

Теми практичних занять

№ з/п	Назва теми	Кількість годин	
		ДФН	ЗФН
1	Статистичні методи в географії: основні поняття і терміни	4	
2	Статистичне дослідження і статистичне спостереження. Усвідомлення географічної та математичної сутності поставленого завдання	4	
3	Статистичні показники: поняття, форми, види. Абсолютні і відносні показники	4	4
4	Середні статистичні показники. Мода і медіана	4	
5	Статистична база даних. Матриці даних	4	
6	Побудова одновимірних статистичних моделей	4	2
7	Побудова двовимірних статистичних моделей. Кореляційний статистичний аналіз	4	2
8	Побудова двовимірних статистичних моделей. Регресійний статистичний аналіз	4	2
9	Побудова багатовимірних статистичних моделей. Кластер-аналіз вихідних даних. Інтерпретація отриманих результатів	4	2
10	Побудова багатовимірних статистичних моделей. Факторний аналіз вихідних даних. Інтерпретація отриманих результатів	4	
11	Використання теорії графів в географічних дослідженнях	4	
12	Прогнозування географічних процесів	4	
	Разом	48	12

Практичні роботи виконуються за індивідуальними завданнями для кожного студента. Індивідуальність завдань забезпечується тим, що студенти працюють з власною інформаційною базою даних з обраної теми дослідження.

Практичне заняття 1-2
СТАТИСТИЧНІ МЕТОДИ В СУСПІЛЬНІЙ ГЕОГРАФІЇ:
ОСНОВНІ ПОНЯТТЯ І ТЕРМІНИ

Питання для обговорення:

1. Сутність і співвідношення понять:
методи, методика і методологія;
статистика і геостатистика;
статистичні методи, математичні методи і математико-статистичні методи;
модель, математична модель, статистична модель;
моделювання, географічне моделювання, математичне моделювання, статистичне моделювання;
географічна інформація (дані) і статистична інформація (дані);
статистичне дослідження і статистичне спостереження.
2. Класифікація статистичних методів.
3. Основні джерела географічної інформації.
4. Класифікація географічної інформації.

Завдання до практичної роботи:

Завдання 1. Побудуйте схеми, на яких відобразить, як співвідносяться поняття:

методи, методика і методологія;
статистика і геостатистика;
статистичні методи, математичні методи і математико-статистичні методи;
модель, математична модель, статистична модель;
моделювання, географічне моделювання, математичне моделювання, статистичне моделювання;
географічна інформація (дані) і статистична інформація (дані);
статистичне дослідження і статистичне спостереження.

На схемах (або нижче як висновок) зробіть відповідні пояснення, які б відображали, сутність понять та чим відрізняються і чим схожі ці поняття.

Завдання 2. Статистичні методи надзвичайно різноманітні. Їх можна класифікувати за різними критеріями. Наведіть приклади класифікації статистичних методів за різними критеріями. Побудуйте відповідні схеми.

Завдання 3. Сучасна людина живе в інформаційному світі. Інформація, в

тому числі географічна, поступає до нас з різних джерел. Візуально покажіть основні джерела географічної інформації (форма візуалізації – довільна).

Завдання 4. Географічну інформацію можна класифікувати за способом її сприйняття людиною, формою подання, суспільним значенням, галуззю застосування тощо. Наведіть приклади класифікації географічної інформації за різними критеріями. Побудуйте відповідні схеми. Наведіть конкретні приклади.

Практичне заняття 3-4
СТАТИСТИЧНЕ ДОСЛІДЖЕННЯ І СТАТИСТИЧНЕ
СПОСТЕРЕЖЕННЯ.
УСВІДОМЛЕННЯ СУСПІЛЬНО-ГЕОГРАФІЧНОЇ ТА
МАТЕМАТИЧНОЇ СУТНОСТІ ПОСТАВЛЕНОГО ЗАВДАННЯ

Питання для обговорення:

1. Статистична інформація і статистичні дані: визначення і сутність понять.
2. Способи отримання статистичної інформації.
3. Статистичне дослідження і статистичне спостереження: сутність і співвідношення понять.
4. Класифікація статистичних досліджень за різними критеріями.
5. Вимоги до статистичних досліджень.
6. Статистичні таблиці, їх види. Підсумок і присудок статистичних таблиць.
7. Етапи проведення статистичного дослідження.

