

Методичні вказівки до лабораторних робіт з курсу «Статистика» для студентів напрямку 6.030601 «Менеджмент». **Застосування MS EXCEL для розв’язання задач статистики** / уклад. Ширяєва Н. В., Білоцерківський О. Б. – Х. : НТУ «ХП», 2015. – 37 с.

Укладачі: Н. В. Ширяєва,
О. Б. Білоцерківський

Рецензент Т. В. Данько

Кафедра менеджменту зовнішньоекономічної діяльності та фінансів

ВСТУП

Електронна таблиця *Excel* є універсальним обчислювальним інструментом, знання якого обов'язкове для студентів економічних спеціальностей. Електронні таблиці відкрили нову епоху у програмуванні – так зване «програмування без мови», завдяки чому складні розрахунки стали доступними непрограмістам.

Вивчення проблем статистики в середовищі *Excel* має певні переваги порівняно з використанням спеціалізованих статистичних пакетів, в яких усі розрахунки виконуються автоматично і користувач отримує відразу готовий результат. *Мета методичних вказівок* – вивчити певні розділи навчального курсу «Статистика», показати, як знаходять ті або інші результати, освоїти особливості обчислювальних алгоритмів, довести теоретичні міркування до числа з подальшою інтерпретацією отриманих результатів і економічних висновків.

У даних методичних вказівках розглянуто основні чисельні методи, які використовуються в курсі «Статистика». Кожен розділ присвячений окремій темі курсу, і всі розділи побудовані однаково. Спочатку викладаються необхідні теоретичні відомості, потім докладно розглядаються розв'язання задач. Наприкінці кожного розділу наведено варіанти індивідуальних домашніх завдань, що вибираються за останньою цифрою номера залікової книжки студента, що розраховуються за формулою, яку наведено у главі 1.

Дані методичні вказівки не замінюють підручники зі статистики. Теоретичні основи викладаються у стислому вигляді. Даються тільки ті відомості, які необхідні безпосередньо для розв'язання задач. Як підручники можна використати також роботи [1–15].

1. ЗАГАЛЬНІ ВІДОМОСТІ ПРО СТАТИСТИЧНІ ФУНКЦІЇ В MICROSOFT EXCEL

Microsoft Excel надає широкі можливості для аналізу статистичних даних. При розв'язанні простих завдань у цій програмі завжди під руками вбудовані функції, такі як *СРЗНАЧ()*, *МЕДИАНА()* й *МОДА()*. Якщо ж їх виявляється недостатньо, варто звернутися до «**Пакет аналіза**».

Надбудова «**Пакет аналіза**» є доповненням і містить набір функцій та інструментів, що розширюють вбудовані аналітичні можливості *Excel*, й дозволяє будувати гістограми, робити випадкові або періодичні вибірки даних і знаходити їхні статистичні характеристики, генерувати нерівномірно розподілені випадкові числа, проводити регресійний аналіз, виконувати перетворення Фур'є та ін.

У статистиці сукупність результатів вимірів називають розподілом. *Microsoft Excel* дає можливість аналізувати розподіли, використовуючи вбудовані статистичні функції, функції аналізу вибірок і генеральної сукупності, а також інструменти «*Описательная статистика*», «*Гистограмма*», «*Ранг*» та «*Персентиль*».

1.1. Основні вбудовані статистичні функції

Функція *СРЗНАЧ()* обчислює середнє арифметичне або просто середнє значення для послідовності чисел: підсумовуються числові значення в інтервалі середніх значень і результат ділиться на їх кількість. Ця функція ігнорує порожні, логічні й текстові середні значення.

Функція *МЕДИАНА()* обчислює медіану множини чисел. Медіана – це число, що є серединою множини: кількість чисел, менші й більші медіани, дорівнюють одна одній. Якщо кількість чисел або середніх значень парне, то результатом буде середнє значення двох чисел у середині множини.

Функція *МОДА()* повертає найбільше, що часто зустрічається, значення у множині чисел.

Функція *МАКС()* повертає найбільше значення серед заданих чисел.

Функція *МИН()* повертає мінімальне значення серед заданих чисел.

Функція *СУММПРОИЗВ()* повертає суму добутків відповідних членів двох і більше масивів-аргументів (але не більше 30 аргументів). Нечислові значення, що зустрічаються в аргументах, інтерпретуються нулями.

Функція *СУММКВ()* повертає суму квадратів аргументів.

1.2. Аналіз вибірок і сукупності

Дисперсія й стандартне відхилення – це статистичні характеристики розподілу наборів або генеральної сукупності даних. Стандартне відхилення визначається як квадратний корінь із дисперсії. Як правило, близько 68 % даних генеральної сукупності з нормальним розподілом перебуває в межах одного стандартного відхилення й близько 95 % – у межах двох стандартних відхилень. При великому значенні стандартного відхилення дані широко розкидані щодо середнього значення, а при маленькому – вони групуються близько до середнього значення.

Чотири статистичні функції *ДИСП()*, *ДИСПР()*, *СТАНДОТКЛОН()* і *СТАНДОТКЛОНП()* – призначені для обчислення дисперсії й стандартного відхилення чисел в інтервалі середніх значень. Перед тим як обчислювати дисперсію й стандартне відхилення набору даних, потрібно визначити, чи являють ці дані генеральну сукупність або вибірку з генеральної сукупності. У випадку вибірки з генеральної сукупності варто використовувати функції *ДИСП()* і *СТАНДОТКЛОН()*, а у випадку генеральної сукупності – функції *ДИСПР()* і *СТАНДОТКЛОНП()*.

Функція *СУММСУММКВ()* обчислює суму сум квадратів відповідних елементів у масивах.

Функція *СУММКВРАЗН()* обчислює суму квадратів різниці відповідних елементів у масивах.

1.3. Інструмент аналізу *Описова статистика*

Цей інструмент дає можливість побудувати таблицю параметрів описової статистики для одного або більше наборів вхідних даних. Для кожного набору вхідних даних у вихідному інтервалі будується таблиця з наступною інформацією: «Среднее», «Стандартная ошибка», «Медиана», «Мода», «Стандартное отклонение», «Дисперсия выборки», «Экцесс», «Асимметричность», «Интервал», «Минимум», «Максимум», «Сумма», «Счет», «Наибольший (k)», «Наименьший (k)» (для будь-якого *k*) й «Уровень надежности» (довірчий інтервал). Статистичній обробці піддається один або кілька наборів даних, які розташовано в інтервалі, посилення на який задається в полі «Входной интервал». Перемикач «Группирование» дає можливість уточнити, як розміщуються дані: стовпцями чи рядками. Якщо стовпці або рядки даних мають мітки, то при встановленому прапорці «Метки в первой строке / Метки в первом столбце» вони використовуються як заголовки стовпців статистичних параметрів вихідної таблиці. Адреса верхньо-

го лівого середнього значення для цієї таблиці задається в полі «*Выходной интервал*». При встановленому прапорці «*Итоговая статистика*» створюється докладна вихідна таблиця, та, встановивши відповідні прапорці, можна помістити в неї додаткові дані.

Як і інші інструменти «*Описательная статистика*» створює таблицю параметрів. Якщо потрібна не таблиця, а окремі статистичні характеристики, то їх можна одержати або за допомогою іншого інструмента «**Пакет аналіза**», або за допомогою функцій Microsoft Excel.

1.4. Інструмент «Гістограма»

Гістограми або лінійчаті діаграми – зручний засіб для обробки результатів вимірів. Область значень величини, що вимірюється, розбивається на декілька інтервалів, які також називаються кишнями, в яких у вигляді стовпців відкладається кількість вимірів, що потрапили в цей інтервал, який називається частотою. Кишені не обов'язково повинні бути рівними, але вони повинні розташовуватися за зростанням границь. Взагалі поле «*Интервал карманов*» можна залишити порожнім. *Microsoft Excel* рівномірно розподілить кишені в інтервалі від мінімального до максимального значення вихідних даних. Число кишень буде дорівнювати квадратному кореню з кількості вихідних значень.

Можна виконати до трьох типів аналізу, встановивши відповідні прапорці: «*Парето*» (відсортована гістограма), «*Интегральный процент*» і «*Вывод графика*».

