

## РЕГРЕСІЙНИЙ ТА КОРЕЛЯЦІЙНИЙ АНАЛІЗ ЗАСОБАМИ OPENOFFICE У НАВЧАННІ СТОХАСТИКИ МАЙБУТНІХ ЕКОНОМІСТІВ

Наталія ШУЛЬГА

*В статті досліджується можливість застосування табличного процесору Calc пакету прикладних програм OpenOffice у процесі навчання стохастики студентів економічних факультетів університетів. Розглянуто особливості застосування вбудованих функцій OpenOffice Calc для проведення регресійного та кореляційного аналізу статистичних даних.*

*The article explores the possibility of using table processor Calc package applications OpenOffice in teaching Stochastics in economic universities. The author reveals the features of the application OpenOffice Calc built-in functions for regression and correlation analysis of statistical data.*

**Постановка проблеми.** На початку XXI сторіччя суспільно-економічний розвиток людства може бути охарактеризовано як перехід до інформаційного суспільства. Основою та рушійною силою даного процесу є інформація, темпи зростання якої прискорюються з кожним роком. Інформатизація суспільства має наслідки як позитивні (збільшення інформації приводить до утворення нових знань, нових технологій, нових продуктів), так і негативні (через велику кількість другорядної інформації зростає час на її опрацювання, збільшується ймовірність втрати більш важливих знань, отже може з'явитись така парадоксальна ситуація, коли через надлишок інформації виникає інформаційний голод). Зростає необхідність у фахівцях, що мають здатність швидко адаптуватись до нових, або суттєво змінених, умов праці, знаходити та використовувати інноваційні технології, що дають можливість швидко отримати результат. Тому, система вищої освіти повинна бути спрямована на підготовку фахівця, у якого сформовано не тільки базу знань, вмінь та навичок, а й готовність до їх самостійного пошуку, оновлення, перетворення, здатність продукувати та втілювати в життя оригінальні, нестандартні ідеї. Великим попитом на сучасному ринку праці користуються фахівці, що здатні знаходити більш раціональні шляхи розв'язання професійних проблем, в тому числі і за рахунок використання новітніх інформаційних технологій. Отже, на одному з чільних місць в навчальному процесі повинна стояти задача формування готовності використання можливостей комп'ютерної техніки в майбутній професійній діяльності.

Крім того, в епоху, коли недетерміновані процеси складають основу соціально-економічного буття, велику роль в підготовці висококваліфікованих фахівців відіграє вивчення такої дисципліни як «Теорія ймовірності та математична статистика». Адже, опанування методами даної дисципліни дозволяє будувати моделі стохастичних процесів та явищ, аналізувати їх, прогнозувати напрямок їх розвитку. В той же час, створене програмне забезпечення, що дозволяє спростити виконання вказаних задач, автоматизувати процес їх розв'язання (таке, наприклад, як Derive, GRAN, MathCad, MatLab, MS Excel, Maple, Statgraph, Statistica та ін.). Отже, виникає необхідність в створенні методичних розробок, спрямованих на розкриття можливостей застосування комп'ютерної техніки в процесі навчання стохастичності у вищій школі.

**Аналіз актуальних досліджень.** Методичні особливості навчання теорії ймовірності та математичної статистики розглянуто в досить великій кількості наукових праць. Значну частину з них присвячено вивченню стохастики в школі. Проблемам вивчення теорії ймовірностей і математичної статистики у вищій школі приділяли увагу Н. Дергунова, Т. Задорожня, О. Лебедева, Н. Паніна, Т. Полякова, Л. Пуханова. Особливостям застосування комп'ютерної техніки у процесі навчання даної дисципліни розкрито в роботах Д. Вауліної, М. Жалдака, Н. Кузьміна, А. Ліпінської, Г. Михаліна, Л. Пуханової, С. Самсонової, М. Суворової, С.Щербатих. Окрім наукових досліджень, останнім часом почали видавати підручники, навчальний матеріал в яких супроводжується прикладами застосування комп'ютерних технологій.

