарій MS Excel і може мати комерційне значення.



Література

1. Efroimson M. A. Multiple regression analysis. — In: *Mathematical Methods for Digital Computers*, Ed. by Ralston A. and Wilf H.S., N.Y., 1960. — P. 191–203.
2. Афифи А. Статистический анализ: Подход с использованием ЭВМ / А. Афифи, С. Эйзен. — М.: Мир, 1982. — 488 с.
3. Вапник В. Н. Восстановление зависимостей по эмпирическим данным / В. Н. Вапник. — М.: Наука, 1979. — 448 с.
4. Кубайчук О. О. Асимптотика оцінки для баєсового порогу // *Вісн. Київ. нац. ун-ту ім. Тараса Шевченка. Математика. Механіка.* — 2008. — № 19. — С. 47–50.
5. Кубайчук О. О. Асимптотична поведінка оцінки методу мінімізації емпіричного ризику для баєсового порога // *Наук. вісті НТУУ “КПІ”.* — 2010. — № 4. — С. 78–85.
6. Кубайчук О. О. Практичні заняття з економетрії в Excel / О. О. Кубайчук, С. А. Теренчук. — К.: Вид-во Європейського ун-ту, 2007. — 118 с.
7. Міхайленко В. М. Теорія ймовірностей, ймовірнісні процеси та математична статистика / В. М. Міхайленко, С. А. Теренчук, О. О. Кубайчук. — К.: Вид-во Європейського ун-ту, 2007. — 163 с.
8. Уокенбах Д. Профессиональное программирование на VBA в Excel 2003 / John Walkenbach; [пер. с англ. И. Василенко]. — М.: Вильямс, 2005. — 800 с.

Розв’язано задачу побудови покрокової регресії в середовищі MS Excel. Для визначення ефективного набору предикторів розроблено надбудови MS Excel, які реалізують стандартну покрокову процедуру з можливістю заміни змінної та метод множинної кореляції.

Решена задача построения пошаговой регрессии в среде MS Excel. Для определения эффективного набора предикторов разработаны надстройки MS Excel, которые реализуют стандартную пошаговую процедуру с возможностью замены переменной и метод множественной корреляции.

Solved the problem of constructing stepwise regression among MS Excel. To determine the effective set of predictors designed MS Excel Add-ins, which implement the standard stepwise algorithm with replaceable variable and the method of multiple correlation.

Надійшла 18 вересня 2014 р.