У стовпці «*Частота*» виводиться число вихідних значень, які більше або дорівнюють лівій границі кишні, але менші ніж ліва границя наступної кишні. Останнім значенням стовпця є кількість вихідних значень, що більше або дорівнюють лівій границі останньої кишні. Стовпець інтервалів для кишень дублюється в стовпець «*Карман*». Це зручно, якщо вихідний інтервал задля результатів аналізу задано в іншому місці, а не поруч із інтервалом кишень. Не можна вихідний інтервал поєднати з вихідним інтервалом кишень. Оскільки інтервал кишень копіюється, то його краще заповнити числовими константами, а не формулами. Якщо все-таки потрібно формули, то в них

обов'язково повинні використовуватися абсолютні посилання, інакше результати копіювання можуть виявитися невірними.

При встановленому прапорці *«Парето»* вихідна таблиця сортується за убутванням частоти, а при встановленому прапорці *«Интегральный процент»* у таблицю додається стовпець із процентним внеском кожного кластера, що накопичується.

Інструмент *«Гистограмма»* генерує таблицю числових констант. У тих випадках, коли є необхідним зв'язок з вихідними значеннями, варто використовувати формули з табличною функцією *«ЧАСТОТА (массив данных; массив кишень)»*.

1.5. Інструмент *«Ранг»* і *«Персентиль»*

У надбудові **«Пакет анализа»** є декілька функцій, які виймають інформацію про ранг та персентиль (у вітчизняній літературі зустрічається термін «процентиль»).

Функція *«ПРОЦЕНТРАНГ»* (массив; x ; розрядність) повертає процентне значення елемента, що задано, з безлічі даних. Дана функція може використовуватися для створення таблиці персентилів, яку пов'язано з вихідними даними. Значення цієї таблиці будуть обновлятися при кожній зміні вхідних величин.

Функція *«ПЕРСЕНТИЛЬ»* (массив, k) використовується для визначення елемента у вхідній множині массив із зазначеним рівнем персентиля k , що задається у вигляді десяткового дробу між 0 і 1.

Функція *«КВАРТИЛЬ»* (массив, частина) аналогічна щодо попередньої. Перший аргумент массив задає вхідний інтервал даних, другий аргумент частина може приймати тільки п'ять таких значень: 0 – найменша величина, 1 – величина 25-го персентиля, 2 – медіана (величина 50 персентиля), 3 – величина 75-го персентиля, 4 – найбільша величина. Замість функції *«КВАРТИЛЬ()»* для одержання найменшого й найбільшого значень можна використовувати функцію *«МИН()»* і *«МАКС()»* відповідно, а для одержання

медіани – функцію «МЕДИАНА()». Ці функції обчислюються швидше, ніж функція «КВАРТИЛЬ()», особливо у випадку великих масивів даних.

Функції «НАИМЕНЬШИЙ» (масив, k) та «НАИБОЛЬШИЙ» (масив, k) обчислюють відповідно k -тий найменший і k -тий найбільший елемент у множині даних, що задається аргументом масив.

1.6. Генерування випадкових чисел

Існує вбудована функція «СЛЧИС()», призначена для генерування рівномірно розподілених випадкових чисел в інтервалі від 0 до 1. Інакше кажучи, при кожному звертанні до функції «СЛЧИС()» вона з рівною ймовірністю повертає випадкове число між 0 і 1.

Засоби генерації випадкових чисел з «Пакету аналіза» дають можливість одержувати й нерівномірні розподіли. Ці масиви випадкових чисел можуть використовуватися при розв'язанні завдань методом Монте-Карло. Реалізовано такі типи розподілів: «Равномерное», «Нормальное», «Бернуллі», «Биномиальное», «Пуассона» й «Дискретное». Сьомий тип – «Модельное» дає можливість заповнити числами масив за певним законом.

1.7. Побудова вибірок з генеральної сукупності

Інструмент аналізу «ВИБОРКА» дає можливість витягати підмножини чисел із заданої великої множини (генеральної сукупності) чисел. Із вхідного інтервалу середніх значень можна вибрати певну кількість величин або випадковим образом, або кожне n -те значення й помістити їх у заданий вихідний інтервал.

1.8. Обчислення ковзної середньої

Ковзна середня – це методика прогнозування, що спрощує аналіз тренду (тенденції), шляхом згладжування флуктуацій результатів вимірів за деякий період часу. Ці флуктуації можуть бути викликані випадковим «шумом», тобто побічним результатом методики виміру. Наприклад, результати

виміру росту дитини будуть змінюватися залежно від похибки лінійки та й від того, чи стоїть дитина прямо, чи сутулячись. Флуктуації у вимірах можуть викликатися й іншими умовами. Це вносить систематичну помилку. Наприклад, виторг за місяць може залежати від кількості робочих днів або від того, що провідний продавець перебував у відпустці.

1.9. Лінійна й експоненціальна регресії

В *Excel* є п'ять функцій для лінійної регресії: «*ЛИНЕЙН()*», «*ТЕНДЕНЦИЯ()*», «*ПРЕДСКАЗ()*», «*НАКЛОН()*» і «*СТОШУХ()*»; та дві функції для експонентної регресії: «*ЛГРФПРИБЛ()*» і «*РОСТ()*». Ці функції вводяться як табличні форми й повертають результат у вигляді інтервалу масиву. Кожна з функцій може мати один або декілька аргументів.

Наприклад, якщо потрібно передбачити обсяг продажів при стабільних продажах або продажах, що слабо змінюються, то можна скористатися регресією як засобом прогнозування. Однак можна одержати незадовільні результати в періоди перед святами.

Лінійна регресія дає можливість щонайкраще провести пряму лінію через точки одномірного масиву.

Експоненціальна регресія дає можливість щонайкраще провести експоненціальну криву по точках даних, які змінюються нелінійно. Наприклад, дані про зростання населення майже завжди краще описуються не прямою лінією, а експоненціальною кривою.

Множинна регресія являє аналіз більш ніж одного набору даних, може бути як лінійною, так і експоненціальною. Наприклад, для оцінки вартості будинку в деякому районі необхідно використовувати дані про його площу, число ванних кімнат, розміру ділянки землі й віку будівлі.

Завдання 1.1. Систематизуйте дані про польоти авіакомпанії Pigs & People Airline (табл. 1.1 та 1.2), які відбувалися протягом останніх 50 днів, та:

1) побудуйте частотний розподіл;

2) знайдіть середню, моду та медіану без застосування функцій MS EXCEL;

3) розрахуйте вбудовані статистичні функції; проаналізуйте такі дані, як: а) вибірку, б) генеральну сукупність; застосуйте надбудову «**Описательная статистика**», що наведено у п.1.1–1.3. Порівняйте отримані результати з розділом 2.

Таблиця 1.1 – Дані про кількість пасажирів, що обрали для перельоту авіакомпанію Pigs & People Airline

68	71	77	83	79	83	84	74	82	97
72	74	57	67	69	77	73	78	93	95
50	60	70	66	76	78	81	79	90	83
70	84	59	75	94	80	84	91	101	86
65	72	85	79	71	93	92	102	80	69

Таблиця 1.2 – Дані про кількість миль прольотів для авіакомпанії Pigs & People Airline

569.3	420.4	468.5	443.9	403.7
519.7	518.7	445.3	459.0	373.4
493.7	505.7	453.7	397.1	463.9
618.3	493.3	477.0	380.0	423.7
391.0	553.5	513.7	330.0	419.8
370.7	544.1	470.0	361.9	483.8
405.7	550.6	504.6	343.3	497.9
453.3	604.3	473.3	393.9	478.4
437.9	320.4	473.3	359.3	568.2
450.0	413.4	469.3	383.7	469.1

Завдання 1.2. За даними про обсяг капіталу й прибутку комерційних банків, що наведено в таблиці 1.3, складіть:

1) комбінаційне угруповання банків за даними ознаками, утворивши по 3 групи з рівними інтервалами;

2) знайдіть середню, моду та медіану без застосування функцій MS Excel;

3) розрахуйте вбудовані статистичні функції; проаналізуйте такі дані, як: а) вибірку, б) генеральну сукупність; застосуйте надбудову «**Описательная статистика**», що наведено у п.1.1–1.3. Порівняйте отримані результати з розділом 2.

4) повторіть дії п.1–3, помноживши числові дані на коефіцієнт:

$$k = \frac{100 + N}{100}, \quad (1.1)$$

де N – номер прізвища студента в журналі групи.

