УДК 378.22:159-051:[004]

***О.О. Карабин**, канд. фіз.- мат. наук, доцент, **О.Ю. Чмир**, канд. фіз.- мат. наук, доцент*
(Львівський державний університет безпеки життєдіяльності)

ВИКЛАДАННЯ БАГАТОФАКТОРНОГО КОРЕЛЯЦІЙНОГО АНАЛІЗУ З ВИКОРИСТАННЯМ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Висвітлюється питання можливості удосконалення методики викладання багатомірного кореляційного аналізу за допомогою використання прикладних математичних пакетів. Основним методом дослідження є порівняння можливостей прикладних математичних пакетів до вирішення задачі здійснення багатомірного регресійного аналізу. Розглянуто можливості Excel та Statistica на прикладі виконання лабораторної роботи. Завдання полягає в тому, щоб визначити вплив факторів на результуючу ознаку. На прикладі виконання одного завдання показано, що пакет STATISTICA має набагато більше графічних можливостей та набагато більше операційних функцій для здійснення багатомірного кореляційного аналізу. Як показує досвід, цим пакетом можна користуватись після оволодіння техніки багатомірного кореляційного аналізу та маючи ґрунтовну теоретичну підготовку. Калькулятор EXCEL дає можливість оволодіти технікою багатомірного кореляційного аналізу, зрозуміти його тонкощі та особливості.

Ключові слова: багатомірний кореляційний аналіз, коефіцієнти регресії, нормовані коефіцієнти регресії, коефіцієнти еластичності.

О.А. Карабын, О.Ю. Чмыр

ПРЕПОДАВАНИЕ МНОГОФАКТОРНОГО КОРРЕЛЯЦИОННОГО АНАЛИЗА С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИОННО-КОМУНИКАЦИОНЫХ ТЕХНОЛОГИЙ

Рассматривается вопрос о возможности совершенствования методики преподавания многомерного корреляционного анализа с помощью использования прикладных математических пакетов. Основным методом исследования является сравнение возможностей прикладных математических пакетов к решению задачи осуществления многомерного регрессионного анализа. Рассмотрены возможности Excel и Statistica на примере выполнения лабораторной работы. Задача состоит в том, чтобы определить влияние факторов на результирующий признак. На примере выполнения одной задачи показано, что пакет STATISTICA имеет гораздо больше графических возможностей и гораздо больше операционных функций для осуществления многомерного корреляционного анализа. Как показывает опыт, этим пакетом можно пользоваться после овладения техники многомерного корреляционного анализа, имея основательную теоретическую подготовку. Калькулятор EXCEL дает возможность овладеть техникой многомерного корреляционного анализа, понять его тонкости и особенности.

Ключевые слова: многомерный кореляционный анализ, коэффициенты регресии, нормированные коэффициенты регресии, коэффициенты эластичности.

O.O. Karabyn, O.Yu. Chmyr

THE ASPECTS OF MULTIDIMENSIONAL CORRELATION ANALYSIS TEACHING

In this article the issue of possibility of improvement in teaching methods in multidimensional correlation analysis with the application of applied statistical packages is described. The main method of research is the comparison of applied statistical packages for solving a problem of the multidimensional regression analysis implementation. The possibilities of Excel and Statistica are described in an example of a laboratory work. The task is to identify an impact of factors on the re-

sulting indication. As shown in the example of a task solution, STASTICA package has much more graphical features and operational functions for performing a multidimensional correlation analysis. Experience shows that the package can be used after mastering the technique of multidimensional correlation analysis and possessing a strong theoretical background. Calculator EXCEL gives an opportunity to master the multidimensional correlation analysis technique and to better understand its' features and properties.

Key words: the multidimensional correlation analysis, regression coefficients, the normalized regression coefficients, the elasticity coefficients.

Постановка проблеми. Зміни, що відбуваються в сучасному суспільстві в першу чергу стосуються молодого покоління. На другий план відходить важка фізична праця і значно зростає роль та попит на інтелектуальну працю. Можна вважати, що тепер ми стоїмо на порозі революційної ситуації в освіті, адже саме високі вимоги до фахівців вимагають зміни системи викладання, перегляду змісту навчальних програм навчальних дисциплін, відходу від застарілих методів навчання до модернових з використанням сучасних технічних засобів, якими в повсякденному житті користуються фактично всі, від малого до старого.

Молоде покоління, по-перше, швидко адаптується до нового середовища, по-друге, саме воно є причиною і активним учасником процесів оновлення суспільства. В таких умовах система освіти, яку ми зараз маємо, абсолютно не задовольняє потреб тих, на кого вона працює. Для того, щоб студенти, а саме вони є тими, на кого спрямована система освіти, були активними учасниками навчального процесу, мали мотивацію до навчання, уявляли собі кінцевий продукт своєї навчальної діяльності, мають відбутися зміни в структурі навчальних планів. Оскільки теперішнє законодавство надає вищим навчальним закладам більше автономії, то саме від навчальних закладів, а отже, і від самих викладачів великою мірою залежить те, чи навчальний процес наповнений потрібними актуальними дисциплінами, чи він наповнений тільки тим, хто що вміє викладати.

