

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДНІПРОПЕТРОВСЬКИЙ ДЕРЖАВНИЙ
АГРАРНО-ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

*Присвячується
95-річчю від заснування ДДАЕУ
та 50-річчю створення кафедри
інформаційних систем і технологій*

Васильєва Н. К., Мироненко О. А.,
Самарець Н. М., Чорна Н. О.

ЕКОНОМЕТРИКА В ЕЛЕКТРОННИХ ТАБЛИЦЯХ

Навчальний посібник

*За загальною редакцією
д-ра екон. наук, проф. Н. К. Васильєвої*

Дніпро
Видавець Біла К. О.
2017

УДК 330.43(075.8)
ББК 65в631я73
Е 45

Рекомендовано Міністерством освіти і науки України, ДУ «НМЦ «Агроосвіта», до видання та використання в освітньому процесі для підготовки здобувачів ступеня вищої освіти «бакалавр» зі спеціальностей 071 «Облік і оподаткування», 072 «Фінанси, банківська справа та страхування», 073 «Менеджмент», 075 «Маркетинг» в аграрних навчальних закладах (№ 128/619 від 16.12.2016)

Рекомендовано до друку вченою радою Дніпропетровського державного аграрно-економічного університету (протокол № 4 від 24.11.2016)

Рецензенти:

Ульянченко О. В., член-кор. НААН України, д-р екон. наук, проф., завідувач кафедри маркетингу, підприємництва і організації виробництва Харківського національного аграрного університету ім. В. В. Докучаєва

Правдюк Н. Л., д-р екон. наук, проф., завідувач кафедри бухгалтерського обліку Вінницького національного аграрного університету

Вініченко І. І., д-р екон. наук, проф., завідувач кафедри економічної теорії та економіки сільського господарства Дніпропетровського державного аграрно-економічного університету

Е 45 Економетрика в електронних таблицях : навч. посіб. / Васильєва Н. К., Мироненко О. А., Самарець Н. М., Чорна Н. О. ; за заг. ред. Н. К. Васильєвої. – Дніпро : Біла К. О., 2017. – 149 с.

ISBN 978-617-645-249-2

Розглянуто основи економетрики, пов'язані з побудовою парних та множинних лінійних регресій, аналізом їх статистичної значущості, обчисленням прогнозів, виявленням мультиколінеарності, встановленням причин та наслідків виникнення автокореляції, створенням найбільш розповсюджених нелінійних моделей. По кожній темі наведено основні теоретичні прийоми побудови та аналізу економетричної моделі, запитання та завдання для контролю засвоєння матеріалів розділу. Навчальне видання ілюстровано прикладами проведення обчислень інструментами електронних таблиць MS Excel, OpenOffice Calc та Gnumeric.

Для студентів, аспірантів та викладачів вищих навчальних закладів і фахівців аграрного сектору, які прагнуть підвищити ефективність діяльності засобами економетрики в комп'ютерному середовищі електронних таблиць.

УДК 330.43(075.8)
ББК 65в631я73

ISBN 978-617-645-249-2

© Авторський колектив, 2017

ЗМІСТ

<i>Вступ</i>	4
Розділ 1. Парна лінійна економетрична модель	10
1.1. Етапи побудови і аналізу однофакторної лінійної регресії.....	10
1.2. Створення парної лінійної економетричної моделі в електронних таблицях	23
<i>Контрольні запитання до розділу 1</i>	41
<i>Завдання для контролю засвоєння матеріалу розділу 1</i>	42
Розділ 2. Множинна лінійна економетрична модель	49
2.1. Поняття множинної регресії. Багатофакторний кореляційний аналіз	49
2.2. Лінійний регресійний аналіз: теорія та інструменти табличних процесорів	58
<i>Контрольні запитання до розділу 2</i>	82
<i>Завдання для контролю засвоєння матеріалу розділу 2</i>	83
Розділ 3. Мультиколінеарність	90
3.1. Поняття та ознаки мультиколінеарності	90
3.2. Тестування алгоритмом Феррара–Глобера та вилучення мультиколінеарності	93
<i>Контрольні запитання до розділу 3</i>	103
<i>Завдання для контролю засвоєння матеріалу розділу 3</i>	104
Розділ 4. Автокореляція	107
4.1. Поняття, причини і наслідки автокореляції	107
4.2. Тест Дарбіна–Уотсона та метод Ейткена при автокореляції залишків	111
<i>Контрольні запитання до розділу 4</i>	123
<i>Завдання для контролю засвоєння матеріалу розділу 4</i>	123
Розділ 5. Нелінійні економетричні моделі	126
5.1. Види нелінійних регресій	126
5.2. Створення нелінійних економетричних моделей в електронних таблицях	134
<i>Контрольні запитання до розділу 5</i>	143
<i>Завдання для контролю засвоєння матеріалу розділу 5</i>	143
<i>Список використаної літератури</i>	147

ВСТУП

Ефективна діяльність у будь-якій сфері економіки потребує від спеціаліста використання сучасних наукових досягнень, знання інформаційних технологій та комп'ютерної техніки. Застосування математичних методів в економіці дає змогу формально описати найважливіші зв'язки складових елементів систем і процесів, а також отримати нові знання про об'єкт розгляду. Економетрика – прикладна економіко-математична дисципліна, яка вивчає динаміку реальних мікро- та макроекономічних явищ для кількісного та якісного аналізу й прогнозування результатів розвитку економічних систем. Відомий економіст, один із колишніх президентів Економетричного товариства Ц. Гріліхес писав: «Економетрика є одночасно нашим телескопом і нашим мікроскопом для вивчення навколишнього економічного світу». Важливою задачею економетрії в ринковій економіці є дослідження кількісних зв'язків між показниками для кращого розуміння господарських явищ і процесів, що дозволяє більш обґрунтовано сформулювати управлінські рішення та дати прогнози на майбутнє. Економетричні моделі широко використовуються для аналізу, вивчення і подальшого вдосконалення економіки розвинутих країн.

Економетрика виникла на базі біометрики та успадкувала її термінологію (регресія, кореляція) та методи аналізу (t -критерій Стьюдента, F -критерій Фішера тощо). Особливі досягнення пов'язані з розвитком економетрики в останні десятиріччя, про що свідчить нагородження Нобелівськими преміями ряду вчених, які внесли суттєвий вклад в розвиток математичних методів та моделей. Так, «За створення та застосування динамічних моделей до аналізу економічних процесів» у 1969 р. Нобелівську премію одержали Р. Фріш і Я. Тінберген. У 1980 р. «За створення економетричних моделей і їх застосування до аналізу коливань економіки та економічної політики» Нобелівську премію одержав Л. Клейн. «За роз'яснення

теорії ймовірностей, що заклала основи економетрики, та дослідження одночасних економічних структур» Нобелівським лауреатом 1989 р. став Т. Гаавельмо. У 2000 р. Нобелівську премію одержали Д. Хекман і Д. Макфадден «За розвиток теорії та методів аналізу дискретних вибірок». Нобелівську премію з економіки 2015 року «За дослідження споживання, бідності та добробуту» одержав фахівець з економетрики А. Дітон.

Ефективно вирішувати задачі економетричного моделювання, які базуються на статистичних даних великого обсягу, стало можливим тільки з появою такого потужного та поширеного інструменту, як комп'ютер. У даний час користувачу доступна певна кількість програмних продуктів, які відповідають основним загальним вимогам до програмного забезпечення для вирішення задач економетрики:

- наявність зручних засобів для роботи з вихідними даними;
- розрахунок статистичних характеристик;
- підтримка методів побудови моделей взаємозв'язків та оцінки адекватності моделей;
- вбудовані методи прогнозування та статистичні критерії;
- забезпечення можливості створення сценарію обчислень;
- візуалізація проміжних і кінцевих результатів дослідження.

Програмне забезпечення, що застосовується в економетричних дослідженнях, можна розділити на наступні групи:

1) програми, які реалізують технологію електронних таблиць, наприклад, *MS Excel*, *OpenOffice Calc*, *Gnumeric*;

2) статистичні пакети загального призначення, наприклад, *SPSS*, *STATISTICA*, *STATGRAPHICS*;

3) програми, орієнтовані на рішення економетричних задач, наприклад, *Econometric Views*, *STADIA*, *Matrixer*.

Табличний процесор *MS Excel* – найпопулярніший сучасний обчислювальний інструмент. Він дозволяє легко оперувати з числовими даними, має зручний інтерфейс, дозволяє будувати різноманітні

графіки, безліч діаграм, які сприяють більш повному способу представлення інформації та засвоєння матеріалу. Основними перевагами *MS Excel* є доступність, наявність великого методичного забезпечення з прикладами використання в економетриці та сумісність з більшістю статистичних пакетів прикладних програм.

Незважаючи на незаперечні істотні переваги *MS Excel*, сучасною тенденцією впровадження інформаційних технологій у різних сферах є перехід від пропрієтарного програмного забезпечення, на яке зберігаються як немайнові, так і майнові авторські права, до вільного, автори і власники якого дозволяють вивчати, модифікувати і поширювати свій продукт на безкоштовних засадах. Крім зменшення витрат на інформатизацію, основними перевагами вільного програмного забезпечення є також усунення залежності від одного постачальника, безпечність, висока швидкість розробки нових релізів та випуску нових поправок і версій програмних продуктів.

Одними зі зручних інструментів статистичної обробки даних вільного програмного забезпечення є *Gnumeric* та *OpenOffice Calc*. *Gnumeric* – це табличний процесор, який випускається під ліцензією GNU General Public License. *Gnumeric* входить у базову поставку *Lubuntu* – одного із офіційних відгалужень ОС *Ubuntu*. Програма має ряд особливостей: не перевантажений інтерфейс, розширені функції для телекомунікацій і фінансів, велика кількість вбудованих інженерних та статистичних функцій, висока якість діаграм і широкі можливості їх експорту в різні графічні формати, відсутність обмежень стосовно кількості аркушів у файлі, можливість форматування окремих елементів тексту в комірках електронних таблиць тощо. *Gnumeric* забезпечує сумісність з *MS Excel* та *OpenOffice Calc* по функціях і форматах таблиць. Зважаючи на компактність та високу швидкодію, *Gnumeric* часто включають до складу різних *LiveCD* – операційних систем для завантаження зі зйомного носія (CD, DVD, USB-накопичувача).

OpenOffice Calc – електронні таблиці з відкритого офісного пакету *Apache OpenOffice* (раніше відомого як *OpenOffice.org*). Це потужна програма з високими вимогами до безпеки, функціональності, гнучкості й сумісності. *Calc* випускається під ліцензією GNU Lesser General Public License та за своїми можливостями приблизно відповідає *MS Excel*. Він є аналогом *LibreOffice Calc*. *OpenOffice Calc* дозволяє створювати електронні таблиці та виконувати в них обчислення, проводити сортування та вибірки даних, створювати діаграми, оформляти та друкувати таблиці, проглядати підсумки роботи за допомогою зведених таблиць. *Calc* може відкривати і зберігати файли в форматі *Excel*, але порівняно з *Excel* у нього є додаткові особливості. Наприклад, *Calc* має систему, яка автоматично визначає послідовності графіків, побудованих на основі даних користувача, підтримує створення таблиць у форматі PDF. Однією з властивостей пакету є можливість застосовувати його без інсталяції.

Економетрика використовує традиційні та спеціально розроблені методи для виявлення й аналізу кількісних взаємозв'язків між економічними показниками. Поряд із мікроекономікою та макроекономікою, економетрика дає фундаментальні основи сучасної економічної освіти. У поєднанні з комп'ютерними технологіями вона стала базовою складовою підготовки спеціалістів практично всіх економічних напрямів та викладається в провідних університетах світу. Засобами економетричного моделювання досліджують проблеми ринку, інвестицій, банківської справи, фінансової і кредитної політики, ціноутворення, попиту та пропозиції. Дисципліну «Економетрика» вивчають студенти таких спеціальностей, як «Облік і оподаткування», «Фінанси, банківська справа та страхування», «Менеджмент», «Маркетинг» тощо. Крім того, економетричні розробки поширені у прикладних дослідженнях з біології, соціології, психології та інших суспільних і природничих наук, де необхідно оцінювати моделі на базі великої кількості статистичних даних.

Економетрика є важливим елементом професійної підготовки фахівців у вищих аграрних навчальних закладах України для наближення вітчизняного змісту підготовки кадрів галузі сільського господарства до європейських і світових стандартів. Враховуючи складну стохастичну природу факторів аграрного виробництва, економетрика надає досліднику ефективні методи та інструменти кореляційно-регресійного аналізу статистичних даних, побудови адекватних моделей і прогнозів. Застосування економетричних моделей в аграрній економіці дає змогу виокремити та формально описати найважливіші зв'язки між показниками виробничої діяльності сільськогосподарських підприємств, оцінити потенційні наслідки застосування різних стратегій керування та оптимізації досліджуваних процесів. Чим точніше прогнозування процесів розвитку аграрного сектору країни, тим ефективніше планування й управління цими процесами в реальних умовах господарювання.

Навчальний посібник «Економетрика в електронних таблицях» призначено для студентів економічних спеціальностей вищих навчальних закладів для організації аудиторної та самостійної роботи. Поданий матеріал дає можливість отримати необхідні теоретичні та практичні знання про використання сучасного програмного забезпечення при розв'язуванні конкретних прикладних задач економетричного моделювання, що відповідають реальним економічним обставинам діяльності аграрних підприємств України. Використані в якості вхідних умов статистичні дані охоплюють різноманітні приклади з тваринництва та рослинництва, виробництва і споживання продуктів харчування на макро- та мікрорівнях вітчизняної економіки.

Навчальне видання складається з 5 розділів, де розглянуто основні економетричні положення, пов'язані з такими питаннями, як побудова парних та множинних лінійних регресійних моделей, аналіз їх статистичної значущості, обчислення прогнозів, дослідження наявності мультиколінеарності між пояснювальними змінними,

встановлення причин і наслідків виникнення автокореляції в економетричних моделях, побудова найбільш розповсюджених нелінійних моделей. По кожній темі наведено основні теоретичні положення та запропоновано методологію обчислень, аналіз за отриманими параметрами відповідної моделі, запитання та завдання для контролю засвоєння матеріалів розділу. Навчальне видання ілюстровано прикладами оформлення розрахункових результатів на аркушах електронних таблиць *MS Excel*, *OpenOffice Calc* та *Gnumeric*.

Розроблений посібник створено колективом авторів кафедри інформаційних систем і технологій Дніпропетровського державного аграрно-економічного університету згідно з програмою курсу «Економетрика» для підготовки бакалаврів галузі знань «Управління та адміністрування». При формуванні матеріалів використано авторські та вітчизняні навчально-методичні розробки, матеріали Інтернет-ресурсів, статистичні дані офіційного веб-сайту Державної служби статистики України. Авторський колектив висловлює щирю вдячність рецензентам – професорам О. В. Ульяновку, Н. Л. Правдюк та І. І. Вініченку за ряд корисних рекомендацій та цінних порад під час розгляду рукопису, що значною мірою посприяли поліпшенню змісту навчального видання.

Навчальний посібник адресовано насамперед студентам, аспірантам та викладачам аграрно-економічних спеціальностей вищих навчальних закладів. Крім того, він може бути корисним фахівцям сільського господарства, які прагнуть підвищити ефективність діяльності засобами економетрики в комп'ютерному середовищі електронних таблиць.

Розробка та видання навчального посібника «Економетрика в електронних таблицях» частково профінансовані за темою БФ-7/2017 «Управління розвитком сільськогосподарських ринків, аграрної та екологічної логістики в системі продовольчої безпеки».

Розділ 1

ПАРНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ

1.1. Етапи побудови і аналізу однофакторної лінійної регресії

Перша принципова проблема, з якою стикається кожний, хто проводить економічні дослідження, – це задача про встановлення взаємозв'язків між економічними величинами. Так, попит на деякий товар, що формується на ринку, залежить від ціни цього товару, ціни конкуруючих товарів, платоспроможності споживачів тощо. Витрати, що пов'язані з виготовленням будь-якої продукції, залежать від обсягу виробництва, технології, цін на основні виробничі ресурси. Можна було б навести ще багато прикладів про взаємозв'язки між економічними показниками, адже вони не є ізольованими, автономними, а мають між собою прямий чи зворотний зв'язок. Звідси, щоб ефективно управляти економічними процесами і явищами, треба вміти вимірювати цей зв'язок кількісно. Зазначену проблему економіки можна вирішити, побудувавши економетричну модель.

Модель – це умовне зображення об'єкта, що віддзеркалює його найістотніші характеристики, які необхідні для проведення дослідження. Економетрична модель – це функція чи система функцій, що описує кореляційно-регресійний зв'язок між економічними показниками, один чи кілька з яких є залежними змінними, інші – незалежними величинами.

У загальному вигляді економетрична модель запишеться так:

$$y = f(x_1, x_2, \dots, x_m, e), \quad (1.1)$$

де y – залежна змінна; $x_j, j = 1..m$ – незалежні змінні; e – стохастична складова.

Незалежні змінні моделі називаються пояснювальними, наперед заданими змінними, або факторами. Залежна змінна називається пояснюваною змінною, показником, результативною ознакою тощо.

Весь масив числових даних, що вивчається, називається генеральною сукупністю, а її частина, що піддається аналізу, спостереженню, – вибіркою або вибірковою сукупністю. Як по генеральній, так і по вибірковій сукупності можна розрахувати середні величини, дисперсії, середні квадратичні відхилення і так далі. Вони, як правило, не співпадають, виникають так звані помилки вибірки, що є різницею між показниками генеральної та вибіркової сукупностей.

Обсяг вибірки n – це кількість наборів даних, включених до неї. Оптимальний обсяг вибірки для побудови економетричної моделі повинний бути в 3–5 разів більше кількості факторів. Вибірки можна умовно розділити на великі й малі. Вважається, що вибірки з обсягом більше 30 можна віднести до великих.

Для вибірових сукупностей $x_i, i = 1..n$, обчислюють наступні основні статистичні характеристики.

1. Середнє арифметичне \bar{x} :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1.2)$$

2. Медіана – число, яке є серединою значень вибірки. Якщо обсяг вибірки парний, то відносно медіани половина значень вибірки більше, а половина – менше. Якщо обсяг вибірки не парний, то медіана співпадає з середнім за зростанням значенням вибірки.

3. Мода – значення показника, що зустрічається в сукупності найчастіше.

4. Незміщена вибіркова дисперсія D^2 :

$$D^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}{n(n-1)}. \quad (1.3)$$

У статистичних дослідженнях доводиться застосовувати малі вибірки. Відомо, що репрезентативність (помилка) вибірки значною мірою залежить від її обсягу.

Випадкові помилки вибірки при досить великому її обсязі підпорядковуються закону нормального розподілу. При цьому поправка на величину зміщення вибіркової дисперсії $(n-1)/n$ дуже близька до одиниці і не враховується в обчисленнях.

5. Вибіркове середнє квадратичне відхилення D :

$$D = \sqrt{D^2}$$

характеризує середнє відхилення значень вибірки від \bar{x} .

6. Ексцес \hat{E} :

$$\hat{E} = \left(\frac{n(n-1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{D} \right)^4 \right) - \frac{3(n-1)^2}{(n-2)(n-3)}. \quad (1.4)$$

Ексцес характеризує загострений або згладжений розподіл ознаки в сукупності у порівнянні з нормальним розподілом. Додатний ексцес означає загострений, від'ємний – згладжений розподіл.

7. Асиметрія \hat{A} :

$$\hat{A} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{D} \right)^3. \quad (1.5)$$

Асиметрія характеризує зміщення ознаки в сукупності відносно її середньої величини. Додатна асиметрія означає зрушення розподілу в бік позитивних відхилень від середньої, від'ємна – у зворотний бік.

Дисперсії ексцесу і асиметрії визначаються за формулами:

$$D_{\hat{E}}^2 = \frac{24n(n-1)^2}{(n+5)(n+3)(n-2)(n-3)}; \quad D_{\hat{A}}^2 = \frac{6n(n-1)}{(n+3)(n+1)(n-2)}.$$

На практиці користуються таким наближеним критерієм згоди для генеральної сукупності:

$$|\hat{E}| \leq 2D_{\hat{E}}; \quad |\hat{A}| \leq 2D_{\hat{A}}.$$

Кореляційно-регресійний аналіз або економетричне моделювання виконують за вибірками з коефіцієнтами ексцесу й асиметрії менше 3.

8. Розмах варіації I :

$$I = x_{\max} - x_{\min}, \quad (1.6)$$

де x_{\max} , x_{\min} – максимальне і мінімальне значення ознаки у вибірці.

9. Середня помилка вибірки S :

$$S = \sqrt{\frac{D^2}{n}}. \quad (1.7)$$

Приклад 1.1. За показниками врожайності (z_1), ц/га, та зібраної площі (z_2), тис. га, картоплі в 2003–2014 рр. по Дніпропетровській області, наведеними у табл. 1.1, обчислити валовий збір картоплі (x), тис. т. Із використанням стандартних функцій електронних таблиць *Gnumeric* визначити основні статистичні характеристики отриманої вибірки валового збору картоплі.

Таблиця 1.1

Вхідні дані прикладу 1.1

№ з/п	z_1	z_2
1	125,2	47,5
2	139,9	48,2
3	123,8	47,1
4	120,5	48,3
5	76,2	49,7
6	119,6	47,6
7	92,3	48,8
8	88,5	49,0
9	113,9	49,3
10	91,1	52,9
11	105,8	50,7
12	126,7	52,0

Розв’язання.

Вхідні дані прикладу введено в діапазон комірок B2:C13 аркуша електронних таблиць *Gnumeric* та обчислено валовий збір картоплі (x) у діапазоні комірок D2:D13.

Основні статистичні показники по стовпцю значень x отримано з використанням інструменту *Описательные статистики* надбудови *Статистика* (рис. 1.1).

A	B
	Столбец 1
Среднее	542,08
Стандартная ошибка	25,93
Медиана	565,41
Мода	#N/A
Стандартное отклонение	89,84
Выборочная дисперсия	8071,15
Эксцесс	-0,55
Асимметрия	-0,36
Диапазон	295,60
Минимум	378,71
Максимум	674,32
Сумма	6504,91
Количество	12

Рис. 1.1. *Розрахункові результати прикладу 1.1 в Gnumeric*

Середнє значення валового збору картоплі складає 542,08 тис. т, медіана дорівнює 565,41 тис. т, отже, половина значень ознаки (а саме, 6 значень) – менше значення медіани, половина – більше. Мода в даній вибірковій сукупності відсутня.

Вибіркова дисперсія становить 8071,15; вибіркове середнє квадратичне відхилення 89,84 тис. т. Значення ексцесу і асиметричності відповідно $-0,55$ та $-0,36$ означають згладжений розподіл, зміщений у бік негативних відхилень від середньої величини.

Мінімальне значення валового збору картоплі дорівнює 378,71 тис. т, максимальне 674,32 тис. т, відповідно розмах варіації складає 295,6 тис. т. Середня помилка вибірки 25,93 тис. т.

Приклад 1.2. За показниками поголів'я корів (z_1), тис. голів, і середньорічного надою молока від однієї корови (z_2), кг, в Україні за 2005–2014 рр. (табл. 1.2) обчислити валовий надій молока (x), тис. т. Із використанням стандартних функцій електронних таблиць *MS Excel* визначити основні статистичні характеристики для вибірки валового надою молока.

Вхідні дані прикладу 1.2

№ з/п	z_1	z_2
1	3635,1	3487
2	3346,7	3652
3	3095,9	3665
4	2856,3	3793
5	2736,5	4049
6	2631,2	4082
7	2582,2	4174
8	2554,3	4361
9	2443,0	4446
10	2262,7	4508

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C11 аркуша електронних таблиць *MS Excel* та обчислено валовий надій молока (x) у діапазоні комірок D2:D11.

Основні статистичні показники по стовпцю значень x отримано з використанням інструменту *Описательная статистика* надбудови *Анализ данных* (рис. 1.2).

	F	G	H
1		<i>Столбец1</i>	
2			
3	Среднее		11187,80
4	Стандартная ошибка		233,27
5	Медиана		10970,83
6	Мода		#Н/Д
7	Стандартное отклонение		737,66
8	Дисперсия выборки		544144,44
9	Эксцесс		0,89
10	Асимметричность		1,11
11	Интервал		2475,34
12	Минимум		10200,25
13	Максимум		12675,59
14	Сумма		111878,04
15	Счет		10
16	Уровень надежности(95,0%)		527,69

Рис. 1.2. **Розрахункові результати прикладу 1.2 у MS Excel**

Середнє значення валового надою молока складає 1187,80 тис. т, медіана дорівнює 10970,83 тис. т, отже, половина значень ознаки (а саме, 5 значень) – менше значення медіани, половина – більше. Мода в даній вибірковій сукупності відсутня.

Вибіркова дисперсія дорівнює 544144,44, вибіркове середнє квадратичне відхилення 737,66 тис. т. Значення ексцесу і асиметричності відповідно 0,89 та 1,11 означають загострений розподіл, зміщений у бік позитивних відхилень від середньої величини.

Мінімальне значення валового надою молока становить 10200,25 тис. т, максимальне 12675,59 тис. т, відповідно розмах варіації 2475,34 тис. т. Середня помилка вибірки дорівнює 233,27 тис. т.

Альтернативними (до розглянутих у *прикладax 1.1* та *1.2*) засобами обчислення статистичних характеристик вибірок слугують такі стандартні функції *MS Excel*: *СРЗНАЧ*, *МЕДИАНА*, *МОДА*, *СТАНДОТКЛОН*, *ДИСП*, *ЭКЦЕСС*, *СКОС*, *МАКС*, *МИН*, *СУММ*, *СЧЁТ*. При застосуванні програми *OpenOffice Calc* використовують аналогічні функції: *AVERAGE*, *MEDIAN*, *MODE*, *STDEV*, *VAR*, *KURT*, *SKEW*, *MAX*, *MIN*, *SUM*, *COUNT*. У табличному процесорі *Gnumeric* відповідні функції прописані так: *average*, *median*, *mode*, *stdev*, *var*, *kurt*, *skew*, *max*, *min*, *sum*, *count*.

Кореляція (лат. *correlatio* – співвідношення, взаємозв'язок) або кореляційна залежність – це статистичний взаємозв'язок двох випадкових величин, причому зміни значень однієї з цих величин приводять до зміни значень іншої.

Регресія (лат. *regressio* – зворотній рух, відхід) – математичний вираз, що відображає залежність між залежною змінною *y* і незалежною змінною *x* за умови, що цей вираз матиме статистичну значущість та економічний зміст.

Побудова будь-якої економетричної моделі здійснюється за наступним планом:

– постановка задачі;

- специфікація моделі;
- параметризація моделі;
- верифікація моделі;
- використання моделі.

Найпростішою серед лінійних економетричних моделей є парна однофакторна лінійна регресія, котра описує зв'язок між двома економічними показниками, один з яких (y) є залежною (пояснюваною), а другий (x) – незалежною (пояснювальною) змінною:

$$\hat{y} = a_0 + a_1x, \quad (1.8)$$

де числовий параметр a_0 називається перетином, числовий параметр a_1 – нахилом.

До найбільш розповсюджених методів знаходження числових значень параметрів економетричної моделі відноситься метод найменших квадратів. У його основу покладено критерій, згідно якого «найкращою» серед усіх можливих вважається така функція регресії, для якої сума квадратів залишків є мінімальною. З математичної точки зору цей критерій має такий вигляд:

$$\sum_{i=1}^n e_i^2 \rightarrow \min. \quad (1.9)$$

Параметри a_0 і a_1 можна визначити за формулами:

$$a_1 = \frac{\text{cov}(x, y)}{D_x^2}; \quad (1.10)$$

$$a_0 = \bar{y} - a_1\bar{x}, \quad (1.11)$$

де $\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ – вибірковий коефіцієнт коваріації;

$D_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ – вибірка дисперсія незалежної змінної x ;

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ – середнє значення незалежної змінної x ;

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ – середнє значення залежної змінної y .

До основних показників якості побудованої економетричної моделі відносяться наступні:

- стандартна похибка рівняння регресії σ_e ;
- коефіцієнт кореляції R ;
- коефіцієнт детермінації R^2 .

Стандартна похибка рівняння регресії σ_e визначається за формулою:

$$\sigma_e = \sqrt{\sigma_e^2}, \quad (1.12)$$

де σ_e^2 – незміщена оцінка дисперсії залишків, яка обчислюється за співвідношенням:

$$\sigma_e^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}. \quad (1.13)$$

Коефіцієнт кореляції є кількісною мірою тісноти лінійного зв'язку між змінними моделі та визначається за формулою:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1.14)$$

де y_i – вибіркові значення залежної змінної; x_i – вибіркові значення незалежної змінної.

Коефіцієнт кореляції набуває значень у діапазоні від -1 до 1 , тобто $-1 \leq R \leq 1$. Знак коефіцієнта кореляції показує напрям кореляційної залежності – прямий або зворотний. Абсолютна величина вказує на силу кореляційного зв'язку – дуже тісний, сильний, середній, помірний чи слабкий (табл. 1.3).

Характер кореляційного зв'язку

Сила зв'язку	Характер зв'язку	
	Прямий (+)	Зворотний (-)
Повний	1	-1
Дуже тісний	від 0,9 до 0,99	від -0,99 до -0,9
Сильний (високий, суттєвий)	від 0,7 до 0,9	від -0,9 до -0,7
Середній (помітний)	від 0,5 до 0,7	від -0,7 до -0,5
Помірний	від 0,3 до 0,5	від -0,5 до -0,3
Слабкий	від 0 до 0,3	від -0,3 до 0
Лінійний зв'язок відсутній	0	0

Коефіцієнт детермінації використовується як критерій адекватності (відповідності) моделі статистичним даним, оскільки він є мірою пояснювальної сили незалежної змінної та показує, яка частина варіації залежної змінної пояснюється лінійною варіацією (змінною) незалежної ознаки, а не іншими випадковими факторами, які акумулюються у стохастичній складовій моделі.

Коефіцієнт детермінації характеризує, наскільки значним є лінійний вплив пояснювальної змінної моделі на залежну. Якщо цей вплив є значним, то побудована модель дійсно описує лінійну залежність між відповідними економічними показниками, і ця залежність є суттєвою. Якщо ж цей вплив є незначним, то модель є неадекватною статистичним даним, і лінійна регресійна залежність між економічними показниками у ній є достатньо сумнівною та неякісною. Коефіцієнт детермінації позначається через R^2 і обчислюється на основі попередньо знайденого коефіцієнта кореляції за формулою:

$$R^2 = (R)^2. \quad (1.15)$$

Коефіцієнт детермінації набуває значень у діапазоні від 0 до 1, тобто $0 \leq R^2 \leq 1$. Чим більше значення коефіцієнта детермінації (чим ближче воно до 1), тим більш вагомим та систематичним є вплив пояснювальної змінної на залежну, і тим більше підстав

стверджувати, що саме варіацією значень пояснювальної змінної пояснюється варіація значення залежної змінної моделі, а не іншими випадковими та неврахованими у моделі факторами. Іншими словами, близьке до 1 значення коефіцієнта детермінації свідчить про високий рівень адекватності побудованої регресійної моделі статистичним даним.

Навпаки, чим менше значення коефіцієнта детермінації (чим ближче воно до 0), тим менш вагомим є вплив пояснювальної змінної на залежну і тим менше підстав стверджувати, що саме зміною значень пояснювальної змінної визначається зміна значення залежної змінної моделі, а не іншими випадковими та неврахованими у моделі факторами. Іншими словами, мале, близьке до 0 значення коефіцієнта детермінації свідчить про низький рівень адекватності побудованої регресійної моделі статистичним даним.

Перевірка статистичної значущості моделі у цілому зводиться до перевірки статистичної значущості коефіцієнта детермінації, тобто перевірки того, чи суттєво його значення за генеральною сукупністю відрізняється від знайденого за вибіркою. Це дає підставу вважати, що відібрані фактори пояснюють варіацію залежної змінної y , тобто побудована модель адекватна фактичним статистичним даним.

Для перевірки статистичної значущості (або адекватності) моделі у цілому використовується F -критерій Фішера:

$$F = \frac{R^2}{1 - R^2} \cdot (n - 2), \quad (1.16)$$

де n – кількість вибірових спостережень.

Розрахункове значення F -критерію знаходиться у вихідних даних інструменту *Регресія* та стандартній функції *ЛИНЕЙН* у *MS Excel*. Аналогічні розрахунки F -критерію можна отримати в *OpenOffice Calc* стандартною функцією *LINEST*, а у програмі *Gnumeric* – за допомогою інструменту *Регресія* або функцією *linest*.

Для перевірки статистичної значущості моделі у цілому слід виконати такі кроки.

1. На основі значення вибіркового коефіцієнта детермінації R^2 за формулою (1.16) обчислити розрахункове значення F -критерію.

2. Визначити критичне значення F -критерію $F(\alpha; \nu_1; \nu_2)$ для заданого рівня значущості α (як правило, $\alpha = 0,05$ або $\alpha = 0,01$) зі ступенями свободи $\nu_1 = 1$ і $\nu_2 = n - 2$ за допомогою стандартних функцій *FPACPOBP* (*MS Excel*) або *FINV* (*OpenOffice Calc* та *Gnumeric*).

3. Якщо виконується умова $F > F(\alpha; \nu_1; \nu_2)$, то це свідчить про статистичну значущість побудованої регресійної моделі у цілому на рівні значущості α та її адекватність генеральній сукупності. У протилежному випадку регресійна модель вважається статистично незначущою у цілому, тобто неадекватною статистичним даним. Довірча ймовірність регресійної моделі для генеральної сукупності складає $p = 1 - \alpha$.

Для обчисленого F -критерію можна знайти точний рівень значущості α за допомогою стандартних функцій *FPACPI* (*MS Excel*) або *FDIST* (*OpenOffice Calc* та *Gnumeric*) зі ступенями свободи $\nu_1 = 1$ і $\nu_2 = n - 2$.

У випадку статистично значущої моделі необхідно визначити, за рахунок яких параметрів забезпечується її загальна статистична значущість. Для цього необхідно перевірити статистичну значущість кожного параметра моделі за двостороннім t -критерієм Стьюдента:

$$t_{a_j} = \frac{a_j}{\sigma_{a_j}}, \quad j = 0, 1, \quad (1.17)$$

де a_j – значення j -го параметра лінійної регресії; σ_{a_j} – його стандартна похибка.

Перевірка статистичної значущості кожного параметра моделі здійснюється за такою схемою.