Таблиця 1.3 – Розподіл комерційних банків за розміром капіталу та прибутку

№ з/п	Розмір капіталу, млн грн	Прибуток, млн грн	№ з/п	Розмір капіталу, млн грн	Прибуток, млн грн
1	2	3	1	2	3
1	6,2	4,6	14	6,2	4,7
2	11,9	8,5	15	8,6	7,2
3	7,6	5,3	16	5,4	4,0
4	10,6	8,8	17	7,0	5,8
5	8,1	6,2	18	9,6	7,8
6	8,3	4,1	19	8,1	6,9
7	12,0	8,2	20	5,2	4,3
8	5,1	3,6	21	7,3	6,0
9	7,8	4,1	22	8,2	6,4
10	5,4	3,3	23	5,4	4,1
11	6,4	5,2	24	3,1	2,7
12	8,3	5,8	25	4,4	3,0
13	5,2	3,3	26	3,0	2,2

2. ГРАФІЧНИЙ АНАЛІЗ СТАТИСТИЧНИХ ДАНИХ У MICROSOFT EXCEL

2.1. Загальні відомості

У статистиці й економіці візуалізація даних використовується найбільш широко для відображення інформації. Основною властивістю візуалізації є її наочність, а людина одержує приблизно 80 % знань про навколишній світ за допомогою органів зору.

Поле діаграми – це простір, у якому розміщаються геометричні або інші графічні знаки, що утворюють діаграму. Границі поля діаграми характеризуються її розміром і співвідношенням сторін. Розмір поля залежить від призначення діаграми, наприклад, для демонстрації на виставці або для наукового звіту. Пропорції сторін повинні не тільки задовольняти естетичним вимогам, але й сприяти найкращому зоровому сприйняттю статистичних даних, що зображуються. На практиці, як правило, застосовують діаграми зі співвідношенням сторін від 1:1,30 до 1:1,50. Найзручнішим є співвідношення сторін «золотий перетин» $1:\sqrt{2}$. Саме таке співвідношення сторін має стандартний аркуш паперу для письма формату А4.

В *Excel* можна побудувати дев'ять типів двовимірних і шість типів тривимірних діаграм, кожна з яких має декілька різновидів. Можна також комбінувати основні типи діаграм. Наприклад, на одній діаграмі можна накласти графік зміни вартості товару на графік зміни обсягів його продажів. Можливе створення «діаграм-картинок», в яких замість звичайних ліній і стовпців використовуються найрізноманітніші графічні образи. Все це в сполученні з користувальницькими форматами дає практично необмежені можливості для побудови діаграм.

Діаграма називається *впровадженою*, якщо розташовується на одному аркуші з вихідними даними.

Ряди даних – це набори значень, які потрібно зобразити на діаграмі. Кожен використовуваний у діаграмі ряд даних не може мати більше 4000

значень. На одній діаграмі може бути зображене до 255 рядів даних, але при цьому загальна кількість крапок даних не може перевищувати 32000.

Категорії служать для впорядкування значень в рядах даних. Наприклад, при розгляді доходу за минулі десять років категоріями є роки, при побудові графіка розподілу прибутку відділами компанії категоріями є назви або коди відділів.

Якщо виникає ускладнення, що вважати рядом, а що – категорією, то застосовується таке правило: ряди даних – це ті значення, які потрібно вивести на діаграмі, а категорії – це «заголовки», над якими ці значення відкладаються. Категорії являють собою значення аргументу, а ряди даних – відповідні значення функції.

2.2. Впроваджені діаграми

Для побудови впровадженої діаграми досить виділити потрібні дані й натиснути кнопку «*Мастер диаграмм*» на стандартній панелі інструментів або відкрити підменю «*Диаграмма*» з меню «*Вставка*». Все, що потрібно зробити далі, – вказати місце впровадження діаграми, її розмір і задати бажані параметри діаграми в діалогових вікнах, що послідовно відкриваються. Для зміни розмірів діаграми виділіть її, а потім перетягніть один з розташованих по її периметру маркерів у потрібному напрямку. Для видалення впровадженої діаграми необхідно спочатку виділити її, а потім виконати команду «*Все*» з підменю «*Очистить*» меню «*Правка*» або натиснути клавішу *Del*. Якщо потрібно побудувати діаграму для двох чи більше несуміжних інтервалів, то спочатку створюється структура даних, приховуються рядки і стовпці, що заважають, а потім будується діаграма.

При виділенні даних для діаграми бажано поряд з рядками даних виділити також назви категорій і назви самих рядів. Однак, якщо діаграма будується за єдиним стовпцем або рядком, заголовок не слід включати у виділений інтервал, тому що в цьому випадку *Майстер диаграм* може не відрізнити заголовок від самих даних.

Аналіз виділених даних в *MS Excel* заснований на тому, що кількість рядків повинна бути менше, ніж кількість категорій. Якщо діаграма будується для інтервалу середніх значень, що має більше стовпців, ніж рядків (або рівно), то рядами даних вважаються рядки. Якщо інтервал середніх значень має більше рядків, ніж стовпців, то рядами даних вважаються стовпці.

Щоб заголовки змінювалися відповідно до зміни вихідних даних, необхідно зв'язати їх з відповідними середніми значеннями після створення діаграми.

Найшвидший спосіб побудови діаграми – це виділити дані й натиснути клавішу *F11*. На окремому аркуші буде побудована діаграма відповідно до прийнятого за замовчуванням типу і форматом.

2.3. Редагування впроваджених діаграм

Діаграма *Microsoft Excel* складається з таких класів елементів, кожен з яких можна виділити та відредагувати: область діаграми, область побудови, основа (тільки для об'ємних діаграм), стінки (тільки для об'ємних діаграм), легенда, осі, текст, стрілки, лінії сітки, перший ряд даних, другий і наступні ряди даних, лінії проєкції, лінії мінімакс рівнів, плюс-маркери, мінус-маркери, лінії рядів.

Елемент, що форматується, виділяється, як правило, натисканням лівої кнопки миші. Щоб одночасно виділити елемент діаграми і відкрити діалогове вікно форматування, потрібно двічі натиснути ліву кнопку миші на цьому елементі.

Стрілки використовуються для залучення уваги до важливої інформації. Щоб намалювати стрілку, виконуються такі дії. Натиснути кнопку «*Рисование*» на «*Стандартной*» панелі інструментів, а потім кнопку «*Стрелка*» на панелі інструментів «*Рисование*». Встановити курсор миші в тому місці екрану, де повинна починатися стрілка, та натиснути ліву кнопку миші, не відпускаючи її. Перетягнути курсор миші в те місце екрану, де стрілка повинна закінчуватися, й відпустити кнопку.

Завдання 2.1. Для вихідних даних (довільних або з попередньої лабораторної роботи) необхідно побудувати кільцеву діаграму, зовнішнє кільце якої показує обсяг продажів деякої фірми по місяцях, перше внутрішнє – за квартали, друге внутрішнє – за півріччя, центр, або наступне кільце, – за рік. Підпишіть назви періодів і частку кожного періоду. Стрілками покажіть найбільше й найменше значення обсягів продажів для помісячної розбивки.

Експеримент за методом Монте-Карло

Очевидно, ніхто точно не знає, чому експеримент за методом Монте-Карло називається саме так. Можливо, ця назва має якесь відношення до відомого казино як символу дії законів випадковості.

У реальній ситуації справжні значення α й β невідомі, а, отже, невідомо, гарні чи погані оцінки дає регресійний аналіз. Експеримент за методом Монте-Карло – це штучний контрольований експеримент, що дає можливість такої перевірки. Найпростіший експеримент за методом Монте-Карло складається із трьох частин.

По-перше: 1) вибираються справжні значення α й β ; 2) задається вектор значень x і число спостережень n , розраховується точне значення $\alpha + \beta x$; 3) використовується генерація випадкових чисел за певним законом для одержання випадкового фактора u в кожному зі спостережень.

По-друге, у кожному спостереженні генерується значення y на основі α, β, x та u .

По-третє, застосовується регресійний аналіз для оцінювання параметрів a і b з використанням отриманих вищевказаним образом значень y для відповідних значень x . Це дозволить бачити, чи є знайдені a і b хорошими оцінками α й β , а також дозволить відчутти придатність методу побудови

регресії. Різницю між справжніми значеннями й оцінкою буде обумовлено впливом випадкового члена u .