Завдання до практичної роботи:

Завдання 1. Користуючись сайтом Державної служби статистики, побудуйте прості, групові та комбінаційні статистичні таблиці. Визначте підмет і присудок побудованих статистичних таблиць.

Завдання 2. Користуючись сайтом Державної служби статистики, розробіть етапи невеликого статистичного дослідження (*тематика – на вибір студента*):

I етап: 1.1. Сформулювати мету і завдання, об'єкт і предмет дослідження.

1.2. Розробити план і програму дослідження.

План дослідження – це свого роду організація запланованого дослідження, вирішення питань організаційного характеру. При складанні плану дослідження необхідно врахувати наступні питання: визначити місце, де буде проводитися дослідження; терміни проведення роботи на всіх етапах; вид статистичного дослідження; хто буде проводити і керувати дослідженням тощо.

Програма дослідження – це перелік чітко сформульованих питань, на які необхідно отримати відповіді на підставі різних видів робіт. Програма дослідження складається з трьох частин:

а) програма збору (визначення одиниці спостереження, виділення всіх статистичних ознак, що підлягають обліку);

- б) програма розробки (групування статистичних ознак; складання макетів статистичних таблиць);
- в) програма аналізу (визначення показників, які будуть розраховуватися під час дослідження).

II етап: 2.1. Зібрати статистичний матеріал (5-10 показників).

III етап: 3.1. Звести статистичний матеріал в розроблені макети статистичних таблиць.

3.2. Статистично обробити отриманий матеріал відповідно до програми аналізу даного дослідження (*за можливості*).

3.3. Графічно відобразити отримані статистичні результати (*для прикладу*).

IV етап: 4.1. На підставі аналізу отриманих даних зробити відповідні висновки, розробити необхідні практичні рекомендації та провести літературне оформлення.

Практичне заняття 5-6
СТАТИСТИЧНІ ПОКАЗНИКИ:
ПОНЯТТЯ, ФОРМИ, ВИДИ.
АБСОЛЮТНІ І ВІДНОСНІ ПОКАЗНИКИ

Питання для обговорення:

1. Поняття «статистичний показник». Якісний, кількісний, часовий і просторовий зміст статистичних показників.
2. Класифікація статистичних показників за різними критеріями.
3. Абсолютні показники, їх види та форми. Одиниці вимірювання.
4. Відносні показники, їх види та форми.
5. Відносні показники динаміки (темпи зростання). Темпи росту і темпи приросту, розраховані ланцюговим і базисним способами.
6. Відносні показники структури і координації.
7. Відносні показники порівняння. Просторове порівняння і порівняння зі стандартом.
8. Відносні показники планового завдання і виконання плану.
9. Відносні показники інтенсивності.
10. Відносні показники диференціації.
11. Індеси та їх використання в географічних дослідженнях.

Завдання до роботи в аудиторії:

Завдання 1. Користуючись статистичними даними табл. 1, розрахуйте показники динаміки експорту товарів України в 2010-2021 рр. (темпи росту і темпи приросту ланцюговим та базисним способами).

Таблиця 1

Динаміка експорту товарів України в 2010-2021 рр.
(за даними Держслужби статистики України)

Роки*	Експорт товарів, млн. дол. США	Темпи росту, %		Темпи приросту, %	
		ланцюговим способом	базисним способом	ланцюговим способом	базисним способом
2010	51405,2				
2011	68394,2				
2012	68830,4				
2013	63320,7				

2014	53901,7				
2015	38127,1				
2016	36361,7				
2017	43264,7				
2018	47335,0				
2019	50054,6				
2020	49191,8				
2021	68072,3				

* З 2024 року – без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Відносні показники динаміки характеризують напрям та інтенсивність зміни показника за часом і визначаються співвідношенням його значень за два періоди або моменти часу. Відносні показники динаміки називають *темпами зростання*.

Розрізняють темпи росту і темпи приросту:

$$TP_{\Pi} = \Pi_2 : \Pi_1 \times 100 \%,$$

де TP_{Π} – темпи росту показника, Π_2 – обсяг показника досліджуваного періоду (року), Π_1 – обсяг показника попереднього або базового періоду (року). Темпи росту більше 100 % свідчать про позитивну динаміку (зростання), менше 100 % – про негативну динаміку (зменшення).

$$TP_{\Pi} = (\Pi_2 : \Pi_1 \times 100 \%) - 100 \% = TP_{\Pi} - 100 \%$$

де TP_{Π} – темпи приросту показника, Π_2 – обсяг показника досліджуваного періоду (року), Π_1 – обсяг показника попереднього або базового періоду (року). Темпи приросту зі знаком «+» свідчать про позитивну динаміку (зростання), зі знаком «-» – про негативну динаміку (зменшення).