Як уже зазначалось вище, на перший план виходить інтелектуальна праця. Інтелектуальна праця в технічних галузях чи навіть гуманітарних неможлива без математичної складової. Як відомо, «математика приводить розум в порядок». В час, коли знання точних наук вийшло на чільне місце, ми зіткнулись з протиріччям. Всі розуміють, що математична складова є обов'язковою в системі підготовки фахівців практично всіх рівнів і галузей, але тим не менше частка математичних дисциплін і їх обсяг в навчальних планах підготовки фахівців зменшується. Вирішити таке протиріччя можна за допомогою самостійної роботи студентів, але такий шлях потребує високої мотивації студентів до навчання, чого на даний час ми на жаль не маємо.

В більшості технічних навчальних закладів курс вищої математики викладають протягом першого та другого років навчання. Одним з розділів вищої математики, на який вже не вистачає часу є теорія ймовірностей і математична статистика. В [1] наголошується, що "Класичний курс вищої математики для інженерних спеціальностей у вишах базується фактично на численні нескінченно малих і принципово мало змінився за останні 100 років". Чисельні методи, а також теорія ймовірностей і математична статистика належать до тих розділів, які даються не достатньою мірою в класичному курсі математики. Проблему можна вирішити введенням спеціальних курсів для магістрів. Таким необхідним курсом для магістрів не лише технічних спеціальностей, але і гуманітарних, є статистичний аналіз. Кожен фахівець, який хоче комфортно почувати себе в сучасному інформаційному суспільстві, повинен володіти елементами статистичного аналізу хоча б на початковому рівні, мати уявлення про ряди розподілу, закони розподілу, числові характеристики випадкових величин, графічне зображення рядів розподілу. Підхід до викладання статистичного аналізу має бути таким, щоб студент бачив можливості використання статистичного апарату в своїй майбутній професійній діяльності. Для здійснення цієї мети обов'язковим є проведення лабораторних робіт із застосуванням прикладних статистичних пакетів та використанням прикладних практичних задач.

Аналіз останніх досліджень та публікацій. Серед наукових публікацій, що стосуються проблем викладання статистичного аналізу, можна виділити роботи Лапача С. М. [1, 2]. Тут запропоновано логічну схему при вивченні статистичних методів та наголошується на тому, що викладання повинно базуватись на розв'язанні спеціально підготовлених задач, які розділяються на дві групи: спеціальні навчальні задачі і реальні задачі.

Одним з надзвичайно важливих розділів статистичного аналізу, який має дуже широке застосування, є багатомірний регресійний аналіз. В роботі [3] акцентується на тому, що для успішного використання багатомірного кореляційного аналізу потрібен системний підхід, який включає теорію планування експерименту і численні перевірки за статистичними критеріями отриманих результатів моделювання.

Дуже мало є доступної навчальної літератури, де детально і зрозуміло висвітлюються тонкощі багатомірного кореляційного аналізу та його здійснення за допомогою програмних прикладних математичних пакетів. В цьому напрямку слід відзначити роботи [4-6], де в доступній формі показано виконання багатомірного кореляційного аналізу.

Постановка завдання. В цій роботі висвітлюється питання можливості удосконалення методики викладання багатомірного кореляційного аналізу за допомогою використання прикладних математичних пакетів.

Виклад основного матеріалу. Кореляцію, за допомогою якої вивчається вплив на результативну ознаку двох та більше взаємозв'язаних факторних ознак, називають множинною.

Багатофакторні регресійні моделі дають змогу оцінювати вплив на досліджувану результативну ознаку кожного окремого із включених у рівняння факторів при фіксованому значенні (на середньому рівні) інших факторів. При цьому важливою умовою множинної кореляції є відсутність функціонального зв'язку між факторами.

Формула лінійного рівняння множинної регресії має такий вигляд:

$$y_x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n ,$$

де y_x – теоретичне значення результативної ознаки; β_0, β_1, β_2, …, β_n – параметри рівняння; x_1, x_2, …, x_n – факторні ознаки.

Окремі коефіцієнти регресії цього рівняння характеризують вплив відповідного фактора на результативний показник при фіксованому значенні інших факторів. Вони показують, наскільки зміниться результативний показник при зміні відповідного фактора на одиницю.

Показниками тісноти зв'язку при множинній кореляції є парні, часткові та множинні (сукупні) коефіцієнти кореляції та множинний коефіцієнт детермінації.

Парні коефіцієнти кореляції використовують для вимірювання тісноти зв'язку між двома досліджуваними ознаками без урахування їх взаємодії з іншими ознаками, які включені у кореляційну модель. Кореляційний зв'язок між факторами в рівнянні множинної регресії називають колінеарністю або мультиколінеарністю. Мультиколінеарність ускладнює дослідження впливу окремих факторів на результативну ознаку, оскільки взаємодія колінеарних факторів у моделі подвоюватиметься та спотворюватиме результати. Чим вища колінеарність, тим менш надійними будуть показники впливу окремих факторів. Допустимою колінеарністю для практичних цілей, що не спотворює результати досліджень, вважають таку, при якій парні коефіцієнти кореляції між факторними та результативною ознаками більші за коефіцієнти кореляції між супутніми факторами.