1. Для кожного параметру a_j за формулою (1.17) обчислити розрахункове значення критерію Стюдента t_{a_j} .

2. Для вибраного рівня значущості α (як правило, $\alpha = 0,05$ або $\alpha = 0,01$) і ступеня свободи $\nu = n - 2$ за допомогою стандартних функцій *СТЬЮДРАСПОБР* (*MS Excel*) або *TINV* (*OpenOffice Calc* та *Gnumeric*) знайти двостороннє критичне значення t -розподілу Стюдента $t(\alpha, \nu)$.

3. Якщо виконується умова $|t_{a_j}| > t(\alpha; \nu)$, то це свідчить про суттєву статистичну значущість a_j для генеральної сукупності. У протилежному випадку параметр a_j вважається статистично незначущим.

Для обчисленого t -критерію можна знайти точний рівень значущості α за допомогою стандартних функцій *СТЬЮДРАСП* (*MS Excel*) або *TDIST* (*OpenOffice Calc* та *Gnumeric*) зі ступенем свободи $\nu = n - 2$.

Для статистично значущих параметрів моделі можна побудувати їх інтервали довіри. Інтервал довіри – це проміжок, у який з наперед заданою ймовірністю потрапляють фактичні, дійсні значення параметрів моделі, тобто параметри побудованої лінійної регресії. Інтервали довіри параметрів моделі являють собою інтервальні оцінки реальних параметрів \hat{a}_j на відміну від точкових a_j , $j = 0, 1$.

Довірчі інтервали мають вигляд:

$$a_j - t(\alpha; \nu) \cdot \sigma_{a_j} \leq \hat{a}_j \leq a_j + t(\alpha; \nu) \cdot \sigma_{a_j}, \quad j = 0, 1, \quad (1.18)$$

де a_j – значення параметра моделі (вибірковий параметр моделі); σ_{a_j} – його стандартна похибка; $t(\alpha; \nu)$ – критичне значення

двостороннього t -критерію Стьюдента для рівня значущості α і ступеня свободи $\nu = n - 2$.

Перевірка статистичної значущості коефіцієнта кореляції зводиться до перевірки того, чи суттєво відрізняються його значення для вибірки та генеральної сукупності. Вона здійснюється за допомогою t -критерію Стьюдента, розрахункове значення якого у цьому випадку обчислюється за формулою:

$$t_R = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}, \quad (1.19)$$

де R – вибірковий коефіцієнт кореляції (1.14).

Процедура перевірки статистичної значущості вибіркового коефіцієнта кореляції аналогічна процедурі перевірки статистичної значущості параметрів моделі. При виконанні умови $|t_R| > t(\alpha; \nu)$ можна зробити висновок, що вибірковий коефіцієнт кореляції є статистично значущим і характеризує дійсний, реально існуючий лінійний зв'язок між залежною та пояснювальною змінною моделі. У протилежному випадку вибірковий коефіцієнт кореляції є статистично не значущим, а лінійний зв'язок між змінними моделі відсутній.

Якщо побудована модель є адекватною в цілому, а її параметри статистично значущі, то її можна застосовувати для прогнозування значень залежної змінної.

1.2. Створення парної лінійної економетричної моделі в електронних таблицях

Серед численних зв'язків між економічними показниками завжди можна виділити такий показник, вплив якого на результативну ознаку є основним, найбільш важливим. Щоб виміряти цей зв'язок кількісно, необхідно побудувати парну регресію. Розглянемо

побудову та аналіз лінійної економетричної моделі з використанням стандартних функцій електронних таблиць на прикладах.

Приклад 1.3. У табл. 1.4 наведено вибірккові статистичні дані по Україні за 2002–2015 рр. щодо врожайності овочів (y), ц/га, та обсягу внесених мінеральних добрив (x), кг/га. Із використанням стандартних функцій електронних таблиць *OpenOffice Calc* побудувати та провести дослідження економетричної моделі у вигляді парної лінійної регресії.

Розв’язання.

Вхідні дані прикладу введено в діапазон комірок B2:C15 аркуша електронних таблиць *OpenOffice Calc*. Економетрична модель у вигляді парної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x.$$

Таблиця 1.4

Вхідні дані прикладу 1.3

№ з/п	y	x
1	103,1	33
2	128,4	41
3	138,6	64
4	154,5	90
5	169,8	94
6	167,2	138
7	211,8	124
8	256,3	134
9	207,0	159
10	280,5	162
11	313,7	196
12	312,3	197
13	346,4	201
14	363,4	194

Коефіцієнт кореляції R розраховано за допомогою стандартної функції в комірці E2 = *CORREL*(B2:B15;C2:C15). Відповідно до

табл. 1.3, значення $R = 0,93$ свідчить про дуже тісний прямий кореляційний зв'язок між ознаками y та x .

Для виявлення економетричної залежності побудовано діаграму розсіювання *Мастером діаграмм*. З контекстного меню на діаграмі обрано *Вставити лінію тренда* → *Линейная* та встановлено два прапорця: *Показати рівняння*, *Показати коефіцієнт детермінації*. Результат, зображений на рис. 1.3, дає візуальне уявлення про існування лінійної залежності між змінними, тобто вхідні дані можна наблизити прямою лінією на кореляційному полі точок.

За допомогою стандартних функцій отримано числові значення параметрів рівняння регресії. Параметр a_0 обчислено в комірці G2 = *INTERCEPT*(B2:B15;C2:C15). Параметр a_1 розраховано в комірці F2 = *SLOPE*(B2:B15;C2:C15).

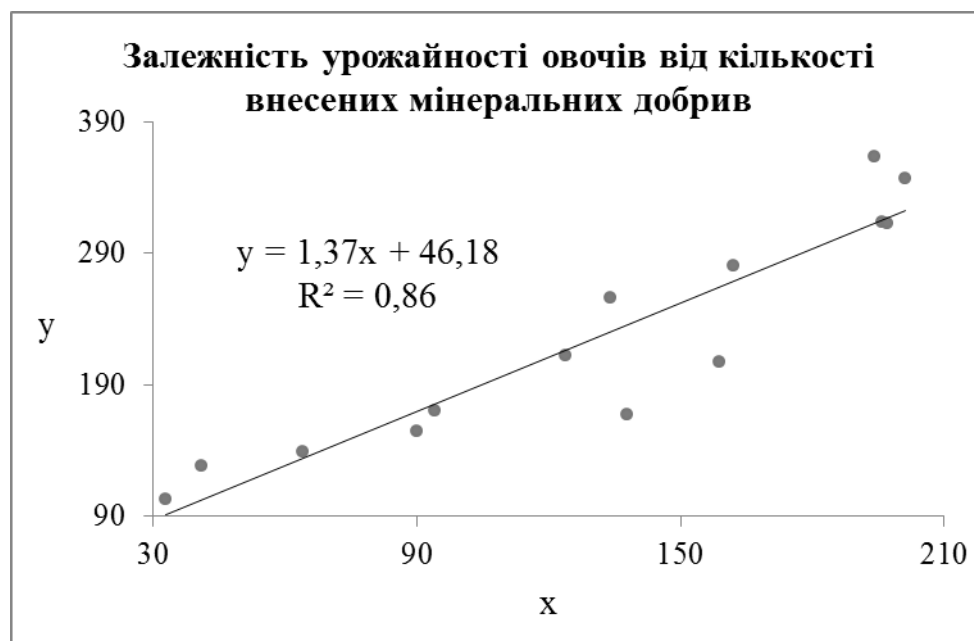


Рис. 1.3. Діаграма розсіювання прикладу 1.3 в *OpenOffice Calc*

Рівняння залежності врожайності овочів від кількості внесених мінеральних добрив збігається за двома способами та має вигляд:

$$\hat{y} = 46,18 + 1,37x.$$

Дамо змістовне тлумачення отриманих результатів. Параметр a_1 характеризує граничну зміну врожайності овочів залежно від зміни кількості внесених мінеральних добрив: якщо кількість внесених добрив зростає на 1 кг, то урожайність овочів у середньому збільшиться на 1,37 ц/га. Числове значення параметра $a_0 = 46,18$ задає врожайність овочів без внесення мінеральних добрив.

Значення коефіцієнта детермінації R^2 , відображене на діаграмі, можна також обчислити, скориставшись відомим значенням коефіцієнта кореляції R в комірці E4 = E2^2, чи функцією в комірці E5 = RSQ(B2:B15;C2:C15). Значення $R^2 = 0,86$ вказує, що варіація врожайності овочів на 86 % визначається лінійною варіацією кількості внесених мінеральних добрив. Величини R та R^2 для парної економетричної моделі свідчать про її достовірність, тому що їх значення наближаються до одиниці.

Значення F -критерію розраховано в комірці E7 = E5/(1-E5)*12 за співвідношенням (1.16). Одержано $F = 76,66$.

Рівень значущості визначеного F -критерію обчислено в комірці F7 = FDIST(E7;1;12) та отримано $\alpha(F) = 1,48E - 06$. Отже, довірчав ймовірність регресійної моделі для генеральної сукупності складає $p \approx 1$. Це свідчить про статистичну значущість побудованої моделі та її адекватність статистичним даним генеральної сукупності.

Приклад 1.4. Із використанням стандартних функцій електронних таблиць *MS Excel* побудувати та провести дослідження достовірності економетричної моделі у вигляді парної лінійної регресії, яка характеризує залежність середніх цін реалізації яєць (y), грн/тис. шт., від їх обсягу виробництва (x), млн шт., у сільськогосподарських підприємствах України за вибіркою 2005–2014 рр. (табл. 1.5).

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C11 аркуша електронних таблиць *MS Excel*.

Вхідні дані прикладу 1.4

№ з/п	y	x
1	251,8	6458,1
2	192,7	7632,8
3	274,4	7604,5
4	377,4	8470,4
5	403,9	9258,2
6	470,6	10249,6
7	521,5	11738,2
8	627,0	11977,4
9	656,7	12234,2
10	782,4	12536,2

Економетрична модель у вигляді парної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x.$$

Парний коефіцієнт кореляції розраховано в комірці E2 = *КОРРЕЛ*(B2:B11;C2:C11). Його значення $R = 0,95$ свідчить про дуже тісний прямий кореляційний зв'язок між ознаками y та x . Таким чином, зі збільшенням значення незалежної змінної x пояснювана змінна y монотонно зростатиме.

Для перевірки статистичної значущості коефіцієнта кореляції у комірці E4 = *E2*КОРЕНЬ*(8)/*КОРЕНЬ*(1-E2^2) обчислено значення $t_R = 8,55$ за формулою (1.19). Точний рівень значущості розраховано в комірці E6 = *СТЬЮДРАСП*(E4;8;2) та отримано $\alpha(t_R) = 2,68E - 05$. Довірча ймовірність коефіцієнта кореляції для генеральної сукупності складає $p \approx 1$, що свідчить про статистичну значущість коефіцієнта кореляції за критерієм Стьюдента.

Для виявлення економетричної залежності побудовано діаграму розсіювання: *Вставка*→*Діаграма*, тип *Точечная*. У контекстному меню на діаграмі вибрано *Добавить линию тренда* і встановлено

прапорці *Показати уравнение на диаграмме* та *Поместить на диаграмму величину достоверности аппроксимации R²*.

Результат, представлений на рис. 1.4, дає візуальне уявлення про існування лінійної залежності між змінними, тобто вхідні дані можна наблизити прямою лінією на кореляційному полі точок.

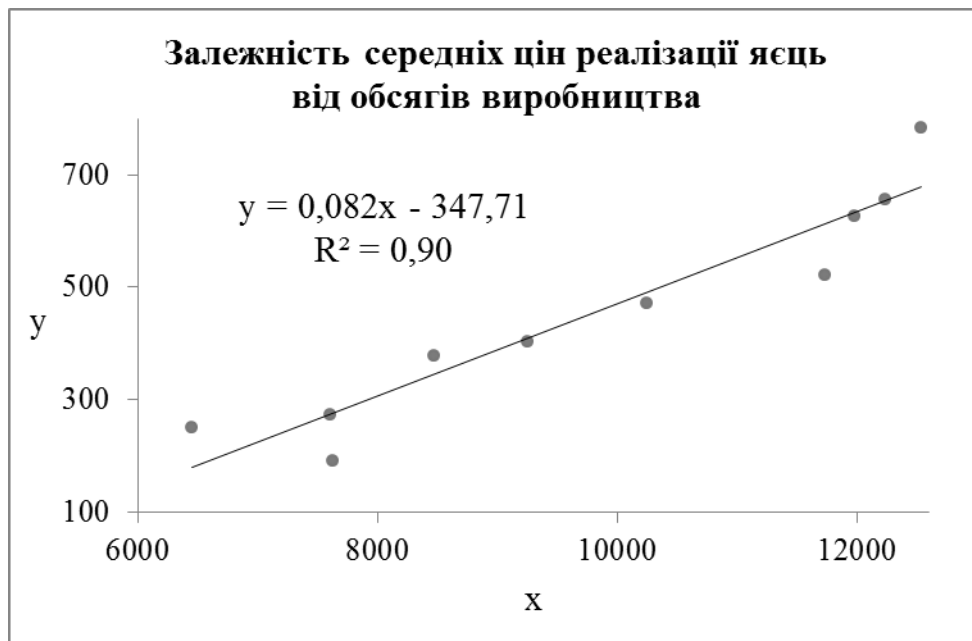


Рис. 1.4. *Діаграма розсіювання прикладу 1.4 у MS Excel*

Для обчислення параметрів лінійної регресії використано стандартну функцію *ЛИНЕЙН*, результати роботи якої розміщено в діапазоні комірок G2:H6 (рис. 1.5).

fx {=ЛИНЕЙН(B2:B11;C2:C11;1;1)}				
	F	G	H	I
1				
2		0,08	-347,71	
3		0,01	96,10	
4		0,90	64,09	
5		73,18	8	
6		300595,69	32861,37	
7				

Рис. 1.5. *Розрахункові результати прикладу 1.4 у MS Excel*

Відповідність одержаних числових значень параметрам моделі показано в табл. 1.6.

Таблиця 1.6

Регресійна статистика функції ЛИНЕЙН

Значення параметра a_1	0,08	-347,71	Значення параметра a_0
Стандартна похибка параметра a_1	0,01	96,10	Стандартна похибка параметра a_0
Коефіцієнт детермінації R^2	0,90	64,09	Стандартна похибка рівняння регресії σ_e^2
F -критерій Фішера	73,18	8	Число ступенів свободи $\nu_2 = n - 2$
Регресійна сума квадратів $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	300595,7	32861,37	Залишкова сума квадратів $\sum_{i=1}^n (y_i - \bar{y})^2$

Розрахована економетрична модель має вигляд:

$$\hat{y} = -347,71 + 0,08x.$$

Параметр $a_1 = 0,08$ показує, що зі збільшенням обсягу виробництва яєць на 1 млн шт. середні ціни їх реалізації зростуть на 0,082 грн/тис. шт. Параметр a_0 для даного прикладу не має економічного сенсу.

Коефіцієнт детермінації $R^2 = 0,90$ свідчить про те, що 90 % варіації результативної ознаки y пояснюються лінійною варіацією незалежної ознаки x , решта 10 % варіації зумовлена дією не включених до економетричної моделі факторів або присутністю в ній нелінійних компонентів.

Для отриманого значення F -критерію обчислено рівень значущості α в комірці $F5 = FPACP(G5;1;8)$. Одержано $\alpha(F) = 2,8E - 05$, що відповідає довірчій імовірності $p \approx 1$ для генеральної сукупності. Це свідчить про статистичну значущість побудованої лінійної парної

моделі у цілому та її адекватність статистичним даним генеральної сукупності.

Порівняння абсолютних значень параметрів регресії a_0 і a_1 з відповідними стандартними похибками вказує, що абсолютні значення параметрів більші, ніж їх стандартні похибки. Даний висновок є одним із свідчень статистичної значущості параметрів.

Значення t -критерію для параметрів регресії a_0 і a_1 розраховано в комірках $G11 = G2/G3$ та $H11 = H2/H3$ за формулою (1.17). Для обчислених t -критеріїв знайдено двосторонні рівні значущості:

$$G13 = \text{СТБЮДРАСП}(\text{ABS}(G11);8;2); \quad \alpha(t_{a_1}) = 2,7\text{E} - 05;$$

$$H13 = \text{СТБЮДРАСП}(\text{ABS}(H11);8;2); \quad \alpha(t_{a_0}) = 0,007.$$

Параметри економетричної моделі є статистично значущими за t -критерієм Стьюдента, про що можна судити зі значень показника довірчої ймовірності $p = 1 - \alpha = 0,993$.

Приклад 1.5. Із використанням інструменту *Регрессия* табличного процесора *Gnumeric* побудувати економетричну модель у вигляді парної лінійної залежності валової продукції рослинництва в розрахунку на 100 га угідь (y), тис. грн, від продуктивності праці на 1 зайнятого в галузі (x), тис. грн, за вибіркою для сільськогосподарських підприємств України 2005–2014 рр. (табл. 1.7).

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C11 аркуша електронних таблиць *Gnumeric*. Коефіцієнт кореляції розраховано в комірці D12 = *correl*(B2:B11;C2:C11). Отримано $R = 0,87$, що свідчить про суттєвий кореляційний зв'язок між ознаками y і x (табл. 1.3).

Із використанням *Мастера диаграмм* побудовано діаграму розсіювання: *Вставка*→*Диаграмма*→*XY*. На вкладці *Настройка диаграммы* обрано *Добавить*→*Линия тренда*→*линейное*→*Уравнение линейная регрессия*.

Вхідні дані прикладу 1.5

№ з/п	y	x
1	229,0	86,19
2	257,6	38,59
3	241,4	36,94
4	363,3	57,98
5	328,4	138,74
6	310,5	133,60
7	433,9	171,14
8	388,6	155,54
9	484,1	202,22
10	522,1	228,88

Результат, представлений на рис. 1.6, дає візуальне уявлення про існування лінійної залежності між змінними, тобто вхідні дані можна наблизити прямою лінією на кореляційному полі точок.

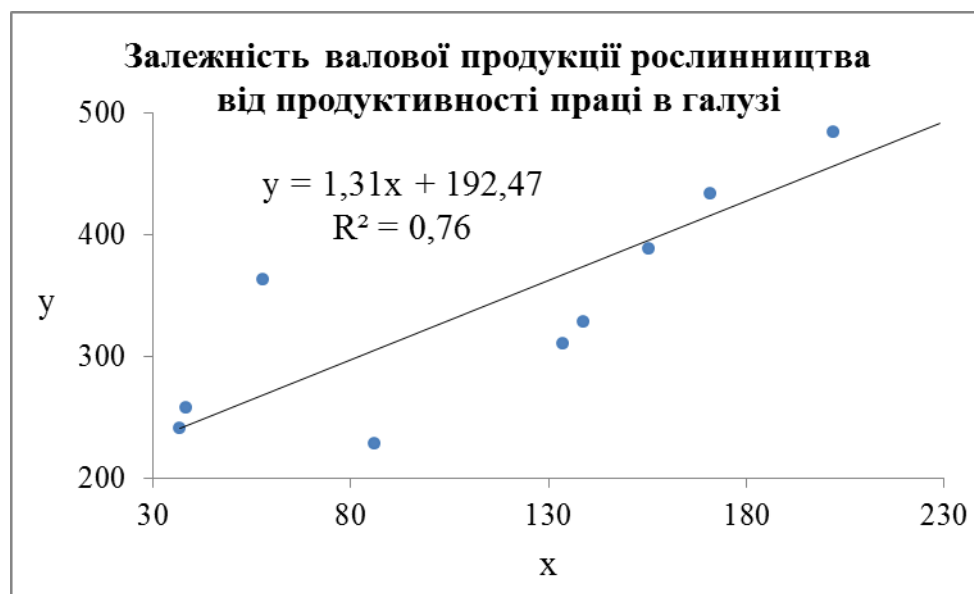


Рис. 1.6. Діаграма розсіювання прикладу 1.5 у *Gnumeric*

Для дослідження економетричної моделі використано надбудову *Статистика*→*Зависимые наблюдения*→*Регрессия*. Результати роботи інструменту *Регрессия* розміщено на новому аркуші електронних

таблиць, починаючи з комірки A1 (рис. 1.7). У стовпці *Коеффіцієнти* розташовані числові значення параметрів моделі a_0 і a_1 . Економетрична модель у вигляді парної лінійної регресії має вигляд:

$$\hat{y} = 192,47 + 1,31x.$$

	A	B	C	D	E	F	G
1	Итоговый вывод		Response Variable	Столбец 3			
2							
3	Регрессионные статистики						
4	Множественная R	0,87					
5	R^2	0,76					
6	Стандартная ошибка	52,86					
7	Вычисленная R	0,73					
8	Наблюдения	10					
9							
10	Дисперсионный анализ						
11		степень свободы	сумма квадратов	Квадрат среднего	F	Значимость F	
12	Регрессия	1	70609,03	70609,03	25,27	0,001018	
13	Остатки	8	22350,86	2793,86			
14	Всего	9	92959,89				
15							
16		Коеффициенты	Стандартная ошибка	t-статистика	p-значение	Ниже 95%	Выше 95%
17	Пересечение	192,47	36,55	5,27	0,00076	108,17	276,76
18	Столбец 2	1,31	0,26	5,03	0,00102	0,71	1,91
19							
20	Ограничение	Столбец 2	Prediction	Столбец 3	Residual	Leverages	Internally studentized
21	1	86,19	305,17	229	-76,17	0,14	-1,55
22	1	38,59	242,93	257,6	14,67	0,28	0,33
23	1	36,94	240,77	241,4	0,63	0,29	0,01
24	1	57,98	268,28	363,3	95,02	0,21	2,02
25	1	138,74	373,88	328,4	-45,48	0,10	-0,91
26	1	133,6	367,16	310,5	-56,66	0,10	-1,13
27	1	171,14	416,25	433,9	17,65	0,15	0,36
28	1	155,54	395,85	388,6	-7,25	0,12	-0,15
29	1	202,22	456,88	484,1	27,22	0,24	0,59
30	1	228,88	491,74	522,1	30,36	0,36	0,72

Рис. 1.7. Розрахункові результати прикладу 1.5 у Gnumeric

Параметр $a_1 = 1,31$ визначає, що зі збільшенням продуктивності праці сільськогосподарських підприємств галузі рослинництва на 1 тис. грн валова продукція рослинництва у розрахунку на 100 га угідь зростає на 1,31 тис. грн. Параметр a_0 для даного прикладу не має економічного сенсу.

Розраховане значення коефіцієнта детермінації $R^2 = 0,76$ свідчить про сильний зв'язок між ознаками, тобто 76 % загальної варіації валової продукції рослинництва сільськогосподарських підприємств обумовлена лінійними змінами продуктивності праці в галузі рослинництва, а 24 % – впливом не включених до регресійної моделі факторів або нелінійних компонентів моделі.

Значення F -критерію Фішера $F = 25,27$ має рівень значущості $\alpha(F) = 0,001$, тобто близький до 0, що свідчить про високу довірчу ймовірність отриманої регресійної моделі для генеральної сукупності.

У стовпці *P-значення* знаходяться рівні значущості α , які відповідають обчисленим t -критеріям: для генеральної сукупності коефіцієнти регресії a_0 і a_1 мають довірчу ймовірність $p = 0,99$.

Стовпці *Нижче 95%* та *Вьше 95%* – відповідно нижні та верхні межі інтервалів дійсних коефіцієнтів регресії для довірчої ймовірності 0,95, тобто:

$$108,17 \leq \hat{a}_0 \leq 276,77; \quad 0,71 \leq \hat{a}_1 \leq 1,91.$$

Побудована економетрична модель у вигляді однофакторної лінійної регресії статистично значуща і адекватна даним згідно зі статистичними критеріями.

Приклад 1.6. У табл. 1.8 наведено вибіркові статистичні дані по Україні за 2005–2014 рр. щодо середньорічного надою молока від однієї корови (y), кг, та середньої ціни реалізації молока та молочних продуктів (x), грн/т. Із використанням інструменту *Регресія* програми *MS Excel* побудувати та провести кореляційно-регресійний аналіз економетричної моделі у вигляді парної лінійної регресії, яка характеризує залежність між указаними ознаками.

Таблиця 1.8

Вхідні дані прикладу 1.6

№ з/п	y	x
1	3487	1126,9
2	3652	1070,2
3	3665	1660,6
4	3793	2065,1
5	4049	1888,8
6	4082	2938,7
7	4174	3041,6
8	4361	2662,2
9	4446	3364,0
10	4508	3588,4

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C11 аркуша електронних таблиць *MS Excel*. Коефіцієнт кореляції розраховано в комірці D13 = *KORPEЛ*(B2:B11;C2:C11). Відповідно до табл. 1.3, значення коефіцієнта кореляції $R = 0,92$ свідчить про дуже тісний прямий кореляційний зв'язок між ознаками y і x .

Для проведення кореляційно-регресійного аналізу лінійної економетричної моделі використано надбудову *Данные*→*Анализ данных*→*Регрессия*.

На новому аркуші електронних таблиць сформовано результати роботи інструменту *Регрессия* (рис. 1.8).

	A	B	C	D	E	F	G
1	Вывод ИТОГОВ						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	0,92					
5	R-квадрат	0,85					
6	Нормированный R-квадрат	0,83					
7	Стандартная ошибка	146,13					
8	Наблюдения	10					
9							
10	<i>Дисперсионный анализ</i>						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
12	Регрессия	1	990503,49	990503,49	46,38	0,000136	
13	Остаток	8	170836,61	21354,58			
14	Итого	9	1161340,10				
15							
16		<i>Коэффициент</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
17	Y-пересечение	3164,56	134,07	23,60	1,10E-08	2855,40	3473,73
18	Переменная X 1	0,37	0,05	6,81	1,36E-04	0,24	0,49

Рис. 1.8. Розрахункові результати прикладу 1.6 у *MS Excel*

У стовпці *Коэффициенты* розташовані числові значення параметрів регресії a_0 і a_1 . Парна лінійна регресія описується рівнянням

$$\hat{y} = 3164,56 + 0,37x,$$

яке вказує, що зростання середньої ціни реалізації молока та молочних продуктів на 1 грн/т стимулює збільшення середньорічного

надою молока від однієї корови на 0,37 кг.

Розраховане значення коефіцієнта детермінації $R^2 = 0,85$ свідчить про сильний зв'язок між ознаками, тобто 85 % загальної варіації середньорічного надою молока обумовлено лінійними змінами середньої ціни реалізації молока та молочних продуктів, а 15 % – впливом не включених до регресійної моделі факторів або нелінійних компонентів моделі.

Значення F -критерію Фішера $F = 46,38$ має рівень значущості $\alpha(F) = 0,00014$, тобто близький до 0, що свідчить про високу довірчу ймовірність отриманої регресійної моделі для генеральної сукупності.

У стовпці *P-значення* знаходяться рівні значущості α , які відповідають обчисленим t -критеріям: для генеральної сукупності коефіцієнт регресії a_0 має довірчу ймовірність $p \approx 1$, а коефіцієнт a_1 – довірчу ймовірність $p = 0,99$.

Стовпці *Нижние 95%* та *Верхние 95%* – відповідно нижні та верхні межі інтервалів реальних коефіцієнтів регресії для довірчої ймовірності 0,95, тобто:

$$2855,40 \leq \hat{a}_0 \leq 3473,73; \quad 0,24 \leq \hat{a}_1 \leq 0,49.$$

Побудована лінійна економетрична модель статистично значуща та адекватна фактичним даним за статистичними критеріями.

Інструментом *Регресия* побудовано таблицю розрахованих за рівнянням регресії значень результативної ознаки \hat{y}_i та їх відхилень $e_i = y_i - \hat{y}_i$, $i = 1..10$, від фактичних статистичних даних (рис. 1.9).

Приклад 1.7. Обсяги виробництва (y) та споживання (x) плодів і ягід в Україні на 1 особу, кг/рік, за 2005–2014 рр. наведено в табл. 1.9. Провести кореляційно-регресійний аналіз лінійної економетричної моделі з використанням стандартних функцій електронних таблиць *OpenOffice Calc*.

	A	B	C
22	ВЫВОД ОСТАТКА		
23			
24	<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>
25	1	3577,23	-90,23
26	2	3556,47	95,53
27	3	3772,67	-107,67
28	4	3920,79	-127,79
29	5	3856,23	192,77
30	6	4240,70	-158,70
31	7	4278,39	-104,39
32	8	4139,45	221,55
33	9	4396,45	49,55
34	10	4478,62	29,38

Рис. 1.9. Обчислені значення результативної ознаки \hat{y}_i та залишків e_i прикладу 1.6

Таблица 1.9

Вхідні дані прикладу 1.7

№ з/п	y	x
1	36	37,1
2	24	34,8
3	32	42,1
4	33	43,5
5	35	45,6
6	38	48,0
7	41	52,6
8	44	53,3
9	50	56,3
10	46	52,3

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C11 аркуша електронних таблиць *OpenOffice Calc*. Коефіцієнт кореляції розраховано в комірці D12 = $CORREL(B2:B11;C2:C11)$. Відповідно до табл. 1.3, одержане значення $R = 0,91$ свідчить про дуже тісний прямий кореляційний зв'язок між ознаками y і x .

Перевірку статистичної значущості коефіцієнта кореляції за t -критерієм Стьюдента відповідно до (1.19) здійснено в комірці $D3 = D12 * SQRT(8) / SQRT(1 - D12^2)$, $t_R = 6,12$. Рівень значущості розраховано в комірці $G3 = TDIST(D3; 8; 2)$, $\alpha(t_R) = 0,00028$. Довірча ймовірність коефіцієнта кореляції для генеральної сукупності складає $p \approx 1$, що свідчить про статистичну значущість коефіцієнта кореляції за t -критерієм Стьюдента.

Числові значення показників моделі для кореляційно-регресійного аналізу обчислено стандартною функцією *LINEST* у діапазоні комірок E1:F5 аналогічно функції *ЛИНЕЙН* (*MS Excel*) за **прикладом 1.4**. Одержана економетрична модель має вигляд:

$$\hat{y} = -6,69 + 0,96x.$$

Параметр a_1 характеризує граничну зміну обсягу виробництва плодів і ягід на 1 особу залежно від їх обсягу споживання на 1 особу: якщо обсяг споживання плодів і ягід на 1 особу зростає на 1 кг, то обсяг виробництва плодів і ягід на 1 особу збільшиться на 0,96 кг. Параметр a_0 для даного прикладу не має економічного сенсу.

Розраховане значення коефіцієнта детермінації $R^2 = 0,82$ свідчить про суттєвий зв'язок між ознаками. Воно вказує, що 82 % загальної варіації обсягу виробництва плодів та ягід на 1 особу обумовлено лінійними змінами обсягу їх споживання, а 18 % – впливом не включених до регресійної моделі факторів або нелінійними залежностями.

Для отриманого значення F -критерію обчислено точний рівень значущості α в комірці $D4 = FDIST(E4; 1; 8)$. Одержаний рівень статистичної значущості регресії $\alpha(F) = 0,00028$ відповідає довірчій імовірності $p \approx 1$ для генеральної сукупності.

За аналогією з **прикладом 1.4** перевірено значущість параметрів регресії a_0 і a_1 . Розрахункові значення t -критерію параметрів регресії

a_0 і a_1 знайдено за формулою (1.17) у комірках E7 = E1/E2, F7 = F1/F2. Для обчислених t -критеріїв визначено точні рівні значущості:

$$E8 = TDIST(ABS(E7);8;2); \quad F8 = TDIST(ABS(F7);8;2),$$

звідки $\alpha(t_{a_1}) = 0,00028$; $\alpha(t_{a_0}) = 0,39$.

Параметр a_1 економетричної моделі є статистично значущим за t -критерієм Стюдента, про що можна судити за високим значенням довірчої ймовірності $p \approx 1$. Параметр a_0 – статистично незначущий за t -критерієм Стюдента згідно з низьким значенням довірчої ймовірності $p = 0,61$.

Для обчислення прогнозних значень пояснюваної змінної із використанням функції *TREND* у комірках C13:C14 введено прогнозні значення пояснювальної ознаки: $\tilde{x}_1 = 54$ та $\tilde{x}_2 = 53,6$. Точкові прогнози $\hat{y}(\tilde{x}_1) = 45,02$ та $\hat{y}(\tilde{x}_2) = 44,64$ для вказаних значень пояснювальної ознаки розраховано в комірках

$$B13:B14 = TREND(B2:B11;C2:C11;C13:C14;1).$$

Приклад 1.8. Засобами стандартних функцій електронних таблиць *Gnumeric* побудувати та провести дослідження економетричної моделі у вигляді лінійної регресії, яка характеризує залежність обсягу реалізація свиней (y), т, від витрачених кормів на їх годівлю (x), тис. ц корм. од., у сільськогосподарських підприємствах областей України за вибіркою 2015 р. (табл. 1.10).

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C25 електронних таблиць *Gnumeric*. Економетрична модель у вигляді парної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x.$$

Для дослідження лінійної економетричної залежності розраховано коефіцієнт кореляції в комірці D27 = *correl*(B2:B25;C2:C25). Відповідно до табл. 1.3, його обчислене значення $R = 0,87$ свідчить про суттєвий прямий кореляційний зв'язок між ознаками y та x .

Вхідні дані прикладу 1.8

№ з/п	Область	у	х
1	Вінницька	9455	811
2	Волинська	12149	494
3	Дніпропетровська	48678	2305
4	Донецька	21011	2548
5	Житомирська	7812	363
6	Закарпатська	3405	189
7	Запорізька	29119	1386
8	Івано-Франківська	44301	1515
9	Київська	54034	2858
10	Кіровоградська	14659	959
11	Луганська	2106	296
12	Львівська	32630	1367
13	Миколаївська	6065	382
14	Одеська	12354	972
15	Полтавська	33334	2159
16	Рівненська	5197	220
17	Сумська	8202	459
18	Тернопільська	24100	1138
19	Харківська	20915	1371
20	Херсонська	13488	627
21	Хмельницька	20277	1085
22	Черкаська	28548	1687
23	Чернівецька	9378	348
24	Чернігівська	10941	680

Перевірку статистичної значущості коефіцієнта кореляції за t -критерієм Стьюдента здійснено за формулою (1.19) у комірці D3 = D27 * sqrt(22) / sqrt(1-D27^2). Для обчисленого t -критерію $t_R = 8,31$ розраховано точний рівень значущості в комірці G3 = tdist(D3;22;2), $\alpha(t_R) = 3,1E - 08$. Довірча ймовірність $p \approx 1$ вказує на статистичну значущість коефіцієнта кореляції за t -критерієм Стьюдента.