Принципи роботи з програмами пакету OpenOffice досить широко висвітлено на спеціальних сайтах в мережі Інтернет, зокрема База знань OpenOffice.org [4], Prosto Pro Office [2], Основи преподавания офісних технологий на базе OpenOffice.org [1] тощо. В роботах І. Акімової [3], В. Клячкіна [5], С. Тімохіної [6], Є. Шереметьєвої [7] представлено методичні розробки щодо застосування офісного додатку OpenOffice в процесі навчання.

**Мета дослідження** полягає в тому, щоб визначити можливості застосування табличного процесору OpenOffice Calc у навчанні стохастики, а саме в процесі вивчення регресійного та кореляційного аналізу студентами економічних спеціальностей університетів.

**Виклад основного матеріалу.** OpenOffice – це повнофункціональний пакет офісних додатків, який було розроблено як альтернативу пакету Microsoft Office як на рівні форматів так і на рівні інтерфейсу. Одним із переваг пакету OpenOffice є те, що він не потребує ніякої плати за ліцензію, в той же час може вільно замінити будь-який з пакетів прикладних офісних програм. Із всіх додатків пакету офісних програм OpenOffice, у навчанні стохастики найбільш широке застосування може отримати табличний процесор Calc, що є аналогом табличного процесору MS Excel. Так само як і MS Excel табличний процесор Calc містить вбудовані функції, що розподілені за наступними категоріями: База даних, Дата і час, Фінансові, Інформація, Логічні, Математичні, Масив, Статистичні, Електронна таблиця, Текст, Додаток.

Розглянемо можливості застосування вбудованих функції із категорій Математичні, Статистичні та Масив для проведення регресійного та кореляційного аналізу економічних статистичних даних.

В стохастичі розглядають три види залежностей:

*Функціональною* називають залежність в якій кожному значенню однієї змінної відповідає досить визначене значення іншої. Функціональна залежність між декількома величинами можлива лише у випадку, якщо на ці величини не впливають випадкові фактори.

*Статистичною* називають залежність в якій кожному значенню однієї змінної відповідає множина можливих значень іншої. Така залежність виникає тому, що на залежну змінну впливає безліч випадкових факторів, які не завжди можна врахувати, а також тому, що вимірювання значень змінних завжди проводиться з випадковою похибкою.

Кореляційною залежністю між двома змінними величинами називають функціональну залежність між значеннями однієї з них та умовним математичним сподіванням іншої.

Статистичні зв'язки між змінними можуть бути вивчені на основі методів дисперсійного, кореляційного та регресійного аналізу. Методами дисперсійного аналізу встановлюється наявність впливу заданого фактору (факторів) на процес, що досліджується. Кореляційний аналіз надає можливість оцінити силу такого зв'язку. Методами регресійного аналізу можна визначити конкретну математичну модель, що описує зв'язок між факторами, та оцінити її адекватність. Таким чином, кореляційний аналіз дає відповіді на питання: Чи існує зв'язок між явищами? Якщо взаємозв'язок між явищами існує, то наскільки він сильний? Регресійний аналіз надає можливість відповісти на питання: Який характер зв'язку між явищами? Як побудувати та дослідити модель даного зв'язку, що називається лінією регресії?

В залежності від того, яка кількість факторів впливає на залежну змінну, регресія може бути парною (в моделі приймають участь два фактори: залежний та незалежний), або множинною (коли на залежну змінну впливають декілька факторів). Характер зв'язку між змінними може бути як лінійним, так і нелінійним.

Рівняння прямої лінії парної регресії  $Y$  на  $X$ , де  $Y$  – залежна змінна, а  $X$  – фактор, що впливає на неї, має вигляд (1):

$$\bar{y}_x = \rho_{yx} x + b \tag{1}$$

Параметри рівняння  $\rho_{yx}$  та  $b$ , частіше за все, визначають на основі методу найменших квадратів.