Відносні показники динаміки обчислюють як відношення досліджуваного рівня до рівня, який взято за базу порівняння. Базою порівняння може бути:

- змінний попередній рівень (відносні показники динаміки *зі змінною базою порівняння*, розраховані *ланцюговим способом*);
- постійний віддалений за часом рівень (відносні показники динаміки *з постійною базою порівняння*, розраховані *базисним способом*)

$$ВП_{динамікиланцюгова} = \frac{показник_{наступного\ періоду}}{показник_{попереднього\ періоду}} \cdot 100\%$$

$$ВП_{динамікибазисна} = \frac{показник_{наступного\ періоду}}{показник_{першого\ періоду}} \cdot 100\%$$

Завдання 2. Користуючись статистичними даними табл. 2, розрахуйте показники структури і координації статевої структури населення України в 2010-2022 рр.

Таблиця 2

Динаміка чисельності постійного населення України за статтю в 2010-2022 рр. (на 1 січня поточного року)
(за даними Держслужби статистики України)

Роки*	Чисельність постійного населення, осіб			Показники структури	Показники координації
	обидві статі	чоловіки	жінки		
2010	45782592	21107067	24675525		
2011	45598179	21032616	24565563		
2012	45453282	20976712	24476570		
2013	45372692	20962744	24409948		
2014	45245894	20918288	24327606		
2015	42759661	19787826	22971835		
2016	42590879	19717881	22872998		
2017	42414905	19644580	22770325		
2018	42216766	19558180	22658586		
2019	41983564	19455272	22528292		
2020	41732779	19343440	22389339		
2021	41418717	19195376	22223341		
2022	40997698	19006979	21990719		

* З 2024 року – без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Відносні показники структури характеризують склад, структуру сукупності за тією чи іншою ознакою і показують внесок складових сукупності

до загальної маси. Вони визначаються відношенням розмірів складових частин сукупності до загального підсумку. Скільки складових, стільки відносних показників структури. Визначаються простим чи десятинним дробом або процентом. Сума часток дорівнює 1 або 100 %.

$$\text{Відносний показник структури} = \frac{\text{Показник, що характеризує частину сукупності}}{\text{Показник по усій сукупності загалом}}$$

або

$$\text{Відносний показник структури} = \frac{\text{Показник, що характеризує частину сукупності}}{\text{Показник по усій сукупності загалом}} * 100 \%$$

Відносні показники координації характеризують структурованість сукупності. Вони дають співвідношення різних структурних одиниць тієї самої сукупності і показують, скільки одиниць однієї частини сукупності припадає на 1, 100, 1000 і більше одиниць іншої, взятої за базу порівняння. Наприклад, скільки чоловіків припадає на 1000 жінок або скільки жінок припадає на 1000 чоловіків у загальній чисельності населення, скільки іноземних інвестицій припадає на одиницю власних коштів, скільки жінок-політиків припадає на 1000 чоловіків-політиків у державному управлінському апараті тощо.

$$\text{Відносний показник координації} = \frac{\text{Показник, що характеризує } i - \text{ту частину сукупності}}{\text{Показник, що хар } - \text{є частину сукупності вибрану для бази порівняння}}$$

В якості бази порівняння вибирається та частина, яка має найбільшу питому вагу або є пріоритетною з економічної, соціальної або будь-якої іншої точки зору.

Завдання 3. Користуючись статистичними даними табл. 3, розрахуйте показники порівняння (просторового порівняння, порівняння зі стандартом) для споживання населенням України м'яса та м'ясопродуктів у 2020 р.

Таблиця 3

Обсяги споживання населенням України м'яса та м'ясопродуктів у 2020 р.
(за даними Держслужби статистики України)

Регіони	Споживання населенням, кг	Показники порівняння зі стандартом		Показники просторового порівняння
		з раціональними нормами	з мінімальними нормами	

Раціональні норми споживання	83			
Мінімальні норми споживання	52			
Україна	53,8			
Вінницька	57,5			
Волинська	55,1			
Дніпропетровська	65,6			
Донецька	48,7			
Житомирська	54,2			
Закарпатська	52,2			
Запорізька	56,1			
Івано-Франківська	49,8			
Київська	60,6			
Кіровоградська	57,8			
Луганська	43,5			
Львівська	53,1			
Миколаївська	57,2			
Одеська	50,1			
Полтавська	50,9			
Рівненська	48,5			
Сумська	51,4			
Тернопільська	47,7			
Харківська	53,5			
Херсонська	55,0			
Хмельницька	53,8			
Черкаська	52,1			
Чернівецька	44,4			
Чернігівська	53,0			

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Відносні показники порівняння у звичайному розумінні характеризують порівняння однойменних показників, що стосуються різних об'єктів, взятих за той самий період чи момент часу. Обчислюються у відносних величинах або процентах.