Числові значення параметрів моделі знайдено стандартною функцією *linest* за аналогією з **прикладом 1.7**. Результати роботи функції розміщено в діапазоні комірок E1:F5. Одержана економетрична модель має вигляд:

$$\hat{y} = 1712,98 + 16,44x.$$

Параметр a_1 характеризує граничну зміну обсягів реалізації свиней залежно від витрат кормів на їх годівлю: якщо обсяги витрат кормів на годівлю свиней збільшаться на 1 тис. ц корм. од., то обсяги реалізації свиней зростуть на 16,44 т. Параметр a_0 для даного прикладу не має економічного сенсу.

Розраховане значення коефіцієнта детермінації $R^2 = 0,76$ свідчить про суттєвий зв'язок між ознаками. Воно вказує, що 76 % загальної варіації обсягів реалізації свиней обумовлено лінійними змінами обсягів витрат кормів на їх годівлю, а 24 % – впливом не включених до регресійної моделі факторів або нелінійними компонентами моделі.

Для отриманого значення F -критерію обчислено рівень значущості в комірці D4 = *fdist*(E4;1;22). Одержано $\alpha(F) = 3,12E - 08$, що відповідає довірчій ймовірності $p \approx 1$ для генеральної сукупності.

Розрахункові значення t -критерію Стьюдента для параметрів регресії a_0 і a_1 знайдено в комірках E7 = E1/E2, F7 = F1/F2 за формулою (1.17).

Для обчислених t -критеріїв визначено рівні значущості у комірках E8 = *tdist*(*abs*(E7);22;2) та F8 = *tdist*(*abs*(F7);22;2). Одержано $\alpha(t_{a_1}) = 3,12E - 08$ і $\alpha(t_{a_0}) = 0,52$.

Параметр економетричної моделі a_1 є статистично значущим за критерієм Стьюдента, про що можна судити за значенням довірчої ймовірності $p \approx 1$, тоді як параметр a_0 статистично незначущий згідно довірчій ймовірності $p = 0,48$.

Для розрахунку прогнозного значення результативної ознаки застосовано стандартну функцію *forecast*. А саме, у комірку C26 введено прогнозне значення незалежної змінної $\tilde{x} = 454,1$. Точковий прогноз $\hat{y}(\tilde{x}) = 9178,77$ отримано для вказаного значення пояснювальної ознаки в комірці B26 = *forecast* (C26;B2:B25;C2:C25).

Контрольні запитання до розділу 1

1. За допомогою яких функцій та інструментів табличного процесора *MS Excel* можна розрахувати основні статистичні характеристики вибірки?
2. Який загальний вигляд має парна лінійна регресія? Що таке нахил і перетин в економетричній моделі?
3. Опишіть способи знаходження параметрів парної лінійної регресії в *OpenOffice Calc*.
4. Перелічіть вбудовані функції для обчислення параметрів однофакторної лінійної регресії в електронних таблицях *Gnumeric*.
5. Як визначається коефіцієнт кореляції? Які його властивості?
6. Яким чином визначається вибірковий коефіцієнт детермінації для однофакторної лінійної регресії в електронних таблицях *MS Excel*, *OpenOffice Calc* та *Gnumeric*?
7. Для чого і як будуються інтервали довіри для параметрів парної лінійної регресії?
8. Як обчислюються прогнози за однофакторною лінійною регресією в табличних процесорах *OpenOffice Calc*, *Gnumeric* та *MS Excel*?
9. Опишіть перевірку адекватності економетричної моделі у вигляді парної лінійної регресії за *F*-критерієм в електронних таблицях *Gnumeric*, *MS Excel* та *OpenOffice Calc*.
10. Яким чином перевіряється значущість параметрів однофакторної лінійної регресії на основі *t*-критерію Стьюдента засобами табличних процесорів *MS Excel*, *Gnumeric* та *OpenOffice Calc*?

Завдання для контролю засвоєння матеріалу розділу 1

Завдання 1.

Із використанням стандартних функцій електронних таблиць *MS Excel* побудуйте та проведіть дослідження економетричної моделі у вигляді парної лінійної регресії, яка характеризує залежність забезпеченості сільськогосподарських підприємств сівалками (y), шт., від площі збирання зернових та зернобобових культур (x), тис. га, областей України за вибіркою 2015 р. (табл. 1.11). Обчисліть точковий прогноз пояснюваної ознаки за середнім значенням пояснювальної змінної функцією *ТЕНДЕНЦІЯ*.

Таблиця 1.11

Вхідні дані завдання 1

№ з/п	Область	y	x
1	Вінницька	4590	819,3
2	Волинська	792	269,7
3	Дніпропетровська	5594	1187,0
4	Донецька	2817	538,6
5	Житомирська	1202	349,2
6	Закарпатська	101	88,5
7	Запорізька	5323	941,7
8	Івано-Франківська	552	153,0
9	Київська	3388	548,4
10	Кіровоградська	5644	807,9
11	Луганська	2433	396,5
12	Львівська	818	300,3
13	Миколаївська	4374	948,0
14	Одеська	4752	1193,0
15	Полтавська	4231	934,0
16	Рівненська	561	245,0
17	Сумська	1940	667,0
18	Тернопільська	1277	442,7
19	Харківська	4571	1063,0
20	Херсонська	2944	780,5
21	Хмельницька	1861	527,2

Продовження табл. 1.11

22	Черкаська	3076	645,0
23	Чернівецька	344	117,0
24	Чернігівська	1934	678,5

Завдання 2.

Із використанням стандартних функцій табличного процесора *OpenOffice Calc* побудуйте та проведіть дослідження економетричної моделі у вигляді парної лінійної регресії, яка характеризує залежність поголів'я ВРХ (y), тис. голів, від посівних площ кормових культур (x), тис. га, на основі вибірки за 2005–2014 рр. для сільськогосподарських підприємств Дніпропетровської області (табл. 1.12). Обчисліть точковий прогноз пояснюваної ознаки за середнім значенням пояснювальної змінної функцією *TREND*.

Таблиця 1.12

Вхідні дані завдання 2

№ з/п	y	x
1	216,2	102,7
2	194,6	85,5
3	166,2	83,5
4	149,2	69,8
5	139,6	69,0
6	134,2	66,6
7	135,5	68,8
8	146,2	89,2
9	144,7	66,8
10	131,1	61,5

Завдання 3.

У табл. 1.13 розміщено обсяги споживання (y) та виробництва (x) овочів на 1 особу, кг/рік, в Україні за 2003–2014 рр. Проведіть економетричне дослідження залежності між вказаними показниками за допомогою інструменту *Регресія* електронних таблиць *Gnumeric*.

Обчисліть точковий прогноз пояснюваної ознаки за середнім значенням пояснювальної змінної функцією *forecast*.

Таблиця 1.13

Вхідні дані завдання 3

№ з/п	<i>y</i>	<i>x</i>
1	101,7	137
2	115,4	147
3	120,2	155
4	126,7	172
5	118,4	147
6	129,2	172
7	137,1	181
8	143,5	177
9	162,8	215
10	163,4	220
11	163,3	217
12	163,2	224

Завдання 4.

Виконайте економетричне дослідження залежності виробництва м'яса у забійній масі (*y*), тис. т, від середньої ціни його реалізації (*x*), грн/т, для сільськогосподарських підприємств України за вибірковими статистичними спостереженнями 2005–2014 рр., наведеними у табл. 1.14. Проведіть лінійний кореляційно-регресійний аналіз зв'язку зазначених показників із використанням стандартних функцій електронних таблиць *OpenOffice Calc*.

Таблиця 1.14

Вхідні дані завдання 4

№ з/п	<i>y</i>	<i>x</i>
1	588,1	6909,9
2	736,5	6307,7
3	918,8	6466,5
4	979,5	10184,3
5	1034,2	10362,9
6	1134,4	10797,1

Продовження табл. 1.14

7	1215,3	11967,2
8	1271,2	13456,9
9	1441,3	12901,3
10	1451,8	15736,9

Завдання 5.

Дані по Україні за 2005–2014 рр., наведені в табл. 1.15, характеризують середньодобові прирости свиней (y), г, та середні ціни реалізації свиней у живій масі (x), грн/т. Із використанням стандартних функцій табличного процесора *MS Excel* обчисліть значення параметрів лінійної регресії, коефіцієнти кореляції та детермінації. Оцініть статистичну значущість отриманих коефіцієнтів відповідно до вказаних вибіркового статистичних даних.

Таблиця 1.15

Вхідні дані завдання 5

№ з/п	y	x
1	281	10171,0
2	294	8416,2
3	307	7610,0
4	361	13178,0
5	375	14152,9
6	375	12477,5
7	414	13667,2
8	448	16068,4
9	474	15765,0
10	481	18446,9

Завдання 6.

Дані по Дніпропетровській області за 2005–2014 рр., наведені в табл. 1.16, характеризують обсяги виробництва (y) і споживання (x) яєць на 1 особу, шт./рік. Проведіть лінійний кореляційно-регресійний аналіз залежності між вказаними показниками з використанням стандартних функцій електронних таблиць *OpenOffice Calc*.

Вхідні дані завдання 6

№ з/п	y	x
1	255	231
2	284	251
3	258	257
4	267	268
5	284	280
6	298	298
7	328	326
8	330	324
9	289	327
10	319	325

Завдання 7.

Дані, наведені в табл. 1.17 за 2005–2014 рр., характеризують обсяги споживання м'яса та м'ясопродуктів на 1 особу (y), кг/рік, та середню заробітну плату населення України (x), грн/міс. Побудуйте рівняння парної лінійної регресії та дайте економічну інтерпретацію отриманої моделі в електронних таблицях *Gnumeric*.

Вхідні дані завдання 7

№ з/п	y	x
1	39,1	806
2	42,0	1041
3	45,7	1351
4	50,6	1806
5	49,7	1906
6	52,0	2239
7	51,2	2633
8	54,4	3026
9	56,1	3265
10	54,1	3480

Завдання 8.

Показники обсягів виробництва плодів і ягід сільськогосподарськими підприємствами (y), тис. т, та середньої заробітної плати у сільському господарстві (x), грн/міс., в Україні за 2005–2014 рр. наведено в табл. 1.18. Із використанням стандартних функцій електронних таблиць *OpenOffice Calc* побудуйте та проведіть аналіз економетричної залежності між наведеними показниками.

Таблиця 1.18

Вхідні дані завдання 8

№ з/п	y	x
1	200,1	437
2	136,2	581
3	198,0	771
4	231,7	1101
5	214,8	1220
6	286,8	1430
7	299,8	1791
8	369,0	2026
9	444,2	2270
10	332,0	2476

Завдання 9.

Із використанням інструменту *Регресія* табличного процесора *Gnumeric* побудуйте та проведіть дослідження економетричної моделі у вигляді парної лінійної регресії за даними 2005–2014 рр. по Україні (табл. 1.19) про обсяг валової продукції тваринництва сільськогосподарських підприємств у розрахунку на 100 га сільськогосподарських угідь (y), тис. грн, і продуктивність праці в тваринництві на 1 зайнятого в сільськогосподарському виробництві (x), тис. грн.

Таблиця 1.19

Вхідні дані завдання 9

№ з/п	y	x
1	70,3	47,9
2	86,0	40,8

Продовження табл. 1.19

3	97,1	52,6
4	103,3	64,9
5	115,2	113,9
6	126,8	130,5
7	136,2	148,8
8	146,5	171,8
9	157,1	198,2
10	165,9	224,1

Завдання 10.

Дані по Україні за 2006–2015 рр., наведені в табл. 1.20, характеризують обсяги виробництва (y), тис. т, та середні ціни реалізації (x), грн/т, плодів і ягід господарствами населення. Із використанням стандартних функцій електронних таблиць *MS Excel* побудуйте рівняння лінійної регресії та дайте економічну інтерпретацію отриманих результатів.

Таблиця 1.20

Вхідні дані завдання 10

№ з/п	y	x
1	978,1	2094,5
2	1271,6	2324,0
3	1272,4	3847,6
4	1403,3	4271,2
5	1459,7	5043,7
6	1596,5	6097,5
7	1639,7	5985,4
8	1851,1	6412,5
9	1667,1	7569,5
10	1741,1	8031,1

Використана література: [1, 2, 4, 7–11, 14–18, 20, 22].

Розділ 2

МНОЖИННА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ

2.1. Поняття множинної регресії. Багатофакторний кореляційний аналіз

На будь-який економічний показник, як правило, впливає не один, а декілька факторів. У подібних випадках маємо справу з множинною моделлю (регресією), що описує взаємний зв'язок між пояснюваною змінною y та пояснювальними факторами x_1, x_2, \dots, x_m . Математично її можна подати у вигляді функціональної залежності результату y від m пояснювальних змінних:

$$y = f(x_1, x_2, \dots, x_j, \dots, x_m).$$

Інформація про значення y, x_1, x_2, \dots, x_m міститься у відповідних вибіркових статистичних даних – n спостереженнях (вимірюваннях) кожного показника $(y_i, x_{1i}, \dots, x_{mi})$, $i = 1..n$. Таким чином, постає задача виявлення статистичного взаємозв'язку між y та x_1, x_2, \dots, x_m .

На підставі якісного теоретичного аналізу залежностей між економічними показниками можна навести клас розрахункових функцій, які можуть описувати ці взаємозв'язки:

1) лінійна регресія

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m; \quad (2.1)$$

2) степенева регресія

$$\hat{y} = a_0x_1^{a_1}x_2^{a_2}\dots x_m^{a_m}; \quad (2.2)$$

3) обернено пропорційна регресія

$$\hat{y} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_m}{x_m}; \quad (2.3)$$

4) квадратична регресія

$$\hat{y} = a_0 + a_1x_1^2 + a_2x_2^2 + \dots + a_mx_m^2, \quad (2.4)$$

де $a_0, a_1, a_2, \dots, a_m$ – числові параметри рівняння множинної регресії, які підлягають визначенню.

При дослідженні конкретного економічного явища першочерговим завданням є пошук найточнішої аналітичної форми опису статистичного зв'язку між його показниками, причому певна форма залежності повинна мати відповідне економічне обґрунтування. Якщо вигляд залежності встановити важко, то за перше наближення до моделі обирають лінійну форму. Оскільки лінійні функції є найбільш дослідженими, то нелінійні функції намагаються передусім звести до лінійних, і використання економетричних методів розглядають, як правило, на базі лінійних моделей. Так, серед наведених видів функцій (2.2)–(2.4) є нелінійними, але за допомогою перетворення змінних ці функції можна звести до лінійних форм.

Предметом досліджень розділу 2 є багатofакторна лінійна регресійна модель (2.1).

У матричній формі економетрична модель має вигляд:

$$\hat{Y} = AX, \quad (2.5)$$

де \hat{Y} – розрахунковий вектор пояснюваної змінної; X – матриця пояснювальних змінних; A – вектор параметрів моделі. Він обчислюється на основі метода найменших квадратів за оператором:

$$A = (X'X)^{-1} X'Y. \quad (2.6)$$

У наведеному співвідношенні X' – транспонована матриця X ; $(X'X)^{-1}$ – обернена матриця до добутку матриць $X'X$; Y – вектор спостережень пояснюваної змінної.

Статистичні закономірності зміни вихідних параметрів економічного явища під впливом діючих на них факторів вивчають за допомогою кореляційного аналізу. Його основними завданнями є:

- вимірювання тісноти лінійного зв'язку двох чи більше ознак;
- відбір чинників, що найбільш істотно впливають на результативну ознаку на підставі вимірювання ступеня зв'язку між явищами;
- виявлення раніше невідомих причинно-наслідкових зв'язків.

До основних засобів аналізу в даному разі належать:

- парні коефіцієнти кореляції;
- частинні коефіцієнти кореляції;
- множинні коефіцієнти кореляції та детермінації.

Кореляція безпосередньо не виявляє зв'язків між явищами, але надає їх кількісну характеристику. За кожною парою ознак можна обчислити вибіркового парний коефіцієнт кореляції, який є показником тісноти лінійного статистичного зв'язку. Позначимо парний коефіцієнт кореляції між пояснюваною (y) та пояснювальними ознаками (x_j) через r_{yx_j} , а між пояснювальними ознаками x_s та x_j – через $r_{x_s x_j}$, $s = 1..m$, $j = 1..m$:

$$r_{yx_j} = \frac{\frac{\sum_{i=1}^n x_{ji} \cdot y_i}{n} - \bar{x}_j \cdot \bar{y}}{\sigma_{x_j} \cdot \sigma_y}; \quad r_{x_s x_j} = \frac{\frac{\sum_{i=1}^n x_{si} \cdot x_{ji}}{n} - \bar{x}_s \cdot \bar{x}_j}{\sigma_{x_s} \cdot \sigma_{x_j}},$$

де \bar{x}_s , \bar{x}_j , \bar{y} – середні значення:

$$\bar{x}_s = \frac{\sum_{i=1}^n x_{si}}{n}; \quad \bar{x}_j = \frac{\sum_{i=1}^n x_{ji}}{n}; \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n};$$

σ_{x_s} , σ_{x_j} , σ_y – середні квадратичні відхилення:

$$\sigma_{x_s} = \sqrt{\frac{\sum_{i=1}^n x_{si}^2}{n} - (\bar{x}_s)^2}; \quad \sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^n x_{ji}^2}{n} - (\bar{x}_j)^2}; \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^n y_i^2}{n} - (\bar{y})^2}.$$

Усі розраховані парні коефіцієнти кореляції зручно згрупувати в кореляційній матриці, що є симетричною відносно головної діагоналі:

$$Q_{m+1} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{x_1 y} & 1 & r_{x_1 x_2} & \dots & r_{x_1 x_m} \\ r_{x_2 y} & r_{x_2 x_1} & 1 & \dots & r_{x_2 x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_m y} & r_{x_m x_1} & r_{x_m x_2} & \dots & 1 \end{pmatrix}.$$

Як правило, парні коефіцієнти кореляції між пояснювальними ознаками відрізняються від нуля. У цьому випадку варіація ознаки y , яка пояснюється при обчисленні r_{yx_j} варіацією фактора x_j , обумовлена впливом декількох корелюючих між собою ознак. Для виявлення «чистого» впливу x_j на y необхідно одержати таку вибірку, в якій всі факторні ознаки, крім x_j , приймають деякі фіксовані значення (наприклад, середні), та визначити частинні коефіцієнти кореляції. Сукупний зв'язок пояснюваної ознаки з усіма пояснювальними факторами оцінюється за допомогою коефіцієнта детермінації R^2 .

Значення парних коефіцієнтів кореляції між змінними можна отримати за допомогою функцій *KORPEL* (*MS Excel*) або *CORREL* (*OpenOffice Calc*). Кореляційну матрицю розраховують із використанням надбудови *Анализ даних* → *Корреляция* табличного процесора *MS Excel*. При застосуванні програми *Gnumeric* необхідно скористатися надбудовою *Статистика* → *Описательные статистики* → *Корреляция*.

Для перевірки статистичної значущості розрахованих парних коефіцієнтів кореляції необхідно обчислити відповідні їм t -критерії Стьюдента зі ступенем свободи $\nu = n - 2$:

$$t_r = \frac{r \cdot \sqrt{\nu}}{\sqrt{1 - r^2}}. \quad (2.7)$$

Розраховане значення критерію порівнюється з двостороннім критичним значенням $t(\alpha; \nu)$: якщо $|t_r| > t(\alpha; \nu)$, то коефіцієнт кореляції має значущість на рівні α і для генеральної сукупності. Критичне значення t -критерію $t(\alpha; \nu)$ можна знайти за допомогою функції *СТЬЮДРАСПОБР* (*MS Excel*) або функції *TINV* (*OpenOffice Calc*).

Для отриманого значення t -критерію точний рівень значущості α обчислюють за допомогою стандартної функції *СТЬЮДРАСП* (*MS Excel*) або *TDIST* (*OpenOffice Calc*). Він означає, що довірча ймовірність коефіцієнта кореляції становить $p = 1 - \alpha$.

Приклад 2.1. У табл. 2.1 наведено вибірові статистичні дані щодо обсягу виробництва сільськогосподарськими підприємствами цукрових буряків (y), тис. т, їх середніх цін реалізації (x_1), грн, та врожайності (x_2), ц/га, по Вінницькій області за 2000–2015 рр. Із використанням стандартних функцій електронних таблиць *MS Excel* обчислити парні коефіцієнти кореляції між ознаками, побудувати кореляційну матрицю та оцінити статистичну значущість отриманих коефіцієнтів.

Таблиця 2.1

Вхідні дані прикладу 2.1

№ з/п	y	x_1	x_2
1	1916,3	122,1	186
2	1910,8	135,1	178
3	2068,9	120,5	201
4	1727,9	135,8	216
5	2487,7	129,9	254
6	2050,4	195,4	252
7	2861,1	191,7	291
8	2894,4	140,9	296
9	2236,5	193,8	360
10	1404,6	277,4	290
11	2456,5	385,0	327
12	3012,7	403,8	418
13	2844,0	443,0	350
14	2562,3	355,2	400
15	3044,2	445,3	500
16	2057,0	660,3	398

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок A2:C17 аркуша електронних таблиць *MS Excel*. Кореляційну матрицю отримано з використанням інструменту *Корреляція*. Значення парних коефіцієнтів кореляції розраховано функцією *KORPEЛ*. Результати проведених обчислень наведено на рис. 2.1.

F9		=КОРРЕЛ(A2:A17;B2:B17)							
	A	B	C	D	E	F	G	H	
1	y	x1	x2						
2	1916,3	122,1	186		Кореляційна матриця				
3	1910,8	135,1	178			y	x1	x2	
4	2068,9	120,5	201		y	1			
5	1727,9	135,8	216		x1	0,296	1		
6	2487,7	129,9	254		x2	0,613	0,786	1	
7	2050,4	195,4	252						
8	2861,1	191,7	291		Коефіцієнти кореляції				
9	2894,4	140,9	296			0,296	0,613	0,786	
10	2236,5	193,8	360						
11	1404,6	27							
12	2456,5	38							
13	3012,7	40							
14	2844,0	44							
15	2562,3	35							
16	3044,2	44							
17	2057,0	66							
18									
19									
20									
21									
22									

Корреляция

Входные данные

Входной интервал:

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Рис. 2.1. Розрахункові результати та вікно інструменту *Корреляция* прикладу 2.1 в MS Excel

Одержано такі значення парних коефіцієнтів кореляції: $r_{yx_1} = 0,296$ – між обсягом виробництва цукрових буряків та їх середніми цінами реалізації; $r_{yx_2} = 0,613$ – між обсягом виробництва цукрових буряків та їх урожайністю; $r_{x_1x_2} = 0,786$ – між середніми цінами реалізації цукрових буряків та їх урожайністю. Таким чином, помітний кореляційний зв'язок спостерігається між пояснюваною змінною y та пояснювальною змінною x_2 . Зв'язок між y та x_1 можна охарактеризувати як слабкий, між пояснювальними змінними x_1 та x_2 – як високий (табл. 1.3).

З метою оцінки статистичної значущості отриманих коефіцієнтів кореляції розраховано відповідні їм t -критерії Стьюдента за

формулою (2.7). Для обчислених t -критеріїв знайдено рівні значущості за допомогою стандартної функції *СТЬЮДРАСП* (рис. 2.2):

$$t_1 = 1,16; \quad t_2 = 2,90; \quad t_3 = 4,76;$$

$$\alpha(t_1) = 0,265; \quad \alpha(t_2) = 0,012; \quad \alpha(t_3) = 0,0003.$$

F14		fx		=СТЬЮДРАСП(F13;14;2)			
	D	E	F	G	H	I	J
7							
8		Коефіцієнти кореляції					
9			0,296	0,613	0,786		
10							
11		V =		14			
12							
13		t	1,16	2,90	4,76		
14		α	0,265	0,012	0,0003		
15		$p = 1 - \alpha$	0,735	0,988	0,9997		

Рис. 2.2. *Перевірка статистичної значущості коефіцієнтів кореляції прикладу 2.1 у MS Excel*

Розрахованим рівням значущості відповідають такі довірчі ймовірності $p = 1 - \alpha$:

$$p_1 = 0,735; \quad p_2 = 0,988; \quad p_3 \approx 1.$$

Аналіз отриманих результатів свідчить про те, що коефіцієнт кореляції r_{yx_1} має достатньо низьку довірчу ймовірність $p_1 = 0,735$ та може вважатися статистично незначущим для генеральної сукупності. Таким чином, обсяг виробництва цукрових буряків не має суттєвої лінійної залежності від середніх цін їх реалізації. Натомість коефіцієнти кореляції r_{yx_2} та $r_{x_1x_2}$ можуть вважатися статистично значущими з високою довірчою ймовірністю $p_2 = 0,988$ та $p_3 \approx 1$ відповідно. Звідси випливає, що середні ціни реалізації цукрових буряків мають високий кореляційний зв'язок із урожайністю, а вплив урожайності цукрових буряків на їх обсяг виробництва характеризується як помітний.

Приклад 2.2. У табл. 2.2 наведено середні ціни реалізації сільськогосподарськими підприємствами Вінницької області зернових та зернобобових культур (y), грн/т, їх посівні площі (x_1), тис. га, та врожайність (x_2), ц/га, за 2002–2013 рр. Із використанням стандартних функцій табличного процесора *Gnumeric* обчислити парні коефіцієнти кореляції між ознаками, побудувати кореляційну матрицю та оцінити статистичну значущість отриманих коефіцієнтів відповідно до вказаних вибіркового статистичних даних.

Таблиця 2.2

Вхідні дані прикладу 2.2

№ з/п	y	x_1	x_2
1	338,6	869,3	28,8
2	572,2	703,8	19,2
3	499,8	803,0	28,6
4	448,8	776,7	27,4
5	537,3	733,5	28,8
6	861,8	771,0	23,7
7	766,1	838,6	41,0
8	798,2	863,9	37,0
9	1109,2	863,3	36,9
10	1344,5	877,8	49,3
11	1543,7	869,1	43,1
12	1283,3	889,3	55,7

Розв’язання.

Вхідні дані прикладу введено в діапазон комірок A2:C13 аркуша табличного процесора *Gnumeric*. Кореляційну матрицю отримано із використанням інструменту *Корреляція*.

Розраховані за допомогою функції *CORREL* значення парних коефіцієнтів кореляції, обчислені за формулою (2.7) t -критерії Стьюдента та знайдені (за допомогою стандартної функції *TDIST*) відповідні рівні значущості наведено на рис. 2.3. А саме, між середніми цінами реалізації зернових та зернобобових культур та їх

посівними площами: $r_{yx_1} = 0,582$, $t_1 = 2,26$, $\alpha(t_1) = 0,047$; між середніми цінами реалізації зернових та зернобобових культур та їх урожайністю: $r_{yx_2} = 0,769$, $t_2 = 3,81$, $\alpha(t_2) = 0,003$; між посівними площами та урожайністю: $r_{yx_2} = 0,807$, $t_3 = 4,32$, $\alpha(t_3) = 0,002$.

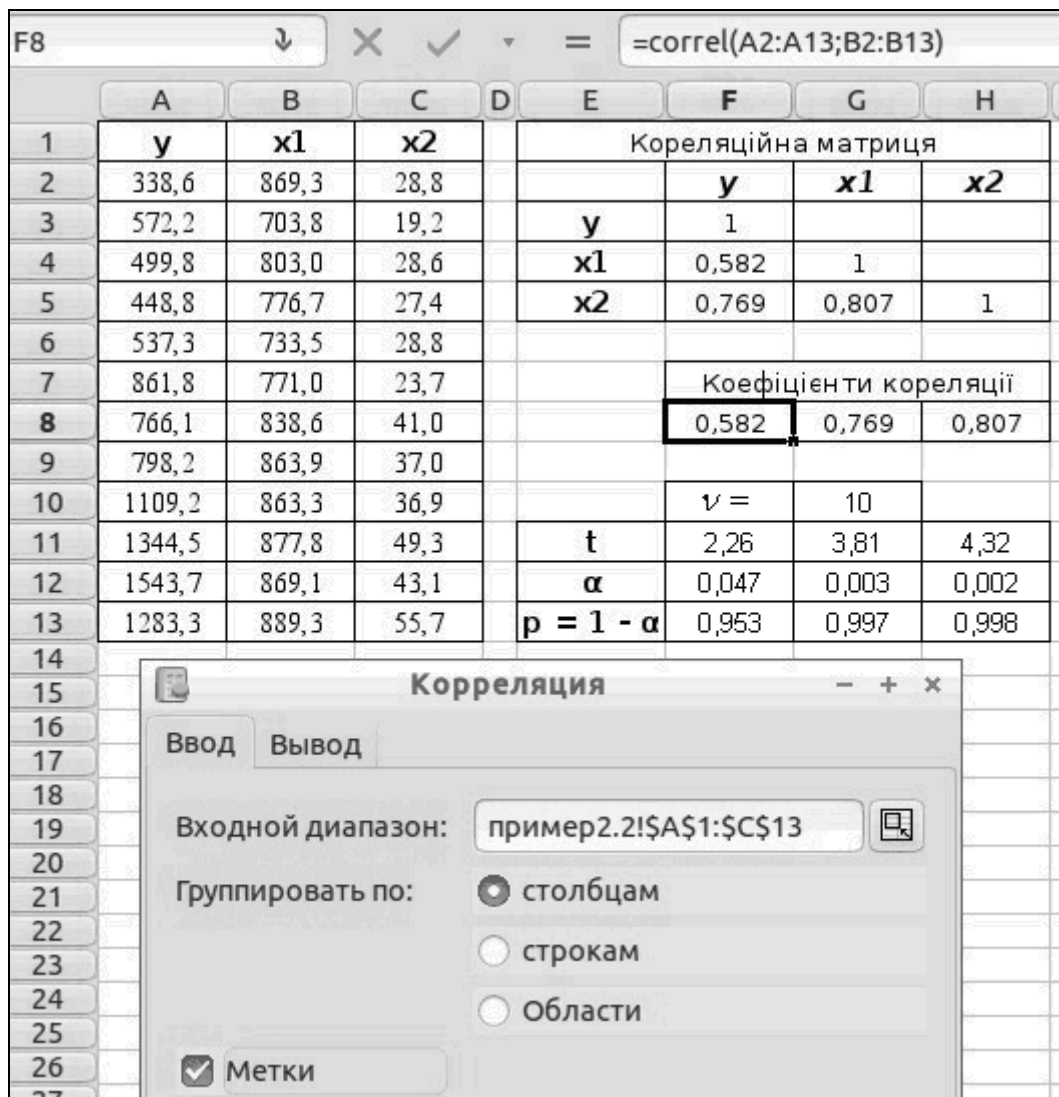


Рис. 2.3. Розрахункові результати та вікно інструменту *Корреляция* прикладу 2.2 в *Gnumeric*

Розрахованим рівням значущості відповідають такі довірчі ймовірності $p = 1 - \alpha$:

$$p_1 = 0,953; \quad p_2 = 0,997; \quad p_3 = 0,998.$$

На підставі аналізу отриманих результатів можна зробити висновок, що коефіцієнт кореляції r_{yx_1} має достатню довірчу ймовірність $p_1 = 0,953$ і вважається статистично значущим для генеральної сукупності, тому середні ціни реалізації зернових та зернобобових культур мають суттєву лінійну залежність від їх посівних площ. Коефіцієнти кореляції r_{yx_2} та $r_{x_1x_2}$ вважаються статистично значущими з високою довірчою ймовірністю $p_2 = 0,997$ та $p_3 = 0,998$ відповідно. Звідси випливає, що спостерігається високий кореляційний зв'язок між середніми цінами реалізації зернових та зернобобових культур і їх врожайністю, а також між посівними площами і врожайністю зернових та зернобобових культур.

2.2. Лінійний регресійний аналіз: теорія та інструменти табличних процесорів

Після встановлення достатньо високих кореляційних зв'язків між факторами та пояснюваною ознакою можна переходити до побудовання багатофакторної лінійної регресійної моделі (2.1). Для її дослідження слід виконати наступні кроки.

1. За даними вибірки обчислити параметри $a_0, a_1, a_2, \dots, a_m$.

Обчислення параметрів рівняння множинної регресії здійснюється за співвідношенням (2.6) чи за допомогою спеціальних надбудов та функцій електронних таблиць *MS Excel* або *OpenOffice Calc*, а також вільного табличного процесора *Gnumeric*. При використанні *MS Excel* застосовується інструмент *Регресия* або функція *ЛИНЕЙН*. В *OpenOffice Calc* для цього призначена функція *LINEST*. У випадку табличного процесора *Gnumeric* застосовують інструмент *Регресия* чи функцію *LINEST*.

Параметри $a_0, a_1, a_2, \dots, a_m$ визначають вплив зміни факторів $x_j, j = 1..m$, на величину y . Якщо модель лінійна, то значення

параметра a_j показує, на скільки одиниць змінюється y при збільшенні x_j на одиницю за умови, що інші фактори є сталими.

2. Для перевірки статистичної значущості моделі обчислити:

– залишки моделі, тобто розбіжності між вибірковими та розрахунковими значеннями залежної змінної:

$$e_i = y_i - \hat{y}_i, \quad i = 1..n; \quad (2.8)$$

– залишкову дисперсію, щоб перевірити загальний вплив пояснювальних змінних на результативну ознаку:

$$\sigma_e^2 = \frac{\sum_{i=1}^n e_i^2}{n - m - 1}; \quad (2.9)$$

– коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}; \quad (2.10)$$

– вибірковий коефіцієнт множинної кореляції:

$$R = \sqrt{R^2}.$$

3. Перевірити статистичну значущість отриманих результатів:

– перевірка статистичної значущості моделі в цілому зводиться до перевірки статистичної значущості коефіцієнта детермінації за допомогою F -критерію Фішера:

$$F = \frac{R^2 / \nu_1}{(1 - R^2) / \nu_2}. \quad (2.11)$$

Обчислити значення F -критерію можна за допомогою функції *ЛИНЕЙН* (*MS Excel*) чи аналогічної функції *LINEST* (*OpenOffice Calc*). Розраховане значення критерію знаходиться також у вихідних даних інструменту *Регрессія* з *MS Excel* або *Gnumeric*. Величина F (2.11) порівнюється з критичним значенням F -критерію $F(\alpha; \nu_1; \nu_2)$ при ступенях свободи $\nu_1 = m$ та $\nu_2 = n - m - 1$ і рівні значущості α . Якщо $F > F(\alpha; \nu_1; \nu_2)$, то регресійна модель статистично значуща.