Параметр  $\rho_{yx}$  називають коефіцієнтом регресії та обчислюють за формулами (2), (3):

$$\rho_{yx} = \frac{n \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - (\sum x)^2}, \text{ для обчислень через суми} \tag{2}$$

$$\rho_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}, \text{ для обчислень через середні значення} \tag{3}$$

Параметр  $b$  називають вільним членом рівняння регресії та обчислюють за формулами (4), (5)

$$b = \frac{\sum x^2 \cdot \sum y - \sum x \cdot \sum xy}{n \sum x^2 - (\sum x)^2}, \text{ для обчислень через суми} \tag{4}$$

$$b = \bar{y} - \rho_{yx} \cdot \bar{x}, \text{ для обчислень через середні значення} \tag{5}$$

Застосування вбудованих функцій OpenOffice Calc дозволяє сформувати навички обчислення значень параметрів, при цьому скоротити час на виконання механічних обчислень. Для обчислень доцільно побудувати розрахункову таблицю, та скористатися наступними функціями:

SUM(Число1;Число2;...) з категорії Математичні для обчислення суми всіх значень із виділеного діапазону;

SUMSQ(Число1;Число2;...) з категорії Математичні для обчислення суми квадратів всіх значень із виділеного діапазону;

SUMPRODUCT(Массив1;Массив2;...) з категорії Масиви для обчислення суми всіх добутків відповідних елементів масиву;

AVERAGE(Число1;Число2;...) з категорії Статистичні, що обчислює середнє значення виділеного діапазону даних;

COUNT(Значение1;Значение2;...) з категорії Статистичні, що визначає кількість даних, які входять до виділеного діапазону.

Крім безпосереднього обчислення параметрів регресії за формулами, до табличного процесору Calc включено функції, що автоматично обчислюють параметри регресії:

SLOPE(Данные Y; Данные X) з категорії Статистичні, що обчислює кут нахилу лінійної регресії;

INTERCEPT(Данные Y; Данные X) з категорії Статистичні, що визначає величину відрізка, який відсікає лінія регресії на осі ординат;

LINEST(Данные Y; Данные X; Тип линии; Статистика) з категорії Масив, що обчислює параметри та інші характеристики рівняння регресії і видає їх у вигляді масиву. Якщо в полі Тип линии немає жодного значення, або встановлено 0, то рівняння регресії буде представлено у вигляді  $\bar{y}_x = \rho_{yx}x$ . Якщо встановлено будь-яке інше значення, будуть обчислені параметри рівняння регресії виду  $\bar{y}_x = \rho_{yx}x + b$ . Якщо в полі Статистика немає жодного значення, або встановлено 0, то результатом обчислення функції LINEST будуть тільки параметри рівняння регресії. Якщо встановлено будь-яке інше значення – крім параметрів, масив даних міститиме й інші характеристики рівняння регресії. Обчислення функції LINEST за допомогою Майстра функцій наведено на рис. 1.

Головною метою регресійного аналізу є побудова моделі, на основі якої можна визначити прогнозні значення залежної змінної для заданих значень незалежного фактору (6).

$$\bar{y}_{i\bar{i}\bar{i}\bar{i}} = \rho_{yx}x_{i\bar{i}\bar{i}\bar{i}} + b \tag{6}$$

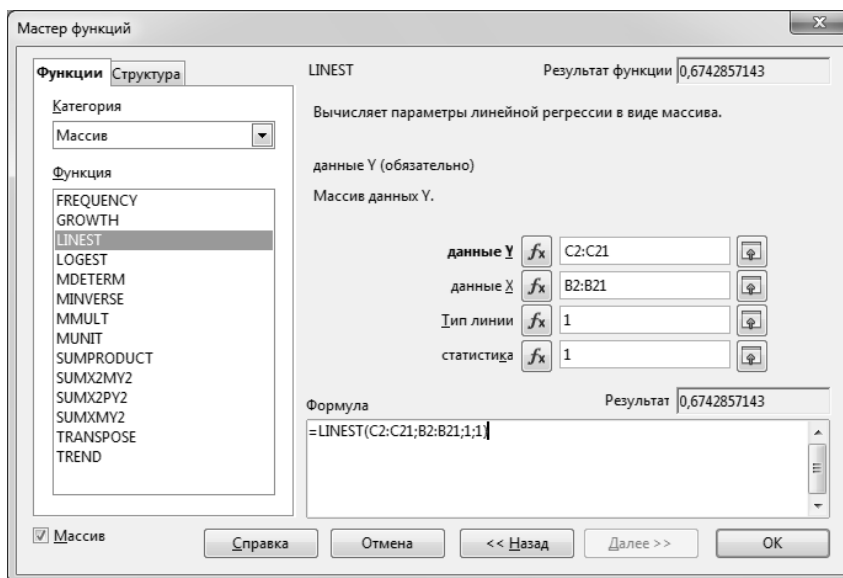


Рис. 1. Обчислення функції LINEST

Для обчислення прогнозних значень, можна використати статистичну функцію FORECAST(Значение; Данные Y; Данные X).