До цього виду відносних показників належать відносні показники просторового порівняння та відносні показники порівняння зі стандартом.

Відносний показник порівняння зі стандартом являє собою порівняння

фактичних значень показників з певним еталоном – стандартом, нормативом, оптимальним рівнем.

Відносний показник просторового порівняння – це відношення розмірів або рівнів однойменних показників за різними територіями чи об'єктами. Найчастіше це регіональні чи міжнародні порівняння показників економічного розвитку або життєвого рівня. Базою порівняння може бути будь-який об'єкт. Головне, щоб методика розрахунку порівнюваних показників була однаковою.

Завдання 4. Користуючись статистичними даними табл. 4, розрахуйте показники планового завдання і виконання плану виробництва зернових і зернобобових культур та урожайності цукрового фабричного буряка в 2021-2022 рр.

Таблиця 4

Виробництво зернових і зернобобових культур та урожайність цукрового фабричного буряка в 2020-2022 рр. (план і факт)

	2020 рік	2021 рік (план)	2021 рік (факт)	2022 рік (план)	2022 рік (факт)
виробництво культур зернових та зернобобових, тис. т	64933	65000	86010	86000	53864
урожайність буряка цукрового фабричного, ц з 1 га	416,2	400	479,1	400	541,2

Відносний показник планового завдання – це відношення величин показника, встановленого на плановий період, до його величини, досягнутого за попередній період, який взято за базу зрівняння.

$$\text{Відносний показник планового завдання} = \frac{\text{Рівень, що планується на } (i + 1) \text{ період}}{\text{Досягнутий рівень в } i - \text{му періоді}}$$

або

$$\text{Відносний показник планового завдання} = \frac{\text{Рівень, що планується на } (i + 1) \text{ період}}{\text{Досягнутий рівень в } i - \text{му періоді}} * 100 \%$$

Відносний показник виконання плану – являє собою відношення фактично досягнутого рівня до планового завдання.

$$\text{Відносний показник виконання плану} = \frac{\text{Досягнутий рівень в (i + 1) – му періоді}}{\text{Рівень, що планувався на (i + 1) період}}$$

або

$$\text{Відносний показник виконання плану} = \frac{\text{Досягнутий рівень в (i + 1) – му періоді}}{\text{Рівень, що планувався на (i + 1) період}} * 100 \%$$

Завдання 5. Користуючись статистичними даними табл. 5, розрахуйте можливі показники інтенсивності.

Таблиця 5

Регіони	Середня чисельність наявного населення у 2021 р., тис. осіб	Площа, тис. км ²	Загальна площа житлових будівель, прийнятих в експлуатацію, тис. м ²	Виробництво культур зернових і зернобобових, тис. т
Україна	41377,8	603,5	7110	53864
Вінницька	1519,3	26,5	404	3564
Волинська	1024,4	20,1	275	1359
Дніпропетровська	3119,3	31,9	182	3271
Донецька	4079,8	26,5	4	544
Житомирська	1187,3	29,8	118	1987
Закарпатська	1247,3	12,8	309	306
Запорізька	1652,5	27,2	19	770
Івано-Франківська	1356,5	13,9	493	844
Київська	1791,8	28,1	1457	2978
Кіровоградська	911,9	24,6	34	3887
Луганська	2112,1	26,7	1	253
Львівська	2487,9	21,8	721	1904
Миколаївська	1100,1	24,6	35	2134
Одеська	2359,7	33,3	694	3062
Полтавська	1361,9	28,8	236	5097
Рівненська	1145,1	20,1	244	1403
Сумська	1044,6	23,8	42	3472
Тернопільська	1026,1	13,8	183	2643
Харківська	2616,4	31,4	43	2445
Херсонська	1009,2	28,5	8	145
Хмельницька	1236,3	20,6	283	3516
Черкаська	1169,5	20,9	93	3659
Чернівецька	893,5	8,1	221	665