Критичні значення $F(\alpha; \nu_1; \nu_2)$ визначаються за допомогою функції *FPACPOBP* (*MS Excel*) або *FINV* (*OpenOffice Calc*).

Для розрахованого значення F -критерію точний рівень значущості α обчислюють за допомогою стандартної функції *FPACPI* (*MS Excel*) або *FDIST* (*OpenOffice Calc*). Він означає, що довірна ймовірність регресійної моделі для генеральної сукупності складає $p = 1 - \alpha$;

– перевірити значущість коефіцієнта множинної кореляції R за двостороннім t -критерієм Стюдента зі ступенем свободи $\nu = n - 2$ та обчислити відповідний рівень значущості за аналогією з методикою, розглянутою в підрозділі 2.2 для парних коефіцієнтів кореляції:

$$t_R = \frac{R \cdot \sqrt{\nu}}{\sqrt{1 - R^2}}; \quad (2.12)$$

– перевірити значущість кожного параметра регресії за допомогою двостороннього t -критерію Стюдента та обчислити відповідний точний рівень значущості за аналогією з методикою, розглянутою в підрозділі 2.1 для парних коефіцієнтів кореляції:

$$t_j = \frac{a_j}{\sigma_{a_j}}; \quad \sigma_{a_j} = \sigma_e \cdot \sqrt{\{(X'X)^{-1}\}_{jj}}, \quad j = 1..m. \quad (2.13)$$

Стандартні похибки σ_{a_j} розраховують із використанням функції *ЛИНЕЙН* (*MS Excel*) чи *LINEST* (*OpenOffice Calc*). Значення критеріїв t_j знаходяться у вихідних даних інструменту *Регрессия* з *MS Excel* або *Gnumeric*.

4. Обчислити та інтерпретувати коефіцієнти еластичності:

$$E_j = \frac{\partial \hat{y}}{\partial x_j} \cdot \frac{\bar{x}_j}{\bar{y}}; \quad E_{x_j} = \frac{\partial \hat{y}}{\partial x_j} \cdot \frac{\tilde{x}_j}{\hat{y}(\tilde{x}_j)}. \quad (2.14)$$

Середні коефіцієнти еластичності E_j обчислюються для середнього значення \bar{x}_j та показують, на скільки відсотків зміниться у відносно свого середнього значення \bar{y} зі зростанням x_j на 1 % відносно середнього значення \bar{x}_j , $j = 1..m$. Величини E_j можна порівнювати

між собою та, відповідно, ранжувати пояснювальні фактори за силою їх впливу на результативну ознаку. Точковий коефіцієнт еластичності E_{x_j} обчислюється для конкретного значення $x_j = \tilde{x}_j$ у діапазоні між $\min_{i=1..n} x_{ji}$ та $\max_{i=1..n} x_{ji}$ і показує, на скільки відсотків зміниться у відносно значення $\hat{y}(\bar{x}_1, \dots, \bar{x}_{j-1}, \tilde{x}_j, \bar{x}_{j+1}, \dots, \bar{x}_m)$ зі зростанням x_j на 1 % від заданої величини \tilde{x}_j .

Загальна середня еластичність E від усіх факторів x_j дорівнює:

$$E = \sum_{j=1}^m E_j.$$

Загальна середня еластичність показує, на скільки відсотків зміниться y , якщо одночасно збільшити на 1 % усі пояснювальні фактори відносно їх середніх значень.

5. Визначити довірчі інтервали реальних параметрів регресії \hat{a}_j при рівні значущості α або довірчій імовірності $p = 1 - \alpha$:

$$a_j - \sigma_{a_j} \cdot t(\alpha; \nu) \leq \hat{a}_j \leq a_j + \sigma_{a_j} \cdot t(\alpha; \nu). \quad (2.15)$$

Фактичні значення довірчих інтервалів параметрів регресії знаходяться у вихідних даних інструменту *Регресия* з *MS Excel* або *Gnumeric*.

6. Обчислити точковий та інтервальний прогноз за значеннями $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m$.

Точковий прогноз $\hat{y}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m)$ одержують при підстановці прогнозних значень пояснювальних змінних $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m$ до рівняння регресії (2.1) або за допомогою функцій *ТЕНДЕНЦИЯ* (*MS Excel*) чи *TREND* (*OpenOffice Calc*).

Інтервальний прогноз вказує нижню та верхню межу проміжку, в якому знаходиться істинне значення прогнозованої результативної ознаки $y(\tilde{x}_1, \dots, \tilde{x}_m)$ при рівні значущості α :

$$\hat{y}(\tilde{x}_1, \dots, \tilde{x}_m) - \sigma_e \cdot t(\alpha; \nu) \leq y(\tilde{x}_1, \dots, \tilde{x}_m) \leq \hat{y}(\tilde{x}_1, \dots, \tilde{x}_m) + \sigma_e \cdot t(\alpha; \nu). \quad (2.16)$$

Приклад 2.3. Із використанням стандартних функцій електронних таблиць *MS Excel* побудувати та провести дослідження економетричної моделі у вигляді двофакторної лінійної регресії, яка характеризує залежність виробництва молока у Дніпропетровській області (y), тис. т, від витрат кормів на виробництво 1 ц молока (x_1), кг корм. од., та поголів'я корів на 1 січня (x_2), тис. голів, за вибіркою 2000–2015 рр. (табл. 2.3).

Таблиця 2.3

Вхідні дані прикладу 2.3

№ з/п	y	x_1	x_2
1	522,4	141	249,3
2	527,6	140	192,6
3	558,3	135	179,3
4	508,0	126	175,1
5	492,6	138	147,7
6	494,0	123	131,1
7	474,3	118	118,1
8	415,7	110	108,2
9	381,4	105	97,2
10	359,2	109	88,3
11	339,8	105	83,4
12	341,7	107	80,5
13	343,4	107	79,8
14	348,0	117	80,2
15	357,2	93	80,2
16	344,6	92	75,9

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок A3:C18 аркуша електронних таблиць *MS Excel*. Економетрична модель у вигляді двофакторної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2.$$

Параметри моделі та коефіцієнти для її подальшого дослідження обчислено за допомогою функції *ЛИНЕЙН* (рис. 2.4). Одержана економетрична модель має вигляд:

$$\hat{y} = 78,90 + 2,07x_1 + 0,85x_2. \quad (2.17)$$

F4				fx {=ЛИНЕЙН(A3:A18;B3:C18;1;1)}				
	A	B	C	D	E	F	G	H
1	Вхідні дані							
2	y	x₁	x₂		Параметри моделі			
3	522,4	141	249,3			a₂	a₁	a₀
4	527,6	140	192,6			0,85	2,07	78,90
5	558,3	135	179,3		σ_a	0,37	1,21	103,43
6	508,0	126	175,1		R² =	0,84	35,32	#Н/Д
7	492,6	138	147,7		F =	33,82	13	#Н/Д
8	494,0	123	131,1		α(F) =	7,1E-06		
9	474,3	118	118,1					
10	415,7	110	108,2		t	2,32	1,72	0,76
11	381,4	105	97,2		α(t)	0,04	0,11	0,46
12	359,2	109	88,3		t(0,05;13) =	2,16		
13	339,8	105	83,4					
14	341,7	107	80,5		Довірчі інтервали	0,06	-0,53	-144,54
15	343,4	107	79,8			1,65	4,68	302,34
16	348,0	117	80,2					
17	357,2	93	80,2		Прогноз			
18	344,6	92	75,9			ŷ	ẋ₁	ẋ₂
19	Середні значення					392,5	110	100,3
20	425,5	116,6	122,9		Довірчі інтервали	316,2		
21	Коефіцієнти еластичності					468,8		
22	E	0,57	0,25					

Рис. 2.4. *Розрахункові результати прикладу 2.3 у MS Excel*

Дамо змістовне тлумачення отриманих результатів. Параметр a_1 характеризує граничну зміну виробництва молока залежно від зміни витрат кормів: якщо витрати кормів на виробництво 1 ц молока зростуть на 1 кг корм. од., то виробництво молока збільшиться на 2,07 тис. т за сталого поголів'я корів. Параметр a_2 характеризує граничну зміну виробництва молока залежно від зміни поголів'я корів. Так, якщо поголів'я корів збільшиться на 1 тис. голів, то

виробництво молока зростає на 0,85 тис. т при сталій величині витрат кормів. Параметр a_0 для даного прикладу не має економічного сенсу.

Розраховане значення коефіцієнта детермінації $R^2 = 0,84$ свідчить про сильний зв'язок між ознаками. Воно вказує, що 84 % загальної варіації виробництва молока обумовлено впливом витрат кормів на виробництво 1 ц молока та поголів'я корів, а 16 % – впливом не включених до регресійної моделі факторів або нелінійними компонентами моделі.

Для отриманого значення F -критерію обчислено рівень значущості α в комірці $F8 = FPACП(F7;2;13)$. Одержано $\alpha(F) = 7,1E - 06$, що відповідає довірчій імовірності $p \approx 1$ для генеральної сукупності.

З метою оцінки статистичної значущості отриманих параметрів регресії у комірках F10:H10 розраховано відповідні їм t -критерії Стьюдента за формулою (2.13). Для обчислених t -критеріїв у комірках F11:H11 знайдено точні рівні значущості за допомогою стандартної функції *СТЬЮДРАСП*, а саме:

$$t_0 = 0,76; \quad t_1 = 1,72; \quad t_2 = 2,32;$$
$$\alpha(t_0) = 0,46; \quad \alpha(t_1) = 0,11; \quad \alpha(t_2) = 0,04.$$

Розрахованим рівням значущості відповідають наступні довірчі ймовірності $p = 1 - \alpha$:

$$p_0 = 0,54; \quad p_1 = 0,89; \quad p_2 = 0,96.$$

Аналіз отриманих результатів свідчить, що параметри регресії a_1, a_2 мають достатню довірчу ймовірність та можуть вважатися статистично значущими для генеральної сукупності; параметр a_0 – статистично незначущий. У підсумку можна констатувати, що виробництво молока має суттєву лінійну залежність від витрат кормів на виробництво 1 ц молока та поголів'я корів.

Значення стандартних похибок σ_{a_j} , необхідні для обчислення довірчих інтервалів параметрів регресії, знаходяться у вихідних даних функції *ЛИНЕЙН* (комірки F5:H5).

Двостороннє критичне значення t -критерію обчислено у комірці F12 = *СТБЮДРАСПОБР*(0,05;13). Довірчі інтервали реальних параметрів регресії розраховано за формулою (2.15):

$$-144,54 \leq \hat{a}_0 \leq 302,34; \quad -0,53 \leq \hat{a}_1 \leq 4,68; \quad 0,06 \leq \hat{a}_2 \leq 1,65.$$

Точковий прогноз $\hat{y}(\tilde{x}_1, \tilde{x}_2) = 392,5$ отримано при підстановці прогнозних значень пояснювальних змінних $\tilde{x}_1 = 110$ та $\tilde{x}_2 = 100,3$ до рівняння регресії (2.17).

Інтервальний прогноз знайдено за формулою (2.16). Необхідне для його обчислення значення стандартної похибки $\sigma_e = 35,32$ знаходиться у вихідних даних функції *ЛИНЕЙН* (комірка G6). Інтервальний прогноз вказує нижню і верхню межу проміжку, в якому знаходиться істинне значення прогнозованої результативної ознаки при рівні значущості $\alpha = 0,05$:

$$316,2 \leq y(\tilde{x}_1, \tilde{x}_2) \leq 468,8.$$

Для економічної інтерпретації зв'язків між змінними в комірках B22 і C22 обчислено середні коефіцієнти еластичності за формулою (2.14):

$$\text{– за витратами кормів} \quad E_1 = \frac{\bar{x}_1}{\bar{y}} \cdot a_1 = 0,57;$$

$$\text{– за поголів'ям корів} \quad E_2 = \frac{\bar{x}_2}{\bar{y}} \cdot a_2 = 0,25.$$

Економічна інтерпретація отриманих значень середніх коефіцієнтів еластичності така. Зі збільшенням витрат кормів на 1 % відносно свого середнього рівня середнє виробництво молока збільшиться на 0,57 % за сталого поголів'я корів. Якщо поголів'я корів збільшиться на 1 % відносно свого середнього рівня, то виробництво молока збільшиться на 0,25 % за сталої величини витрат кормів.

Приклад 2.4. Із використанням стандартних функцій електронних таблиць *OpenOffice Calc* побудувати та провести дослідження економетричної моделі у вигляді двофакторної лінійної регресії, яка

характеризує залежність обсягу виробництва м'яса в Україні на 1 умовну голову великої рогатої худоби (y), кг, від середньодобових приростів (x_1), г, та витрат кормів на приріст ВРХ (x_2), ц корм. од./ц м'яса, за вибіркою 2000–2014 рр. (табл. 2.4).

Таблиця 2.4

Вхідні дані прикладу 2.4

№ з/п	y	x_1	x_2
1	110	255	16,73
2	113	318	15,27
3	119	322	15,72
4	122	298	15,50
5	130	356	14,33
6	138	392	15,89
7	149	401	15,53
8	165	416	14,80
9	176	449	14,49
10	185	469	15,23
11	185	456	16,12
12	191	477	15,05
13	203	500	14,92
14	211	504	15,07
15	215	525	15,05

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок A3:C17 аркуша електронних таблиць *OpenOffice Calc*. Економетрична модель у вигляді двофакторної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2.$$

Параметри моделі та коефіцієнти для її подальшого дослідження обчислено функцією *LINEST*. Результати роботи функції *LINEST* розміщено в діапазоні комірок F3:H6 (рис. 2.5).

Одержана економетрична модель має вигляд:

$$\hat{y} = -93,32 + 0,45x_1 + 4,59x_2. \quad (2.18)$$

F3				fx Σ = {=LINEST(A3:A17;B3:C17;1;1)}				
	A	B	C	D	E	F	G	H
1	Вхідні дані					Параметри моделі		
2	y	x ₁	x ₂			a ₂	a ₁	a ₀
3	110	255	16,73			4,59	0,45	-93,32
4	113	318	15,27		σ _a	4,01	0,03	68,04
5	119	322	15,72		R ² =	0,96	8,34	#Н/Д
6	122	298	15,50		F =	132,81	12	#Н/Д
7	130	356	14,33		α(F) =	6,5E-09		
8	138	392	15,89					
9	149	401	15,53		t	1,14	14,93	1,37
10	165	416	14,80		α(t)	0,27	4,1E-09	0,20
11	176	449	14,49		t(0,05;12) =	2,18		
12	185	469	15,23					
13	185	456	16,12		Довірчі інтервали	-4,15	0,38	-241,57
14	191	477	15,05			13,33	0,51	54,92
15	203	500	14,92					
16	211	504	15,07			Прогноз		
17	215	525	15,05			ŷ	ẋ ₁	ẋ ₂
18	Середні значення					137	350	16
19	161	409	15,31		Довірчі інтервали	119		
20	Коефіцієнти еластичності					156		
21	E	1,14	0,44					

Рис. 2.5. Розрахункові результати прикладу 2.4 в OpenOffice Calc

Дамо змістовне тлумачення отриманих результатів. Параметр a_1 характеризує граничну зміну обсягу виробництва м'яса на 1 умовну голову ВРХ залежно від середньодобових приростів ВРХ: якщо прирости зростуть на 1 г, то обсяг виробництва м'яса збільшиться на 0,45 кг у разі сталих витрат кормів.

Параметр a_2 характеризує граничну зміну обсягу виробництва м'яса на 1 умовну голову ВРХ залежно від зміни витрат кормів на приріст ВРХ. Так, якщо витрати кормів збільшаться на 1 ц корм. од., то обсяг виробництва м'яса на 1 умовну голову ВРХ збільшиться на 4,59 кг за сталої величини середньодобових приростів ВРХ. Параметр a_0 для даного прикладу не має економічного сенсу.

Розраховане значення коефіцієнта детермінації $R^2 = 0,96$ свідчить про сильний зв'язок між ознаками. Воно вказує, що 96 % загальної варіації обсягу виробництва м'яса на 1 умовну голову ВРХ обумовлено впливом середньодобових приростів та витрат кормів, а 4 % – впливом не включених до регресійної моделі факторів або нелінійними компонентами моделі.

Для отриманого значення F -критерію обчислено рівень статистичної значущості регресії α у комірці $F7 = FDIST(F6;2;12)$. Одержано $\alpha(F) = 6,5E - 09$, що відповідає довірчій ймовірності $p \approx 1$ для генеральної сукупності.

З метою оцінки статистичної значущості параметрів регресії розраховано відповідні їм t -критерії Стюдента, знайдено рівні значущості за допомогою стандартної функції $TDIST$ та визначено довірчі ймовірності за аналогією з **прикладом 2.3** (рис. 2.5).

Аналіз результатів свідчить, що параметр регресії a_1 має високу довірчу ймовірність $p_1 \approx 1$ та може вважатися статистично значущим для генеральної сукупності. Таким чином, обсяг виробництва м'яса на 1 умовну голову ВРХ має суттєву лінійну залежність від середньодобових приростів ВРХ. Натомість параметр регресії a_2 має низьку довірчу ймовірність $p_2 = 0,73$ та може вважатися статистично незначущим для генеральної сукупності, тобто обсяг виробництва м'яса не має суттєвої лінійної залежності від витрат кормів на приріст ВРХ.

Значення стандартних похибок σ_{a_j} , необхідні для обчислення довірчих інтервалів параметрів регресії, знаходяться у вихідних даних функції $LINEST$ (комірки F4:H4). Двостороннє критичне значення t -критерію обчислено у комірці $F11 = TINV(0,05;12)$. Довірчі інтервали реальних параметрів регресії розраховано за формулою (2.15):

$$- 241,57 \leq \hat{a}_0 \leq 54,92; \quad 0,38 \leq \hat{a}_1 \leq 0,51; \quad - 4,15 \leq \hat{a}_2 \leq 13,33.$$

Точковий прогноз $\hat{y}(\tilde{x}_1, \tilde{x}_2) = 137$ отримано при підстановці прогнозних значень пояснювальних змінних $\tilde{x}_1 = 350$ та $\tilde{x}_2 = 16$ до

рівняння регресії (2.18). Інтервальний прогноз знайдено за формулою (2.16). Необхідне для його обчислення значення стандартної похибки $\sigma_e = 8,34$ знаходиться у вихідних даних функції *LINEST* (комірка G5). Інтервальний прогноз вказує нижню і верхню межу проміжку, в якому знаходиться істинне значення прогнозованої результативної ознаки при рівні значущості $\alpha = 0,05$:

$$119 \leq y(\tilde{x}_1, \tilde{x}_2) \leq 156.$$

Для економічної інтерпретації зв'язків між змінними в комірках B21 і C21 обчислено середні коефіцієнти еластичності за середньодобовими приростами та витратам кормів за аналогією з **прикладом 2.3**. Економічна інтерпретація отриманих значень середніх коефіцієнтів еластичності така. Зі збільшенням середньодобових приростів ВРХ на 1 % відносно свого середнього рівня середній обсяг виробництва м'яса збільшиться на 1,14 % за сталих витрат кормів. Якщо витрати кормів збільшаться на 1 % відносно свого середнього рівня, то середній обсяг виробництва м'яса збільшиться на 0,44 % за сталої величини середньодобових приростів ВРХ.

Приклад 2.5. Із використанням функції *ТЕНДЕНЦІЯ* електронних таблиць *MS Excel* розрахувати прогнозні значення пояснюваної змінної $\hat{y}_i, i = 1..15$, точковий прогноз та оцінити статистичну значущість економетричної моделі у вигляді двофакторної лінійної регресії, яка характеризує залежність обсягу виробництва свинини в Україні на 1 голову свиней (y), кг, від їх середньодобових приростів (x_1), г, та витрат кормів на приріст свиней (x_2), ц корм. од./ц свинини, за вибіркою 2000–2014 рр. (табл. 2.5).

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:D16 аркуша електронних таблиць *MS Excel*. Економетрична модель у вигляді двофакторної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2.$$

Вхідні дані прикладу 2.5

№ з/п	y	x_1	x_2
1	67	120	17,90
2	77	173	13,82
3	72	198	11,64
4	69	189	11,32
5	76	246	9,70
6	76	281	8,97
7	75	294	8,12
8	79	307	7,62
9	84	361	6,50
10	81	375	6,59
11	81	373	5,99
12	85	413	5,43
13	92	486	5,03
14	96	472	4,63
15	96	481	4,57

На рис. 2.6 показано розрахункові результати проведених обчислень із використанням функції *ТЕНДЕНЦІЯ*. Прогнозні значення пояснюваної змінної $\hat{y}_i, i = 1..15$, одержано в діапазоні комірок E2:E16 = *ТЕНДЕНЦІЯ*(B2:B16;C2:D16;;1).

Точковий прогноз $\hat{y}(\tilde{x}_1, \tilde{x}_2) = 83,6$ розраховано для прогнозних значень пояснювальних змінних $\tilde{x}_1 = 370$ та $\tilde{x}_2 = 6,5$ у комірці E17 = *ТЕНДЕНЦІЯ*(B2:B16;C2:D16;C17:D17;1). Розбіжності між вибірковими та розрахунковими значеннями залежної змінної $e_i, i = 1..15$, обчислено за формулою (2.8).

Для перевірки загального впливу пояснювальних змінних на результативну ознаку знайдено коефіцієнт детермінації R^2 за співвідношенням $R^2 = r_{y\hat{y}}^2$, де $r_{y\hat{y}}$ – парний коефіцієнт кореляції між фактичними та розрахованими регресією значеннями результативної ознаки,

обчислений у комірці K4 = КОРРЕЛ(B2:B16;E2:E16). Одержано $R^2 = 0,91$, що свідчить про сильний лінійний зв'язок між ознаками.

E17		fx {=ТЕНДЕНЦИЯ(B2:B16;C2:D16;C17:D17;1)}											
	A	B	C	D	E	F	G	H	I	J	K		
1		y	x ₁	x ₂	\hat{y}	e							
2	1	67	120	17,90	70,2	-3,2							
3	2	77	173	13,82	71,1	5,9							
4	3	72	198	11,64	71,3	0,7				$r_{xy} =$	0,95		
5	4	69	189	11,32	69,9	-0,9				$R^2 =$	0,91		
6	5	76	246	9,70	74,1	1,9				F =	57,59		
7	6	76	281	8,97	77,0	-1,0				$\alpha(F) =$	7,1E-07		
8	7	75	294	8,12	77,4	-2,4				F(0,05;2;12) =	3,89		
9	8	79	307	7,62	78,2	0,8							
10	9	84	361	6,50	82,6	1,4				a ₂	a ₁	a ₀	
11	10	81	375	6,59	84,2	-3,2				1,15	0,11	36,89	
12	11	81	373	5,99	83,3	-2,3				σ_a	0,61	0,02	11,36
13	12	85	413	5,43	86,9	-1,9				$R^2 =$	0,91	2,96	#Н/Д
14	13	92	486	5,03	94,2	-2,2				F =	57,59	12	#Н/Д
15	14	96	472	4,63	92,3	3,7							
16	15	96	481	4,57	93,1	2,9							
17	Прогноз		370	6,5	83,6								

Рис. 2.6. Розрахункові результати прикладу 2.5 у MS Excel

Перевірку статистичної значущості одержаних результатів зроблено за допомогою порівняння F -критерію Фішера, обчисленого за співвідношенням (2.11) у комірці K6, та критичного значення $F(0,05;2;12)$, розрахованого за допомогою функції *FPАСПОБР* у комірці K8. А саме, $F > F(0,05;2;12)$, тому регресійна модель статистично значуща при $\alpha = 0,05$. Точний рівень значущості α для F -критерію обчислено у комірці K7 = *FPАСП*(K6;2;12). Його значення вказує на високу довірчу ймовірність $p = 1 - \alpha$ регресійної моделі для генеральної сукупності.

Контроль правильності розрахунків та обчислення параметрів моделі здійснено за допомогою функції *ЛИНЕЙН*, підсумки роботи якої розміщено в діапазоні комірок I11:K14. Порівняння результатів

обчислень за функціями *ТЕНДЕНЦИЯ* та *ЛИНЕЙН* вказує на їх повну тотожність.

Лінійна економетрична модель залежності обсягу виробництва свинини на 1 голову від середньодобових приростів свиней та витрат кормів на виробництво 1 ц свинини в Україні описується регресійним рівнянням:

$$\hat{y} = 36,89 + 0,11x_1 + 1,15x_2.$$

Для обчислення параметрів рівняння регресії за допомогою метода найменших квадратів, аналізу впливу на пояснювану змінну однієї або декількох пояснювальних змінних, побудови довірчих інтервалів, перевірки значущості рівняння регресії тощо використовується інструмент аналізу *Регресия*. Для виклику інструменту *Регресия* при використанні *MS Excel* необхідно виконати команди *Сервис*→*Анализ данных*→*Регресия*. У випадку табличного процесора *Gnumeric* застосовують команди *Статистика*→*Зависимые наблюдения*→*Регресия*.

Розглянемо побудову та аналіз множинної економетричної моделі з використанням інструменту *Регресия* на прикладах.

Приклад 2.6. Із використанням інструменту *Регресия* електронних таблиць *MS Excel* побудувати та провести дослідження економетричної моделі у вигляді двофакторної лінійної регресії, яка характеризує залежність обсягу виробництва яєць у Київській області (y), млн. шт., від середньої річної несучості однієї курки-несучки (x_1), шт., та середньої заробітної плати у сільському господарстві (x_2), грн/міс., за вибіркою 2000–2014 рр. (табл. 2.6).

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок A3:C17 аркуша електронних таблиць *MS Excel*. Економетрична модель у вигляді двофакторної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2.$$

Вхідні дані прикладу 2.6

№ з/п	y	x_1	x_2
1	787,2	206	145
2	795,8	184	198
3	1165,1	247	235
4	1187,0	266	283
5	1377,7	277	389
6	1462,6	277	553
7	1704,1	275	737
8	1784,9	282	997
9	1832,8	274	1434
10	2048,3	273	1593
11	2007,1	268	1820
12	2332,7	272	2165
13	2479,0	276	2433
14	2545,3	276	2713
15	2712,0	258	2994

На рис. 2.7 показано вхідні дані прикладу та діалогове вікно інструменту *Регрессія*, в якому задаються наступні параметри.

1. *Входной интервал Y* – діапазон адрес комірок зі значеннями y_i , $i = 1..15$, та відповідним заголовком.

2. *Входной интервал X* – діапазон адрес комірок зі значеннями x_{1i} , x_{2i} , $i = 1..15$, та відповідним заголовком.

3. *Метки* – опцію необхідно включити, якщо перший рядок вхідних даних містить заголовок.

4. *Уровень надежности* – при увімкненні цієї опції задається ймовірність p при побудові довірчих інтервалів.

5. *Константа – ноль* – опцію необхідно включити, якщо параметр регресії $a_0 = 0$.

6. *Выходной интервал* – при увімкненні активізується поле, до якого необхідно ввести адресу лівої верхньої комірки вихідного діапазону з результатами обчислень інструменту *Регрессія*.

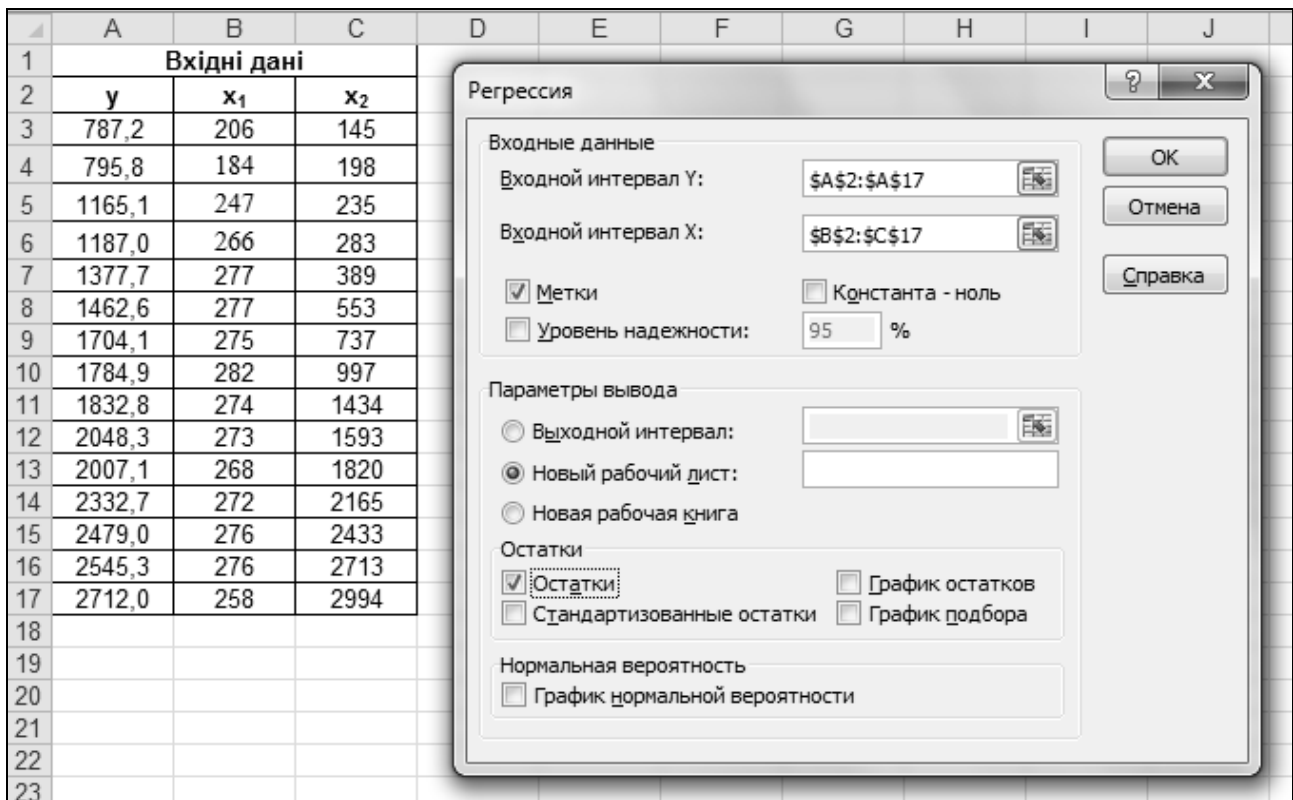


Рис. 2.7. Вхідні дані прикладу 2.6 та вікно інструменту *Регрессия* в MS Excel

7. *Новый рабочий лист* – при увімкненні цієї опції відкривається новий аркуш, до якого, починаючи з комірки A1, вносяться результати роботи інструменту *Регрессия*.

8. *Новая рабочая книга* – при увімкненні відкривається нова книга, на першому аркуші якої, починаючи з комірки A1, розташовуються результати роботи інструменту *Регрессия*.

9. *Остатки* – при увімкненні цієї опції обчислюється стовпець із залишками моделі $e_i = y_i - \hat{y}_i$, $i = 1..15$.

10. *Стандартизованные остатки* – при увімкненні обчислюється стовпець зі стандартизованими залишками.

11. *График остатков* – при увімкненні виводяться точкові графіки залишків $e_i = y_i - \hat{y}_i$, $i = 1..15$, залежно від значень пояснювальних змінних x_j , $j = 1..2$. Кількість графіків дорівнює кількості змінних x_j .

12. *График подбора* – при увімкненні виводяться точкові графіки розрахованих побудованою регресією значень \hat{y}_i залежно від значень пояснювальних змінних $x_j, j=1..2$. Кількість графіків дорівнює кількості змінних x_j .

На рис. 2.8 наведено розрахункові результати, одержані за допомогою інструменту *Регрессия*.

E21		fx		=B15+B16*F21+B17*G21			
	A	B	C	D	E	F	G
1	ВЫВОД ИТОГОВ						
2	Регрессионная статистика						
3	Множественный R	0,99					
4	R-квадрат	0,99					
5	Нормированный R-кв	0,99					
6	Стандартная ошибка	72,6					
7	Наблюдения	15		t(0,05;12) =	2,18		
8	Дисперсионный анализ						
9		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значим. F</i>	
10	Регрессия	2	5248676	2624338	498,1	2,8E-12	
11	Остаток	12	63225	5269			
12	Итого	14	5311901				
13							
14		<i>Кoeffициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
15	Y-пересечение	-413,3	188,3	-2,19	0,049	-823,7	-3,0
16	x1	5,76	0,76	7,61	6,2E-06	4,11	7,41
17	x2	0,53	0,02	24,41	1,3E-11	0,48	0,58
18	ВЫВОД ОСТАТКА						
19	<i>Наблюдение</i>	<i>Предсказанное y</i>	<i>Остатки</i>		Прогноз		
20	1	850,3	-63,12		\hat{y}	\tilde{x}_1	\tilde{x}_2
21	2	751,6	44,20		1820,7	250	1500
22	3	1134,2	30,94				
23	4	1269,0	-82,03		Довірчі інтервали		
24	5	1388,5	-10,79		1662,5	1978,8	
25	6	1475,3	-12,65				

Рис. 2.8. *Розрахункові результати прикладу 2.6 у MS Excel*

Розглянемо показники, об'єднані назвою *Регрессионная статистика*.

Множественный R – коефіцієнт множинної кореляції $R = 0,99$, що означає сильний зв'язок між усіма змінними.

R-квадрат – коефіцієнт детермінації R^2 . Для даного прикладу $R^2 = 0,99$, тобто 99 % варіації y визначаються лінійною варіацією x_1 та x_2 , а 1 % варіації y пояснюється дією не включених до економетричної моделі факторів або присутністю в ній нелінійних компонентів.

Нормированный R-квадрат – приведений коефіцієнт детермінації \widehat{R}^2 вигляду:

$$\widehat{R}^2 = 1 - \frac{n-1}{n-m-1} \cdot (1 - R^2).$$

Стандартная ошибка – значення середньоквадратичного відхилення σ_e .

Наблюдения – кількість спостережень n . Для *прикладу 2.6* $n = 15$.

Перейдемо до показників, об'єднаних назвою *Дисперсионный анализ*.

Стовпець *df* – число ступенів свободи. Для рядка *Регрессия* показник дорівнює числу пояснювальних змінних: $\nu_1 = m$; для рядка *Остаток* – $\nu_2 = n - m - 1$; для рядка *Итого* – $\nu_1 + \nu_2$. Для *прикладу 2.6* $\nu_1 = 2$; $\nu_2 = 12$; $\nu_1 + \nu_2 = 14$.