Розглянемо особливості проведення регресійного аналізу засобами на наступному прикладі: Проводиться дослідження залежності собівартості одиниці продукції від затрат на матеріальні ресурси. За даними отриманої в результаті дослідження вибірки (табл. 1) знайти рівняння прямої лінії регресії  $Y$  на  $X$ .

Таблиця 1

x	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65
y	1,21	1,32	1,44	1,18	1,22	1,55	1,53	1,61	1,62	1,52
x	0,7	0,75	0,8	0,85	0,9	0,95	1	1,05	1,1	1,15
y	1,66	1,73	1,73	1,64	1,68	1,81	1,85	1,79	1,84	1,87

Обчислити прогнозне значення величини собівартості одиниці продукції, якщо затрати на матеріальні ресурси складатимуть 1,2 у.о.

Для обчислення параметрів регресії, побудуємо розрахункову таблицю (рис. 2). В таблиці можна обчислювати або тільки суми значень, або тільки середні значення, в залежності від того, яка формула буде застосована.

На основі даних, отриманих в розрахунковій таблиці, проводять обчислення параметрів рівняння регресії.

	A	B	C	D	E		
1		x	y	x <sup>2</sup>	xy		
2		0,2	1,21	0,04	0,242		
3		0,25	1,32	0,0625	0,33		
4		0,3	1,44	0,09	0,432		
5		0,35	1,18	0,1225	0,413		
6		0,4	1,22	0,16	0,488		
18		1	1,85	1	1,85		
19		1,05	1,79	1,1025	1,8795		
20		1,1	1,84	1,21	2,024		
21		1,15	1,87	1,3225	2,1505		
22	Сума значень	13,5	31,8	10,775	22,586		
23	Формули для обчислень	=SUM(B2:B21)	=SUM(C2:C21)	=SUM(D2:D21)	=SUM(E2:E21)		
24				=SUMSQ(B2:B21)	=SUMPRODUCT(B2:B21;C2:C21)		
25	Середнє значення	0,675	1,59	0,53875	1,1293		
26	Формули для обчислень	=AVERAGE(B2:B21)	=AVERAGE(C2:C21)	=AVERAGE(D2:D21)	=AVERAGE(E2:E21)		
27					=SUM(B2:B21)/COUNT(B2:B21)	=SUM(C2:C21)/COUNT(C2:C21)	=SUM(D2:D21)/COUNT(D2:D21)

Рис. 2. Розрахункова таблиця

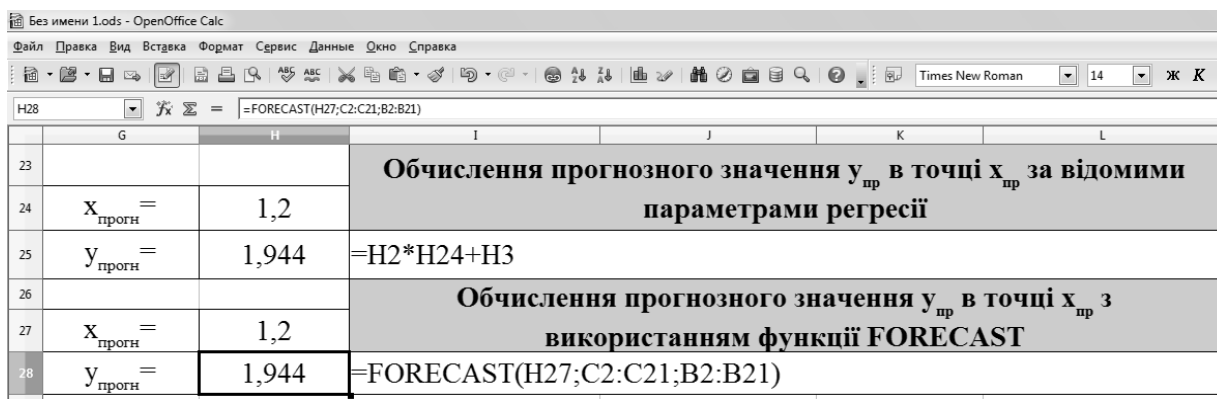
Для того, щоб студенти розуміли сутність даних параметрів, доцільно для їх обчислення застосовувати статистичні функції SLOPE, яка прямо вказує, що коефіцієнт регресії визначає величину нахилу прямої регресії відносно осі абсцис, та INTERCEPT, яка зазначає, що вільний член рівняння регресії визначає точку на осі ординат, через яку проходить лінія регресії.