Чернігівська	968,0	31,9	103	3956
м. Київ	2957,2	0,8	908	

Відносні показники інтенсивності характеризують ступінь поширення явища у певному середовищі. Це іменовані показники, які характеризують співвідношення різноманітних, але зв'язаних між собою абсолютних показників. У чисельнику відносного показника інтенсивності стоїть величина явища, ступень поширення якого вивчається, у знаменнику – величина того середовища, в якому це явище поширюється. Наприклад, щільність населення на 1 км²; вихід валової продукції на одну гривню основних виробничих фондів або середньорічного працівника; кількість квадратних метрів житла, що припадають в середньому на одного жителя; виробництво продукції в розрахунку на душу населення тощо.

Відносний показник інтенсивності =

$$= \frac{\text{Показник, що характеризує явище А}}{\text{Показник, що характеризує середовище розповсюдження явища А}}$$

Якщо обсяги явища незначні відносно обсягів середовища, то їх співвідношення збільшуються у 100, 1000, 10000 і більше разів. Наприклад, показники народжуваності, смертності, шлюбності розраховується на 1000 осіб населення, забезпеченість населення лікарями – на 10000 осіб населення, захворюваність та злочинність – на 100000 осіб населення.

Відносний показник інтенсивності =

$$= \frac{\text{Показник, що характеризує явище А}}{\text{Показник, що характеризує середовище розповсюдження явища А}} * \begin{matrix} 100 \\ \text{або} \\ 1000 \\ \text{або} \\ 10000 \\ \dots\dots\dots \end{matrix}$$

Завдання до самостійної роботи :

Завдання 1. Розгляньте класифікацію статистичних показників за різними критеріями – способом одержання, часовою ознакою, суттю досліджуваних явищ, ступенем агрегування досліджуваних явищ, об'єктом дослідження, взаємозалежністю, можливістю узагальнення (сумування). Користуючись сайтом Державної служби статистики – розділом «Статистична інформація» або розділом «Публікації» (наприклад, Україна у цифрах 2022: статистичний збірник / Відп. за випуск О. А. Вишневська. – К., 2023. – 34 с. або Статистичний щорічник України 2021 / За ред. І. Є. Вернера. – К., 2022. – 447 с.), наведіть по 3-4 приклади *конкретних статистичних показників* кожного виду. Відповідь оформіть у вигляді таблиці (за прикладом).

Таблиця 1

Критерій класифікації статистичних показників	Види статистичних показників за вказаним критерієм	Приклади статистичних показників
часова ознака	моментні	а) чисельність наявного населення України на 01. 01. 2022 р. = 41,2 млн. осіб б) середній розмір місячної пенсії пенсіонерів України на кінець 2022 року = 4622,6 грн.
	інтервальні	

Завдання 2. Оберіть із заповненої Вами таблиці (табл. 1 із попереднього завдання) п'ять будь-яких статистичних показників. Виділіть якісний, кількісний, часовий і просторовий зміст обраних статистичних показників (за прикладом).

а) чисельність наявного населення України на 01. 01. 2022 р. = 41,2 млн. осіб
якісний зміст – чисельність наявного населення;

кількісний зміст – 41,2 млн. осіб;

часовий зміст – на 01. 01. 2022 р.;

просторовий зміст – Україна;

б) середній розмір місячної пенсії пенсіонерів України на кінець 2022 року = 4622,6 грн.

якісний зміст – середній розмір місячної пенсії пенсіонерів;

кількісний зміст – 4622,6 грн.;

часовий зміст – на кінець 2022 року;
просторовий зміст – Україна.

Завдання 3. Розрахуйте відносні статистичні показники, якщо відомо наступне:

1) розрахуйте показники динаміки (темпи росту і приросту ланцюговим і базисним способами) за варіантами (на вибір студента). Побудуйте графіки.