Стовпець *SS* – сума квадратів відхилень. Для рядка *Регрессия* показник дорівнює: $SS_r = \sum_{i=1}^n (\widehat{y}_i - \bar{y})^2$;

для рядка *Остаток*: $SS_e = \sum_{i=1}^n (\widehat{y}_i - y_i)^2$;

для рядка *Итого*: $SS = SS_r + SS_e$.

Стовпець *MS* – дисперсії, обчислені за формулою: $MS = \frac{SS}{df}$,

тобто дисперсія на один ступінь свободи.

Стовпець F – значення F -критерію Фішера, обчислене за формулою (2.11). Для **прикладу 2.6** одержано $F = 498,1$.

Стовпець *Значимость* F – рівень значущості α , який відповідає обчисленому F -критерію. Так, для **прикладу 2.6** значущість F дорівнює $2,8E - 12$, що дозволяє зробити висновок про високу довірчу ймовірність $p \approx 1$ побудованої регресійної моделі для генеральної сукупності.

Перейдемо до наступної групи показників.

Стовпець *Коэффициенты* – обчислені значення параметрів $a_0, a_1, a_2, \dots, a_m$, розташовані зверху вниз. Для **прикладу 2.6** одержана економетрична модель (2.1) має такий вигляд:

$$\hat{y} = -413,3 + 5,76x_1 + 0,53x_2. \quad (2.19)$$

Дамо економічне тлумачення параметрів регресії. Параметр a_1 характеризує граничну зміну обсягу виробництва яєць залежно від зміни середньої річної несучості однієї курки-несучки: якщо несучість зросте на 1 шт., то обсяг виробництва яєць збільшиться на 5,76 млн шт. Параметр a_2 характеризує граничну зміну обсягу виробництва яєць залежно від зміни середньої заробітної плати у сільському господарстві: якщо заробітна плата зросте на 1 грн, то обсяг виробництва яєць збільшиться на 530 тис. шт. Параметр a_0 для даного прикладу не має економічного сенсу.

Стовпець *Стандартная ошибка* – значення стандартного квадратичного відхилення σ_j параметра регресії $a_j, j = 0..2$.

Стовпець *t-статистика* – значення статистик Стьюдента t_j для параметрів регресії $a_j, j = 0..2$.

Стовпець *P-значение* – рівень значущості α , що відповідає обчисленому t -критерію. Для **прикладу 2.6** одержано, що коефіцієнти регресії значущі на рівнях:

$$\alpha(t_0) = 0,049; \quad \alpha(t_1) = 6,2E - 06; \quad \alpha(t_2) = 1,3E - 11.$$

Таким чином, можна вважати, що параметри моделі мають довірчу ймовірність p не менше 0,951 і для генеральної сукупності.

Стовпці *Нижние 95%* та *Верхние 95%* – відповідно нижні та верхні межі інтервалів реальних коефіцієнтів регресії для довірчої ймовірності 0,95, тобто:

$$-823,7 \leq \hat{a}_0 \leq -3,0; \quad 4,11 \leq \hat{a}_1 \leq 7,41; \quad 0,48 \leq \hat{a}_2 \leq 0,58.$$

Перейдемо до групи показників *Вывод остатков*.

Стовпець *Наблюдение* – містить номери спостережень.

Стовпець *Предсказанное Y* – значення \hat{y} , обчислені за розрахованим рівнянням регресії.

Стовпець *Остатки* – значення похибок $e_i = y_i - \hat{y}_i$, $i = 1..15$.

Точковий прогноз $\hat{y}(\tilde{x}_1, \tilde{x}_2) = 1820,7$ отримано при підстановці прогнозних значень пояснювальних змінних $\tilde{x}_1 = 250$ та $\tilde{x}_2 = 1500$ до рівняння регресії (2.19).

Інтервальний прогноз знайдено за формулою (2.16). Необхідне для його обчислення значення стандартної похибки $\sigma_e = 72,6$ знаходиться у вихідних даних інструменту *Регрессия* (комірка B6). Двостороннє критичне значення t -критерію обчислено в комірці F7 = *СТБЮДРАСПОБР*(0,05;12). Інтервальний прогноз указує нижню та верхню межу проміжку, в якому знаходиться істинне значення прогнозованої результативної ознаки $y(\tilde{x}_1, \tilde{x}_2)$ при рівні значущості $\alpha = 0,05$, а саме:

$$1662,5 \leq y(\tilde{x}_1, \tilde{x}_2) \leq 1978,8.$$

Приклад 2.7. Із використанням інструменту *Регрессия* табличного процесора *Gnumeric* побудувати та провести дослідження економетричної моделі у вигляді двофакторної лінійної регресії, яка характеризує залежність обсягу споживання плодів і ягід на 1 особу (y), кг/рік, від обсягу їх виробництва на 1 особу (x_1), кг/рік, та середньої заробітної плати (x_2), грн/міс., за вибіркою статистичних даних України 2000–2014 рр. (табл. 2.7).

Вхідні дані прикладу 2.7

№ з/п	y	x_1	x_2
1	29	30	230
2	26	23	311
3	28	25	376
4	33	35	462
5	34	35	590
6	37	36	806
7	35	24	1041
8	42	32	1351
9	44	33	1806
10	46	35	1906
11	48	38	2250
12	53	41	2648
13	53	44	3041
14	56	50	3282
15	52	46	3480

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок A3:C17 аркуша табличного процесора *Gnumeric*.

Економетрична модель у вигляді двофакторної лінійної регресії описується рівнянням:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2.$$

На рис. 2.9 показано вхідні дані прикладу та діалогове вікно інструменту *Регресія*, в якому задано наступні параметри.

1. *Переменные X* – діапазон адрес комірок зі значеннями x_{1i} , x_{2i} , $i = 1..15$, та відповідним заголовком.

2. *Переменные Y* – діапазон адрес комірок зі значеннями y_i , $i = 1..15$, та відповідним заголовком.

3. *Метки* – включається, якщо перший рядок вхідних даних містить заголовок.

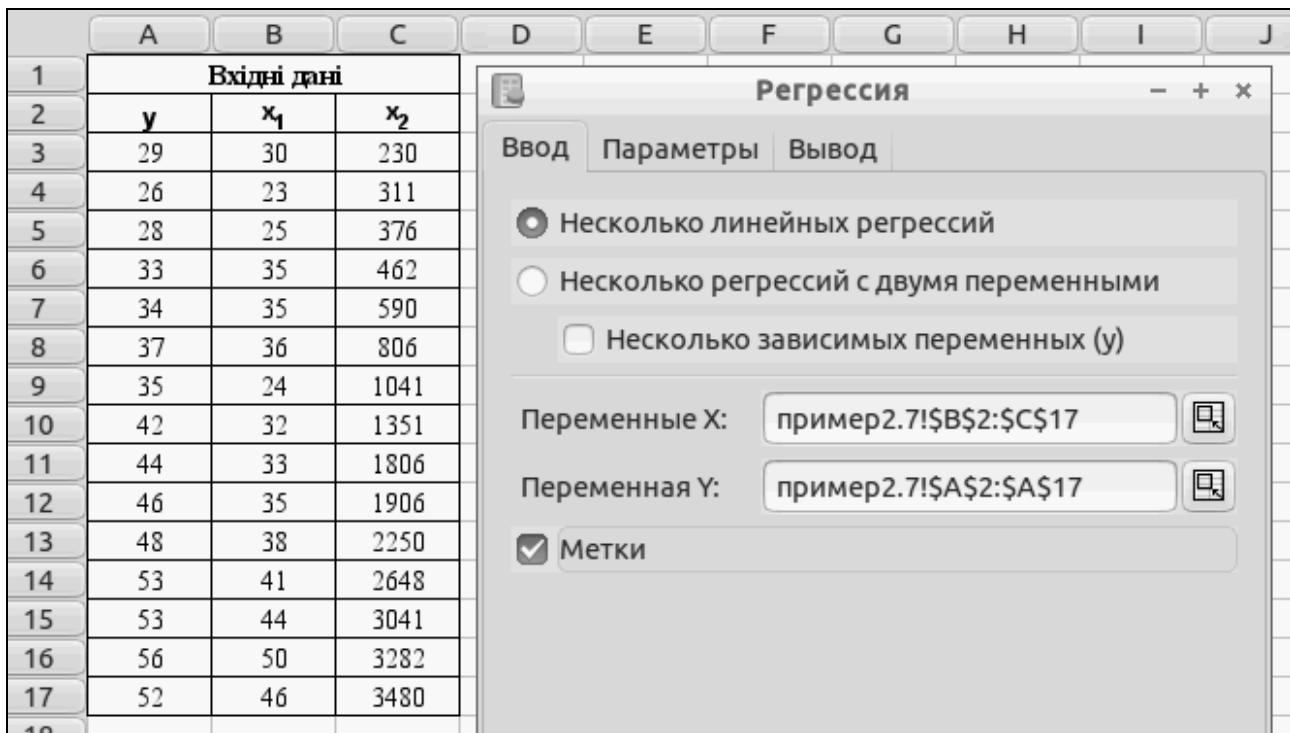


Рис. 2.9. Вхідні дані прикладу 2.7 та вікно інструменту

Регрессия в Gnumeric

Результати роботи інструменту *Регрессия* вносяться на новий аркуш, починаючи з комірки A1.

На рис. 2.10 наведено одержані розрахункові результати за допомогою інструменту *Регрессия*.

Коефіцієнт множинної кореляції $R = 0,97$, що означає сильний зв'язок між усіма змінними. Коефіцієнт детермінації $R^2 = 0,95$, тобто 95 % варіації y визначаються лінійною варіацією x_1 та x_2 , а 5 % – дією не включених до економетричної моделі факторів або присутністю в ній нелінійних компонентів.

Обчислений F -критерій Фішера $F = 104,5$ має рівень значущості $\alpha(F) = 2,6E - 08$, тобто близький до 0, що свідчить про високу довірчу ймовірність отриманої регресійної моделі для генеральної сукупності.

Економетрична модель *прикладу 2.7* має такий вигляд:

$$\hat{y} = 21,75 + 0,23x_1 + 0,01x_2. \quad (2.20)$$

	A	B	C	D	E	F	G
1	Итоговый вывод		Response Vari	y			
2							
3	Регрессионные статистики						
4	Множественная R	0,97					
5	R^2	0,95					
6	Стандартная ошибка	2,54					
7	Вычисленная R	0,94					
8	Наблюдения	15					
9							
10	Дисперсионный анализ						
11		степень св	сумма квадр.	Квадрат сред	F	Значимость F	
12	Регрессия	2	1343,8	671,9	104,5	2,6E-08	
13	Остатки	12	77,2	6,4			
14	Всего	14	1420,9				
15							
16		Коэфф.	Стандартная	t-статистика	p-значение	Ниже 95%	Выше 95%
17	Пересечение	21,75	4,32	5,04	2,9E-04	12,34	31,16
18	x1	0,23	0,16	1,43	0,18	-0,12	0,58
19	x1	0,01	0,00	6,47	3,0E-05	0,005	0,010
20							
21	Ограничение	x1	x1	Prediction	y	Residual	Leverages
22	1	30	230	30	29	-1,3	0,2
23	1	23	311	29	26	-3,3	0,2
24	1	25	376	30	28	-2,2	0,2
25	1	35	462	33	33	-0,1	0,3
26	1	35	590	34	34	0,0	0,2
27	1	26	806	26	27	1,2	0,2

Рис. 2.10. Розрахункові результати прикладу 2.7 в Gnumeric

Дамо економічне тлумачення параметрів регресії. Параметр a_1 характеризує граничну зміну обсягу споживання плодів і ягід залежно від зміни обсягу їх виробництва: якщо обсяг річного виробництва на 1 особу зросте на 1 кг, то обсяг споживання збільшиться на 0,23 кг. Параметр a_2 характеризує граничну зміну обсягу споживання залежно від зміни середньої заробітної плати: якщо заробітна плата підвищиться на 1 грн, то обсяг споживання плодів і ягід збільшиться на 0,01 кг. Параметр a_0 для даного прикладу задає обсяг споживання плодів і ягід імпортного походження.

У стовпці *P-значення* знаходяться рівні значущості α , які відповідають обчисленим *t*-критеріям. Для *прикладу 2.7* отримано, що коефіцієнти регресії значущі на таких рівнях:

$$\alpha(t_0) = 0,00029; \quad \alpha(t_1) = 0,18; \quad \alpha(t_2) = 0,00003.$$

Це означає, що коефіцієнти регресії a_0 та a_2 мають довірчу ймовірність $p \approx 1$ для генеральної сукупності, а коефіцієнт a_1 – довірчу ймовірність $p = 0,82$.

Стовпці *Нижче 95%* та *Вище 95%* – відповідно нижні та верхні межі інтервалів реальних коефіцієнтів регресії для довірчої ймовірності 0,95, тобто:

$$12,34 \leq \hat{a}_0 \leq 31,16; \quad -0,12 \leq \hat{a}_1 \leq 0,58; \quad 0,005 \leq \hat{a}_2 \leq 0,010.$$

У комірках D22:D36 стовпця *Prediction* містяться значення \hat{y} , обчислені за рівнянням регресії. У комірках F22:F36 стовпця *Residual* розташовано значення похибок $e_i = y_i - \hat{y}_i$, $i = 1..15$.

Контрольні запитання до розділу 2

1. Дайте визначення парного коефіцієнта кореляції. Опишіть вигляд кореляційної матриці.
2. За допомогою яких функцій та інструментів табличного процесора *MS Excel* можна розрахувати парні коефіцієнти кореляції між змінними та кореляційну матрицю?
3. Як оцінити статистичну значущість коефіцієнтів кореляції?
4. Назвіть інструменти електронних таблиць *OpenOffice Calc* та *Gnumeric*, за допомогою яких отримують парні коефіцієнти кореляції та кореляційну матрицю.
5. Наведіть формули для обчислення критеріїв Стюдента і Фішера.
6. У чому полягає оцінка значущості параметрів регресії?
7. Назвіть стандартні функції табличного процесора *MS Excel* для обчислення критичного значення t -критерію та рівня значущості розрахованого критерію Стюдента.
8. Поясніть економічний зміст коефіцієнта детермінації.
9. Як перевіряється статистична значущість множинної лінійної регресії?

10. У чому полягає економічний зміст та як обчислити коефіцієнти еластичності?

11. Назвіть стандартні функції електронних таблиць *OpenOffice Calc* для обчислення критичного значення t -критерію та рівня значущості розрахованого критерію Стюдента.

12. Які стандартні функції табличного процесора *MS Excel* використовують для обчислення критичного значення F -критерію та рівня значущості розрахованого критерію Фішера?

13. За допомогою яких інструментів електронних таблиць *MS Excel* знаходять параметри лінійної регресії, коефіцієнт детермінації та F -критерій?

14. Що таке залишки моделі та які інструменти табличного процесора *MS Excel* зручно використовувати для їх обчислення?

15. Дайте визначення поняття точкового прогнозу та назвіть методи його розрахунку в електронних таблицях *MS Excel*.

16. Назвіть інструменти табличних процесорів *OpenOffice Calc* та *Gnumeric* для обчислення параметрів лінійної регресії, коефіцієнта детермінації та F -критерію.

17. У вихідних даних яких інструментів електронних таблиць *Gnumeric* та *MS Excel* знаходяться довірчі інтервали параметрів регресії?

18. Дайте визначення залишків моделі та вкажіть інструменти табличного процесора *OpenOffice Calc* для їх обчислення.

19. Укажіть методи розрахунку точкового прогнозу в електронних таблицях *OpenOffice Calc* та *Gnumeric*.

Завдання для контролю засвоєння матеріалу розділу 2

Завдання 1.

У табл. 2.8 наведено статистичні дані щодо обсягів споживання (y) та виробництва (x_1) овочів на 1 особу, кг/рік, а також середньої

заробітної плати (x_2), грн/міс., за 2000–2014 рр. по Дніпропетровській області. Проведіть аналіз взаємозв'язків між ознаками із використанням інструменту *Корреляція* електронних таблиць *MS Excel*.

Таблиця 2.8

Вхідні дані завдання 1

№ з/п	y	x_1	x_2
1	111	119	273
2	109	120	370
3	116	122	438
4	129	140	526
5	135	143	667
6	139	143	913
7	145	153	1139
8	132	138	1455
9	144	160	1876
10	149	165	1963
11	154	167	2369
12	165	203	2790
13	168	203	3138
14	166	182	3336
15	181	216	3641

Завдання 2.

Із використанням інструменту *Корреляція* табличного процесора *Gnumeric* проведіть аналіз взаємозв'язків між обсягом споживання плодів і ягід на 1 особу (y), кг/рік, обсягом їх виробництва на 1 особу (x_1), кг/рік, і середньою заробітною платою (x_2), грн/міс., за статистичними даними Чернівецької області 2000–2014 рр. (табл. 2.9).

Таблиця 2.9

Вхідні дані завдання 2

№ з/п	y	x_1	x_2
1	43	57	157
2	38	42	218
3	42	59	271

4	49	70	344
5	50	75	441
6	46	70	621
7	46	64	819
8	48	69	1051
9	47	69	1402
10	48	97	1523
11	47	107	1772
12	54	132	1985
13	55	155	2329
14	55	179	2484
15	55	188	2578

Завдання 3.

У табл. 2.10 наведено статистичні дані стосовно обсягу виробництва м'яса на 1 умовну голову ВРХ (y), кг, середньодобових приростів (x_1), г, і витрат кормів у розрахунку на 1 умовну голову ВРХ (x_2), ц корм. од./рік, по Дніпропетровській області за 2000–2014 рр. Із використанням стандартних функцій електронних таблиць *MS Excel* побудуйте економетричну модель у вигляді двофакторної лінійної регресії та проведіть дослідження на предмет залежності між указаними ознаками.

Таблиця 2.10

Вхідні дані завдання 3

№ з/п	y	x_1	x_2
1	113	217	20,7
2	112	313	25,8
3	126	343	26,2
4	144	282	23,2
5	164	335	22,4
6	209	394	24,1
7	227	397	23,6
8	256	403	25,4

9	323	445	22,3
10	324	467	23,6
11	318	463	21,7
12	346	480	22,3
13	344	490	22,3
14	345	501	22,9
15	351	566	21,7

Завдання 4.

Із використанням стандартних функцій табличного процесора *OpenOffice Calc* побудуйте та проведіть дослідження економетричної моделі у вигляді двофакторної лінійної регресії, яка характеризує залежність виробництва молока (y), тис. т, від посівних площ кормових культур (x_1), тис. га, та поголів'я корів (x_2), тис. голів, за вибірковими статистичними даними України 2000–2014 рр. (табл. 2.11).

Таблиця 2.11

Вхідні дані завдання 4

№ з/п	y	x_1	x_2
1	12658	2359	7063
2	13444	2709	6375
3	14142	2873	5858
4	13661	2887	5074
5	13710	3185	4243
6	13714	3487	3738
7	13287	3652	3277
8	12262	3665	3028
9	11761	3793	2752
10	11610	4049	2658
11	10896	4096	2599
12	10751	4177	2658
13	11067	4372	2475
14	11191	4455	2289
15	11133	4508	2101

Завдання 5.

Із використанням функції *TREND* електронних таблиць *OpenOffice Calc* обчисліть прогностні значення пояснюваної змінної $\hat{y}_i, i = 1..15$, точковий прогноз за середніми значеннями пояснювальних змінних та оцініть статистичну значущість економетричної моделі у вигляді двофакторної лінійної регресії, яка характеризує залежність виробництва яєць в Україні (y), млрд шт., від середньої несучості однієї курки-несучки (x_1), шт./рік, та середньої заробітної плати у сільському господарстві (x_2), грн/міс., за вибіркою 2000–2014 рр. (табл. 2.12).

Таблиця 2.12

Вхідні дані завдання 5

№ з/п	y	x_1	x_2
1	8,81	213	111
2	9,67	235	151
3	11,31	246	178
4	11,48	254	210
5	11,96	271	295
6	13,05	274	415
7	14,23	271	553
8	14,06	275	733
9	14,96	278	1077
10	15,91	280	1206
11	16,24	282	1422
12	17,90	286	1786
13	18,36	292	2024
14	19,09	289	2269
15	19,59	276	2476

Завдання 6.

У табл. 2.13 наведено вибіркові статистичні дані про посівні площі кормових культур (y), тис. га, середні ціни реалізації молока сільськогосподарськими підприємствами (x_1), грн/т, та поголів'я

корів (x_2), тис. голів, у Полтавській області за 2000–2015 рр. Побудуйте множинну лінійну регресійну модель залежності площі кормових культур від середніх цін реалізації молока та поголів'я корів і проведіть її аналіз за допомогою інструменту *Регресія* електронних таблиць *MS Excel*.

Таблиця 2.13

Вхідні дані завдання 6

№ з/п	y	x_1	x_2
1	398	526	245
2	367	609	233
3	334	536	233
4	298	721	216
5	244	899	196
6	217	1203	186
7	185	1094	179
8	171	1823	165
9	158	2146	157
10	158	1944	148
11	164	3173	142
12	156	3057	137
13	157	2575	136
14	141	3403	134
15	136	3627	132
16	126	4454	129

Завдання 7.

Із використанням інструменту *Регресія* електронних таблиць *Gnumeric* побудуйте та проведіть дослідження економетричної моделі у вигляді двофакторної лінійної регресії, яка характеризує залежність виробництва кукурудзи на зерно (y), тис. т, від її урожайності (x_1), ц/га, та наявності кукурудзозбиральних комбайнів у сільськогосподарських підприємствах (x_2), шт., за вибірковими статистичними даними Полтавської області 2000–2014 рр. (табл. 2.14).

Вхідні дані завдання 7

№ з/п	y	x_1	x_2
1	381	33	358
2	324	28	321
3	470	34	306
4	783	39	285
5	1156	40	262
6	991	48	244
7	836	41	236
8	1271	47	214
9	1876	57	199
10	1760	59	174
11	1654	44	159
12	3505	79	143
13	2547	47	152
14	4081	69	128
15	3380	58	127

Використана література: [1–4, 7–13, 15–23].

Розділ 3

МУЛЬТИКОЛІНЕАРНІСТЬ

3.1. Поняття та ознаки мультиколінеарності

Однією з умов, які необхідні для оцінювання параметрів загальної лінійної моделі методом найменших квадратів, є умова, що стосується матриці вхідних даних X . Ця матриця має розміри $n \times t$ і повинна мати ранг t , тобто серед пояснювальних змінних моделі не може бути лінійно залежних. Проте оскільки економічні показники, які входять до економетричної моделі як пояснювальні змінні, на практиці дуже часто пов'язані між собою, то це може стати перешкодою для оцінювання параметрів моделі методом найменших квадратів та істотно вплинути на якість економетричного моделювання. Відтак в економетричних дослідженнях важливо з'ясувати, чи існують між пояснювальними змінними взаємозв'язки, які називають мультиколінеарністю.

Мультиколінеарність означає існування тісної лінійної залежності або кореляції між двома чи більше пояснювальними змінними. Вона негативно впливає на кількісні характеристики економетричної моделі або робить її побудову взагалі неможливою. Так, мультиколінеарність пояснювальних змінних призводить до зміщення оцінок параметрів моделі, через що з їх допомогою не можна зробити коректні висновки про результати взаємозв'язку пояснюваної та пояснювальних змінних. Зокрема, коли між пояснювальними змінними існує функціональний зв'язок, оцінити їх вплив на пояснювану ознаку взагалі неможливо.

Нехай зв'язок між пояснювальними змінними не функціональний, проте статистично суттєвий. Тоді попри те, що оцінити параметри методом найменших квадратів теоретично можливо, знайдена оцінка може призвести до таких помилкових значень параметрів, що сама модель стане беззмістовною.

Основні наслідки мультиколінеарності:

1. Падає точність оцінювання, яка виявляється в тому, що:

- а) помилки деяких конкретних оцінок стають занадто великими;
- б) ці помилки досить корельовані одна з одною;
- в) дисперсії оцінок параметрів різко збільшуються.

2. Оцінки параметрів деяких змінних моделі можуть бути незначущими через наявність їх взаємозв'язку з іншими ознаками, а не тому, що вони не впливають на залежну змінну. У такому разі множина вибіркового даних не дає змоги виявити цей вплив.

3. Оцінки параметрів стають досить чутливими до обсягів сукупності спостережень. Збільшення сукупності спостережень іноді може призвести до істотних змін в оцінках параметрів.

З огляду на перелічені наслідки мультиколінеарності, при побудові економетричної моделі потрібно мати інформацію про те, що між пояснювальними змінними не існує мультиколінеарності.

Розглянемо основні її ознаки.

1. Коли серед значень парних коефіцієнтів кореляції пояснювальних змінних є такі, що наближаються до 1, то це означає можливість існування мультиколінеарності.

Інформацію про лінійну залежність може дати симетрична матриця парних коефіцієнтів кореляції (або кореляції нульового порядку) між пояснювальними змінними:

$$r_{xx} = \begin{pmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & r_{x_2x_3} & \dots & r_{x_2x_m} \\ r_{x_3x_1} & r_{x_3x_2} & 1 & \dots & r_{x_3x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_2} & r_{x_mx_3} & \dots & 1 \end{pmatrix}. \quad (3.1)$$

Проте коли до моделі входять більш як дві пояснювальні змінні, вивчення питання про мультиколінеарність не може обмежуватись аналізом тільки матриці (3.1), адже явище мультиколінеарності в

жодному разі не зводиться лише до існування парної кореляції між незалежними змінними.

Більш загальна перевірка передбачає знаходження визначника (детермінанта) матриці r_{xx} , який називається детермінантом кореляції та позначається $\det r_{xx}$. Його числові значення задовольняють умову: $\det r_{xx} \in [0, 1]$.

2. Якщо $\det r_{xx} = 0$, то існує повна мультиколінеарність, а коли $\det r_{xx} = 1$ – мультиколінеарність відсутня. Чим ближче $\det r_{xx}$ до 0, тим певніше можна стверджувати, що між пояснювальними змінними існує мультиколінеарність. Незважаючи на те, що на числове значення $\det r_{xx}$ впливає дисперсія пояснювальних змінних, цей показник можна вважати точковою мірою рівня мультиколінеарності.

3. Коли частинний коефіцієнт детермінації R_j^2 , який обчислено для регресійних залежностей між однією пояснювальною змінною та іншими, має значення, яке близьке до 1, то можна стверджувати про наявність мультиколінеарності.

4. Якщо в економетричній моделі знайдено мале значення параметра a_j при високому рівні частинного коефіцієнта детермінації R_j^2 і при цьому F -критерій істотно відрізняється від 0, то це також свідчить про наявність мультиколінеарності.

5. Нехай при побудові економетричної моделі на основі покрокової регресії введення нової пояснювальної змінної істотно змінює оцінку параметрів моделі при незначному підвищенні (або зниженні) коефіцієнтів кореляції чи детермінації. Тоді ця змінна, очевидно, перебуває у лінійній залежності від інших, які було введено до моделі раніше.

Усі ці ознаки мультиколінеарності мають один спільний недолік: жодна з них чітко не розмежовує випадки, коли мультиколінеарність суттєва або коли нею можна знехтувати.

3.2. Тестування алгоритмом Феррара–Глобера та вилучення мультиколінеарності

Найповніше дослідити мультиколінеарність можна за допомогою алгоритму Феррара–Глобера. Він включає три етапи згідно з видами статистичних критеріїв, за якими перевіряється мультиколінеарність: всього масиву пояснювальних змінних (χ^2 – «хі»-квадрат), кожної пояснювальної змінної з рештою змінних (F -критерій), кожної пари пояснювальних змінних (t -критерій). Усі ці критерії при порівнянні з їх критичними значеннями дають змогу робити конкретні висновки щодо наявності чи відсутності мультиколінеарності пояснювальних факторів.

Перший етап алгоритму – перевірка на загальну мультиколінеарність – здійснюється у наступній послідовності.

1. Обчислити кореляційну матрицю парних коефіцієнтів кореляції (3.1). Для цього застосовують надбудови *Анализ данных* → *Корреляция* електронних таблиць *MS Excel* або *Статистика* → *Описательные статистики* → *Корреляция* табличного процесора *Gnumeric*. Значення парних коефіцієнтів кореляції між пояснювальними змінними можна отримати також за допомогою функції *KORPEL* (*MS Excel*) або *CORREL* (*OpenOffice Calc*) та сформувати з них кореляційну матрицю.

2. Розрахувати детермінант кореляції $\det r_{xx}$ із використанням функції *МОПРЕД* (*MS Excel*) або *MDETERM* (*OpenOffice Calc*).

3. Визначити критерій χ^2 :

$$\chi^2 = -[n - 1 - (2m + 5)/6] \cdot \ln(\det r_{xx}). \quad (3.2)$$

Значення цього критерію порівнюється з критичним значенням $\chi^2(\alpha; \gamma)$ при $\gamma = m \cdot (m - 1) / 2$ ступенях свободи та рівні значущості α . Якщо $\chi^2 > \chi^2(\alpha; \gamma)$, то в масиві пояснювальних змінних існує

мультиколінеарність. У протилежному випадку ($\chi^2 \leq \chi^2(\alpha; \gamma)$) із довірчою ймовірністю $p = 1 - \alpha$ вона відсутня, тому необхідності у подальших розрахунках за алгоритмом Феррара–Глобера немає. Для визначення критичного значення $\chi^2(\alpha; \gamma)$ зручно застосувати функцію *XI2OBP* (*MS Excel*) або *CHIINV* (*OpenOffice Calc*).

Другий етап дослідження за алгоритмом Феррара–Глобера – це перевірка мультиколінеарності кожної пояснювальної змінної з рештою. Для цього необхідно скористатися інформацією, яку надає обернена матриця $C = r_{xx}^{-1}$. Матриця C розраховується з використанням функції *МОБП* (*MS Excel*) чи *MINVERSE* (*OpenOffice Calc*). Далі слід обчислити F -критерії за співвідношенням:

$$F_j = (c_{jj} - 1) \cdot \frac{n - m}{m - 1}, \quad (3.3)$$

де c_{jj} — діагональні елементи матриці C , $j = 1..m$.

Фактичні значення критеріїв порівнюються з критичним значенням F -критерію $F(\alpha; \nu_1; \nu_2)$ при $\nu_1 = m - 1$ та $\nu_2 = n - m$ ступенях свободи та рівні значущості α . Якщо $F_j > F(\alpha; \nu_1; \nu_2)$, то відповідна j -та незалежна змінна мультиколінеарна з іншими. Критичне значення F -критерію знаходять за допомогою функції *FPАСПОБП* (*MS Excel*) або *FINV* (*OpenOffice Calc*).

Якщо коефіцієнт детермінації для кожної змінної $R_j^2 = 1 - 1/c_{jj}$ має значення, яке близьке до 1, то можна стверджувати про наявність мультиколінеарності.

Третій етап дослідження за алгоритмом Феррара–Глобера перевіряє мультиколінеарність кожної пари пояснювальних змінних та потребує:

- по-перше, знаходження частинних коефіцієнтів кореляції

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}, \quad (3.4)$$

де c_{kj} – елемент матриці C , що міститься в k -му рядку та j -му стовпці;
 c_{kk} і c_{jj} – діагональні елементи матриці C ; $j \neq k$, $j = 1..m$, $k = 1..m$;

- по-друге, обчислення t -критеріїв

$$t_{kj} = \frac{r_{kj} \cdot \sqrt{\nu}}{\sqrt{1 - r_{kj}^2}}, \quad (3.5)$$

де $\nu = n - m$, $j \neq k$, $j = 1..m$, $k = 1..m$.

Фактичні значення критеріїв порівнюються з критичним значенням $t(\alpha; \nu)$ при ν ступенях свободи та рівні значущості α . Якщо $|t_{kj}| > t(\alpha; \nu)$, то між незалежними змінними x_k і x_j існує мультиколінеарність. Критичне значення двостороннього t -критерію можна знайти із застосуванням функції *СТЬЮДРАСПОБР* (*MS Excel*) або *TINV* (*OpenOffice Calc*).

Відкинувши одну зі змінних мультиколінеарної пари, можна найпростіше позбутися мультиколінеарності в економетричній моделі. Якщо $F_j > F(\alpha; \nu_1; \nu_2)$, тобто коли x_j залежить від усіх інших пояснювальних факторів, то необхідно вирішувати питання про її вилучення з переліку змінних. Якщо $|t_{kj}| > t(\alpha; \nu)$, то змінні x_k і x_j тісно пов'язані між собою. Звідси, аналізуючи рівень критеріїв Фішера та Стьюдента, можна зробити обґрунтований висновок про те, яку зі змінних необхідно вилучити з дослідження або замінити іншою. Але на практиці вилучення якогось чинника часто суперечить логіці економічних зв'язків, і заміна масиву незалежних змінних завжди має узгоджуватись з економічною доцільністю, що впливає з мети дослідження.

Якщо виявлено мультиколінеарність, то потрібно певним чином перетворити вхідну інформацію, щоб уникнути мультиколінеарності, а далі знов оцінити параметри моделі методом найменших квадратів. Пояснювальні змінні моделі можна скоригувати в такий спосіб:

- взяти відхилення від середньої;
- абсолютні значення замінити на абсолютний приріст;
- використати темп зміни показників тощо.

Розглянемо застосування алгоритму Феррара–Глобера для тестування наявності мультиколінеарності пояснювальних змінних та можливий метод її усунення на конкретних прикладах.

Приклад 3.1. У табл. 3.1 наведено статистичні дані по Україні 2000–2014 рр. щодо обсягу виробництва соняшнику (y), тис. т, його врожайності (x_1), ц/га, середньої потужності двигуна трактора у сільськогосподарських підприємствах (x_2), кВт, та середньої заробітної плати у сільському господарстві (x_3), грн/міс. Із використанням стандартних функцій електронних таблиць *MS Excel* провести дослідження на наявність мультиколінеарності пояснювальних змінних та побудувати економетричну модель у вигляді лінійної регресії, яка характеризує залежність виробництва соняшнику від означених факторів.

Розв’язання.

Вхідні дані прикладу введено в діапазон комірок A3:D17 аркуша електронних таблиць *MS Excel* (рис. 3.1).

Матрицю парних коефіцієнтів кореляції (3.1) обчислено в комірках G3:I5 з використанням інструменту *Корреляція*:

$$r_{xx} = \begin{pmatrix} 1 & 0,901 & 0,924 \\ 0,901 & 1 & 0,983 \\ 0,924 & 0,983 & 1 \end{pmatrix}.$$

Аналізуючи отримані парні коефіцієнти кореляції, можна дійти висновку, що між змінними x_1 , x_2 , x_3 існує сильний лінійний зв’язок.