Завдання 2.2. Проведіть перші дві частини експерименту за методом Монте-Карло для функції $y = \alpha + 1,5x + u$, де α дорівнює номеру варіанта, а u нормально розподілена випадкова величина з нульовим середнім і одиничною дисперсією. $n = 20, x_1 = 11, 1 \leq x_{i+1} - x_i \leq 1,5$ для $\forall i$. Розрахуйте за допомогою вбудованої функції *KORPEL()* коефіцієнт кореляції між точним значенням функції й значенням функції з помилкою. Отримане значення повинне бути не менш 0,97. Для виконання завдання згідно зі своїм номером варіанта, необхідно помножити числові дані на коефіцієнт, який наведено у формулі (1.1).

3. СТАТИСТИЧНЕ ВИВЧЕННЯ ВЗАЄМОЗВ'ЯЗКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ЯВИЩ

Ціль роботи: розв'язання найпростіших завдань парної та множинної регресії за допомогою *MS Excel*.

Завдання 3.1. Опишіть регресійною залежністю такі дані.

Щороку дев'ятикласники південного заходу штату Огайо (США) здають кваліфікаційний тест. У таблиці 3.1 наведено дані про 47 шкіл цього штату за 2008/09 навчальний рік. Позначення: *School District* – найменування району школи; *Percentage Passing* – кількість студентів, що пройшли тест (у відсотках); *Percentage Attendance* – середня кількість присутніх школярів у день (у відсотках); *Salary* – середня зарплата вчителя (у доларах США); *Spending* - витрати на навчання учня (у доларах США).

Порядок виконання роботи

1. Побудуйте точкову залежність кількості студентів, що здали тест y , від кількості присутніх студентів x .

2. Попередньо ознайомившись із довідкою про вбудовану функцію *ЛІНЕЙН()* визначте за її допомогою коефіцієнти парної регресії a й b для моделі $y = \alpha + \beta x + u$, а також статистику з регресії. Для цього виділіть масив з п'яти рядків і двох стовпців, починаючи з лівого верхнього кута (число стовпців дорівнює числу потрібних коефіцієнтів у рівнянні регресії, число рядків завжди дорівнює п'яти), у який буде поміщено статистику з регресії. Вкажіть аргументи для функції *ЛІНЕЙН()*. Натисніть на кнопку *ОК*, потім на функціональну клавішу *F2*, а далі *SHIFT+CTRL+ENTER*.

3. Поясніть зміст коефіцієнтів парної регресії, виходячи зі специфіки проблеми.

4. Нанесіть на графік, що побудовано у розділі 1, лінію регресії, обчисливши її за формулою $\hat{y} = a + bx$. Якщо редагування графіка викликає труднощі, побудуйте все заново. При обчисленні вектора \hat{y} перейдіть до абсолютної адресації за a й b , поставивши перед змінюваною частиною адреси посилання символ \$, або скористайтеся функціональною клавішею $F4$.

5. Розрахуйте стандартну помилку оцінки $S_{YX} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}}$ та порівняйте її з обчисленою в статистиці з регресії.

6. Розрахуйте коефіцієнт детермінації R^2 , використовуючи статистику з регресії.

7. Розрахуйте коефіцієнт кореляції та інтерпретуйте його. Порівняйте три способи розрахунку коефіцієнта кореляції: 1) як квадратний корінь із коефіцієнта детермінації; 2) як коефіцієнт кореляції між x і y (за допомогою вбудованої функції $KORPEL()$), 3) як коефіцієнт кореляції між y і \hat{y} .

8. Виконайте аналіз залишків e ($e = y - \hat{y}$) та візуально визначте адекватність моделі за гистограмою розподілу залишків, скориставшись інструментом «Гистограмма».

9. Для рівня значущості 0,05 перевірте існування лінійної залежності між незалежною й залежною змінними (як нульова гіпотеза – гіпотеза про дорівненість нулю коефіцієнта кореляції, що еквівалентно гіпотезі $\beta = 0$). t -

статистика має вигляд: $t = \frac{R}{\sqrt{\frac{1-R^2}{n-2}}}$, $t_{крит}$ знайдіть за допомогою вбудованої

функції $СТЬЮДРАСПОБР()$.

10. Знайдіть довірчий інтервал для ймовірності 0,95 коефіцієнта нахилу лінії регресії ($b \pm \sigma_b t_{крит}$) та інтерпретуйте його. Значення σ_b визначене в статистиці з регресії.

11. Повторіть пункти 1–10, розглядаючи середню зарплату вчителя як незалежну змінну.

12. Повторіть пункти 1–10, розглядаючи витрати на навчання як незалежну змінну.

13. Яка із трьох моделей краще прогнозує відсоток школярів, що здали тест? Чому?

14. Визначіть параметри в моделі множинної регресії. Порівняйте коефіцієнт детермінації з кращим з пп.13. Яка з моделей краще? Чому?

15. Повторіть пункти 1–14 за допомогою інструмента «Регресія» надбудови «Пакет аналіза». Порівняйте отримані результати.

16. Повторіть дії п.п. 1–14, помноживши числові дані на коефіцієнт (див. формулу (1.1)).

Таблиця 3.1 – Розрахункові дані

School	Passing, %	Attendance, %	Salary, \$	Spending, \$
Edgewood	70	93,84	33635	2044,09
Fairfield	67	94,76	34697	2681,96
Hamilton	49	92,22	33052	2057,44
Lakota	82	96,07	34924	2314,5
Madison	60	93,93	28835	1952,53
Middletown	42	92,21	35717	2321,55
New Miami	22	93,22	32194	2219,1
Ross	55	94,83	33189	2155,05
Talawanda	66	94,54	32979	2057,39
Batavia	67	94,03	31746	2313,28
Bethel-Tate	59	93,96	33797	1992,31
Clermont	38	91,93	30544	1843,6
Felicity-Franklin	49	94,2	31902	1888,67
Goshen	43	93,05	30073	1879,21
Milford	68	94,42	33068	1954
New Richmond	48	93,06	37632	2629,04
West Clermont	53	92,82	34067	1995,02

Продовження таблиці 3.1

№	School	Passing, %	Attendance, %	Salary, \$	Spending, \$
18	Williamsburg	71	94,16	31014	2102,11
19	Cincinnati	26	88,67	39862	3266,03
20	Deer Park	65	94,4	36463	2499,16
21	Finneytown	78	94,36	36452	2681,05
22	Forest Hill	78	95,93	39902	2604
23	Indian Hill	92	96,13	41301	3366,87
24	Lockland	60	92,08	39080	2533,45
25	Loveland	38	94,93	37824	2386,29
26	Madeira	88	95,66	36595	2752,51
27	Mariemont	83	96,33	39761	3492,31
28	Mount Healthy	40	92,87	38948	2615,28
29	North College Hill	44	94,04	34898	2135,64
30	Northwest	59	93,95	37013	2061,55
31	Norwood	58	92,63	36613	2474,4
32	Oak Hills	72	95,77	38010	2317,03
33	Reading	69	94,4	40142	1319,19
34	Southwest	60	94,5	32970	2115,67
35	St.Bernard-Elmwood Place	58	93,71	38722	2752,77
36	Sycamore	84	95,76	43748	3427,84
37	Three Rivers	45	93,67	39398	2557,77
38	Winton Woods	46	94,25	40320	2635,29
39	Wyoming	93	96,84	38817	3407,06
40	Carlisle	58	94,97	32753	2264,21
41	Franklin	59	93,39	34507	2138,21
42	Kings	66	94,71	32227	1930,2
43	Lebanon	69	94,66	34911	2092,51
44	Little Miami	62	94,72	34775	2157,53
45	Mason	84	95,56	35259	2325,48
46	Springboro	74	95,76	35483	2200,18
47	Wayne	79	95,68	32675	2316,78

4. РОЗВ'ЯЗАННЯ ЗАВДАНЬ МНОЖИННОЇ РЕГРЕСІЇ ЗА ДОПОМОГОЮ MICROSOFT EXCEL

Ціль роботи: розв'язання завдань множинної регресії за допомогою *MS Excel*.

Завдання 4.1. Виберіть змінні та опишіть регресійною залежністю пробіг автомобіля на одному галоні палива.