Якщо необхідно сформулювати навички аналізу параметрів рівняння, порівняння їх з тими, що отримані на основі інших вибірок, то для того, щоб скоротити час, який витрачається на обчислення, доцільно застосувати функцію LINEST, що видає параметри рівняння регресії у масиві з іншими даними.

Обчислити прогнозне значення можна як на основі отриманих параметрів регресії, так і з використанням вбудованої функції FORECAST (рис. 4)

В кореляційному аналізі використовують наступні основні прийоми:

- побудову кореляційного поля та побудову кореляційної таблиці;
- обчислення вибірових коефіцієнтів кореляції та їх аналіз;
- перевірку статистичних гіпотез про значущість зв'язку між досліджуваними випадковими величинами.



	G	H	I	J	K	L
23			<b>Обчислення прогнозного значення <math>y_{пр}</math> в точці <math>x_{пр}</math> за відомими параметрами регресії</b>			
24	$x_{прогн} =$	1,2				
25	$y_{прогн} =$	1,944	=H2*N24+H3			
26			<b>Обчислення прогнозного значення <math>y_{пр}</math> в точці <math>x_{пр}</math> з використанням функції FORECAST</b>			
27	$x_{прогн} =$	1,2				
28	$y_{прогн} =$	1,944	=FORECAST(H27;C2:C21;B2:B21)			

Рис. 4. Обчислення прогнозного значення

Кореляційним полем називають графічне представлення взаємозалежності двох факторів. Для цього на координатній площині точками зображують всі пари чисел  $(x_i, y_i)$ . За формою та розташуванням точок в кореляційному полі можна зробити попередні припущення щодо виду кореляційної залежності між досліджуваними величинами. Для побудови кореляційного поля в OpenOffice Calc застосовують Майстер діаграм, який може бути викликаний за допомогою команд Вставка – Діаграма. На вкладці «Тип діаграммы» потрібно вибрати «Діаграма XY». Якщо на побудованій діаграмі, виділити точки, що задають дані, то на площину діаграми може бути додано лінію регресії (за командами Вставка – Лінія тренда). Тип регресії може бути лінійним, логарифмічним, експоненціальним, або степеневим. Можна, також, вказати, щоб на діаграмі було відображене рівняння тренду та величина коефіцієнту детермінації (встановивши прапорці у відповідні поля). Коефіцієнт детермінації  $R^2$  є величиною, що надає можливість оцінити якість побудованого рівняння, оскільки його значення вказує на ту долю (у %, або в відносних одиницях) зміни залежної величини  $Y$ , що відбувається під впливом фактору  $X$ . Чим ближчий коефіцієнт детермінації до 1 (або 100%), тим більш якісною вважається побудована модель регресії.

Математичною мірою, що характеризує тісноту взаємозв'язку між випадковими величинами є коефіцієнт кореляції, а його статистичною оцінкою є вибіровий коефіцієнт кореляції, що може бути обчислений за формулою(7):

$$r_{YX} = \frac{cov(x, y)}{\sigma_x \cdot \sigma_y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}}, \text{ де} \tag{7}$$

$cov(x, y)$  – кореляційний момент, або коваріація змінних  $X$  та  $Y$ ;

$\sigma_x, \sigma_y$  – вибірові середньоквадратичні відхилення відповідно фактору  $X$  та залежної змінної  $Y$ .

Коефіцієнт кореляції може змінюватись в проміжку [-1; 1]. При чому знак коефіцієнта вказує на вид рівняння кореляції: якщо коефіцієнт від'ємний, то пряма лінія

регресії має обернено пропорційний зв'язок між змінними; якщо коефіцієнт кореляції додатний – то прямо пропорційний. Чим ближче величина коефіцієнта кореляції до  $|1|$ , тим більш тісний зв'язок існує між змінними  $X$  та  $Y$ . Приклади графічного дослідження тісноти та виду зв'язку наведено на рис. 5 – 7.