Роки	Варіант 1. Обсяг іноземних інвестицій в Україну, на початок року, млн. дол. США	Варіант 2. Обсяг іноземних інвестицій з України, на початок року, млн. дол. США	Варіант 3. Обсяг експорту послуг України, млн. дол. США	Темпи росту (зменшення), %		Темпи приросту (зменшення), %	
				ланцюговим способом	базисним способом	ланцюговим способом	базисним способом
2000	3281,8	98,5	3655,1				
2001	3875,0	170,3	3731,9				
2002	4555,3	155,7	4303,8				
2003	5471,8	144,3	4524,9				
2004	6794,4	166,0	5612,7				
2005	9047,0	198,6	6443,2				
2006	16890,0	219,5	7791,8				
2007	21607,3	243,3	9435,1				
2008	29542,7	6196,6	12260,1				
2009	35616,4	6203,1	10129,7				
2010 ¹	38992,9	5760,5	11936,3				
2011 ¹	45370,0	6402,8	14180,3				
2012 ¹	48197,6	6435,4	14096,2				
2013 ¹	51705,3	6568,1	14233,2				
2014 ¹	53704,0	6702,9	11520,8				
2015 ¹	40725,4	6456,2	9736,6				
2016 ¹	32122,5	6315,2	9868,0				
2017 ¹	31230,3	6346,3	10790,3				
2018 ¹	31606,4	6322,0	11679,9				
2019 ¹	32905,1	6294,4	15660,9				
2020 ¹	35809,6	6272,7	11547,2				

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

2) розрахуйте показники структури і координації за варіантами (на вибір студента). Графічно візуалізуйте отримані результати.

Варіант 1

Гендерний склад ВРУ протягом років незалежності країни

Номер скликання ВРУ	Кількість народних депутатів		
	всього	з них жінок	з них чоловіків
I (1992-1993 рр.)	475	12	463
II (1994-1997 рр.)	436	18	418
III (1998-2001 рр.)	477	38	439
IV (2002-2005 рр.)	509	28	481
V (2006-2007 рр.)	483	42	441
VI (2008-2012 рр.)	541	42	499
VII (2013-2014 рр.)	478	46	432
VIII (2015-2019 рр.)	468	56	412
IX (2020-2021 рр.)	440	92	348

Варіант 2

Вікова структура населення України на 01. 01. 2022 р.

	Усього, осіб	з них у віці		
		0–14 років	15–64 роки	65 років і старше
Україна¹	40997698	6119886	27646706	7231106
Вінницька	1502430	228250	1004375	269805
Волинська	1018628	196450	685096	137082
Дніпропетровська	3093176	467034	2081587	544555
Донецька	4046487	423519	2674678	948290
Житомирська	1179801	190552	790179	199070
Закарпатська	1241643	242204	846345	153094
Запорізька	1637673	232973	1102443	302257
Івано-Франківська	1349096	224384	927622	197090
Київська	1789300	316611	1201941	270748
Кіровоградська	897297	131792	597961	167544
Луганська	2098324	189316	1396732	512276
Львівська	2459763	396357	1688299	375107
Миколаївська	1091106	164901	737530	188675
Одеська	2340332	393810	1569062	377460
Полтавська	1344445	188739	912376	243330
Рівненська	1140724	229140	764199	147385

Сумська	1033580	132578	702848	198154
Тернопільська	1018462	156857	699887	161718
Харківська	2583325	349863	1782295	451167
Херсонська	1000166	158536	673166	168464
Хмельницька	1225666	190583	819510	215573
Черкаська	1157115	157088	776993	223034
Чернівецька	887392	151873	606264	129255
Чернігівська	950773	125440	631852	193481
м. Київ	2910994	481036	1973466	456492

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Варіант 3
Робоча сила за статтю, типом місцевості
та віковими групами у 2021 році¹

	всього	жінки	чоловіки	міська місцевість	сільська місцевість
населення віком 15 років і старше	17405,0	8293,0	9112,0	11948,7	5456,3
з нього					
15-70 років	17321,6	8248,2	9073,4	11906,7	5414,9
20-64 роки	17018,3	8091,9	8926,4	11716,4	5301,9
працездатного віку	16666,8	7911,6	8755,2	11483,3	5183,5
за віковими групами					
15-24 роки	1128,9	502,8	626,1	671,0	457,9
25-29 років	1839,0	748,9	1090,1	1217,1	621,9
30-34 роки	2459,9	1056,6	1403,3	1695,6	764,3
35-39 років	2664,2	1206,4	1457,8	1978,8	685,4
40-49 років	4723,3	2353,8	2369,5	3332,5	1390,8
50-59 років	3851,5	2043,1	1808,4	2588,3	1263,2
60-70 років	654,8	336,6	318,2	423,4	231,4
71 рік і старше	83,4	44,8	38,6	42,0	41,4

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

3) розрахуйте показники порівняння (просторового порівняння,

порівняння зі стандартом) для споживання населенням України якогось продукту харчування (на вибір студента, окрім м'яса та м'ясопродуктів) у 2020 р. Проранжируйте регіони України за отриманими показниками. Побудуйте гістограми.