Часткові коефіцієнти кореляції характеризують тісноту зв'язку результативної ознаки з однією факторною ознакою за умови, що інші факторні ознаки перебувають на постійному рівні.

Коефіцієнт множинної (сукупної) детермінації показує, яка частка зміни досліджуваного результативного показника зумовлена впливом факторів, включених у рівняння множинної регресії. Він може мати значення від 0 до + 1. Чим ближчий коефіцієнт множинної детермінації до одиниці, тим більшою є зміна результативного показника під впливом відібраних факторів. Коефіцієнт множинної детермінації визначають за такою формулою: $R^2 = \frac{\sigma^2_{calc}}{\sigma^2_r}$, де σ^2_{calc} – дисперсія результативного показника, обчислена за рівнянням множинної регресії; $\sigma^2_{calc} = \overline{y^2_x} - \overline{y}^2$; σ^2_r – загальна дисперсія результативного показника $\sigma^2_r = \overline{y^2} - \overline{y}^2$. Середню помилку вибіркового коефіцієнта множинної детермінації визначають за такою формулою: $m_R = \frac{1-R^2}{\sqrt{n-m-1}}$, де n – кількість спостережень; m – кількість факторів. Число $n-m-1$ визначає кількість ступенів свободи. Перевірку гіпотези про вірогідність коефіцієнта множинної детермінації здійснюють за допомогою критерію Стьюдента. Фактичне значення критерію Стьюдента обчислюють за формулою

$$t_{\text{емп}} = \frac{R}{m_R}. \tag{1}$$

Якщо емпіричне значення критерію перевищує табличне при заданому рівні значущості і числі ступенів свободи $k = n - m - 1$, то можна зробити висновок про вірогідність коефіцієнта множинної детермінації.

Важливими показниками кореляційного аналізу є коефіцієнти еластичності та нормовані коефіцієнти регресії. Потреба в їх застосуванні зумовлена тим, що коефіцієнти регресії, маючи різні фізичний зміст і одиниці вимірювання, не дають чіткого уявлення про те, які фактори мають найбільший вплив на результативну ознаку, тобто коефіцієнти регресії не можна безпосередньо порівнювати між собою. Позбавлені такого недоліку коефіцієнти еластичності та нормовані коефіцієнти регресії.

Нормовані коефіцієнти регресії показують, на скільки середніх квадратичних відхилень змінюється результативний показник при зміні відповідного фактора на одне значення середнього квадратичного відхилення. Вони характеризують вплив окремих факторів на результативну ознаку. Їх визначають за формулою

$$a_i = \beta_i \frac{\sigma_{x_i}}{\sigma_y}, \tag{2}$$

де β_i – коефіцієнт регресії на i-му факторі; σ_{x_i} – середнє квадратичне відхилення i-го фактора; σ_y – середнє квадратичне відхилення результативного показника.

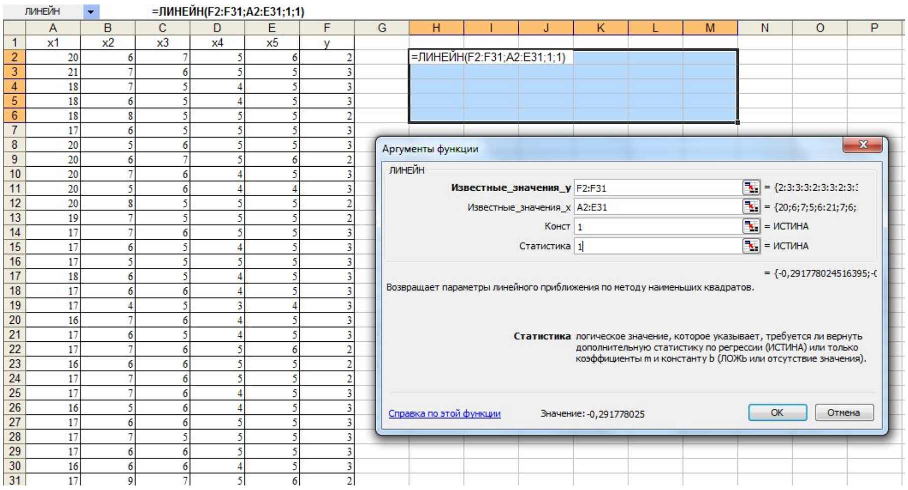
Коефіцієнти еластичності E_i показують, на скільки відсотків змінюється результативна ознака при зміні факторної ознаки на 1%. Їх обчислюють за формулою

$$E_i = \beta_i \frac{\overline{x}_i}{\overline{y}}, \tag{3}$$

де β_i – коефіцієнт регресії при i-ому факторі, $\overline{x}_i$ – середнє значення i-го фактора; $\overline{y}$ – середнє значення результативної ознаки.