Із використанням функції *МОПРЕД* у комірці G7 отримано значення детермінанта кореляції $\det r_{xx} = 0,005$. Внаслідок того, що $\det r_{xx}$ близьке до 0, можна стверджувати, що між пояснювальними змінними існує мультиколінеарність.

Вхідні дані прикладу 3.1

№ з/п	y	x_1	x_2	x_3
1	3457	12,2	64,7	111
2	2251	9,4	65,5	151
3	3271	12,0	65,9	178
4	4254	11,2	66,2	210
5	3050	8,9	67,4	295
6	4706	12,8	68,2	415
7	5324	13,6	69,7	553
8	4174	12,2	71,1	733
9	6526	15,3	72,9	1077
10	6364	15,2	74,0	1206
11	6735	15,0	83,3	1422
12	8614	18,4	86,4	1786
13	8313	16,5	86,2	2024
14	10941	21,7	88,7	2269
15	10134	19,4	90,4	2476

Значення критерію $\chi^2 = 65,12$ розраховано за формулою (3.2) у комірці F11. Критичне значення χ^2 для ступеня свободи $\nu = 3$ та рівня значущості $\alpha = 0,05$ обчислено в комірці G11 = *XI2OBR*(0,05;3). Одержано $\chi^2(0,05;3) = 7,81$. Оскільки $\chi^2 > \chi^2(0,05;3)$, то в масиві пояснювальних змінних існує мультиколінеарність.

Обернену матрицю $C = r_{xx}^{-1}$ знайдено з використанням функції *МОБР* (діапазон комірок F14:H16):

$$C = \begin{pmatrix} 6,93 & 1,69 & -8,06 \\ 1,69 & 30,87 & -31,92 \\ -8,06 & -31,92 & 39,84 \end{pmatrix}.$$

Використовуючи діагональні елементи матриці C , за співвідношенням (3.3) у комірках G19:H19 обчислено F -критерії:

$$F_1 = 35,6; \quad F_2 = 179,2; \quad F_3 = 233,0.$$

B25		fx			{=ЛИНЕЙН(A3:A17;B3:C17;1;0)}				
	A	B	C	D	E	F	G	H	I
1	Вхідні дані				Кореляційна матриця				
2	y	x ₁	x ₂	x ₃		x ₁	x ₂	x ₃	
3	3457	12,2	64,7	111		x ₁	1	0,901	0,924
4	2251	9,4	65,5	151		x ₂	0,901	1	0,983
5	3271	12,0	65,9	178		x ₃	0,924	0,983	1
6	4254	11,2	66,2	210					
7	3050	8,9	67,4	295		det r _{xx}	0,005	ln(det r _{xx})	-5,35
8	4706	12,8	68,2	415					
9	5324	13,6	69,7	553		Критерій χ^2			
10	4174	12,2	71,1	733		χ^2	$\chi^2(0,05;3)$		
11	6526	15,3	72,9	1077		65,12	7,81		
12	6364	15,2	74,0	1206					
13	6735	15,0	83,3	1422		Обернена матриця			
14	8614	18,4	86,4	1786		6,93	1,69	-8,06	
15	8313	16,5	86,2	2024		1,69	30,87	-31,92	
16	10941	21,7	88,7	2269		-8,06	-31,92	39,84	
17	10134	19,4	90,4	2476					
18						F(0,05;2;12) =	3,89		
19		Часткові коефіцієнти кореляції				F	35,6	179,2	233,0
20		r ₁₂	r ₁₃	r ₂₃					
21		-0,115	0,485	0,910		t(0,05;12) =	2,18		
22						t	0,40	1,92	7,61
23		Параметри моделі							
24		a ₂	a ₁	a ₀					
25		100,8	480,8	-8509,5					

Рис. 3.1. Розрахункові результати прикладу 3.1 в MS Excel

Для рівня значущості $\alpha = 0,05$ і ступенів свободи $\nu_1 = 2$ та $\nu_2 = 12$ критичне значення критерію Фішера знайдено в комірці H18 = $F_{PACPOBR}(0,05;2;12)$. Одержано $F(0,05;2;12) = 3,89$.

Оскільки

$$F_1 > F(0,05;2;12), \quad F_2 > F(0,05;2;12), \quad F_3 > F(0,05;2;12),$$

то кожна незалежна змінна мультиколінеарна з іншими.

Скориставшись елементами матриці C згідно (3.4), у комірках B21:D21 обчислено частинні коефіцієнти кореляції:

$$r_{12} = -0,115; \quad r_{13} = 0,485; \quad r_{23} = 0,910.$$

Частинні коефіцієнти кореляції r_{12} та r_{13} значно нижчі, ніж відповідні парні коефіцієнти кореляції $r_{x_1x_2}$ та $r_{x_1x_3}$, адже при їх розрахунку виключено вплив інших ознак.

На основі частинних коефіцієнтів кореляції за формулою (3.5) у комірках G22:I22 визначено t -критерії:

$$t_{12} = 0,40; \quad t_{13} = 1,92; \quad t_{23} = 7,61.$$

Для рівня значущості $\alpha = 0,05$ і ступеня свободи $\nu = 12$ двостороннє критичне значення критерію Стюдента знайдено в комірці H21 = *СТЬЮДРАСПОБР*(0,05;12). Одержано $t(0,05;12) = 2,18$. Обчислені t -критерії t_{12} та t_{13} менші за критичне значення $t(0,05;12)$. Тому відповідні пари пояснювальних змінних не є мультиколінеарними. Обчислений критерій t_{23} більше критичного значення $t(0,05;12)$. Відтак змінні x_2 та x_3 тісно взаємопов'язані.

Для усунення мультиколінеарності в досліджуваній економетричній моделі необхідно відкинути одну зі змінних мультиколінеарної пари. У даному випадку можна позбутися змінної x_3 , для якої розрахований F -критерій F_3 має найбільше значення. Таким чином, економетрична модель матиме вигляд двофакторної лінійної регресії:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2.$$

Для обчислення її параметрів застосовано функцію *ЛИНЕЙН*, звідки одержано таке регресійне рівняння:

$$\hat{y} = -8509,5 + 480,8x_1 + 100,8x_2.$$

Приклад 3.2. У табл. 3.2 наведено статистичні дані щодо обсягів споживання (y) і виробництва (x_1) овочів на 1 особу, кг/рік, середньої заробітної плати у цілому по країні (x_2), грн/міс., та середніх цін реалізації овочів сільськогосподарськими підприємствами (x_3), грн/т, в Україні за 2000–2014 рр. Із використанням стандартних функцій

електронних таблиць *Gnumeric* провести дослідження на наявність мультиколінеарності пояснювальних змінних та побудувати економетричну модель у вигляді лінійної регресії, яка характеризує залежність обсягу споживання овочів від означених факторів.

Таблиця 3.2

Вхідні дані прикладу 3.2

№ з/п	y	x_1	x_2	x_3
1	102	118	230	572
2	105	121	311	749
3	108	121	376	865
4	114	137	462	1013
5	115	147	590	1225
6	120	155	806	1462
7	127	172	1041	1547
8	118	147	1351	1995
9	129	172	1806	2060
10	137	181	1906	1790
11	144	178	2250	2489
12	163	217	2648	2090
13	164	223	3041	1924
14	164	218	3282	2304
15	163	224	3480	2514

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок A3:D17 аркуша табличного процесора *Gnumeric* (рис. 3.2).

Матрицю парних коефіцієнтів кореляції (3.1) обчислено в комірках G3:I5 із використанням інструменту *Корреляція*:

$$r_{xx} = \begin{pmatrix} 1 & 0,961 & 0,844 \\ 0,961 & 1 & 0,892 \\ 0,844 & 0,892 & 1 \end{pmatrix}.$$

Аналізуючи отримані парні коефіцієнти кореляції, можна дійти висновку, що між змінними x_1 , x_2 , x_3 існує сильний лінійний зв'язок.

	A	B	C	D	E	F	G	H	I
1	Вхідні дані					Кореляційна матриця			
2	y	x ₁	x ₂	x ₃			x ₁	x ₂	x ₃
3	102	118	230	572		x ₁	1	0,961	0,844
4	105	121	311	749		x ₂	0,961	1	0,892
5	108	121	376	865		x ₃	0,844	0,892	1
6	114	137	462	1013					
7	115	147	590	1225		det r _{xx}	0,015	ln(det r _{xx})	-4,18
8	120	155	806	1462					
9	127	172	1041	1547		Критерій χ^2			
10	118	147	1351	1995		χ^2	$\chi^2(0,05;3)$		
11	129	172	1806	2060		50,84	7,81		
12	137	181	1906	1790					
13	144	178	2250	2489		Обернена матриця			
14	163	217	2648	2090		13,36	-13,63	0,88	
15	164	223	3041	1924		-13,63	18,80	-5,26	
16	164	218	3282	2304		0,88	-5,26	4,95	
17	163	224	3480	2514					
18						F(0,05;2;12) =	3,89		
19		Часткові коефіцієнти кореляції				F	74,2	106,8	23,7
20		r ₁₂	r ₁₃	r ₂₃					
21		0,860	-0,109	0,546		t(0,05;12) =	2,18		
22						t	5,85	0,38	2,26
23		Параметри моделі							
24		a ₃	a ₁	a ₀					
25		0,001	0,57	32,28					

Рис. 3.2. Розрахункові результати прикладу 3.2 в Gnumeric

Із використанням функції *MDETERM* у комірці G7 отримано значення детермінанта кореляції $\det r_{xx} = 0,015$. Внаслідок того, що $\det r_{xx}$ близьке до 0, можна стверджувати, що між пояснювальними змінними існує мультиколінеарність.

Значення критерію $\chi^2 = 50,84$ розраховано за формулою (3.2) у комірці F11. Критичне значення χ^2 для ступеня свободи $\gamma = 3$ та рівня значущості $\alpha = 0,05$ обчислено в комірці G11 = *CHIINV*(0,05;3). Одержано $\chi^2(0,05;3) = 7,81$. Оскільки $\chi^2 > \chi^2(0,05;3)$, то в масиві пояснювальних змінних існує мультиколінеарність.

Обернену матрицю $C = r_{xx}^{-1}$ знайдено в діапазоні комірок F14:H16 із використанням функції *MINVERSE*:

$$C = \begin{pmatrix} 13,36 & -13,63 & 0,88 \\ -13,63 & 18,80 & -5,26 \\ 0,88 & -5,26 & 4,95 \end{pmatrix}.$$

Використовуючи діагональні елементи матриці C , за співвідношенням (3.3) у комірках G19:I19 обчислено F -критерії:

$$F_1 = 74,2; \quad F_2 = 106,8; \quad F_3 = 23,7.$$

Для рівня значущості $\alpha = 0,05$ і ступенів свободи $\nu_1 = 2$ та $\nu_2 = 12$ за допомогою функції *FINV* у комірці H18 визначено критичне значення $F(0,05;2;12) = 3,89$. Оскільки

$$F_1 > F(0,05;2;12), \quad F_2 > F(0,05;2;12), \quad F_3 > F(0,05;2;12),$$

то кожна незалежна змінна мультиколінеарна з іншими.

Застосовуючи елементи матриці C згідно (3.4), у комірках B21:D21 обчислено частинні коефіцієнти кореляції:

$$r_{12} = 0,860; \quad r_{13} = -0,109; \quad r_{23} = 0,546.$$

Частинні коефіцієнти кореляції r_{13} та r_{23} значно нижчі, ніж відповідні парні коефіцієнти кореляції $r_{x_1x_3}$ та $r_{x_2x_3}$, бо при їх розрахунку виключено вплив інших ознак.

На основі обчислених частинних коефіцієнтів кореляції за формулою (3.5) у комірках G22:I22 визначено t -критерії:

$$t_{12} = 5,85; \quad t_{13} = 0,38; \quad t_{23} = 2,26.$$

Для рівня значущості $\alpha = 0,05$ і ступеня свободи $\nu = 12$ двостороннє критичне значення критерію Стьюдента знайдено в комірці H21 = *TINV*(0,05;12). Одержано $t(0,05;12) = 2,18$. Обчислені t -критерії t_{12} та t_{23} більші за їх критичне значення $t(0,05;12)$. Звідси випливає, що відповідні пари пояснювальних змінних є мультиколінеарними. Обчислений критерій t_{13} менше критичного значення $t(0,05;12)$. Тому змінні x_1 та x_3 не взаємопов'язані.

Для усунення мультиколінеарності в досліджуваній економетричній моделі необхідно відкинути змінну x_2 . Таким чином, економетрична модель матиме вид двофакторної лінійної регресії:

$$\hat{y} = a_0 + a_1x_1 + a_3x_3.$$

Щоб обчислити її параметри, застосовано функцію *LINEST*, звідки одержано наступне регресійне рівняння:

$$\hat{y} = 32,28 + 0,57x_1 + 0,001x_3.$$

Контрольні запитання до розділу 3

1. Дайте визначення мультиколінеарності.
2. Наведіть ознаки мультиколінеарності.
3. Які наслідки має існування мультиколінеарності?
4. Назвіть можливі шляхи усунення мультиколінеарності.
5. Які статистичні критерії використовуються для виявлення мультиколінеарності?
6. Дайте коротку характеристику алгоритму Феррара–Глобера.
7. Розшифруйте призначення критеріїв χ^2 , Фішера та Стьюдента в алгоритмі Феррара–Глобера.
8. Назвіть основні інструменти електронних таблиць *MS Excel* при визначенні мультиколінеарності змінних.
9. Опишіть реалізацію алгоритму Феррара–Глобера засобами табличного процесора *OpenOffice Calc*.
10. За допомогою яких функцій електронних таблиць *Gnumeric* виконуються обчислення за алгоритмом Феррара–Глобера?

Завдання для контролю засвоєння матеріалу розділу 3

Завдання 1.

Обчисліть інструментами електронних таблиць *MS Excel* *F*-критерії для дослідження мультиколінеарності трьох пояснювальних змінних, якщо вибірку обсягом $n = 20$ характеризує матриця:

$$C = r_{xx}^{-1} = \begin{pmatrix} 1,60 & -0,41 & 0,61 \\ -0,41 & 1,50 & 0,83 \\ 0,61 & 0,81 & 1,47 \end{pmatrix}.$$

Завдання 2.

Використовуючи елементи матриці *C* завдання 1, обчисліть засобами табличного процесора *Gnumeric* коефіцієнти детермінації змінних. Що вони характеризують?

Завдання 3.

Використовуючи матрицю *C* завдання 1, обчисліть інструментами електронних таблиць *MS Excel* критерій χ^2 для оцінювання мультиколінеарності в масиві пояснювальних змінних.

Завдання 4.

Для наведених значень парних коефіцієнтів кореляції між пояснювальними змінними $r_{x_1x_2} = 0,43$; $r_{x_1x_3} = -0,45$; $r_{x_2x_3} = -0,23$ запишіть матрицю кореляції нульового порядку та обчисліть її визначник засобами табличного процесора *Gnumeric*.

Завдання 5.

У табл. 3.3 наведено вибірккові статистичні дані по Україні 2000–2014 рр. щодо споживання (y) і виробництва (x_1) м'яса на 1 особу, кг/рік, середньої заробітної плати (x_2), грн/міс., та середніх цін реалізації м'яса сільськогосподарськими підприємствами (x_3), грн/т.

Із використанням стандартних функцій електронних таблиць *MS Excel* проведіть дослідження на наявність мультиколінеарності пояснювальних змінних та побудуйте економетричну модель у вигляді лінійної регресії, яка характеризує залежність споживання м'яса від означених факторів.

Таблиця 3.3

Вхідні дані завдання 5

№ з/п	y	x_1	x_2	x_3
1	33	34	230	2358
2	31	31	311	4176
3	33	34	376	3644
4	35	36	462	3481
5	39	34	590	5093
6	39	34	806	6910
7	42	37	1041	6308
8	46	41	1351	6467
9	51	41	1806	10184
10	50	42	1906	10363
11	52	44	2250	10769
12	51	46	2648	11953
13	54	48	3041	13412
14	56	52	3282	12868
15	54	55	3480	15737

Завдання 6.

У табл. 3.4 наведено вибіркві статистичні дані по Україні за 2000–2014 рр. щодо споживання (y) та виробництва (x_1) яєць на 1 особу, шт./рік, середньої заробітної плати (x_2), грн/міс., та середніх цін реалізації яєць сільськогосподарськими підприємствами (x_3), грн/тис. шт. Із використанням стандартних функцій електронних таблиць *Gnumeric* проведіть дослідження на наявність мультиколінеарності пояснювальних змінних та побудуйте економетричну модель у вигляді лінійної регресії, яка характеризує залежність споживання яєць від означених факторів.

Вхідні дані завдання 6

№ з/п	y	x_1	x_2	x_3
1	166	179	230	192
2	180	199	311	210
3	209	235	376	168
4	214	240	462	193
5	220	252	590	238
6	238	277	806	252
7	251	304	1041	193
8	252	302	1351	274
9	260	323	1806	377
10	272	345	1906	404
11	289	373	2250	470
12	310	413	2648	521
13	308	425	3041	629
14	310	443	3282	658
15	310	456	3480	782

Використана література: [1–3, 6–11, 13–17, 20–22].

Розділ 4

АВТОКОРЕЛЯЦІЯ

4.1. Поняття, причини і наслідки автокореляції

Економетричну модель можна побудувати, використавши два типи вихідних даних, що характеризують:

- сукупність різних об'єктів у певний момент (період) часу;
- один об'єкт за ряд послідовних моментів (періодів) часу.

Моделі, побудовані за даними першого типу, називаються просторовими моделями. Моделі, побудовані за даними другого типу, називають моделями часових рядів.

Часовий ряд – це сукупність значень будь-якого показника за декілька послідовних моментів (періодів) часу. Кожен рівень часового ряду формується під впливом великої кількості факторів, які умовно можна поділити на ті, що:

- 1) формують тенденцію ряду;
- 2) формують циклічні коливання ряду;
- 3) є випадковими.

За різних сполучень цих факторів залежність рівнів ряду від часу може набувати різних форм.

По-перше, відмінною особливістю більшості часових рядів економічних показників є сукупний довготривалий вплив факторів на динаміку результативної ознаки. Ці фактори, взяті окремо, можуть спричиняти різноспрямовану дію на пояснювану змінну.

По-друге, для пояснюваної ознаки можуть бути характерними циклічні коливання. Вони можуть носити сезонний характер, оскільки економічна діяльність ряду галузей залежить від періоду року (наприклад, ціни на сільськогосподарську продукцію, обсяги її пропозиції на ринку тощо).

У більшості випадків фактичний рівень часового ряду можна подати як суму або добуток трендової, циклічної та випадкової компонентів. Модель, де часовий ряд представлений як сума перелічених компонентів, називається адитивною. Модель, де часовий ряд представлений як добуток указаних компонентів, називається мультиплікативною.

Основна задача економетричного дослідження окремого часового ряду – виявлення і надання кількісного виразу кожній з перерахованих компонент з тим, щоб використовувати отриману інформацію для прогнозування майбутніх значень ряду або для подальшої побудови моделей взаємозв'язку. Вплив фактора часу буде виражений у кореляційній залежності між значеннями залишків e_i за поточний і попередній моменти часу, яка називається автокореляцією залишків. Таким чином, автокореляція залишків – це взаємозв'язок послідовних елементів часового чи просторового ряду даних. В економетричних дослідженнях дисперсія залишків у даному разі є сталою.

Найпростішим варіантом автокореляції залишків є випадок, коли залишки e_i задовольняють схему авторегресії першого порядку, тобто залежать тільки від залишків попереднього періоду, а саме:

$$e_i = \rho \cdot e_{i-1} + u_i, \quad (4.1)$$

причому $|\rho| < 1$. Величина ρ – коефіцієнт автокореляції залишків першого порядку, який характеризує рівень взаємозв'язку кожного наступного значення з попереднім. Якщо $\rho > 0$, то автокореляція залишків є позитивною, якщо $\rho < 0$, то автокореляція залишків є негативною. Негативна автокореляція в економетричних моделях спостерігається дуже рідко. При $\rho = 0$ автокореляція залишків відсутня.

Графічно випадки позитивної та негативної автокореляції залишків, а також її відсутності представлено на рис. 4.1.

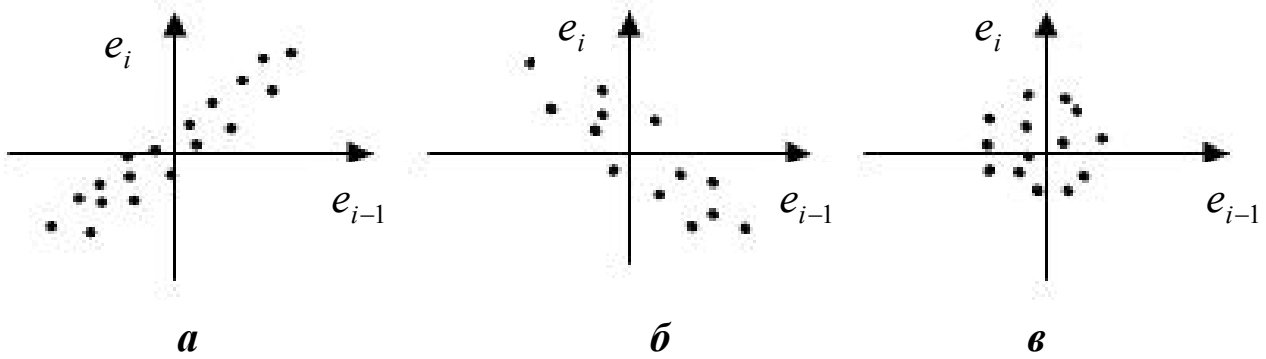


Рис. 4.1. *Випадки автокореляції залишків:*
а – позитивної; б – негативної; в – відсутності

Розглянемо економетричне рівняння множинної лінійної регресії вигляду:

$$y_i = a_0 + \sum_{j=1}^m a_j \cdot x_{ji} + e_i, \quad i = 1..n, \quad (4.2)$$

де m – кількість пояснювальних змінних моделі; n – кількість вибіркового спостережень; $y_i, x_{1i}, \dots, x_{mi}$ – i -й набір вибіркового ознак.

Для кожного моменту часу $i = 1..n$ значення компоненти e_i визначається як

$$e_i = y_i - \hat{y}_i \quad (4.3)$$

або

$$e_i = y_i - \left(a_0 + \sum_{j=1}^m a_j \cdot x_{ji} \right). \quad (4.4)$$

Розглядаючи послідовність залишків як часовий ряд, можна побудувати графік їх залежності від часу. Відповідно до передумов методу найменших квадратів, залишки повинні бути випадковими. Проте при моделюванні часових рядів залишки демонструють певну тенденцію або циклічні коливання. Це свідчить про те, що кожне наступне значення e_i залежить від попередніх. У такому випадку кажуть про наявність автокореляції залишків.

Автокореляція залишків може бути викликана декількома причинами. По-перше, іноді вона пов'язана з вхідними даними та викликана наявністю помилок вимірювання значень результативної

ознаки. По-друге, причина автокореляції залишків може бути у помилковому формулюванні моделі. Дуже часто причиною автокореляції є фактор часу. Крім того, в якості таких суттєвих факторів можуть виступати лагові значення змінних, включених до моделі. Або до моделі не включено декілька другорядних факторів, що мають значний сумісний вплив на результат через збіг тенденцій їх зміни або фаз циклічних коливань.

Як числову характеристику автокореляції використовують коефіцієнт автокореляції ρ . Він обчислюється за безпосередніми даними рядів динаміки, коли фактичні рівні одного ряду розглядаються як значення факторної ознаки, а рівні цього ж самого ряду зсувом на один період приймаються за результативну ознаку.

Коефіцієнт автокореляції ρ має декілька важливих властивостей. По-перше, згідно його побудови, він характеризує тісноту тільки лінійного зв'язку поточного і попередніх рівнів ряду. Тому за коефіцієнтом автокореляції можна судити лише про наявність лінійної (або близької до лінійної) тенденції. Для деяких часових рядів, що мають сильну нелінійну тенденцію (наприклад, квадратичну або експонентційну), коефіцієнт автокореляції рівнів вхідного ряду може наближатися до нуля.

По-друге, за знаком коефіцієнта автокореляції ρ неможливо зробити висновок про зростаючу або спадаючу тенденцію в рівнях ряду. Більшість часових рядів економічних даних містять позитивну автокореляцію рівнів, однак при цьому вони можуть мати загальну спадаючу тенденцію.

Кількість періодів, за яким розраховується коефіцієнт автокореляції, називається лагом. Зі збільшенням лагу кількість пар значень, за якими обчислюється коефіцієнт автокореляції, зменшується.

У загальному випадку залежність між значеннями стохастичної складової e_i у випадку автокореляції залишків можна подати формулою:

$$e_i = \rho_1 e_{i-1} + \rho_2 e_{i-2} + \dots + \rho_s e_{i-s} + u_i, \quad (4.5)$$

де $\rho_1, \rho_2, \dots, \rho_s$ – коефіцієнти автокореляції 1, 2 та s -го порядків відповідно; u_i – помилка.

Послідовність коефіцієнтів автокореляції 1, 2 та s -го порядків називають автокореляційною функцією часового ряду.

Якщо не звернути увагу на автокореляцією залишків і оцінити параметри моделі методом найменших квадратів, то можна отримати такі небажані наслідки:

- вибіркові дисперсії параметрів регресії можуть бути невиправдано великими;

- статистичні критерії t - і F -статистики, які знайдено для лінійної моделі, практично не можуть бути використані в дисперсійному аналізі при автокореляції;

- параметри економетричної моделі виявляються незначущими, через що отримані прогнози будуть неефективними з дуже великою вибірковою дисперсією.

Один із можливих шляхів усунення проблеми автокореляції залишків полягає в застосуванні узагальненого метода найменших квадратів до оцінки параметрів моделі (методу Ейткена).

4.2. Тест Дарбіна–Уотсона та метод Ейткена при автокореляції залишків

Відомо два найбільш поширених методи визначення автокореляції залишків. Перший метод – це побудова графіка залежності залишків від часу та візуальне визначення наявності або відсутності автокореляції (рис. 4.2). Другий метод – використання статистичного тесту Дарбіна–Уотсона.

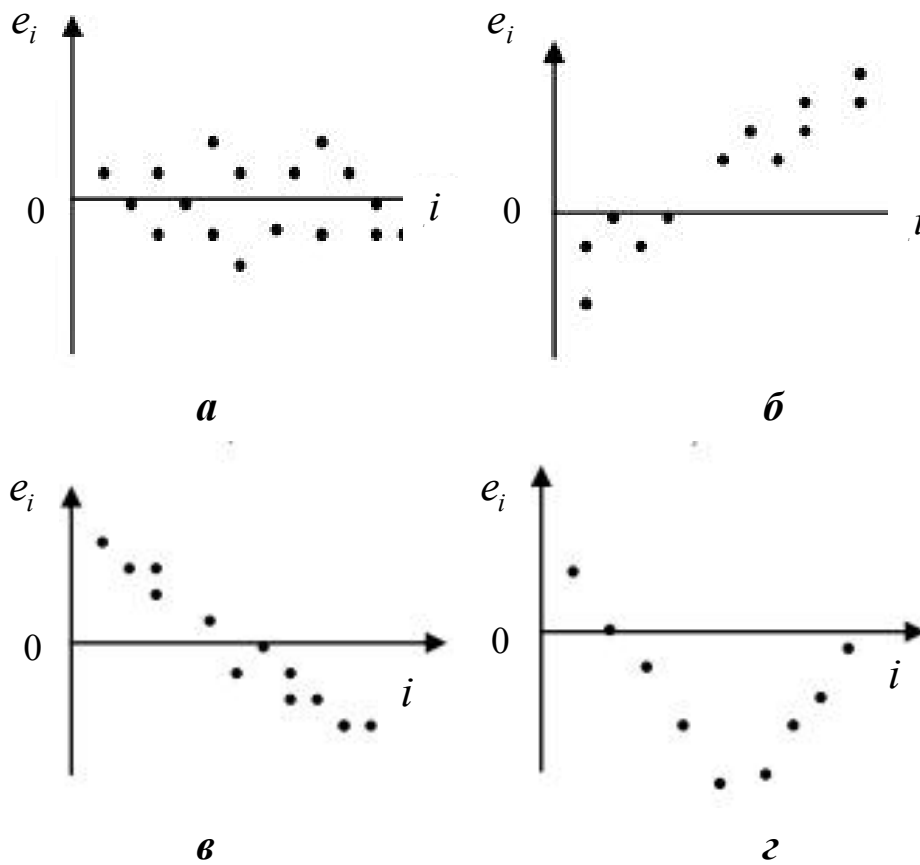


Рис. 4.2. *Моделі залежності залишків від часу:*
а – випадкові залишки; б – зростаюча тенденція;
в – спадаюча тенденція; г – циклічні коливання

Коефіцієнт автокореляції ρ у виразі (4.1) не може бути визначеним безпосередньо, адже неможливо визначити фактичні (у генеральній сукупності спостережень) значення випадкової величини e_i . Але їх можна оцінити методом найменших квадратів на основі відомих залишків для вибірки, звідки отримаємо формулу:

$$\rho = \frac{\text{cov}(e_{i-1}, e_i)}{\sigma_e^2} \approx \frac{\sum_{i=2}^n e_{i-1} \cdot e_i}{\sum_{i=1}^n e_i^2}. \quad (4.6)$$

На практиці замість (4.6) частіше обчислюють оцінку коефіцієнта автокореляції ρ за співвідношенням:

$$\rho = \frac{\frac{1}{n-1} \sum_{i=2}^n e_{i-1} \cdot e_i}{\frac{1}{n} \sum_{i=1}^n e_i^2} = \frac{n}{n-1} \frac{\sum_{i=2}^n e_{i-1} \cdot e_i}{\sum_{i=1}^n e_i^2}. \quad (4.7)$$

Оцінку (4.7) називають циклічним коефіцієнтом автокореляції.

Автокореляція також може бути наслідком помилкової специфікації економетричної моделі. Крім того, наявність автокореляції залишків може означати, що необхідно ввести до моделі нову незалежну змінну. У загальному випадку до моделі включають лише деякі з суттєвих факторів, а вплив ознак, які відсутні в моделі, має позначитися на зміні залишків.

Найбільш відомий і поширений кількісний тест для перевірки моделі на наявність автокореляції залишків ґрунтується на критерії Дарбіна–Уотсона. Цей тест використовується для авторегресійних схем першого порядку в такому вигляді.

1. За припущенням про відсутність автокореляції залишків методом найменших квадратів створюється економетрична модель та обчислюються її залишки e_i , $i = 1..n$.

2. Розраховується критерій Дарбіна–Уотсона за формулою:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}. \quad (4.8)$$

Критерій Дарбіна–Уотсона та коефіцієнт автокореляції залишків першого порядку пов'язані співвідношенням:

$$DW \approx 2 \cdot (1 - \rho). \quad (4.9)$$

Якщо в залишках існує повна позитивна автокореляція і $\rho = 1$, то $DW = 0$. Якщо в залишках є повна негативна автокореляція, то $\rho = -1$ та, відповідно, $DW = 4$. Якщо автокореляція залишків відсутня, то $\rho = 0$ і $DW = 2$. Отже, $0 \leq DW \leq 4$.

3. За обраного рівня значущості α і заданих m факторах моделі та n вибірових спостереженнях за статистичними таблицями DW -розподілу Дарбіна–Уотсона визначаються граничні значення – нижня

d_L та верхня d_U межі критерію Дарбіна–Уотсона, як показано в табл. 4.1 при $\alpha = 0,05$, $m = 1..3$, $n = 6..20$.

Таблиця 4.1

**Таблиця DW-розподілу Дарбіна–Уотсона
при рівні значущості $\alpha = 0,05$**

n	$m=1$		$m=2$		$m=3$	
	d_L	d_U	d_L	d_U	d_L	d_U
6	0,610	1,400	-	-	-	-
7	0,700	1,356	0,467	1,896	-	-
8	0,763	1,332	0,559	1,777	0,368	2,287
9	0,824	1,320	0,629	1,699	0,455	2,128
10	0,879	1,320	0,697	1,641	0,525	2,016
11	0,927	1,324	0,658	1,604	0,595	1,928
12	0,971	1,331	0,812	1,579	0,658	1,864
13	1,010	1,340	0,861	1,562	0,715	1,816
14	1,045	1,350	0,905	1,551	0,767	1,779
15	1,077	1,361	0,946	1,543	0,814	1,750
16	1,106	1,371	0,982	1,539	0,857	1,728
17	1,133	1,381	1,015	1,536	0,897	1,710
18	1,158	1,391	1,046	1,535	0,933	1,696
19	1,180	1,401	1,074	1,536	0,967	1,685
20	1,201	1,411	1,100	1,537	0,998	1,676

4. Формуються зони автокореляційного зв'язку.

Шкалу визначення автокореляції на основі порівняння фактично розрахованого критерію Дарбіна–Уотсона та його критичних значень зображено на рис. 4.3.



Рис. 4.3. Шкала визначення автокореляційного зв'язку

5. На основі розрахованого значення DW робиться висновок про наявність або відсутність автокореляції залишків:

- якщо $0 < DW < d_L$, то існує позитивна автокореляція;
- якщо $4 - d_L < DW < 4$, то існує негативна автокореляція;
- якщо $d_L \leq DW \leq d_U$ або $4 - d_U \leq DW \leq 4 - d_L$, то неможливо

зробити висновок про наявність або відсутність автокореляції залишків;

- якщо $d_U < DW < 4 - d_U$, то автокореляція залишків відсутня.

Оцінку параметрів моделі з автокорельованими залишками можна виконати на основі методу Ейткена. Його доцільно застосовувати у випадках, коли залишки описуються авторегресійною моделлю першого ступеня (4.1).