Регіони	Споживання населенням, 2020 р. ¹									
	м'яса та м'ясо- продуктів, кг	молока та молоко- продуктів, кг	яєць, шт.	хлібних продуктів, кг	картоплі, кг	овоче-баштанних культур, кг	плодів, ягід, винограду, кг	риби та рибо- продуктів, кг	цукру, кг	олії, кг
Раціональні норми споживання	83	380	290	101	124	161	90	20	38	13
Мінімальні норми споживання	52	341	231	94	96	105	68	12	32	8
Україна	53,8	201,9	228	96,6	134	164	56,5	12,4	27,8	12,3
Вінницька	57,5	200,9	299	111,5	168	174,5	63,6	14,7	30,1	13,3
Волинська	55,1	209,2	285	110,1	178,5	164,5	52,3	13,5	31,4	12,5
Дніпропетровська	65,6	196,4	303	88,8	115,7	168,1	63,7	13	27,3	13,1
Донецька	48,7	171,2	251	93,3	92,5	140,4	42,3	12,5	26,8	10,8
Житомирська	54,2	202,8	318	103,7	183,5	169,2	49,8	14,4	26,6	11,5
Закарпатська	52,2	232,8	281	116,7	153,5	170,1	52,2	8	28,4	12,6
Запорізька	56,1	180,4	275	93,2	96,2	172,6	52,5	13,2	27,4	12,2
Івано-Франківська	49,8	300,6	276	112,8	187,3	154,7	54,4	9,4	31,5	12,5
Київська	60,6	209,4	301	75,9	118,9	173,3	77,5	16,1	23,5	11,9
Кіровоградська	57,8	229,2	327	100,4	144,8	186,6	49,4	13,4	32,6	11,9
Луганська	43,5	150,6	217	90,1	110,9	106,5	43	9,2	27,9	11,6
Львівська	53,1	225,9	278	93,6	180,1	182,8	58,5	9,6	29,2	13,6
Миколаївська	57,2	205,4	266	105,8	123,5	172,4	60	13,4	28,3	13,6
Одеська	50,1	180,2	263	98,2	96,7	165,3	59,5	15,1	26,3	14
Полтавська	50,9	195,5	284	97,2	137,7	179	58,3	12,2	27,9	11,2
Рівненська	48,5	190,5	278	90,5	158,8	145,5	44,7	11	27,6	10,6
Сумська	51,4	180,4	275	98,5	168	163,4	43,3	10,6	29,3	10,7
Тернопільська	47,7	237,8	280	93,6	156,5	162,8	52,1	9,4	24,8	13,1
Харківська	53,5	202,6	272	91,9	103,5	166,4	50,7	9,5	23,6	11,2
Херсонська	55	195,8	255	99,6	135,6	181,5	50,6	14,5	30,6	12,9
Хмельницька	53,8	204,2	289	115,4	169,2	162,6	62,8	11,1	28,9	12,7
Черкаська	52,1	222,2	275	113,2	164,4	168,3	60,9	14,2	33,8	12,5
Чернівецька	44,4	220,1	281	109,2	143,7	175,7	65,7	10,3	29,9	12,9

Чернігівська	53	207,5	269	114,1	164,1	177,5	55,7	12,8	38,2	13,4
--------------	----	-------	-----	-------	-------	-------	------	------	------	------

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

4) розрахуйте показники планового завдання і виконання плану за варіантами (на вибір студента):

Варіант 1

Номер супермаркетів	Реалізація продукції, млн. грн.		
	фактично за минулий рік	планове завдання на звітний рік	фактично у звітному році
1	22,1	22,4	23,5
2	32,4	32,5	35,7
3	41,6	42,7	40,2

Варіант 2

	Урожайність сільськогосподарських культур, ц/га		
	фактично за 2021 р.	планове завдання на 2022 р.	фактично у 2022 р.
зернових та зернобобових	53,9	54,6	45,8
картоплі	166,4	170,0	173,5
соняшнику	24,6	24,0	21,6

Варіант 3

	Виробництво чавуну на металургійному комбінаті, тис. т		
	фактично за минулий рік	планове завдання на звітний рік	фактично у звітному році
Передільний	42	45	48
Ливарний	40	42	43
Хромонікелевий	37	37	37
Феромарганцевий	30	28	25
Ферофосфорний	22	25	24