У таблицях 4.1, 4.2 містяться дані про 106 моделей автомобілів, які випущені в 2000 році. Позначення: *Drive Type (Front, Rear)* – тип авто (передньопривідний, задньопривідний); *Mileage (mpg)* – пробіг на одному галоні палива (миль на галон); *Fuel Type (Regular, Premium)* – тип палива (звичайне, поліпшене); *Capacity (gal)* - обсяг паливного баку (галонів); *Length (inches)* – довжина (дюйми); *Wheel Base* – колісна база; *Width* – ширина; *Door Top* – висота дверей; *Luggage* – місткість багажника; *Front Shoulder* – місце для плечей попереду; *Front Leg Room* – місце для ніг попереду; *Front Head Room* – місце для голови попереду; *Rear Shoulder* – місце для плечей позаду; *Rear Leg Room* – місце для ніг позаду; *Rear Head Room* – місце для голови позаду.

Порядок виконання роботи

1. Перейдіть до фіктивних змінних для якісних значень вихідних даних. Попередньо скопіюйте вихідні дані на новий аркуш.
2. Обчисліть кореляційну матрицю. Для зручності розрахунків доцільно стовпець із значеннями функції розташувати перед стовпцями аргументів.
3. Виберіть у першому рядку кореляційної матриці не більш як три аргументи, що мають найбільші значення коефіцієнтів кореляції зі значеннями функції та найменші значення коефіцієнтів взаємної кореляції.
4. Складіть із обраним набором аргументів всі можливі рівняння регресії (три – з одним аргументом, три – із двома аргументами й одне з трьома аргументами).
5. Виберіть кращі моделі за критерієм максимуму коефіцієнта детермінації.

6. Для кращих моделей розрахуйте середнє значення коефіцієнта варіації.

7. Повторіть п.п.4 для кращих моделей ще раз, використавши тільки 80% вихідних даних.

8. Розрахуйте суму квадратів залишків на перевірній послідовності за тими 20 % вихідними даними, що залишилися.

9. Обчисліть коефіцієнт кореляції для кращих моделей і інтерпретуйте його. Порівняйте два способи обчислення коефіцієнта кореляції: 1) як квадратний корінь з коефіцієнта детермінації; 2) як коефіцієнт кореляції між y і \hat{y} .

10. Сформулюйте нульову гіпотезу й перевірте значущість рівняння в цілому на рівні 0,05 за допомогою F -тесту (розподіл Фішера). $F_{\text{дана}}$ обчисліть за допомогою вбудованої функції « $F_{\text{РАСПОБР}}$ ». Кількість ступенів волі 1 дорівнює кількості коефіцієнтів у рівнянні регресії, кількість ступенів волі 2 дорівнює кількості спостережень без одиниці, зменшеному на кількість ступенів волі 1.

11. Виберіть кращу модель на підставі порівняння коефіцієнтів детермінації, середнього значення коефіцієнта варіації, суми квадратів залишків на перевірній послідовності й значення F -тесту. Обґрунтуйте свій вибір.

12. Повторіть дії п.п.1–11 помноживши числові дані на коефіцієнт (див. формулу (1.1)).

Таблиця 4.1 – Розрахункові дані

Name	Drive Type	Mileage (mpg)	Fuel Type	Capacity (gal)	Length (inches)	Wheel Base (inches)	Width (inches)	Turning Circle (feet)
1	2	3	4	5	6	7	8	9
Acura Integra	Front	31	Regular	13,2	172	101	67	39
Acura TL	Front	22	Premium	17,1	193	108	70	40
Buick Century	Front	22	Regular	17,5	195	109	73	40
Buick LeSabre	Front	20	Regular	17,5	200	112	74	42
Buick Park Avenue	Front	21	Regular	18,5	207	114	75	43
Buick Regal	Front	21	Regular	17,5	196	109	73	40
Chevrolet Cavalier	Front	26	Regular	15,2	181	104	68	38
Chevrolet Impala	Front	20	Regular	17,0	200	111	73	40

Продовження таблиці 4.1

1	2	3	4	5	6	7	8	9
Chevrolet Lumina	Front	22	Regular	16,6	201	108	73	44
Chevrolet Malibu	Front	24	Regular	15,2	190	107	69	40
Chevrolet Prizm	Front	31	Regular	13,2	175	97	67	34
Chevrolet Venture	Front	19	Regular	25,0	201	120	72	42
Chrysler 300M	Front	21	Regular	17,0	198	113	74	40
Chrysler Cirrus	Front	22	Regular	16,0	186	108	72	39
Chrysler Concorde	Front	21	Regular	17,0	209	113	74	41
Chrysler Sebring Convertible	Front	21	Regular	16,0	193	106	70	42
Chrysler Town & Country	Front	18	Regular	20,0	200	119	77	41
Daewoo Lenganza	Front	24	Regular	15,8	184	105	70	41
Daewoo Nubira	Front	24	Regular	13,7	175	101	67	37
Dodge Intrepid	Front	22	Regular	17,0	204	113	75	40
Dodge Stratus	Front	22	Regular	16,0	186	108	72	39
Ford Contour	Front	22	Regular	14,5	185	107	69	40
Ford Escort	Front	28	Regular	12,7	175	98	66	36
Ford Taurus	Front	21	Regular	16,0	198	109	73	42
Ford Windstar	Front	18	Regular	26,0	201	121	77	42
Honda Accord	Front	25	Regular	17,1	189	107	70	40
Honda CR-V	Front	24	Regular	15,3	178	103	69	38
Honda Odyssey	Front	19	Regular	20,0	201	118	76	41
Hyundai Tiburon	Front	27	Regular	14,5	171	97	68	37
Infiniti G20	Front	25	Regular	15,9	178	102	67	41
Infiniti I30	Front	22	Premium	18,5	194	108	70	40
Infiniti QX4	Front	16	Regular	21,1	184	106	72	42
Jeep Wrangler	Front	15	Regular	19,0	152	93	67	36
Kia Sephia	Front	26	Regular	13,2	174	101	67	37
Land Rover Discovery	Front	14	Premium	24,1	185	100	74	47
Lexus ES300	Front	22	Premium	18,5	190	105	71	40
Lexus RX300	Front	19	Regular	17,2	180	103	72	43
Mazda 626	Front	24	Regular	16,9	187	105	69	39
Mazda MPV	Front	18	Regular	18,5	187	112	72	41
Mazda Protégé	Front	27	Regular	13,2	174	103	67	38
Mercedes-Benz M- Class	Front	19	Premium	19,0	181	111	72	35
Mercury Cougar	Front	24	Regular	15,0	185	106	70	39
Mercury Sable	Front	21	Regular	16,0	198	109	73	42
Mercury Villager	Front	19	Regular	20,0	195	112	75	41
Mitsubishi Diamante	Front	20	Premium	19,0	194	107	70	41
Mitsubishi Galant	Front	24	Regular	16,3	188	104	69	40

Продовження таблиці 4.1

1	2	3	4	5	6	7	8	9
Mitsubishi Mirage	Front	27	Regular	13,2	174	98	67	36
Nissan Altima	Front	22	Regular	15,7	186	103	69	40
Nissan Maxima	Front	21	Premium	18,5	191	108	70	39
Nissan Quest	Front	19	Regular	20,0	195	112	75	41
Oldsmobile Alero	Front	23	Regular	15,0	187	107	70	39
Oldsmobile Intrigue	Front	21	Regular	17,0	196	109	74	40
Oldsmobile Silhouette	Front	19	Regular	25,0	201	120	72	42
Pontiac Bonneville	Front	20	Regular	18,5	203	112	74	43
Pontiac Grand Am	Front	23	Regular	15,0	186	107	70	39
Pontiac Grand Prix	Front	21	Regular	18,0	197	111	73	40
Pontiac Montana	Front	18	Regular	25,0	201	120	73	43
Pontiac Sunfire	Front	26	Regular	15,2	182	104	68	38
Saab 9-3	Front	22	Regular	17,0	182	103	67	36
Saab 9-5	Front	23	Regular	19,8	189	106	71	38
Saturn L-Series	Front	23	Regular	13,1	190	107	69	39
Saturn S-Series	Front	29	Regular	12,1	177	102	67	41
Saturn SC	Front	28	Regular	12,1	180	102	68	40
Subaru Forester	Front	22	Regular	15,9	175	99	68	39
Subaru Impreza	Front	23	Regular	15,9	172	99	67	36
Subaru Legacy	Front	22	Regular	16,9	184	104	69	39
Toyota Avalon	Front	21	Regular	18,5	192	107	72	40
Toyota Camry	Front	25	Regular	18,5	189	105	70	40
Toyota Camry Solara	Front	22	Regular	18,5	190	105	71	40
Toyota Corolla	Front	30	Regular	13,2	174	97	67	34
Toyota Land Cruiser	Front	14	Premium	25,4	193	112	76	42
Toyota RAV4	Front	22	Regular	15,3	163	95	67	38
Toyota Sienna	Front	19	Regular	21,0	194	114	73	44
Volkswagen Jetta	Front	21	Regular	14,5	172	99	68	36
Volkswagen New Beetle	Front	29	Regular	14,5	161	99	68	36
Volkswagen Passat	Front	24	Premium	16,4	184	106	69	38
Volvo S40/V40	Front	22	Regular	15,8	176	100	68	38
Volvo S70/V70	Front	21	Regular	18,5	186	105	69	38
Volvo S80	Front	21	Regular	21,1	190	110	72	40
Audi A4	Front	22	Premium	15,9	178	103	68	36
Audi A6	Front	20	Premium	18,5	192	109	71	39
BMW 3-Series	Rear	24	Premium	16,6	176	107	69	36