Методи Кочрена–Оркатта і Дарбіна застосовують для оцінки параметрів економетричної моделі тоді, коли залишки описуються авторегресійною моделлю більш високого ступеня:

$$e_i = \rho_1 \cdot e_{i-1} + \rho_2 \cdot e_{i-2} + u_i;$$

$$e_i = \rho_1 \cdot e_{i-1} + \rho_2 \cdot e_{i-2} + \rho_3 \cdot e_{i-3} + u_i.$$

Розглянемо створення лінійної економетричної моделі методом Ейткена. У цьому випадку оператор визначення числових коефіцієнтів лінійної економетричної моделі задається матричною формулою:

$$A = (X'S^{-1}X)X'S^{-1}Y, \quad (4.10)$$

де матриця S^{-1} має вигляд:

$$S^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \quad \dim S^{-1} = n \times n. \quad (4.11)$$

За допомогою матричних перетворень визначаються:

– матриця значень незалежних змінних моделі X :

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{mn} \end{pmatrix}, \quad \dim X = n \times (m+1); \quad (4.12)$$

– транспонована до X матриця X' розраховується за допомогою функції *TRANSPOSE* (*OpenOffice Calc*) або *ТРАНСП* (*MS Excel*):

$$X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}, \quad \dim X' = (m+1) \times n; \quad (4.13)$$

– добуток матриць $X'S^{-1}$ обчислюється функцією *MMULT* (*OpenOffice Calc*) або *МУМНОЖ* (*MS Excel*);

– обчислюється добуток матриць $X'S^{-1}X$;

– обернена матриця $(X'S^{-1}X)^{-1}$ визначається за допомогою функції *MINVERSE* (*OpenOffice Calc*) або *МОБР* (*MS Excel*);

– розраховується добуток матриць $X'S^{-1}Y$;

– вектор значень параметрів A узагальненої моделі визначається функцією *MMULT* (*OpenOffice Calc*) або *МУМНОЖ* (*MS Excel*) за формулою (4.10).

Розглянемо побудову економетричних моделей з автокорельованими залишками за допомогою електронних таблиць на прикладах.

Приклад 4.1. У табл. 4.2 наведено вибіркові статистичні дані щодо обсягу виробництва (y), тис. т, та посівних площ (x), тис. га, зернових і зернобобових культур в Україні за 2002–2015 рр. Із використанням інструментів табличного процесора *OpenOffice Calc* побудувати економетричну модель до вказаних ознак, перевіривши автокореляцію залишків за тестом Дарбіна–Уотсона.

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C15 аркуша електронних таблиць *OpenOffice Calc*. Значення результативного показника за лінійною економетричною моделлю розраховано функцією *TREND*(B2:B15;C2:C15;C2:C15;1) і розташовано в діапазоні

комірок D2:D15. Залишки економетричної моделі за стовпцями B2:B15 та D2:D15 внесено в діапазон комірок E2:E15.

Таблиця 4.2

Вхідні дані прикладу 4.1

№ з/п	<i>y</i>	<i>x</i>
1	38804,3	15447,6
2	20233,9	12494,6
3	41808,8	15433,5
4	38015,5	15004,8
5	34258,3	14515,2
6	29294,9	15114,9
7	53290,1	15636,3
8	46028,3	15837,3
9	39270,9	15090,0
10	56746,8	15723,8
11	46216,2	15449,0
12	63051,3	16209,9
13	62285,3	15681,6
14	63859,3	14800,8

Значення критерію Дарбіна–Уотсона обчислено в комірці

$$F16 = \text{SUMXMY2}(E3:E15;E2:E14) / \text{SUMSQ}(E2:E15)$$

за формулою (4.8). Одержано $DW = 0,96$.

Його граничні значення при рівні значущості $\alpha = 0,05$, $m = 1$, $n = 14$, згідно з табл. 4.1, дорівнюють: $d_L = 1,045$ – нижня межа, $d_U = 1,35$ – верхня межа, звідки $4 - d_U = 2,65$ та $4 - d_L = 2,96$.

Відповідно до шкали визначення автокореляційного зв'язку (рис. 4.3) розраховане значення критерію Дарбіна–Уотсона знаходиться в зоні позитивної автокореляції залишків.

За припущенням, що залишки описуються авторегресійною схемою першого порядку, за формулою (4.7) в комірці

$$G16 = \text{SUMPRODUCT}(E3:E15;E2:E14) / \text{SUMSQ}(E2:E15) * (14/13)$$

обчислено коефіцієнт автокореляції $\rho = 0,3$.

Для створення узагальненої економетричної моделі методом Ейткена, згідно (4.12), у діапазоні комірок K2:L15 сформовано матрицю X . За допомогою матричних перетворень (4.10)–(4.13) отримано параметри узагальненого лінійного рівняння регресії (рис. 4.4).

	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1															
2	X'	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3		15447,6	12494,6	15433,5	15004,8	14515,2	15114,9	15636,3	15837,3	15090,0	15723,8	15449,0	16209,9	15681,6	14800,8
4															
5	S^{-1}	1	-0,30	0	0	0	0	0	0	0	0	0	0	0	0
6		-0,30	1,09	-0,30	0	0	0	0	0	0	0	0	0	0	0
7		0	-0,30	1,09	-0,30	0	0	0	0	0	0	0	0	0	0
8		0	0	-0,30	1,09	-0,30	0	0	0	0	0	0	0	0	0
9		0	0	0	-0,30	1,09	-0,30	0	0	0	0	0	0	0	0
10		0	0	0	0	-0,30	1,09	-0,30	0	0	0	0	0	0	0
11		0	0	0	0	0	-0,30	1,09	-0,30	0	0	0	0	0	0
12		0	0	0	0	0	0	-0,30	1,09	-0,30	0	0	0	0	0
13		0	0	0	0	0	0	0	-0,30	1,09	-0,30	0	0	0	0
14		0	0	0	0	0	0	0	0	-0,30	1,09	-0,30	0	0	0
15		0	0	0	0	0	0	0	0	0	-0,30	1,09	-0,30	0	0
16		0	0	0	0	0	0	0	0	0	0	-0,30	1,09	-0,30	0
17		0	0	0	0	0	0	0	0	0	0	0	-0,30	1,09	-0,30
18		0	0	0	0	0	0	0	0	0	0	0	0	-0,30	1
19															
20	$X'S^{-1}$	0,70	0,49	0,49	0,49	0,49	0,49	0,49	0,49	0,49	0,49	0,49	0,49	0,49	0,70
21		11666,7	4294,0	8525,3	7316,2	6730,0	7375,0	7701,8	7989,6	6921,2	7922,4	7200,4	8274,0	7733,6	10055,5
22		$X'S^{-1}X$			$(X'S^{-1}X)^{-1}$			$X'S^{-1}Y$		A					
23		7,23	109705		20,25	-0,0013		329611		-84282,62					
24		109705,4	1,7E+09		-0,0013	8,7E-08		5,1E+09		8,56					

Рис. 4.4. Розрахункові результати прикладу 4.1 в OpenOffice Calc

Економетрична модель, побудована методом Ейткена, набуває вигляду:

$$\hat{y} = -84282,62 + 8,56x.$$

Приклад 4.2. У табл. 4.3 наведено вибіркові статистичні дані щодо обсягу виробництва (y), тис. т, та врожайності ріпаку (x), ц/га, в сільськогосподарських підприємствах України 2003–2014 рр. За допомогою інструментів табличного процесора *Gnumeric* перевірити автокореляцію залишків економетричної моделі, розрахованої

методом найменших квадратів, та побудувати методом Ейткена узагальнену парну лінійну регресію до вказаних ознак.

Таблиця 4.3

Вхідні дані прикладу 4.2

№ з/п	<i>y</i>	<i>x</i>
1	50,1	9,4
2	147,4	13,9
3	279,5	14,7
4	600,3	15,7
5	1031,6	13,1
6	2784,2	21,0
7	1827,1	18,7
8	1394,5	17,0
9	1389,0	17,4
10	1191,4	22,1
11	2291,9	23,7
12	2149,5	25,6

Розв’язання.

Вхідні дані прикладу введено в діапазон комірок B2:C13 аркуша електронних таблиць *Gnumeric*. Для отримання залишків лінійної економетричної моделі в програмі *Gnumeric* використано надбудову *Статистика*→*Зависимые наблюдения*→*Регрессия*. Результати сформовано на новому аркуші електронних таблиць.

Згідно з (4.8), у комірці

$$D37 = \text{sumxmy}2(\text{C26:C36};\text{C25:C35}) / \text{sumsq}(\text{C25:C36})$$

обчислено критерій Дарбіна–Уотсона $DW = 1,23$. Його граничні значення при рівні значущості $\alpha = 0,05$, $m = 1$, $n = 12$, згідно з табл. 4.1, дорівнюють: $d_L = 0,97$ – нижня межа, $d_U = 1,33$ – верхня межа, звідки $4 - d_U = 2,67$ та $4 - d_L = 3,03$.

Відповідно до шкали визначення автокореляційного зв’язку (рис. 4.3) розраховане значення критерію Дарбіна–Уотсона

знаходиться в зоні невизначеності, тому неможливо зробити висновок про наявність або відсутність автокореляції залишків.

За припущення, що залишки описуються авторегресійною схемою першого порядку, за формулою (4.7) в комірці

$$E37 = \text{sumproduct}(C26:C36;C25:C35) / \text{sumsq}(C25:C36)*(12/11)$$

обчислено коефіцієнт автокореляції $\rho = 0,4$.

Для створення узагальненої економетричної моделі методом Ейткена, згідно з (4.12), у діапазоні комірок K2:L13 сформовано матрицю X . За допомогою матричних перетворень (4.10)–(4.13) отримано параметри узагальненої лінійної регресії (рис. 4.5).

	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1													
2	X'	1	1	1	1	1	1	1	1	1	1	1	1
3		9,4	13,9	14,7	15,7	13,1	21,0	18,7	17,0	17,4	22,1	23,7	25,6
4													
5	S^{-1}	1	-0,40	0	0	0	0	0	0	0	0	0	0
6		-0,40	1,16	-0,40	0	0	0	0	0	0	0	0	0
7		0	-0,40	1,16	-0,40	0	0	0	0	0	0	0	0
8		0	0	-0,40	1,16	-0,40	0	0	0	0	0	0	0
9		0	0	0	-0,40	1,16	-0,40	0	0	0	0	0	0
10		0	0	0	0	-0,40	1,16	-0,40	0	0	0	0	0
11		0	0	0	0	0	-0,40	1,16	-0,40	0	0	0	0
12		0	0	0	0	0	0	-0,40	1,16	-0,40	0	0	0
13		0	0	0	0	0	0	0	-0,40	1,16	-0,40	0	0
14		0	0	0	0	0	0	0	0	-0,40	1,16	-0,40	0
15		0	0	0	0	0	0	0	0	0	-0,40	1,16	-0,40
16		0	0	0	0	0	0	0	0	0	0	-0,40	1
17													
18	$X'S^{-1}$	0,60	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,60
19		3,82	6,46	5,18	7,07	0,47	11,62	6,46	5,24	4,50	9,16	8,37	16,08
20		$X'S^{-1}X$		$(X'S^{-1}X)^{-1}$			$X'S^{-1}Y$		A				
21		4,78	84,44		2,03	-0,10		5948,98		-1284,78			
22		84,44	1663,71		-0,10	0,01		129687,96		143,16			

Рис. 4.5. Розрахункові результати прикладу 4.2 у *Gnumeric*

Відтак узагальнена економетрична модель, побудована методом Ейткена, набуває вигляду:

$$\hat{y} = -1284,78 + 143,16x.$$

Приклад 4.3. У табл. 4.4 наведено вибірові статистичні дані щодо обсягу виробництва (y), тис. т, та врожайності (x), ц/га, овочів у сільськогосподарських підприємствах Дніпропетровської області за 2001–2014 рр. Із використанням стандартних функцій та інструменту *Регрессія* електронних таблиць *MS Excel* дослідити залишки на наявність автокореляції та побудувати методом Ейткена узагальнену лінійну економетричну модель за вказаними ознаками.

Таблиця 4.4

Вхідні дані прикладу 4.3

№ з/п	y	x
1	430,6	132,8
2	433,9	138,0
3	492,1	167,4
4	498,7	168,7
5	493,9	170,0
6	525,1	177,2
7	469,3	158,4
8	541,2	174,8
9	554,5	185,7
10	560,3	183,9
11	674,1	205,7
12	671,6	207,0
13	600,0	179,2
14	709,6	201,9

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C15 аркуша електронних таблиць *MS Excel*. Для отримання залишків лінійної економетричної моделі використано надбудову *Данные→Анализ данных→Регрессія*. Результати сформовано на новому аркуші електронних таблиць.

Згідно з (4.8), у комірці

$$D39 = \text{СУММКВРАЗН}(C26:C38;C25:C37) / \text{СУММКВ}(C25:C38)$$

розраховано значення критерію Дарбіна–Уотсона $DW = 0,54$. Його граничні значення при рівні значущості $\alpha = 0,05$, $m = 1$, $n = 14$,

згідно з табл. 4.1 дорівнюють: $d_L = 1,045$ – нижня межа, $d_U = 1,35$ – верхня межа, звідки $4 - d_U = 2,65$ та $4 - d_L = 2,96$.

Відповідно до шкали визначення автокореляційного зв'язку (рис. 4.3) розраховане значення критерію Дарбіна–Уотсона знаходиться в зоні позитивної автокореляції залишків.

За припущенням, що залишки описуються авторегресійною схемою першого порядку, за формулою (4.7) у комірці
 $E39 = \text{СУММПРОИЗВ}(C26:C38;C25:C37) / \text{СУММКВ}(C25:C38)*(14/13)$
 обчислено коефіцієнт автокореляції $\rho = 0,55$.

Для створення узагальненої економетричної моделі методом Ейткена, згідно з (4.12), у діапазоні комірок K2:L15 сформовано матрицю X . За допомогою матричних перетворень (4.10)–(4.13) отримано параметри узагальненого лінійного рівняння регресії (рис. 4.6).

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1															
2	X'	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3		132,8	138,0	167,4	168,7	170,0	177,2	158,4	174,8	185,7	183,9	205,7	207,0	179,2	201,9
4															
5	S^{-1}	1	-0,55	0	0	0	0	0	0	0	0	0	0	0	0
6		-0,55	1,30	-0,55	0	0	0	0	0	0	0	0	0	0	0
7		0	-0,55	1,30	-0,55	0	0	0	0	0	0	0	0	0	0
8		0	0	-0,55	1,30	-0,55	0	0	0	0	0	0	0	0	0
9		0	0	0	-0,55	1,30	-0,55	0	0	0	0	0	0	0	0
10		0	0	0	0	-0,55	1,30	-0,55	0	0	0	0	0	0	0
11		0	0	0	0	0	-0,55	1,30	-0,55	0	0	0	0	0	0
12		0	0	0	0	0	0	-0,55	1,30	-0,55	0	0	0	0	0
13		0	0	0	0	0	0	0	-0,55	1,30	-0,55	0	0	0	0
14		0	0	0	0	0	0	0	0	-0,55	1,30	-0,55	0	0	0
15		0	0	0	0	0	0	0	0	0	-0,55	1,30	-0,55	0	0
16		0	0	0	0	0	0	0	0	0	0	-0,55	1,30	-0,55	0
17		0	0	0	0	0	0	0	0	0	0	0	-0,55	1,30	-0,55
18		0	0	0	0	0	0	0	0	0	0	0	0	-0,55	1
19															
20	$X'S^{-1}$	0,45	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,45
21		56,93	14,67	49,38	34,20	31,22	50,21	12,76	38,46	44,62	24,30	52,97	57,96	8,56	103,38
22		$X'S^{-1}X$			$(X'S^{-1}X)^{-1}$			$X'S^{-1}Y$		A					
23		3,333	579,62		8,21	-0,05		7654,9		806,73					
24		579,6	104624		-0,05	0,0003		1364001		8,57					

Рис. 4.6. Розрахункові результати прикладу 4.3 у MS Excel

Звідки узагальнена економетрична модель, побудована методом Ейткена, набуває вигляду:

$$\hat{y} = 806,73 + 8,57x.$$

Контрольні запитання до розділу 4

1. Дайте визначення коефіцієнта автокореляції.
2. Назвіть причини виникнення та наслідки автокореляції залишків.
3. У чому полягає тест Дарбіна–Уотсона?
4. За допомогою яких функцій та інструментів табличних процесорів *MS Excel*, *OpenOffice Calc* та *Gnumeric* можна обчислити розрахункове значення *DW*-критерію?
5. Назвіть функції електронних таблиць *OpenOffice Calc*, *Gnumeric* та *MS Excel* для обчислення коефіцієнта автокореляції ρ .
6. Опишіть загальну схему методу Ейткена у випадку автокореляції залишків.
7. Які функції матричних перетворень використовують для реалізації методу Ейткена в табличних процесорах *MS Excel*, *Gnumeric* та *OpenOffice Calc*?

Завдання для контролю засвоєння матеріалу розділу 4

Завдання 1.

У табл. 4.5 наведено вибіркові статистичні дані про обсяги виробництва молока (y), тис. т, і поголів'я корів (x), тис. голів, у Дніпропетровській області за 2001–2014 рр. Із використанням стандартних функцій електронних таблиць *Gnumeric* проведіть дослідження автокореляції залишків та побудуйте методом Ейткена узагальнену лінійну економетричну модель до вказаних ознак.

Вхідні дані завдання 1

№ з/п	y	x
1	527,6	192,6
2	558,3	179,3
3	508,0	175,1
4	492,6	147,7
5	494,0	131,1
6	474,3	118,1
7	415,7	108,2
8	381,4	97,2
9	359,2	88,3
10	339,8	83,4
11	341,7	80,5
12	343,4	79,8
13	348,0	80,2
14	357,2	80,2

Завдання 2.

У табл. 4.6 наведено вибірккові статистичні дані стосовно обсягів споживання (y) та виробництва (x) яєць на 1 особу, шт./рік, в Україні за 2001–2014 рр. Із використанням стандартних функцій електронних таблиць *OpenOffice Calc* проведіть дослідження автокореляції залишків та побудуйте методом Ейткена узагальнену лінійну економетричну модель до вказаних ознак.

Таблиця 4.6

Вхідні дані завдання 2

№ з/п	y	x
1	180	199
2	209	235
3	214	240
4	220	252
5	238	277
6	251	304
7	252	302
8	260	323

Продовження табл. 4.6

9	272	345
10	289	373
11	310	413
12	308	425
13	310	443
14	310	456

Завдання 3.

У табл. 4.7 наведено вибіркові статистичні дані щодо виробництва (y), тис. т, та врожайності (x), ц/га, кукурудзи на зерно в сільськогосподарських підприємствах Полтавської області за 2001–2014 рр. Із використанням стандартних функцій електронних таблиць *MS Excel* проведіть дослідження автокореляції залишків та знайдіть методом Ейткена параметри узагальненої лінійної економетричної моделі до вказаних ознак.

Таблиця 4.7

Вхідні дані завдання 3

№ з/п	y	x
1	324,0	28,1
2	470,3	34,3
3	782,7	38,6
4	1155,5	40,3
5	991,4	47,8
6	835,7	40,8
7	1271,4	47,2
8	1875,5	56,8
9	1759,6	59,0
10	1654,2	43,8
11	3505,0	78,9
12	2547,2	47,2
13	4081,3	68,9
14	3380,1	58,1

Використана література: [1, 3, 5, 9–11, 14–17, 20].

Розділ 5

НЕЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

5.1. Види нелінійних регресій

У багатьох практичних випадках моделювання економічних процесів лінійними економетричними моделями дає цілком задовільний результат і може використовуватися для аналізу та прогнозування. Однак внаслідок різноманіття і складності економічних явищ обмежитися застосуванням тільки лінійних моделей неможливо. Як свідчить економічна теорія, багато економічних залежностей не є лінійними по суті, тому їх моделювання лінійними функціями не дасть адекватний результат. У цьому випадку необхідно використовувати нелінійні економетричні моделі, тобто регресії, які встановлюють нелінійну залежність між економічними показниками, один з яких є пояснюваною змінною, а інші – пояснювальними змінними, або факторами. Рівняння до досліджуваного процесу має бути за можливості простим, щоб сутність зв'язку між змінними проявлялась досить чітко, а параметри рівняння піддавались певному економічному тлумаченню. Із декількох регресійних залежностей зв'язок ознак краще відображає та, якій відповідає більше значення коефіцієнта детермінації R^2 (2.10).

За методами оцінювання параметрів усі нелінійні економетричні моделі поділяються на два типи: 1) нелінійні за факторами, але лінійні за параметрами – квазілінійні моделі; 2) нелінійні за факторами та за параметрами.

Будь-яка квазілінійна модель у загальному випадку може бути представлена в такому вигляді:

$$\hat{y} = a_0 + a_1 \cdot f(x_1) + a_2 \cdot f(x_2) + \dots + a_m \cdot f(x_m),$$

де $f(x_j)$ – нелінійні функції від x_j , $j = 1..m$.

Квазілінійні моделі є більш простим і привабливим варіантом нелінійних моделей. Їх привабливість пояснюється тим, що оцінки

параметрів таких моделей можуть бути отримані методами лінійного регресійного аналізу. Прикладом указаної моделі може бути модель попиту, для якої функція попиту має вигляд:

$$Q = a_0 + a_1 \cdot P + a_2 \cdot P^2,$$

де Q – попит; P – ціна; a_0, a_1, a_2 – параметри моделі.

Нелінійні за факторами і за параметрами економетричні моделі є більш складними, тому оцінювання їх параметрів, як правило, виконується методами нелінійного регресійного аналізу. Прикладом може бути модель, що описує залежність між обсягом надходжень до бюджету і податковою ставкою на основі кривої Лаффера:

$$\hat{y} = a_0 \cdot e^{a_1(x-a_2)^2},$$

де \hat{y} – податкові надходження; x – податкова ставка; a_0, a_1, a_2 – параметри моделі.

Порівнюючи лінійні та нелінійні економетричні моделі, можна зробити висновок, що лінійна регресійна модель, відповідні методи оцінювання її параметрів, тестування і прогнозування в цілому теоретично краще обгрунтовані, ніж нелінійні, та мають відносно простий обчислювальний апарат. За необхідності застосування в аналізі нелінійних моделей практикують такі підходи:

1) замість складної нелінійної функціональної залежності навіть з невеликою кількістю факторів застосувати лінійну регресійну функцію з великою кількістю факторів;

2) за допомогою математичних співвідношень перетворити нелінійну функцію на лінійну.

Процес приведення нелінійної регресійної моделі до лінійного вигляду називається лінеаризацією, а сама модель – лінеаризованою. Для вказаних моделей повністю зберігається методологія економетричного дослідження для загальної лінійної регресійної моделі. Можливість лінеаризації моделі залежить від типу нелінійної моделі.

Розглянемо нелінійні економетричні моделі, які зводяться до лінійних.

1. Степенева однофакторна модель описується функцією регресії такого вигляду:

$$\hat{y} = a_0 \cdot x^{a_1}, \quad (5.1)$$

де a_0, a_1 – параметри моделі.

Нелінійна модель на основі степеневі функції регресії є однією з найпоширеніших у практиці моделей і описує достатньо широкий спектр економічних явищ і процесів, таких як процес виробництва (виробничі функції), попит на товари різних категорій (криві Енгеля), для опису кривих байдужості тощо.

При моделюванні економічних явищ і процесів мають сенс тільки випадки, коли $a_0 \geq 0$, що є типовим для економічних процесів. Якщо значення параметру a_1 – не ціле число, то розглядається лише випадок, коли $x \geq 0$. При цьому залежно від знака та значення параметру a_1 модель може описувати різні економічні процеси: прискорене зростання, уповільнене зростання і спад (рис. 5.1).

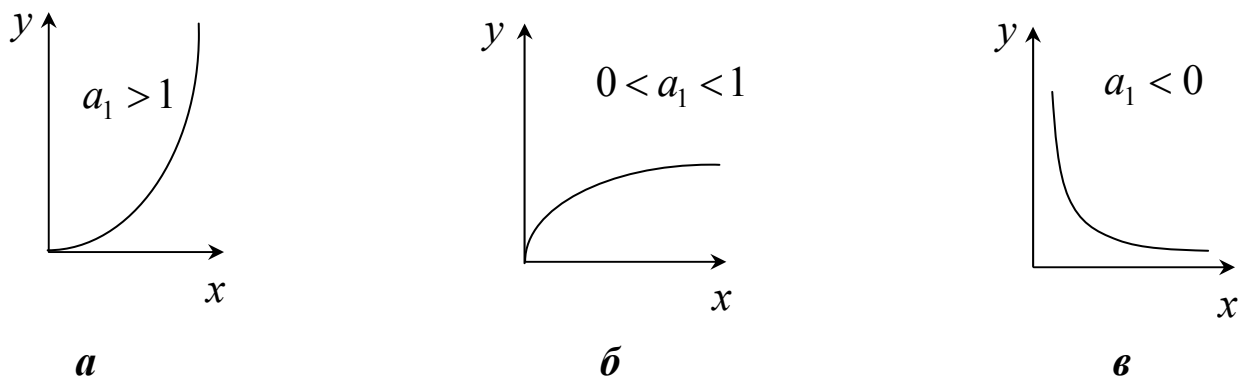


Рис. 5.1. *Графіки степеневі функції для економічних процесів: а – прискореного зростання; б – уповільненого зростання; в – спаду*

Лінеаризація функції (5.1) здійснюється у два кроки. Спочатку виконується логарифмування її лівої та правої частин:

$$\ln \hat{y} = \ln a_0 + a_1 \cdot \ln x,$$

а потім здійснюється така заміна:

$$y_1 = \ln \hat{y}, \quad b_0 = \ln a_0, \quad b_1 = a_1, \quad x_1 = \ln x.$$

У результаті нелінійна функція (5.1) зводиться до лінійної форми:

$$y_1 = b_0 + b_1 \cdot x_1.$$

Параметр a_1 у степеневій моделі характеризує еластичність змінної y за змінною x , тобто цей параметр фактично дорівнює коефіцієнту еластичності y за x . Тому часто степеневу модель ще називають моделлю сталої еластичності, що вказує на можливі напрями її застосування.

2. Показникова однофакторна модель регресії має декілька форм:

$$\hat{y} = a_0 \cdot a_1^x; \quad (5.2)$$

$$\hat{y} = a_0 \cdot e^{a_1 x}; \quad (5.3)$$

$$\hat{y} = a_0 \cdot (1 - a_1)^x; \quad (5.4)$$

$$\hat{y} = e^{a_0 + a_1 x}, \quad (5.5)$$

де a_0, a_1 – параметри моделі.

Експоненціальна модель (5.3) використовується для опису швидкозростаючих або швидкоспадаючих економічних процесів. Найбільш типовим її застосуванням є ситуація, коли аналізується зміна результату y з постійним темпом приросту в часі.

Лінеаризація показникових функцій (5.2)–(5.5) виконується, як і для степеневі, шляхом застосування операції логарифмування та подальшої заміни змінних, тобто:

$$\left. \begin{array}{l} \hat{y} = a_0 \cdot a_1^x \quad \rightarrow \quad \ln \hat{y} = \ln a_0 + x \cdot \ln a_1 \\ \hat{y} = a_0 \cdot e^{a_1 \cdot x} \quad \rightarrow \quad \ln \hat{y} = \ln a_0 + a_1 \cdot x \\ \hat{y} = a_0 \cdot (1 - a_1)^x \quad \rightarrow \quad \ln \hat{y} = \ln a_0 + x \cdot \ln(1 - a_1) \\ \hat{y} = e^{a_0 + a_1 \cdot x} \quad \rightarrow \quad \ln \hat{y} = a_0 + a_1 \cdot x \end{array} \right\} \Rightarrow y_1 = b_0 + b_1 \cdot x.$$

3. Обернено пропорційна однофакторна функція регресії має вигляд:

$$\hat{y} = a_0 + a_1 \cdot \frac{1}{x}, \quad (5.6)$$

де a_0, a_1 – параметри моделі.

Особливістю функції (5.6) є те, що при $x \rightarrow \infty$ залежна змінна моделі y прямує до свого граничного значення a_0 , тобто функція має асимптоту a_0 . Внаслідок цього функція (5.6) застосовується для моделювання економічних явищ і процесів, обмежених зверху або знизу. Найбільш поширеним використанням таких моделей є кількісний опис залежності попиту на блага y від доходу x (криві Торнквіста), а також залежності зміни заробітної плати y від рівня безробіття x (крива Філіпса).

Графік обернено пропорційної функції значною мірою залежить від значень і знаків параметрів a_0 та a_1 . Вигляд графіків указаної функції за різних значень і знаків параметрів a_0 та a_1 наведено на рис. 5.2.

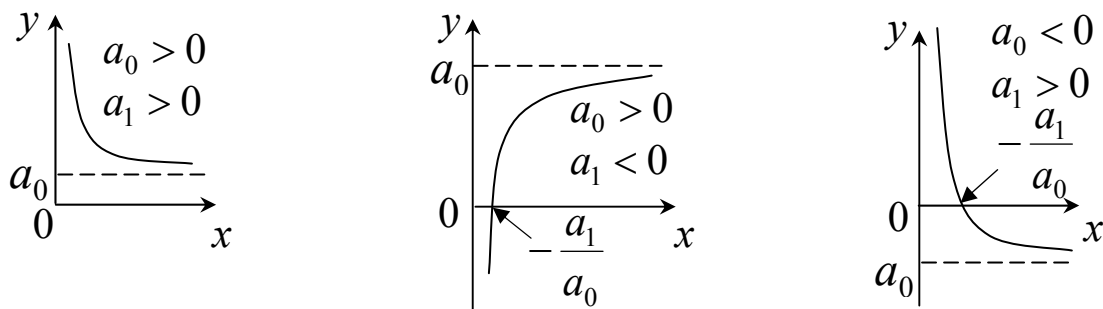


Рис. 5.2. Графіки обернено пропорційної функції

Обернено пропорційна модель відноситься до квазілінійних моделей. У зв'язку з цим її лінеаризація здійснюється простою заміною змінної: $z = \frac{1}{x}$. Після підстановки в (5.6) нової змінної отримуємо таку парну лінійну регресію:

$$\hat{y} = a_0 + a_1 \cdot z.$$

4. Квадратичні моделі використовуються для опису дуже широкого спектра економічних процесів завдяки їхнім універсальним можливостям. У загальному випадку квадратична модель має вигляд:

$$\hat{y} = a_0 + a_1 \cdot x + a_2 \cdot x^2, \quad (5.7)$$

де a_0, a_1, a_2 – параметри моделі.

Залежно від знаків і значень параметрів моделі вона може відображати еволюцію – дуже різну на різних проміжках інтервалу зміни пояснювальної ознаки x : зростання, спад, зростання з подальшим спадом, спад з подальшим зростанням. Так, квадратична модель може описувати кількісну залежність між обсягом випуску і середніми або граничними видатками, між витратами на рекламу та прибутками тощо. На рис. 5.3 наведено можливі варіанти графіків квадратичної функції.

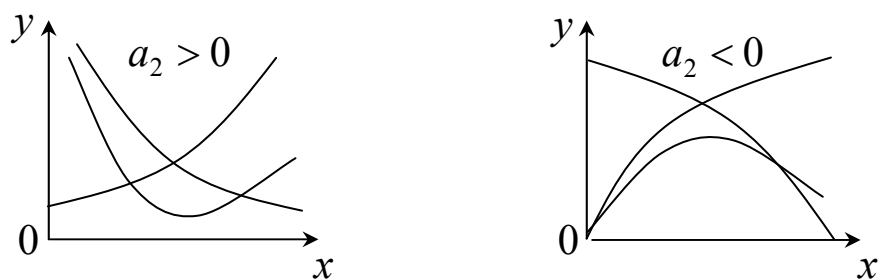


Рис. 5.3. *Графіки квадратичної функції*

Квадратична модель, як і обернено пропорційна, також відноситься до квазілінійних. Її лінеаризація виконується шляхом елементарної заміни факторів: $x_1 = x, x_2 = x^2$ та зведенням її до двофакторної лінійної регресії:

$$\hat{y} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2,$$

параметри якої розраховуються за методикою для лінійних економічних моделей.

5. Крім однофакторної степеневі функції (5.1) широке розповсюдження, особливо при побудові виробничих функцій, набули багатфакторні мультиплікативні моделі вигляду:

$$\hat{y} = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \dots \cdot x_m^{a_m},$$

лінеаризація яких виконується так само, як і для розглянутого випадку однієї пояснювальної змінної.

Класичним прикладом економетричного мультиплікативного моделювання є виробнича функція Кобба–Дугласа, що описує залежність між обсягом виробленої продукції та витраченими для цього ресурсами. Уперше поняття виробничої функцію запропонував шведський економіст К. Вікселль. У 1928 р. виробничу функцію перевірено на статистичних даних Ч. Коббом і П. Дугласом у роботі «Теорія виробництва», де зроблена спроба емпіричним шляхом визначити вплив витрат капіталу та праці на обсяг продукції, що випускається в обробній промисловості США.

Загальний вид функції Кобба–Дугласа такий:

$$P = a_0 \cdot K^{a_1} \cdot L^{a_2}, \quad (5.8)$$

де P – обсяг випуску продукції; K – основний капітал (капітальні витрати); L – затрати праці (чисельність зайнятих у галузі); a_0, a_1, a_2 – параметри моделі.

Таким чином, рівняння (5.8) визначає обсяг випуску (P) як функцію від загальної факторної продуктивності (a_0), витрат капіталу (K), витрат праці (L) і часток капіталу та праці як факторів виробництва в обсязі випуску (відповідно a_1 і a_2). Приріст a_0, K або L спричиняє приріст в обсязі випуску P . У той час як капітал і праця – матеріальні чинники, загальна продуктивність – це скоріше нематеріальний фактор, що характеризує рівень розвитку людського капіталу.

Параметри a_1 і a_2 визначають еластичність випуску продукції за витратами капіталу $E_K = a_1$ та праці $E_L = a_2$.

На їх основі оцінюється вплив кожного з перелічених ресурсів на обсяг випуску продукції.

Сума параметрів $(a_1 + a_2)$ називається ступенем однорідності функції Кобба–Дугласа. Економічний сенс цього показника – віддача від масштабу виробництва. Зокрема, розрізняють такі випадки:

- $a_1 + a_2 = 1$ – постійна віддача, яка означає, що зі збільшенням обох виробничих ресурсів на одиницю обсяг продукції також зросте на одиницю. Отже, ефективність ресурсів у такому разі є сталою;

- $a_1 + a_2 < 1$ – спадна віддача, за якої темпи росту обсягу продукції нижчі за темпи збільшення витрат ресурсів;

- $a_1 + a_2 > 1$ – зростаюча віддача, за якої темпи росту обсягу продукції вищі за темпи збільшення витрат виробничих ресурсів.

Припустимо, що рівень кожного виробничого ресурсу збільшився на s %, тоді величини їх дорівнюватимуть $K\left(1 + \frac{s}{100}\right)$ і $L\left(1 + \frac{s}{100}\right)$.