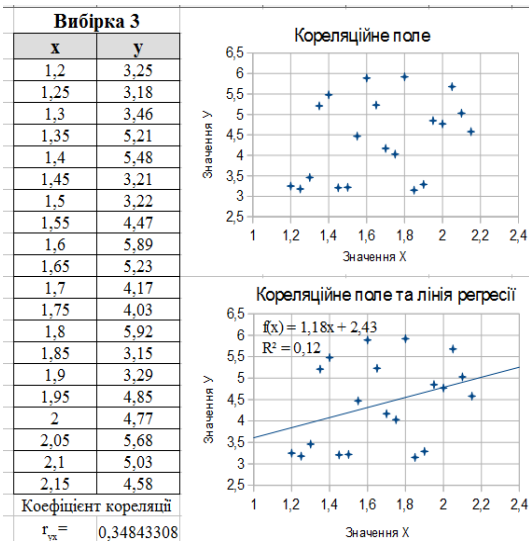


Рис. 5. Зв'язок між змінними практично відсутній

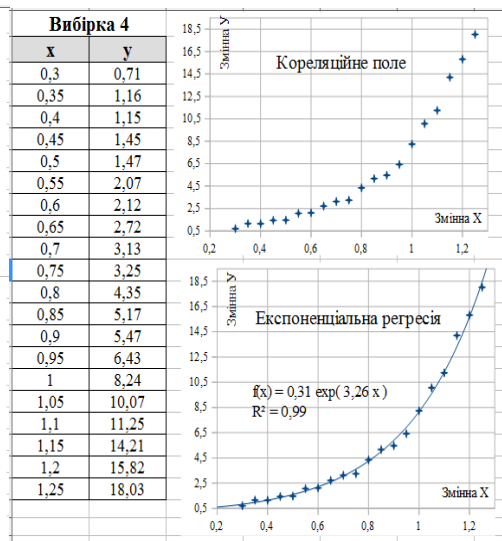


Рис. 6. Нелінійний зв'язок між змінними

Для обчислення коефіцієнта кореляції за формулами використовують вбудовані функції OpenOffice Calc SUMPRODUCT, COUNT, AVERAGE, SUMSQ, STDEVP, які були розглянуті вище. Обчислення кореляційного моменту (коваріації) також можна проводити з використанням статистичної функції COVAR(Данние 1; Данние 2). Коефіцієнт детермінації обчислюється за допомогою статистичної функції RSQ(Данние Y; Данние X). Існує, також, функція для безпосереднього обчислення коефіцієнта кореляції CORREL (Данние 1; Данние 2).