Продовження таблиці 4.1

1	2	3	4	5	6	7	8	9
BMW 5-Series	Rear	20	Premium	18,5	188	111	71	38
Cadillac Catera	Rear	20	Premium	18,0	192	108	70	36
Chevrolet Blazer	Rear	15	Regular	18,0	183	107	68	46
Ford Crown Victoria	Rear	19	Regular	19,0	212	115	78	44
Ford Explorer	Rear	16	Regular	21,0	191	112	70	40
Ford F-150	Rear	15	Regular	25,0	226	139	80	49
Ford Ranger	Rear	18	Regular	20,0	202	126	70	45
GMC Jimmy	Rear	15	Regular	18,0	184	107	68	46
Honda Passport	Rear	18	Regular	21,1	177	106	70	41
Jeep Cherokee	Rear	16	Regular	20,0	168	101	68	38
Jeep Grand Cherokee	Rear	15	Regular	20,5	182	106	72	40
Kia Sportage	Rear	19	Regular	15,8	170	104	68	37
Lexus GS300/GS400	Rear	23	Premium	19,8	189	110	71	39
Lincoln LS	Rear	23	Premium	18,1	194	115	73	39
Lincoln Town Car	Rear	19	Premium	19,0	215	118	78	45
Mercedes-Benz C-Class	Rear	24	Premium	16,4	177	106	68	35
Mercedes-Benz E-Class	Rear	21	Premium	21,1	190	112	71	37
Mercury Grand Marquis	Rear	19	Regular	19,0	212	115	78	44
Mercury Mountaineer	Rear	16	Regular	20,8	190	112	70	41
Suzuki Vitara	Rear	19	Regular	17,4	165	98	70	37
Chevrolet Tracker	Rear	21	Regular	17,4	163	98	67	37
Ford Expedition	Rear	13	Regular	30,0	205	119	79	42
Isuzu Rodeo	Rear	18	Regular	21,1	177	106	70	41
Lincoln Navigator	Rear	13	Regular	30,0	205	119	80	44

Таблиця 4.2 – Розрахункові дані

Weight (lbs)	Door Top (inches)	Luggage (cu. ft.)	Front Shoulder (inches)	Front Leg Room (inches)	Front Head Room (inches)	Rear Shoulder (inches)	Rear Leg Room (inches)	Rear Head Room (inches)
10	11	12	13	14	15	16	17	18
2 635	47,5	13,0	51,5	41,5	2,0	46,5	25,0	0,5
3 460	50,0	14,0	56,0	40,5	2,5	55,0	27,0	2,0
3 350	50,5	17,0	58,0	44,0	5,0	57,0	30,0	2,0

Продовження таблиці 4.2

10	11	12	13	14	15	16	17	18
3 640	51,0	18,0	59,5	40,5	5,0	58,0	27,5	4,5
3 880	51,0	19,0	59,5	42,5	6,0	58,5	31,5	3,0
3 325	50,5	17,0	58,0	43,0	5,0	56,5	29,0	2,5
2 795	50,5	14,0	54,5	41,5	4,5	53,5	28,0	2,0
3 495	51,5	18,0	58,5	42,0	5,0	58,5	29,0	2,5
3 350	51,0	16,0	58,5	42,5	4,5	57,0	30,0	2,0
3 075	50,5	17,0	55,0	43,0	4,5	54,0	30,5	2,0
2 480	48,0	12,0	52,5	40,5	4,0	52,0	26,5	2,0
3 890	62,0	76,0	60,0	41,0	6,0	61,5	30,0	5,5
3 620	49,5	17,0	59,0	40,0	3,0	57,5	30,0	2,5
3 170	49,0	16,0	55,0	42,0	4,0	54,5	31,0	1,5
3 535	49,0	19,0	58,5	42,0	4,0	57,0	34,0	2,0
3 445	48,5	11,0	54,5	41,5	5,0	43,5	27,0	2,0
4 050	60,5	76,0	62,5	41,0	5,5	64,0	29,5	5,0
3 185	50,0	14,0	55,5	42,5	2,5	52,5	30,0	3,5
2 815	50,0	13,0	53,5	41,0	4,5	52,5	27,5	3,0
3 455	49,0	18,0	58,5	43,0	4,5	57,0	32,5	2,5
3 170	49,0	16,0	55,0	42,0	4,0	54,5	31,0	1,5
3 115	50,5	14,0	53,5	41,5	4,5	53,0	28,0	1,0
2 585	49,0	13,0	51,5	40,5	4,0	52,0	27,5	1,5
3 340	50,5	17,0	58,0	41,0	6,0	57,0	29,5	3,0
4 150	61,5	67,5	0,0	40,0	4,5	63,0	30,0	6,0
3 105	50,5	14,0	56,5	40,5	2,0	54,5	30,0	2,5
3 155	59,0	31,0	56,0	39,5	6,0	55,0	27,0	4,5
4 245	62,0	67,0	63,0	40,0	6,5	64,5	29,0	5,0
2 705	46,5	13,0	53,0	41,5	1,5	46,0	25,0	0,0
3 005	50,5	14,0	53,5	41,0	3,0	53,5	26,5	2,0
3 410	51,0	14,0	56,0	40,0	3,0	56,0	29,0	2,0
4 355	60,0	34,5	56,5	40,0	2,5	56,5	28,0	3,0
3 510	64,5	17,0	51,5	41,0	6,5	57,0	25,0	4,5
2 650	50,0	10,0	55,0	41,5	5,0	54,0	29,0	2,5
4 735	65,0	46,5	58,0	39,0	5,5	57,0	27,0	5,0
3 390	50,0	13,0	55,0	41,0	2,0	53,5	27,5	1,5
4 055	59,5	37,5	58,0	42,0	3,0	57,5	31,5	4,5
2 910	49,5	14,0	56,0	41,5	3,5	56,0	30,5	2,5
3 720	60,5	57,0	60,0	39,5	7,0	61,0	26,5	4,5
2 645	50,0	13,0	54,0	41,0	2,5	52,5	27,5	2,0
4 405	64,5	45,5	58,5	41,0	3,5	57,5	29,5	4,5
3 140	47,5	12,0	53,5	40,5	1,5	48,0	26,5	0,5
3 340	50,5	17,0	58,0	41,0	6,0	57,0	29,5	3,0