Без застосування сучасних програмних засобів здійснити виконання багатомірного кореляційного аналізу дуже складно у зв'язку з великою кількістю громіздких обчислень. Розглянемо можливості Excel та Statistica на прикладі виконання лабораторної роботи. Спочатку розглянемо виконання в пакеті Excel. Завдання полягає в тому, щоб визначити вплив факторів x_1, x_2, x_3, x_4, x_5 на ознаку y.

Множиний регресійний аналіз в пакеті Excel здійснюється за допомогою функції =**ЛИНЕЙН.** Перед викликом функції потрібно виділити блок комірок $(m+1)\times 5$, де m – кількість факторів, і в діалоговому вікні ввести дані задачі, як показано на рис. 1.



	A	B	C	D	E	F
1	x1	x2	x3	x4	x5	y
2	20	6	7	5	6	2
3	21	7	6	5	5	3
4	18	7	5	4	5	3
5	18	6	5	4	5	3
6	18	8	5	5	5	2
7	17	6	5	5	5	3
8	20	5	6	5	5	3
9	20	6	7	5	6	2
10	20	7	6	4	5	3
11	20	5	6	4	4	3
12	20	8	5	5	5	2
13	19	7	5	5	5	2
14	17	7	6	5	5	3
15	17	6	5	4	5	3
16	17	5	5	5	5	3
17	18	6	5	4	5	3
18	17	6	6	4	5	3
19	17	4	5	3	4	3
20	16	7	6	4	5	3
21	17	6	5	4	5	3
22	17	7	6	5	6	2
23	16	6	6	5	5	2
24	17	7	6	5	5	2
25	17	7	6	4	5	3
26	16	5	6	4	5	3
27	17	6	6	5	5	3
28	17	7	5	5	5	3
29	17	6	6	5	5	3
30	16	6	6	4	5	3
31	17	9	7	5	6	2

***Рисунок 1** – Вхідні дані задачі та діалогове вікно функції ЛИНЕЙН*

Натиснувши комбінацію клавіш **Ctrl+Shift+Enter** отримуємо результати регресійного аналізу у вигляді таблиці в порядку, наведеному на рис. 2.

β_5	β_4	β_3	β_2	β_1	β_0
σ_{β_5}	σ_{β_4}	σ_{β_3}	σ_{β_2}	σ_{β_1}	σ_{β_0}
R^2	Стандартна похибка				
Критерій Фішера	Ступені свободи				
Сума квадратів відхилень, що пояснюється регресією	Сума квадратів відхилень, що пояснюється похибкою				

***Рисунок 2** – Розташування регресійної статистики*

На рис. 3 показано виконання регресійного аналізу для даних задачі. За допомогою рис. 2 бачимо, що означає кожне число в результатах регресійного аналізу.

H	I	J	K	L	M
-0,29178	-0,22139	-0,03139	-0,10105	-0,02842	6,510216
0,230341	0,153455	0,133247	0,080986	0,049972	1,143039
0,443124	0,382335	#Н/Д	#Н/Д	#Н/Д	#Н/Д
3,819506	24	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,791678	3,508322	#Н/Д	#Н/Д	#Н/Д	#Н/Д

***Рисунок 3** – Результати регресійного аналізу*

В першу чергу звертаємо увагу на коефіцієнт множинної детермінації, що становить 0,44. Це означає, що 44% зміни результуючого показника *y* зумовлені впливом факторів x_1, x_2, x_3, x_4, x_5, включених в регресійну модель. Подальшим кроком є перевірка гіпотези про значимість коефіцієнта множинної детермінації, що дасть змогу зробити висновок про адекватність створеної лінійної регресійної моделі. Формулюємо статистичні гіпотези **H_0: коефіцієнт детермінації не значимий**; **H_1: коефіцієнт детермінації значимий.**

Емпіричне значення критерію Стьюдента $t_{емп}$ обчислюємо за формулою (1). Критичну точку шукаємо за допомогою функції **=СТЬЮДРАСПОБР ( )**, вказавши рівень значущості та число ступенів свободи (рис. 4). Як бачимо, емпіричне значення критерію перевищує критичну точку, тому приймаємо гіпотезу H_1 і робимо висновок про значимість коефіцієнта детермінації, а отже про адекватність нашої регресійної моделі.

Адекватність регресійної моделі можна також перевірити за допомогою критерію Фішера. Статистичні гіпотези формулюємо так: **H_0: модель неадекватна (всі коефіцієнти регресії дорівнюють нулю); H_1: модель адекватна (хоча б один з коефіцієнтів регресії відмінний від нуля)**. Критична область є правобічною. Емпіричне значення критерію Фішера маємо в таблиці (рис. 4) $f_{емп}$ = 3,8195. Критичну точку шукаємо за допомогою функції **= FРАСПОБР ( )** аргументами якої є рівень значущості та кількості ступенів свободи $k_1 = m+1$, $k_2 = n-m-1$, де m – кількість факторів, n – кількість експериментів: $f_{крит}$ = 2,508 (рис. 4). Оскільки емпіричне значення критерію є більшим за критичну точку, то приймаємо альтернативну гіпотезу, що ще раз засвідчує адекватність регресійної моделі.