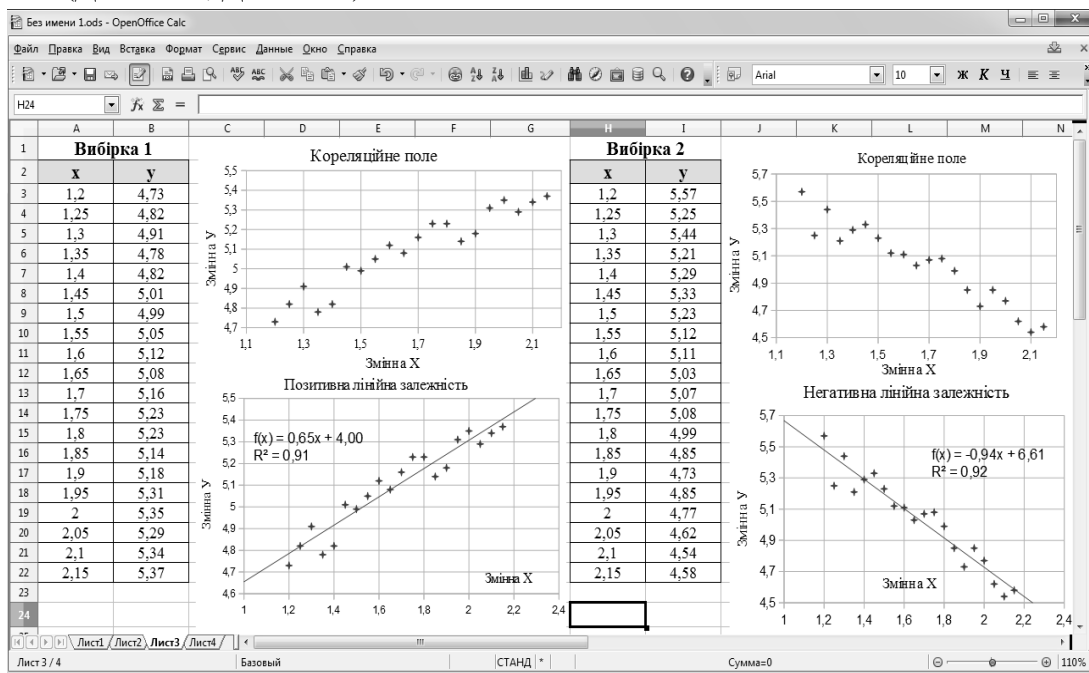


Рис. 7. Тісний лінійний зв'язок

**Висновки та перспективи подальших досліджень.** Таким чином, застосування пакету прикладних офісних програм OpenOffice в процесі стохастичної підготовки майбутніх економістів надає можливість розв'язання значної кількості методичних задач: формування практичних навичок обчислення статистичних величин, формування масивів даних та їх візуальне представлення, графічний аналіз інформації, формування алгоритму обчислення. Крім того, формуються навички застосування комп'ютерних засобів для спрощення розрахунків, оптимального використання часу, необхідного для аналізу та візуалізації інформації, пов'язаної з професійною діяльністю.

#### БІБЛІОГРАФІЯ

1. MDV101: Основы преподавания офисных технологий на базе OpenOffice.org [Електронний ресурс]. – Режим доступу : <[http://www.raspo.ru/files/files/prepodavanie\\_000.pdf](http://www.raspo.ru/files/files/prepodavanie_000.pdf)> – Загол. з екрану. – Мова рос.
2. Prosto Pro Office / Просто Про Office [Електронний ресурс]. – Режим доступу : <<http://prostoprooffice.com/openoffice/>> – Загол. з екрану. – Мова рос.
3. Акимова И. В. Методика работы в СУБД OPEN OFFICE BASE / И. В. Акимова // Известия Пензенского государственного педагогического университета им. В. Г. Белинского. - 2011. - №26. - С. 347 - 357
4. База знаний OpenOffice.org [Електронний ресурс]. – Режим доступу: <[https://wiki.openoffice.org/wiki/RU/knowledge\\_base](https://wiki.openoffice.org/wiki/RU/knowledge_base)> – Загол. з екрану. – Мова рос.
5. Клячкин В. Н. Практикум по статистике, контролю качества и расчетам надежности в OpenOffice.org Calc: учебное пособие / В. Н. Клячкин – Ульяновск : УлГТУ, 2009. – 133 с.
6. Тимохина С. Д. Сравнительный анализ работы программного обеспечения офисных приложений Microsoft Office и OpenOffice.org / С. Д. Тимохина [Електронний ресурс]. – Режим доступу: <[http://mioo.edu.ru/attachments/article/247/Тимохина%20СД%20Сравнительный\\_анализ\\_работы\\_ПО\\_офисных\\_приложений.pdf](http://mioo.edu.ru/attachments/article/247/Тимохина%20СД%20Сравнительный_анализ_работы_ПО_офисных_приложений.pdf)> – Загол. з екрану. – Мова рос.
7. Шереметьева Е. Г. Основы работы с офисным пакетом OpenOffice.org 3.2 : Работа с электронной таблицей OpenOffice.org Calc / Е. Г. Шереметьева [Електронний ресурс]. – Режим доступу: <<http://window.edu.ru/resource/931/72931/files/stup590.pdf>> – Загол. з екрану. – Мова рос.

#### ВІДОМОСТІ ПРО АВТОРА

**Шульга Наталія Вікторівна** – докторант Черкаського національного університету ім. Б. Хмельницького.

*Коло наукових інтересів:* методика навчання стохастики студентів економічних спеціальностей.