Продовження таблиці 4.2

10	11	12	13	14	15	16	17	18
3 995	60,0	56,5	63,0	40,0	4,5	64,0	28,5	5,5
3 575	49,5	14,0	56,0	41,0	3,5	57,0	27,0	1,5
3 110	50,0	14,0	54,0	40,0	2,5	54,0	28,0	2,0
2 525	49,0	11,0	53,0	41,0	5,0	52,0	26,0	2,0
3 050	50,5	14,0	55,0	40,5	4,5	54,0	27,0	3,0
3 315	51,0	15,0	57,0	41,5	3,0	56,0	29,5	2,0
3 995	60,0	56,5	63,0	40,0	4,5	64,0	28,5	5,5
3 100	50,0	15,0	53,5	40,5	3,5	50,5	28,0	2,5
3 490	51,0	16,0	58,0	40,5	3,0	56,0	27,0	2,0
3 890	62,0	76,0	60,0	41,0	6,0	6,5	30,0	5,5
3 640	51,0	18,0	58,5	40,5	2,5	57,0	27,0	2,0
3 095	49,5	15,0	53,0	41,0	3,0	51,0	28,0	2,0
3 400	49,5	16,0	58,0	42,5	4,0	56,0	29,5	2,5
4 005	62,5	76,0	60,0	41,0	6,0	61,5	30,5	6,0
2 795	50,5	13,0	54,5	41,5	4,5	53,5	28,0	2,0
3 130	51,5	24,0	52,5	41,0	3,0	52,0	26,0	3,0
3 475	50,5	16,0	56,5	43,5	5,5	56,0	31,0	3,0
3 020	51,0	13,0	55,5	41,5	3,0	55,0	28,5	3,5
2 465	48,5	12,0	52,0	41,0	5,5	52,5	26,5	2,5
2 420	47,5	11,0	52,5	41,5	4,0	50,0	25,5	0,5
3 195	54,5	35,5	54,0	41,5	6,0	54,0	28,0	5,5
2 895	50,5	19,5	52,5	42,0	4,5	51,0	27,0	2,5
3 300	50,0	12,0	54,0	40,5	4,0	52,5	29,0	2,0
3 455	53,0	16,0	59,0	41,0	4,5	58,0	30,5	3,0
3 285	50,5	14,0	57,0	41,0	5,0	56,5	28,0	2,0
3 285	50,5	14,0	55,5	41,0	4,5	52,5	27,5	1,5
2 530	49,0	12,0	53,0	40,0	4,0	52,0	24,5	1,0
5 435	66,5	50,0	61,5	40,0	3,5	61,0	29,0	4,5
3 000	59,0	28,0	53,0	39,5	3,5	52,5	24,5	4,5
3 990	60,5	63,5	60,0	41,5	6,5	62,0	33,0	4,5
2 940	51,0	13,0	54,0	41,0	4,0	52,5	25,5	1,5
2 780	52,5	12,0	52,0	42,5	8,0	48,0	27,5	0,5
3 175	51,0	15,0	55,5	42,5	6,5	53,5	29,0	3,0
2 960	50,0	13,0	54,5	41,0	4,5	53,5	26,5	2,5
3 305	50,0	15,0	56,0	42,5	3,0	56,5	30,0	3,0
3 630	51,5	14,0	58,5	41,5	3,5	56,5	28,5	4,0
3 345	50,5	14,0	54,0	40,5	3,5	52,0	24,5	2,0
3 785	51,0	15,0	55,0	42,0	4,0	55,0	30,0	3,5
3 265	50,0	11,0	55,0	41,5	3,5	54,0	26,0	3,0
3 585	51,0	11,0	57,0	43,0	3,5	55,0	29,0	3,5
3 770	51,0	14,0	55,5	41,0	5,0	55,5	30,5	3,5

Продовження таблиці 4.2

10	11	12	13	14	15	16	17	18
4 225	60,5	40,0	56,5	42,5	4,5	57,0	30,0	4,0
3 985	52,0	21,0	61,0	42,0	5,0	60,0	29,5	4,0
4 325	62,0	52,0	56,5	43,0	5,0	56,5	30,5	4,5
5 210	70,5	NA	63,5	42,0	7,0	60,0	24,5	2,5
3 870	60,0	NA	54,0	42,5	5,0	18,0	22,5	2,0
4 225	60,5	40,0	56,5	42,5	4,5	57,0	30,0	4,0
3 935	60,0	39,5	56,0	41,0	4,0	56,0	30,0	3,5
3 540	58,5	36,5	54,5	40,5	3,5	54,5	26,0	4,0
4 190	60,5	37,0	58,0	40,0	3,5	57,0	27,0	2,5
3 365	60,5	30,0	53,5	40,0	4,0	54,0	25,0	3,0
3 745	51,0	15,0	58,0	42,5	2,5	56,0	29,0	2,0
3 655	50,0	14,0	58,0	41,0	4,5	57,0	29,0	3,0
4 050	51,5	21,0	61,0	42,5	5,0	60,0	32,0	3,5
3 320	51,0	13,0	55,0	43,0	3,0	54,0	27,5	3,0
3 930	53,0	15,0	56,5	42,0	4,5	56,0	28,0	3,5
3 985	52,0	21,0	61,0	42,0	5,0	60,0	29,5	4,0
4 440	61,5	50,5	56,0	41,5	3,0	56,5	29,0	5,0
3 230	60,5	28,0	53,0	39,0	4,5	52,5	24,0	4,5
3 055	60,0	28,0	53,0	39,0	4,5	52,5	24,0	4,5
5 290	69,5	70,0	63,5	42,0	5,5	63,5	33,5	5,5
3 935	60,0	39,5	56,0	41,0	4,0	56,0	30,0	3,5
5 850	68,0	67,0	63,5	41,5	6,0	63,5	32,0	5,5

5. АНАЛІЗ РЯДІВ ДИНАМІКИ ЗА ДОПОМОГОЮ MICROSOFT EXCEL

Ціль роботи: аналіз рядів динаміки за допомогою *MS Excel*.

Завдання 5.1. Оцініть дохід компанії Філіп Морріс у 2000 і 2001 роках, опираючись на дані про дохід цієї компанії за попередні 25 років.

У таблиці 5.1 містяться дані про доходи компанії з 1975 по 1999 роки в млрд доларів. Позначення: *Year* – рік; *Code Yr* – змінна, яка може використовуватися в якості незалежної; *Revenues* – дохід.

Порядок виконання роботи

1. Побудуйте лінійний тренд виду $\hat{y} = a + bx$.
2. Побудуйте квадратичний тренд виду $\hat{y} = a + b_1x + b_2x^2$.
3. Побудуйте експонентний тренд виду $\hat{y} = ab^x$.
4. Побудуйте авторегресійну модель третього порядку й перевірте її значущість на рівні 0,05: $y_i = a + b_1y_{i-1} + b_2y_{i-2} + b_3y_{i-3}$. $H_0 : \beta_3 = 0$, $H_1 : \beta_3 \neq 0$.
 $t = (b_3 - \beta_3) / \sigma_{\beta_3}$.
5. Якщо це необхідно, побудуйте авторегресійну модель другого порядку й перевірте її значущість на рівні 0,05: $y_i = a + b_1y_{i-1} + b_2y_{i-2}$.
6. Якщо це необхідно, побудуйте авторегресійну модель першого порядку й перевірте її значущість на рівні 0,05: $y_i = a + b_1y_{i-1}$.
7. Виберіть кращу модель та за її допомогою оцініть дохід компанії в 2000 і 2001 роках.
8. Побудуйте в тих самих координатних осях графіки вихідних даних і отриманої регресійної залежності.
9. Повторіть дії п.п.1 – 3, помноживши числові дані на коефіцієнт.

Таблиця 5.1 - Розрахункові дані

Year	CodeYr	Revenues	Year	CodeYr	Revenues
1975	0	3,6	1988	13	31,7
1976	1	4,3	1989	14	44,8

Продовження таблиці 5.1

Year	CodeYr	Revenues	Year	CodeYr	Revenues
1977	2	5,2	1990	15	51,3
1978	3	6,6	1991	16	56,5
1979	4	8,1	1992	17	59,1
1980	5	9,6	1993	18	60,9
1981	6	10,7	1994	19	65,1
1982	7	11,6	1995	20	66,1
1983	8	13	1996	21	69,2
1984	9	13,8	1997	22	72
1985	10	16	1998	23	74,4
1986	11	25,9	1999	24	78,6
1987	12	28,2			

6. АНАЛІЗ ТРЕНДУ, ЦИКЛІЧНОЇ Й СЕЗОННОЇ СКЛАДОВИХ РЯДУ ДИНАМІКИ ЗА ДОПОМОГОЮ MICROSOFT EXCEL

Ціль роботи: аналіз тренду, циклічної й сезонної складових часового ряду за допомогою *MS Excel*.

Завдання 6.1. Знайдіть тренд, циклічну й сезонну складові часового ряду. У таблиці 6.1 містяться дані про обсяги продажів деякого продукту з 01.01.2003 по 21.04.03 р. у кілограмах.