Для встановлення фактора, що найбільше впливає на зміну результуючого показника обчислюємо нормовані коефіцієнти регресії за формулою (2) та коефіцієнти еластичності за формулою (3). Для їх обчислення потрібно знайти середні значення та середньоквадратичні відхилення факторних ознак та результуючого показника. Середньоквадратичні відхилення обчислюємо за допомогою функції **= СТАНДОТКЛОН ( )**, а середні значення за допомогою функції **= СРЗНАЧ ( )**. Результати обчислень видно з рис. 4. Абсолютні величини коефіцієнтів еластичності та нормованих коефіцієнтів регресії показують, що найбільший вплив на результуючий показник y чинить фактор x_5. Його коефіцієнт еластичності становить – 0,54753, а його нормований коефіцієнт регресії становить – 0,28156. Від'ємний знак коефіцієнтів свідчить про те, що із зростанням величини фактора x_5, величина результуючого показника y зменшується. Порівняння абсолютних величин коефіцієнтів еластичності та нормованих коефіцієнтів регресії дають змогу визначити послідовність факторів за спаданням величини їх впливу на результуючий показник.

H	I	J	K	L	M
-0,29178	-0,22139	-0,03139	-0,10105	-0,02842	6,510216
0,230341	0,153455	0,133247	0,080986	0,049972	1,143039
0,443124	0,382335	#Н/Д	#Н/Д	#Н/Д	#Н/Д
3,819506	24	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,791678	3,508322	#Н/Д	#Н/Д	#Н/Д	#Н/Д

Критерій	Емпіричне значення	Критична точка	
Стьюдента	темп	ткрит	
	5,85610886	2,063898547	
Фішера	fемп	fкрит	
	3,819506	2,508189	
Середня помилка коефіцієнта детермінації mR			0,113672

Коефіцієнти регресії					
β_1	β_2	β_3	β_4	β_5	
-0,02842	-0,10105	-0,03139	-0,22139	-0,29178	
Середньоквадратичні відхилення					
σx_1	σx_2	σx_3	σx_4	σx_5	σy
1,47157	1,033352	0,651259	0,571346	0,449776	0,466092
	Середні значення				
x_1	x_2	x_3	x_4	x_5	y
17,8000	6,3667	5,7000	4,5333	5,0667	2,7000
	Коефіцієнти еластичності				
$E1$	$E2$	$E3$	$E4$	$E5$	
-0,18739	-0,23827	-0,06627	-0,37172	-0,54753	
Нормовані коефіцієнти регресії					
$a1$	$a2$	$a3$	$a4$	$a5$	
-0,08974	-0,22403	-0,04386	-0,27138	-0,28156	

Рисунок 4 *– Інтерпретація результатів багатомірного регресійного аналізу*

Для повноти картини обчислимо парні коефіцієнти кореляції за допомогою функції **=КОРРЕЛ** та подамо результати обчислень у вигляді кореляційної матриці (рис. 5).

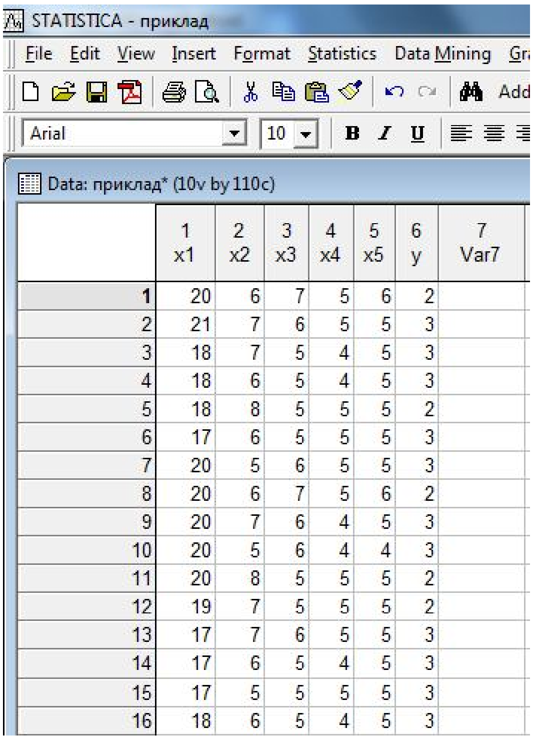
Кореляційна матриця

	x1	x2	x3	x4	x5	y
x1	1	0,072564	0,151118	0,213268	0,072938	-0,19104
x2	0,072564	1	0,11785	0,416626	0,464937	-0,47969
x3	0,151118	0,11785	1	0,259482	0,541514	-0,30672
x4	0,213268	0,416626	0,259482	1	0,527796	-0,55926
x5	0,072938	0,464937	0,541514	0,527796	1	-0,55926

***Рисунок 5** – Кореляційна матриця*

З кореляційної матриці бачимо, що між факторами x_1, x_4 тісніший кореляційний зв’язок, ніж між фактором x_1 та результуючим показником. Фактор x_3, тісніше зв’язаний з фактором x_5, ніж з результуючим показником.