Обсяг продукції на основі виробничої функції запишеться так:

$$P_1 = a_0 \left[K \left(1 + \frac{s}{100} \right) \right]^{a_1} \left[L \left(1 + \frac{s}{100} \right) \right]^{a_2} =$$

$$= a_0 K^{a_1} L^{a_2} \left(1 + \frac{s}{100} \right)^{a_1} \left(1 + \frac{s}{100} \right)^{a_2} = P \left(1 + \frac{s}{100} \right)^{a_1 + a_2}.$$

Звідси при $a_1 + a_2 > 1$ обсяг продукції зростає більш ніж на s %; при $a_1 + a_2 < 1$ – менш ніж на s %; при $a_1 + a_2 = 1$ обсяг продукції збільшиться рівно на s %.

Як і у випадку лінійних моделей, на основі нелінійних економетричних регресій можна обчислити такі показники:

– середню ефективність впливу пояснювальних змінних на залежну

$$A_j = \frac{\bar{y}}{\bar{x}_j}, \quad j = 1..m;$$

– граничну ефективність впливу пояснювальних змінних на залежну

$$M_j = \frac{\partial \hat{y}}{\partial x_j}, j = 1..m;$$

– відносний вплив пояснювальних змінних на залежну – коефіцієнти еластичності

$$E_j = \frac{M_j}{A_j}, j = 1..m.$$

Формули розрахунків показників граничної ефективності та коефіцієнтів еластичності для основних нелінійних регресійних моделей наведено в табл. 5.1.

Таблиця 5.1

Основні показники для нелінійних економетричних моделей

Загальний вигляд нелінійної регресії	M	E
$\hat{y} = a_0 \cdot x^{a_1}$	$a_1 \cdot \frac{\bar{y}}{\bar{x}}$	a_1
$\hat{y} = a_0 \cdot a_1^x$	$\bar{y} \cdot \ln a_1$	$\bar{x} \cdot \ln a_1$
$\hat{y} = a_0 + a_1 \cdot \frac{1}{x}$	$-\frac{a_1}{\bar{x}^2}$	$-\frac{a_1}{\bar{x} \cdot \bar{y}}$
$\hat{y} = a_0 + a_1 \cdot x + a_2 \cdot x^2$	$a_1 + 2 \cdot a_2 \cdot \bar{x}$	$\frac{a_1 \cdot \bar{x} + 2 \cdot a_2 \cdot \bar{x}^2}{\bar{y}}$

5.2. Створення нелінійних економетричних моделей в електронних таблицях

Розглянемо побудову та проаналізуємо нелінійні економетричні моделі в електронних таблицях на конкретних прикладах.

Приклад 5.1. У табл. 5.2 наведено статистичні дані про обсяги споживання (y) і виробництва (x) яєць на 1 особу, шт./рік, в Україні за 2000–2014 рр. Із використанням стандартних функцій програми

OpenOffice Calc побудувати нелінійні однофакторні регресії та обґрунтувати, що найбільш адекватною з них є квадратична модель.

Таблиця 5.2

Вхідні дані прикладу 5.1

№ з/п	<i>y</i>	<i>x</i>
1	166	179
2	180	199
3	209	235
4	214	240
5	220	252
6	238	277
7	251	304
8	252	302
9	260	323
10	272	345
11	289	373
12	310	413
13	308	425
14	310	443
15	310	456

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C16 аркуша електронних таблиць *OpenOffice Calc*.

Для отримання нелінійних однофакторних моделей побудовано діаграму розсіювання *Мастером діаграмм*: *Вставка*→*Діаграма*→*Тип діаграми*→*Діаграма XY*. Далі з контекстного меню на діаграмі обрано пункти *Вставити лінію тренда*→*Полиномиальная (Степенная, Экспоненциальная)* і поставлено два прапорця: *Показати рівняння* та *Показати коефіцієнт детермінації*.

Результат, зображений на рис. 5.4, дає візуальне уявлення про існування нелінійної залежності між змінними.

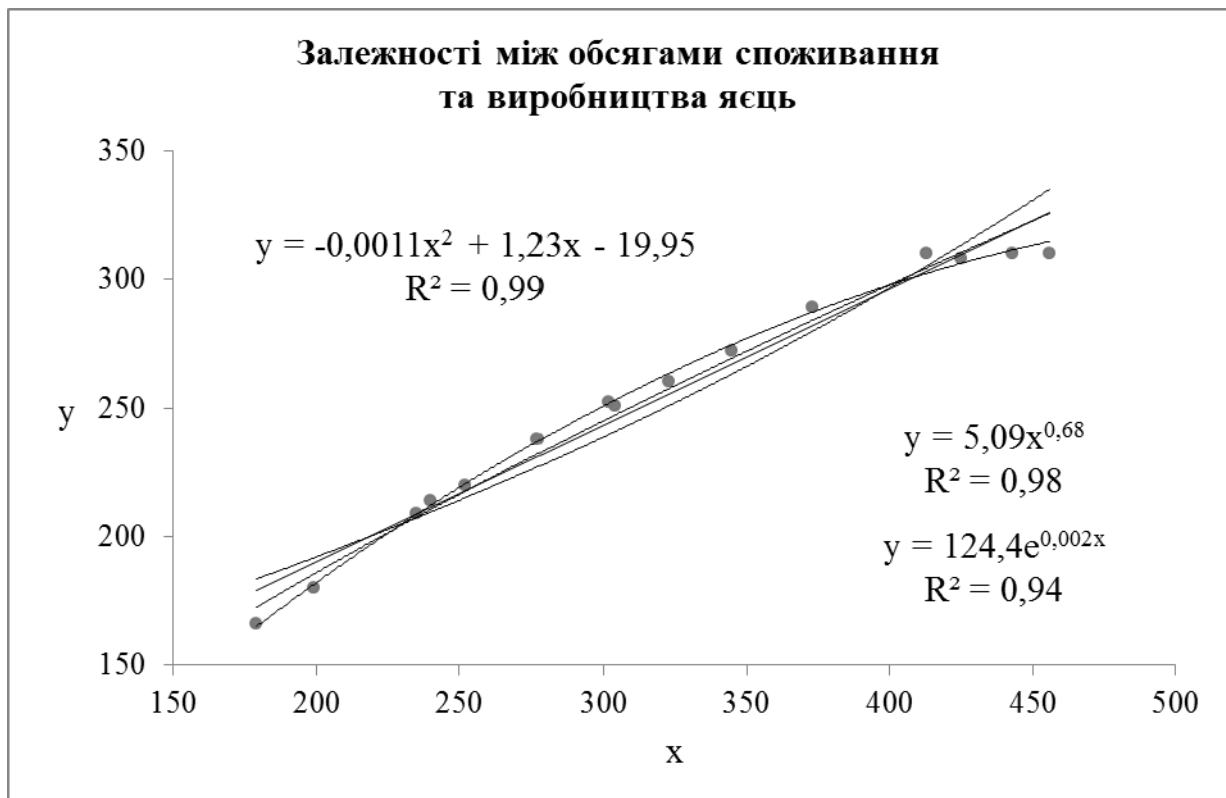


Рис. 5.4. *Розрахункові результати прикладу 5.1 в OpenOffice Calc*

Апроксимацію вхідних даних прикладу на кореляційному полі здійснено:

- квадратичною однофакторною регресією вигляду

$$\hat{y} = -19,95 + 1,23x - 0,0011x^2$$

з коефіцієнтом детермінації $R^2 = 0,99$;

- степеневою однофакторною регресією вигляду

$$\hat{y} = 5,09 \cdot x^{0,68}$$

з коефіцієнтом детермінації $R^2 = 0,98$;

- показниковою однофакторною регресією вигляду

$$\hat{y} = 124,4 \cdot e^{0,002x}$$

з коефіцієнтом детермінації $R^2 = 0,94$.

Із трьох регресійних залежностей зв'язок ознак точніше відображає модель з більшим значенням коефіцієнта детермінації, тобто $R^2 = 0,99$. Тому побудована квадратична однофакторна модель

залежності обсягу споживання від обсягу виробництва яєць набуває вигляду:

$$\hat{y} = -19,95 + 1,23x - 0,0011x^2.$$

Приклад 5.2. У табл. 5.3 наведено вибірккові статистичні дані про обсяги виробництва (y), тис. т, та врожайність (x), ц/га, кукурудзи на зерно в Полтавській області за 2000–2014 рр. Із використанням інструментів табличного процесора *MS Excel* побудувати однофакторні регресії та обґрунтувати, що найбільш адекватною з них є степенева модель.

Таблиця 5.3

Вхідні дані прикладу 5.2

№ з/п	y	x
1	381,0	32,6
2	324,0	28,1
3	470,3	34,3
4	782,7	38,6
5	1155,5	40,3
6	991,4	47,8
7	835,7	40,8
8	1271,4	47,2
9	1875,5	56,8
10	1759,6	59,0
11	1654,2	43,8
12	3505,0	78,9
13	2547,2	47,2
14	4081,3	68,9
15	3380,1	58,1

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C16 аркуша електронних таблиць *MS Excel*. Для створення нелінійних однофакторних регресій побудовано діаграму розсіювання *Мастером діаграмм*: *Вставка*→*Діаграма*, тип *Точечная*. У контекстному меню на діаграмі вибрано *Добавить линию тренда*→*Степенная*

(Полиномиальная, Экспоненциальная) і встановлено прапорці *Показать уравнение на диаграмме* та *Поместить на диаграмму величину достоверности аппроксимации R²*.

Аналіз результатів на рис. 5.5 дає візуальне уявлення про існування нелінійної залежності між змінними прикладу. Апроксимацію вхідних даних на кореляційному полі здійснено:

- квадратичною однофакторною регресією вигляду

$$\hat{y} = -2553,7 + 98,71x - 0,21x^2$$

з коефіцієнтом детермінації $R^2 = 0,79$;

- степеневою однофакторною регресією вигляду

$$\hat{y} = 0,06 \cdot x^{2,60}$$

з коефіцієнтом детермінації $R^2 = 0,85$;

- показниковою однофакторною регресією вигляду

$$\hat{y} = 110,32 \cdot e^{0,05x}$$

з коефіцієнтом детермінації $R^2 = 0,79$.

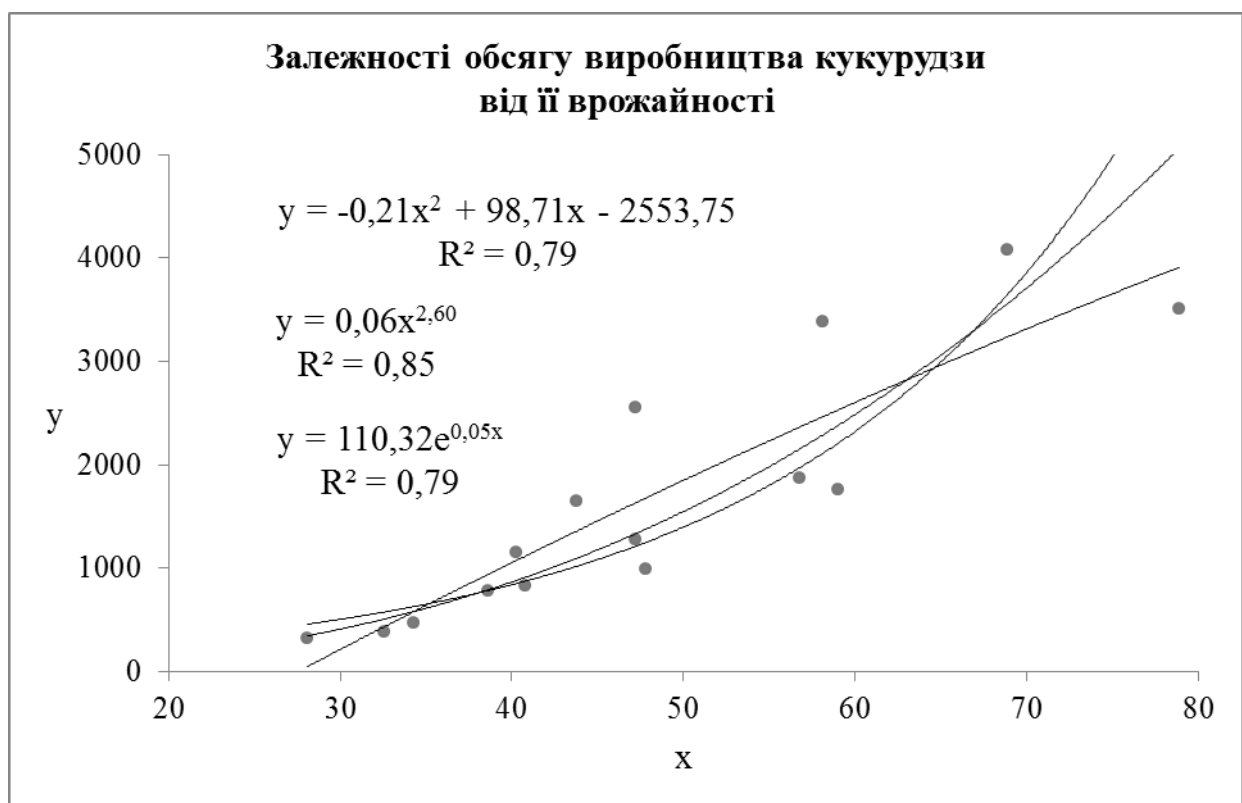


Рис. 5.5. Розрахункові результати прикладу 5.2 у MS Excel

Із трьох регресійних залежностей зв'язок ознак точніше відображає та, якій відповідає більше значення коефіцієнта детермінації, тобто $R^2 = 0,85$. Таким чином, побудована степенева однофакторна модель залежності обсягу виробництва кукурудзи на зерно від її урожайності набуває вигляду:

$$\hat{y} = 0,06 \cdot x^{2,60}.$$

Приклад 5.3. У табл. 5.4 наведено статистичні дані стосовно обсягів виробництва (y) і споживання (x) яєць на 1 особу, шт./рік, за 2005–2014 рр. по Дніпропетровській області. Із використанням інструментів електронних таблиць *Gnumeric* побудувати нелінійні однофакторні регресії та обґрунтувати, що найбільш адекватною з них є показникова модель.

Таблиця 5.4

Вхідні дані прикладу 5.3

№ з/п	y	x
1	255	231
2	284	251
3	258	257
4	267	268
5	284	280
6	298	298
7	328	326
8	330	324
9	289	327
10	319	325

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C11 аркуша електронних таблиць *Gnumeric*. Для отримання нелінійних однофакторних регресій побудовано діаграму розсіювання: *Вставка*→*Діаграма*→*тип XY*. У пункті меню *Настройка диаграммы* вибрано *Диаграмма 1*→*Добавить*→*Линия тренда*→*экспоненциальная (полиномиальная, степенная)*; *Добавить*→*уравнения экспоненциальная (полиномиальная, степенная) регрессия*. Коефіцієнт детермінації R^2

виводиться в разі встановлення відповідного прапорця для кожного виду регресії.

Результат, зображений на рис. 5.6, дає візуальне уявлення про існування нелінійної залежності між змінними прикладу.

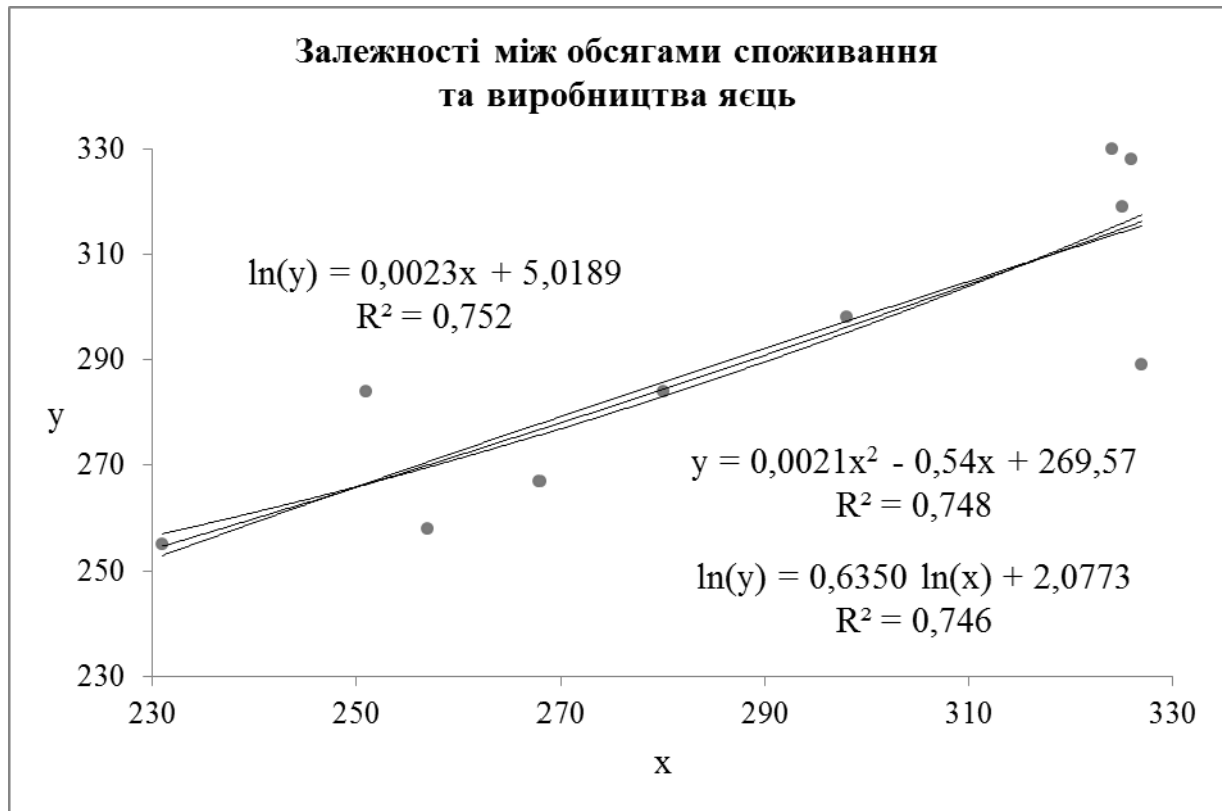


Рис. 5.6. Розрахункові результати прикладу 5.3 у Gnumeric

Апроксимацію вхідних даних на кореляційному полі здійснено:

- показниковою однофакторною регресією вигляду

$$\ln(\hat{y}) = 0,0023x + 5,0189,$$

або після потенціювання з основою e

$$\hat{y} = 151,24e^{0,0023x}$$

з коефіцієнтом детермінації $R^2 = 0,752$;

- квадратичною однофакторною регресією вигляду

$$\hat{y} = 269,57 - 0,54x + 0,0021x^2$$

з коефіцієнтом детермінації $R^2 = 0,748$;

- степенною однофакторною регресією вигляду

$$\ln(\hat{y}) = 0,6350 \ln(x) + 2,0773,$$

або після потенціювання з основою e

$$\hat{y} = 7,98x^{0,635}$$

з коефіцієнтом детермінації $R^2 = 0,746$.

Із трьох регресійних залежностей зв'язок ознак точніше відображає та, якій відповідає більше значення коефіцієнта детермінації, тобто $R^2 = 0,752$. Таким чином, побудована показникова однофакторна модель залежності обсягу виробництва від споживання яєць на 1 особу в рік набуває вигляду:

$$\hat{y} = 151,24e^{0,0023x}$$

Приклад 5.4. У табл. 5.5 наведено статистичні дані щодо обсягів виробництва (y) та споживання плодів і ягід на 1 особу (x), кг/рік, в Україні за 2005–2014 рр. Із використанням стандартних функцій табличного процесора *MS Excel* побудувати поліноміальні однофакторні регресії та вибрати ту з них, що точніше відображає нелінійну залежність між указаними ознаками.

Таблиця 5.5

Вхідні дані прикладу 5.4

№ з/п	y	x
1	36	37,1
2	24	34,8
3	32	42,1
4	33	43,5
5	35	45,6
6	38	48,0
7	41	52,6
8	44	53,3
9	50	56,3
10	46	52,3

Розв'язання.

Вхідні дані прикладу введено в діапазон комірок B2:C11 аркуша електронних таблиць *MS Excel*. Для отримання поліноміальних однофакторних регресій побудовано діаграму розсіювання *Мастером*

діаграмм. У контекстному меню на діаграмі вибрано *Добавить линию тренда* → *Полиномиальная, Степень 2 (3, 4)* і встановлено прапорці *Показать уравнение на диаграмме* та *Поместить на диаграмму величину достоверности аппроксимации R²*. Результат, зображений на рис. 5.7, дає візуальне уявлення про існування нелінійної залежності між змінними прикладу в вигляді полінома.

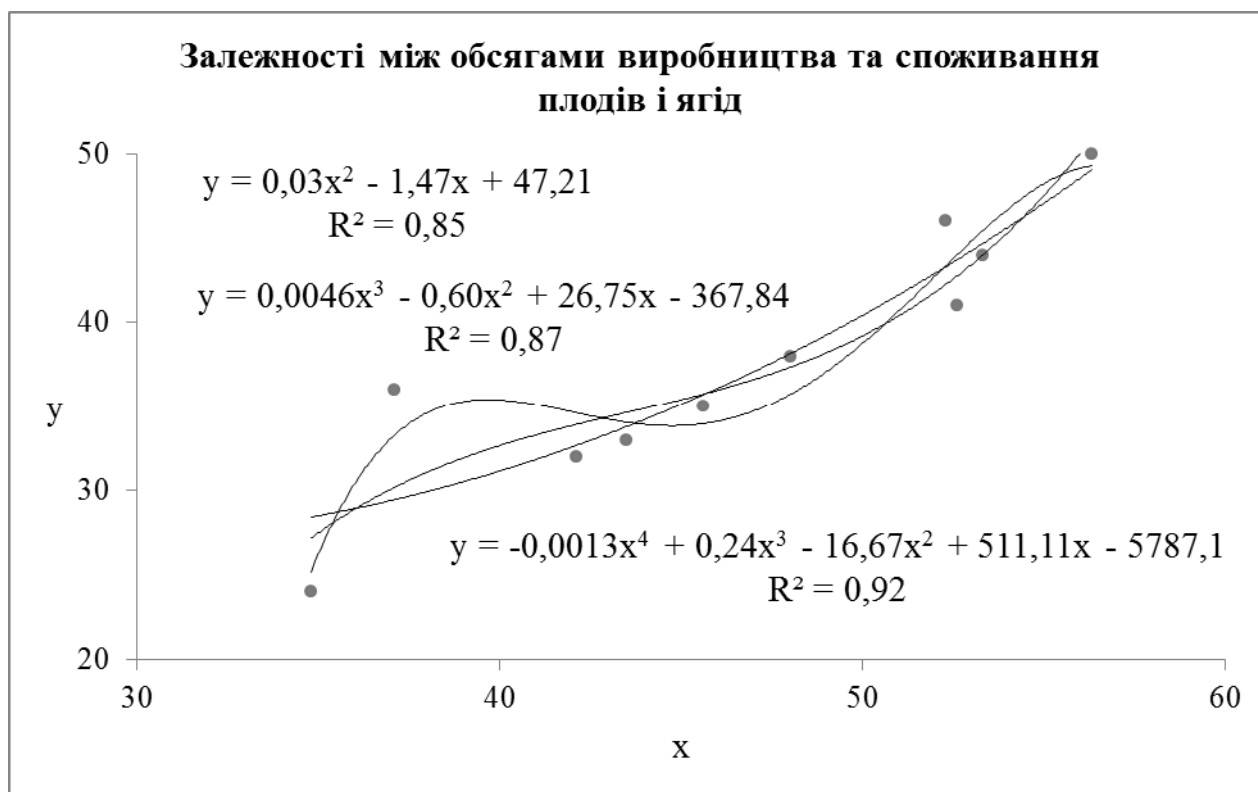


Рис. 5.7. *Розрахункові результати прикладу 5.4 у MS Excel*

Апроксимацію вхідних даних на кореляційному полі здійснено:

- поліноміальною однофакторною регресією 2-го ступеня

$$\hat{y} = 47,21 - 1,47x + 0,03x^2$$

з коефіцієнтом детермінації $R^2 = 0,85$;

- поліноміальною однофакторною регресією 3-го ступеня

$$\hat{y} = -367,84 + 26,75x - 0,60x^2 + 0,0046x^3$$

з коефіцієнтом детермінації $R^2 = 0,87$;

- поліноміальною однофакторною регресією 4-го ступеня

$$\hat{y} = -5787,1 + 511,11x - 16,67x^2 + 0,24x^3 - 0,0013x^4$$

з коефіцієнтом детермінації $R^2 = 0,92$.

Із трьох поліноміальних моделей зв'язок ознак найбільш точно відображає та, якій відповідає більше значення коефіцієнта детермінації, тобто $R^2 = 0,92$. Побудована однофакторна поліноміальна модель 4-го ступеня щодо залежності обсягу виробництва від споживання плодів і ягід на 1 особу набуває вигляду:

$$\hat{y} = -5787,1 + 511,11x - 16,67x^2 + 0,24x^3 - 0,0013x^4.$$

Контрольні запитання до розділу 5

1. Опишіть вигляд, особливості та способи лінеаризації степеневі однофакторної моделі.
2. Наведіть форми показникової однофакторної моделі та способи її лінеаризації.
3. Який вигляд має обернено пропорційна модель та в який спосіб здійснюється її перетворення у лінійну форму?
4. Де застосовується квадратична парна регресія та в який спосіб відбувається її лінеаризація?
5. Наведіть вигляд та властивості виробничої функції Кобба–Дугласа.
6. Опишіть алгоритм та назвіть інструменти табличного процесора *MS Excel* для побудування нелінійних регресій.
7. Які засоби електронних таблиць *OpenOffice Calc* призначені для створення однофакторних нелінійних економетричних моделей?
8. Із використанням яких інструментів електронних таблиць *Gnumeric* можна побудувати нелінійні парні регресії?

Завдання для контролю засвоєння матеріалу розділу 5

Завдання 1.

Дані, наведені в табл. 5.6, характеризують обсяг виробництва (y), тис. т, та середні ціни реалізації (x), грн/т, плодів і ягід господарствами населення України за 2006–2015 рр. За допомогою

електронних таблиць *MS Excel* побудуйте нелінійні однофакторні регресії та обґрунтуйте, що найбільш адекватною з них є квадратична модель.

Таблиця 5.6

Вхідні дані завдання 1

№ з/п	y	x
1	978,1	2094,5
2	1271,6	2324,0
3	1272,4	3847,6
4	1403,3	4271,2
5	1459,7	5043,7
6	1596,5	6097,5
7	1639,7	5985,4
8	1851,1	6412,5
9	1667,1	7569,5
10	1741,1	8031,1

Завдання 2.

Дані, наведені в табл. 5.7, характеризують середні ціни реалізації соняшнику (y), грн/т, і його посівні площі (x), тис. га, у сільськогосподарських підприємствах Дніпропетровській області за 2001–2014 рр. Із використанням табличного процесора *MS Excel* побудуйте поліноміальні однофакторні регресії та виберіть ту, що найбільш адекватно відображає нелінійну залежність між указаними ознаками.

Таблиця 5.7

Вхідні дані завдання 2

№ з/п	y	x
1	669,6	192,6
2	604,0	179,3
3	786,2	175,1
4	952,5	147,7
5	1253,0	131,1
6	1187,4	118,1
7	1734,2	108,2
8	2332,2	97,2

Продовження табл. 5.7

9	2068,3	88,3
10	3195,6	83,4
11	3165,0	80,5
12	2867,1	79,8
13	3563,5	80,2
14	3796,2	80,2

Завдання 3.

Засобами електронних таблиць *OpenOffice Calc* побудуйте нелінійні регресійні моделі залежності обсягу реалізації свиней (y), т, від витрачених кормів на їх годівлю (x), тис. ц корм. од., за вибірковими статистичними даними областей України 2015 р. та обґрунтуйте, що найбільш адекватною з них є степенева модель (табл. 5.8).

Таблиця 5.8

Вхідні дані завдання 3

№ з/п	Область	y	x
1	Вінницька	9455	811
2	Волинська	12149	494
3	Дніпропетровська	48678	2305
4	Донецька	21011	2548
5	Житомирська	7812	363
6	Закарпатська	3405	189
7	Запорізька	29119	1386
8	Івано-Франківська	44301	1515
9	Київська	54034	2858
10	Кіровоградська	14659	959
11	Луганська	2106	296
12	Львівська	32630	1367
13	Миколаївська	6065	382
14	Одеська	12354	972
15	Полтавська	33334	2159
16	Рівненська	5197	220
17	Сумська	8202	459
18	Тернопільська	24100	1138
19	Харківська	20915	1371

20	Херсонська	13488	627
21	Хмельницька	20277	1085
22	Черкаська	28548	1687
23	Чернівецька	9378	348
24	Чернігівська	10941	680

Завдання 4.

Із використанням табличного процесора *MS Excel* побудуйте нелінійні однофакторні регресії та обґрунтуйте, що найбільш адекватною з них є показникова залежність обсягів споживання м'яса та м'ясопродуктів на 1 особу (y), кг/рік, від середньої заробітної плати населення України (x), грн/міс., за вибіркою 2005–2014 рр. (табл. 5.9).

Таблиця 5.9

Вхідні дані завдання 4

№ з/п	y	x
1	39,1	806
2	42,0	1041
3	45,7	1351
4	50,6	1806
5	49,7	1906
6	52,0	2239
7	51,2	2633
8	54,4	3026
9	56,1	3265
10	54,1	3480

Використана література: [1–3, 6–11, 13, 15–17, 20–23].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Апатова Н. В.* Інформатика для економістів : підручник / *Н. В. Апатова, О. М. Гончарова, Ю. Ю. Дюлічева.* – К. : Центр навч. літератури, 2011. – 456 с.
2. *Бредюк В. І.* Курс лекцій з дисципліни «Економетрія» / *В. І. Бредюк.* – Рівне : НУВГП, 2006 – 154 с.
3. Вільна енциклопедія Вікіпедія [Електронний ресурс] // Режим доступу : <http://uk.wikipedia.org>
4. *Глинський Я. М.* Практикум з інформатики / *Я. М. Глинський.* – Львів : СПД Глинський, 2010. – 304 с.
5. *Доля В. Т.* Економетрія : навчальний посібник / *В. Т. Доля.* – Харків : ХНАМГ, 2010. – 171 с.
6. *Дугерти К.* Введение в эконометрику : учебник / *К. Дугерти;* пер. с англ. – М. : Инфра-М, 2009. – 465 с.
7. Економіко–математичне моделювання : навчальний посібник / За ред. *О. Т. Іващука.* – Тернопіль : ТНЕУ «Економічна думка», 2008. – 704 с.
8. Економетрика : підручник / За ред. *О. І. Черняка.* – Миколаїв : МНАУ, 2015. – 414 с.
9. *Здрок В. В.* Економетрія : підручник / *В. В. Здрок, Т. Я. Лагоцький.* – К. : Знання, 2014. – 541 с.
10. Інформатика в LINUX-середовищі : навчальний посібник / За ред. *Н. К. Васильєвої.* – Дніпропетровськ : Біла К. О., 2016. – 268 с.
11. *Казарезов А. Я.* Економіко-математичне моделювання : навчальний посібник / *А. Я. Казарезов, О. О. Ципліцька.* – Миколаїв : Вид-во ЧДУ ім. Петра Могили, 2009. – 248 с.
12. *Козьменко О. В.* Економіко-математичні методи та моделі (економетрика) : навчальний посібник / *О. В. Козьменко, О. В. Кузьменко.* – Суми : Університетська книга, 2014. – 406 с.

13. *Кузьмичов А. І.* Економетрія. Моделювання засобами MS Excel : навчальний посібник / *А. І. Кузьмичов, М. Г. Медведєв.* – К. : Ліра–К, 2011. – 212 с.
14. *Назаренко О. М.* Основи економетрики : підручник / *О. М. Назаренко.* – К. : Центр навч. літератури, 2004. – 392 с.
15. *Наконечний С. І.* Економетрія : підручник / *С. І. Наконечний, Т. О. Терещенко.* – К. : КНЕУ, 2006. – 528 с.
16. *Овсянникова М. М.* Компьютерный практикум по эконометрике / *М. М. Овсянникова.* – Глазов : Глазовский инженерно-экономический институт, 2011. – 64 с.
17. Офіційний веб-сайт Головного управління статистики у Дніпропетровській області [Електронний ресурс] // Режим доступу : <http://dnprstat.gov.ua/>
18. Офіційний веб-сайт Державної служби статистики України [Електронний ресурс] // Режим доступу : <http://ukrstat.gov.ua/>
19. *Руська Р. В.* Економетрика : навчальний посібник / *Р. В. Руська.* – Тернопіль : Тайп, 2012. – 224 с.
20. *Симонович С. В.* Информатика. Базовый курс : учебник / *С. В. Симонович.* – СПб. : Питер, 2011. – 640 с.
21. *Хахаев И. А.* Gnumeric: электронная таблица для всех : учебник / *И. А. Хахаев.* – М. : ALT Linux, 2011. – 192 с.
22. Эконометрика : учебник для бакалавров / Под ред. *И. И. Елисеевой.* – М. : Проспект, 2013. – 288 с.
23. Эконометрика : учебник для студентов высших учебных заведений / Под ред. *Н. Ш. Кремера.* – М. : ЮНИТИ, 2013. – 328 с.

Навчальне видання

Васильєва Наталя Костянтинівна

Мироненко Ольга Андріївна

Самарець Наталія Миколаївна

Чорна Ніна Олександрівна

ЕКОНОМЕТРИКА В ЕЛЕКТРОННИХ ТАБЛИЦЯХ

Навчальний посібник

*За загальною редакцією
д-ра екон. наук, проф. Н. К. Васильєвої*

Редактор: М. П. Гончаренко

Комп'ютерний набір та верстка: В. О. Олексенко

Дніпропетровський державний аграрно-економічний університет

49600, м. Дніпро, вул. Сергія Єфремова, 25

Тел.: +38 (056) 713-51-75, 713-51-36

e-mail: redviddday@i.ua, info@dsau.dp.ua

www.dsau.dp.ua, ojs.dsau.dp.ua

Підписано до друку 03.03.2017. Формат 60x84/16.

Обл.-вид. арк. 7,9. Ум. друк. арк. 7,8. Зам. № 0317-03/1.

Тираж 100 пр. Папір офсетний

Видавець та виготовлювач СПД Біла К. О.

Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи ДК № 3618 від 06.11.2009

Надруковано на поліграфічній базі видавця Білої К. О.

49000, м. Дніпро, пр. Д. Яворницького, 111, оф. 2

тел. +38 (067) 972-90-71

www.impact.dp.ua

e-mail: impact.dnepr@gmail.com