Порядок виконання роботи

1. Пронумеруйте всі дати одну по одній та за допомогою вбудованої функції «*ДЕНЬНЕД()*» визначте день тижня для кожної дати.
2. Усередніть вихідні дані за кожні сім днів методом ковзної середньої.
3. Побудуйте тренд оптимальної складності залежності усереднених обсягів продажів від числа днів з початку року, обчисліть його значення й залишки.
4. Нанесіть на графік вихідні дані й тренд.
5. Знайдіть усереднене значення циклічної компоненти для кожного з днів тижня, починаючи з понеділка. Для формування вибірки за днями тижня скористайтеся вбудованою функцією «*ИНДЕКС()*».
6. Зменшіть залишки на величину циклічної компоненти й побудуйте графік залишків.
7. Підберіть модель для опису сезонної компоненти й розрахуйте її параметри. Якщо це необхідно, розділіть вихідні дані на декілька інтервалів і проведіть розрахунки для кожного з інтервалів.
8. Елімініуйте вплив сезонної компоненти й візуально оцініть автокореляцію залишків.
9. Розрахуйте значення критерію Дарбіна–Уотсона за формулою

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

10. Порівняйте середні значення й стандартні відхилення залишків без виділення сезонної й циклічної компонент та з їх виділенням.

11. Повторіть дії п.п.1–3, помноживши числові дані на коефіцієнт свого варіанту.

Таблиця 6.1 – Розрахункові дані

Дата	Обсяг продажів, кг	Дата	Обсяг продажів, кг	Дата	Обсяг продажів, кг	Дата	Обсяг продажів, кг
01.01	0	20.01	543	08.02	53	27.02	149
02.01	0	21.01	1 101	09.02	0	28.02	150
03.01	0	22.01	334	10.02	280	01.03	82
04.01	123	23.01	278	11.02	612	02.03	0
05.01	377	24.01	340	12.02	383	03.03	414
06.01	362	25.01	17	13.02	359	04.03	335
07.01	0	26.01	0	14.02	139	05.03	487
08.01	372	27.01	295	15.02	197	06.03	343
09.01	178	28.01	386	16.02	0	07.03	5
10.01	411	29.01	402	17.02	314	08.03	0
11.01	53	30.01	366	18.02	494	09.03	0
12.01	0	31.01	223	19.02	340	10.03	148
13.01	336	01.02	23	20.02	392	11.03	193
14.01	456	02.02	0	21.02	302	12.03	15
15.01	414	03.02	233	22.02	45	13.03	0
16.01	602	04.02	312	23.02	0	14.03	28
17.01	285	05.02	457	24.02	123	15.03	63
18.01	124	06.02	441	25.02	196	16.03	0
19.01	0	07.02	298	26.02	279	17.03	180

Продовження табл. 6.1

Дата	Обсяг продажів, кг	Дата	Обсяг продажів, кг	Дата	Обсяг продажів, кг	Дата	Обсяг продажів, кг
18.03	23	27.03	318	05.04	41	14.04	404
19.03	57	28.03	234	06.04	0	15.04	300
20.03	208	29.03	151	07.04	178	16.04	182
21.03	352	30.03	5	08.04	322	17.04	217
22.03	66	31.03	312	09.04	133	18.04	501
23.03	0	01.04	348	10.04	245	19.04	119
24.03	206	02.04	303	11.04	267	20.04	302
25.03	255	03.04	541	12.04	70	21.04	362
26.03	344	04.04	392	13.04	0		

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Поясніть зміст поняття статистичного спостереження.
2. Які вимоги ставлять до статистичного спостереження?
3. Які існують форми статистичного спостереження?
4. Назвіть види статистичного спостереження.
5. Охарактеризуйте способи статистичного спостереження.
6. Які бувають помилки статистичного спостереження?
7. Охарактеризуйте логічний та арифметичний контроль даних.
8. Розкрийте поняття статистичного зведення та назвіть його види.
9. Поясніть зміст поняття статистичного групування та наведіть його види.
10. Яка суть поняття інтервалу групування та які існують його види?
11. Розкрийте поняття ряду розподілу та назвіть його елементи.
12. Наведіть класифікацію варіаційних рядів.
13. Поясніть зміст поняття та наведіть види статистичних показників.
14. Що таке абсолютні величини?
15. Що таке відносні величини?
16. Що таке середні величини?

17. Охарактеризуйте середню гармонічну та наведіть формули для її обчислення.
18. Охарактеризуйте середню квадратичну та подайте формули для її обчислення.
19. Охарактеризуйте середню геометричну та наведіть формули для її обчислення.
20. Що таке мода та медіана ряду розподілу?
21. Поясніть суть варіації.
22. Які існують види основних показників варіації?
23. Наведіть правило додавання дисперсій та складові загальної дисперсії.
24. Поясніть зміст поняття вибіркового спостереження.
25. Які існують види вибірки в статистичних дослідженнях?
26. Назвіть узагальнюючі характеристики в генеральній і вибірковій сукупностях.
27. Розкрийте суть поняття факторних та результативних ознак.
28. Які є види зв'язків між ознаками?
29. Охарактеризуйте поняття кореляційно-регресійного аналізу.
30. Охарактеризуйте поняття парної та множинної регресії, наведіть їх рівняння.
31. Поясніть суть методу найменших квадратів.
32. Опишіть використання методу найменших квадратів для визначення параметрів лінійної парної регресії.
33. Охарактеризуйте коефіцієнт кореляції та його властивості.
34. Наведіть поняття, елементи та види рядів динаміки.
35. Розкрийте поняття аналітичних показників рядів динаміки та їх види.
36. Поясніть суть показника абсолютного приросту, темпу зростання, темпу приросту та коефіцієнтів прискорення.
37. Які існують види середніх показників?
38. Які є методи обчислення середніх рівнів динамічних рядів?
39. Назвіть середні аналітичні показники.
40. Опишіть способи і методи вирівнювання рядів динаміки.

СПИСОК ЛІТЕРАТУРИ

1. Герасименко С. С. Статистика : підручник / С. С. Герасименко, А. В. Головач та ін. – К. : КНЕУ, 2000. – 468 с.
2. Вашків П. Г. Теорія статистики : навч. посіб. / П. Г. Вашків, П. І. Пастер, В. П. Сторожук та ін. – К. : Либідь, 2001. – 320с.
3. Захожай В.Б. Статистика: підручник / В. Б. Захожай, І. І. Попов. – К. : МАУП, 2006. – 536 с.
4. Попов І. І. Теорія статистики. Практикум : навч. посіб. / І. І. Попов. – К. : КНТЕУ, 2006. – 290 с.
5. Статистика : учеб. пособ. / под ред. М. Р. Ефимовой. – М. : ИНФРА – М., 2002.
6. Мармоза А. Т. Практикум з теорії статистики : навч. посіб. / А. Т. Мармоза. – К. : Ельга-Ніка Центр, 2003. – 344с.
7. Елисеева И. И. Общая теория статистики: учеб. / И. И. Елисеева, М. М. Юзбашев. – М. : Финансы и статистика, 2004. – 544 с.
8. Общая теория статистики. Статистическая методология в изучении коммерческой деятельности / под ред. О. Э. Башиной, А. А. Спирина. – М. : Финансы и статистика, 2000. – 440 с.
9. Єріна А. М. Теорія статистики : практикум / А. М Єріна, З. О. Пальян. – К. : Знання, 2004. – 25 с.
10. Попов І. І. Статистика : практикум /І. І. Попов. – К. : КНТЕУ, 2009. – 149 с.
11. Практикум по теории статистики : учеб. пособ. / под ред. Р. А. Шмойловой. – М. : Финансы и статистика, 2002. – 416 с.
12. Сигел Э. Практическая бизнес-статистика : пер. с англ. / Э. Сигел. – М. : Вильямс, 2002. – 1021 с.
13. Статистика : структурно-логічні схеми та задачі : навч. посіб. – КНЕУ, 2007. – 304 с.
14. Тарасенко І. О. Статистика : навч. посіб. / І. О. Тарасенко. – К. : Центр навч. л-ри, 2006. – 344 с.
15. Опря А. Т. Статистика : навч. посіб. / А. Т. Опря. – К. : Центр навч. л-ри, 2005. – 472 с.

ЗМІСТ

Вступ.....	3
1. Загальні відомості про статистичні функції в <i>Microsoft Excel</i>	4
2. Графічний аналіз статистичних даних у <i>Microsoft Excel</i>	12
3. Статистичне вивчення взаємозв'язку соціально-економічних явищ..	17
4. Розв'язання завдань множинної регресії за допомогою <i>Microsoft Excel</i>	21
5. Аналіз рядів динаміки за допомогою <i>Microsoft Excel</i>	29
6. Аналіз тренду, циклічної й сезонної складових ряду динаміки за допомогою <i>Microsoft Excel</i>	31
Контрольні запитання.....	33
Список літератури.....	35