Розглянемо виконання цього завдання за допомогою прикладного пакета STATISTICA. Вводимо дані задачі в робочий лист (рис. 6).



	1 x1	2 x2	3 x3	4 x4	5 x5	6 y	7 Var7
1	20	6	7	5	6	2	
2	21	7	6	5	5	3	
3	18	7	5	4	5	3	
4	18	6	5	4	5	3	
5	18	8	5	5	5	2	
6	17	6	5	5	5	3	
7	20	5	6	5	5	3	
8	20	6	7	5	6	2	
9	20	7	6	4	5	3	
10	20	5	6	4	4	3	
11	20	8	5	5	5	2	
12	19	7	5	5	5	2	
13	17	7	6	5	5	3	
14	17	6	5	4	5	3	
15	17	5	5	5	5	3	
16	18	6	5	4	5	3	

***Рисунок 6** – Фрагмент даних задачі в робочому листі пакета STATISTICA*

В пакеті STATISTICA мультифакторний аналіз можна виконати зайшовши в **меню Statistics – Advanced Linear / Nonlinear Models**. Вибрати в якості типу аналізу **Multiple regression** і в якості методу вирішення **Quick specs dialog**. Після цього натиснути **OK** для входу в діалогове вікно множинної регресії (рис. 7).

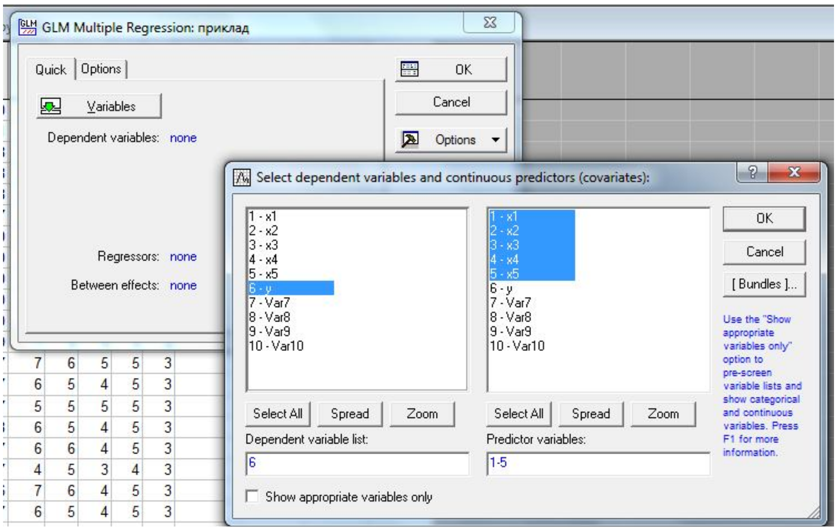
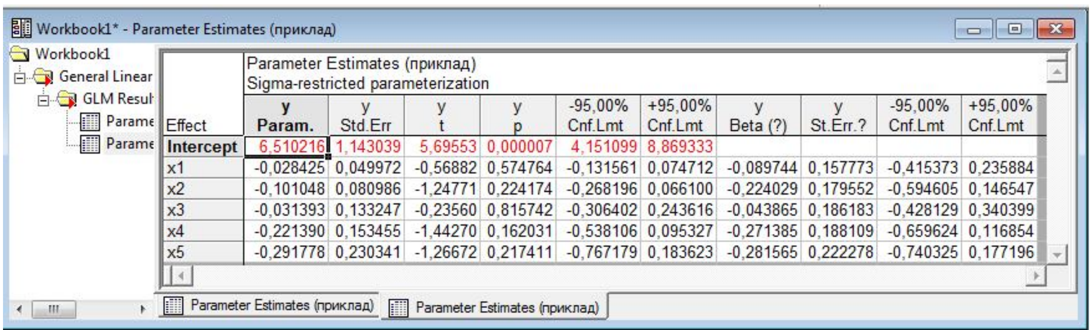


***Рисунок 7** – Діалогове вікно множинної регресії*

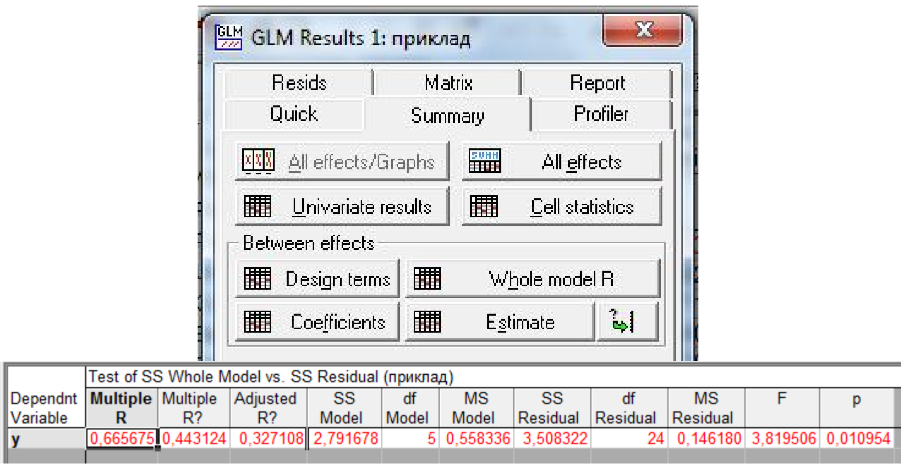
Натиснувши OK в діалоговому вікні множинної регресії, переходимо до вікна результатів регресійного аналізу, де при виділенні опції **Summary** потрібно натиснути клавішу **Coefficients** для відображення обчислених коефіцієнтів регресії (рис. 8).



Effect	y Param.	y Std.Err	y t	y p	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt	y Beta (?)	y St.Err.?	-95,00% Cnf.Lmt	+95,00% Cnf.Lmt
Intercept	6,510216	1,143039	5,69553	0,000007	4,151099	8,869333				
x1	-0,028425	0,049972	-0,56882	0,574764	-0,131561	0,074712	-0,089744	0,157773	-0,415373	0,235884
x2	-0,101048	0,080986	-1,24771	0,224174	-0,268196	0,066100	-0,224029	0,179552	-0,594605	0,146547
x3	-0,031393	0,133247	-0,23560	0,815742	-0,306402	0,243616	-0,043865	0,186183	-0,428129	0,340399
x4	-0,221390	0,153455	-1,44270	0,162031	-0,538106	0,095327	-0,271385	0,188109	-0,659624	0,116854
x5	-0,291778	0,230341	-1,26672	0,217411	-0,767179	0,183623	-0,281565	0,222278	-0,740325	0,177196

***Рисунок 8** – Коефіцієнти регресії*

Коефіцієнти регресії отримуємо у першому стовпчику, а нормовані коефіцієнти регресії маємо у стовпчику **Beta.** Для відображення коефіцієнта множинної детермінації потрібно в меню **GLM Results** відкрити вкладку **Summary**, а в ній **– Whole model R.** В результаті дістанемо табличку з множинним коефіцієнтом детермінації, ступенями свободи та емпіричним значенням критерію Фішера (рис. 9).



Dependnt Variable	Multiple R	Multiple R?	Adjusted R?	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
y	0,665675	0,443124	0,327108	2,791678	5	0,558336	3,508322	24	0,146180	3,819506	0,010954

***Рисунок 9** – Коефіцієнт множинної детермінації*

Для відображення парних коефіцієнтів кореляції в меню **GLM Results** відкрити вкладку **Matrix**, а в ній вкладку **Correlation.** На рис. 10 зображено кореляційну матрицю, аналогічну до кореляційної матриці рис. 5.

Correlations of Vectors in Design Matrix X (приклад)
Correlation matrix for the vectors
in the design matrix X

Effect	Level	Column	Effect (F/R)	Col. 1 Intercpt	Col. 2 x1	Col. 3 x2	Col. 4 x3	Col. 5 x4	Col. 6 x5	Col. 7 y
Intercept		1	Fixed							
x1		2	Fixed		1,000000	0,072564	0,151118	0,213268	0,072938	-0,191044
x2		3	Fixed		0,072564	1,000000	0,117850	0,416626	0,464937	-0,479686
x3		4	Fixed		0,151118	0,117850	1,000000	0,259482	0,541514	-0,306719
x4		5	Fixed		0,213268	0,416626	0,259482	1,000000	0,527796	-0,543852
x5		6	Fixed		0,072938	0,464937	0,541514	0,527796	1,000000	-0,559259
y		7			-0,191044	-0,479686	-0,306719	-0,543852	-0,559259	1,000000

***Рисунок 10 –** Кореляційна матриця в пакеті STATISTICA*

Пакет STATISTICA має можливості побудови діаграм для візуального аналізу залишків з метою виявлення викидів, що перевищують ±3 sigma. Множинна лінійна регресія потребує наявності лінійних співвідношень між змінними і нормального розподілу залишків. Якщо ці вимоги порушені, то остаточний висновок може бути неправильним. Якщо спостережувані залишки є нормально розподіленими, то всі значення повинні вкладатись вздовж прямої лінії. В іншому випадку точки, що зображають залишки, будуть відхилятись від прямої лінії. Для аналізу залишків у вікні діалогу **GLM Results** потрібно натиснути кнопку **More results**, після чого виділити вкладку **Residuals 1** щоб розглянути різні види залишків. Як правило, аналізують стандартизовані залишки. Для цього вибирають опцію **Standardized** в полі **Resids for default plots** і натиснути кнопку **Case no & res** для побудови графіка (рис. 11).

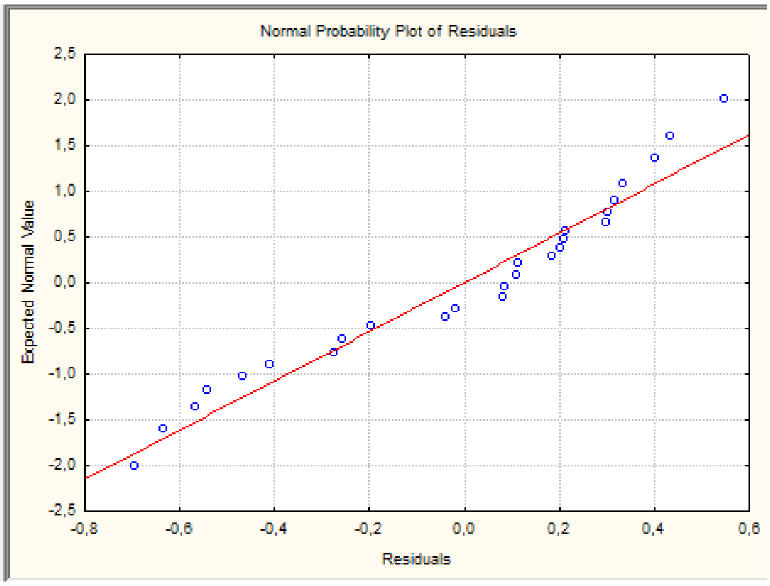


***Рисунок 11 –** Графічний аналіз залишків*

Висновки. Як бачимо пакет STATISTICA має набагато більше графічних можливостей та набагато більше операційних функцій для здійснення багатомірного кореляційного аналізу. Як показує досвід, цим пакетом можна користуватись після оволодіння технікою багатомірного кореляційного аналізу та маючи ґрунтовну теоретичну підготовку. Калькулятор Excel дає можливість оволодіти технікою багатомірного кореляційного аналізу, зрозуміти його тонкощі та особливості.

Список літератури:

1. Лапач С. М. Конфлікт класичного і модернового у викладанні математики у вищій школі [Електронний ресурс]/ Сергій М. Лапач// Математика в сучасному технічному університеті: Збірник науково-методичних праць/ Національний технічний університет України "КПІ". – Київ, 2015. – Вип. 1. – С 162 - 167.
2. Лапач С. М. Проблеми застосування статистичних методів дослідниками: що вчити у ВНЗ? [Електронний ресурс]/ Сергій М. Лапач// Математика в сучасному технічному університеті: Збірник науково-методичних праць/ Національний технічний університет України "КПІ". – Київ, 2015. – Вип. 1. – С 41 - 47.
3. Радченко С. Г. Системное обеспечение получения многофакторных статистических моделей [Електронний ресурс]/ Станислав Т. Радченко// Математика в сучасному технічному університеті: Збірник науково-методичних праць/ Національний технічний університет України "КПІ". – Київ, 2015. – Вип. 1. – С 66 - 71.
4. Лупан І. В., Авраменко О. В. Комп'ютерні статистичні пакети: навчально-методичний посібник. – Кіровоград, 2010. – 218 с.
5. Лупан І. В., Халецька З. П., Чеча В. О. Використання канонічного кореляційного аналізу у педагогічних дослідженнях/ Наукові записки НДУ ім. М. Гоголя. Психолого-педагогічні науки. – 2011. – №10, 63. – 68 с.
6. Тьюки Дж. У. Анализ результатов наблюдений. Разведочный анализ/ Дж. У. Тьюки пер. с англ.; под ред. В. Ф. Писаренко. – Москва: Мир, 1981. – 695 с.

References:

1. Lapach, S. M. (2015) *Conflict classic and modern in the teaching of mathematics in high school [electronic resource].* / Sergiy M. Lapach // Mathematics in the modern Technical University: Collection of scientific and methodological papers / National Technical University of Ukraine "KPI". – Kyiv. – Iss. 1. – pp. 162 - 167.
2. Lapach, S. M. (2015) *Problems of using statistical methods researchers: what teach in universities? [Electronic resource]* / Sergiy M. Lapach // Mathematics in the modern Technical University: Collection of scientific and methodological papers / National Technical University of Ukraine "KPI". – Kyiv. – Iss. 1 – pp. 41 - 47.
3. Radchenko, S. G. (2015) *System provision obtaining multidimensional correlation models [Electronic resource]* / Stanislav T. Radchenko // Mathematics in the modern Technical University: Collection of scientific and methodological papers / National Technical University of Ukraine "KPI". – Kyiv. – Iss. 1 – pp. 66 - 71.
4. Lupan, I. V. and Avramenko, O. V. (2010) *Computer statistical packages*: Textbook. – Kirovograd. – 218 p.
5. Lupan, I. V., Haletska, Z. P. and Checha, V. O. (2011) *Using canonical correlation analysis in pedagogical researches* / Scientific notes NDU named M. Gogol. Psychology-pedagogical science. – no. 10, 63. – 68 p.
6. Tukey, J. W. (1981) *Analysis of results of observations* / J. W. Tukey/ Moscow: Mir. – 695 p.

❖❖❖