5) розрахуйте показники інтенсивності за варіантами (на вибір студента),

використовуючи наступні дані:

Регіони	Середня чисельність наявного населення у 2021 р., тис. осіб	Площа, тис. км ²	Використання електроенергії, тис. кВт*год.	Обсяг валового регіонального продукту України в 2021 р., млн. грн.
Україна	41377,8	603,5	83888551	5450849¹
Вінницька	1519,3	26,5	1728791	173531
Волинська	1024,4	20,1	683202	92535
Дніпропетровська	3119,3	31,9	22012015	582363
Донецька	4079,8	26,5	8901484	283326
Житомирська	1187,3	29,8	1268438	113919
Закарпатська	1247,3	12,8	443351	75626
Запорізька	1652,5	27,2	7454972	228906
Івано-Франківська	1356,5	13,9	2577862	119680
Київська	1791,8	28,1	3014648	291519
Кіровоградська	911,9	24,6	2284442	99564
Луганська	2112,1	26,7	1161036	52135
Львівська	2487,9	21,8	2302941	296182
Миколаївська	1100,1	24,6	3299320	124162
Одеська	2359,7	33,3	2320087	271669
Полтавська	1361,9	28,8	3822701	266694
Рівненська	1145,1	20,1	2678332	88859
Сумська	1044,6	23,8	1196932	105254
Тернопільська	1026,1	13,8	419851	81485
Харківська	2616,4	31,4	3596405	319796
Херсонська	1009,2	28,5	997488	88182
Хмельницька	1236,3	20,6	1710960	119876
Черкаська	1169,5	20,9	1773689	131154
Чернівецька	893,5	8,1	1900689	54582
Чернігівська	968,0	31,9	786969	113474
м. Київ	2957,2	0,8	5551947	1276376

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

б) користуючись нижче наведеними статистичними даними щодо наукової діяльності в Україні в 2017-2020 рр., розрахуйте всі можливі види відносних показників та проаналізуйте за їх допомогою наукову діяльність

України в ці роки (за варіантами на вибір студента):

Наукова діяльність України в 2017-2020 рр.

	всього				з них жінки			
	2017	2018	2019	2020	2017	2018	2019	2020
ВВП, млрд. грн.	2981,2	3560,3	3977,2	4222,0				
Витрати на виконання наукових досліджень і розробок, млн. грн.	13379,3	16773,7	17254,6	17022,4				
Кількість працівників, задіяних у виконанні наукових досліджень і розробок	94274	88128	79262	78860	44173	41323	36989	37501
Варіант 1								
з них за рівнем освіти								
мають вищу освіту	81783	76455	68744	68060	38671	36202	32364	32277
доктори наук	6942	7043	6526	7060	1883	1884	1766	2053
доктори філософії (кандидати наук)	19219	18806	16929	17949	9030	8837	7900	8530
магістри (спеціалісти)	46612	43291	39148	37406	22940	21586	19489	18670
бакалаври (молодші бакалаври, молодші спеціалісти)	9010	7315	6141	5645	4818	3895	3209	3024
Інші рівні освіти	12491	11673	10518	10800	5502	5121	4625	5224
Варіант 2								
з них за галузями наук								
природничі	22140	21805	21305	21106	10611	10138	10039	10064
інженерія та технології	48985	43423	39033	36837	18769	16625	14982	14286
медичні науки та науки про здоров'я	5228	5461	4192	4914	3618	3789	2993	3384
сільськогосподарські та ветеринарні науки	7451	7428	6508	6212	4352	4208	3711	3611
суспільні	7239	6968	6096	7187	4787	4655	3994	4565

гуманітарні науки та мистецтво	3231	3043	2128	2604	2036	1908	1270	1591
Варіант 3								
з них за категоріями персоналу								
дослідники	59392	57630	51121	51427	26533	25780	22649	23338
техніки	9144	8553	7470	7117	5368	4994	4511	4137
допоміжний персонал	25738	21945	20671	20316	12272	10549	9829	10026

Практичне заняття 7-8
СЕРЕДНІ СТАТИСТИЧНІ ПОКАЗНИКИ.
МОДА і МЕДІАНА

Питання для обговорення:

1. Середні показники: визначення поняття, види та форми.
2. Умови застосування середніх показників. Переваги і недоліки середніх показників.
3. Середня арифметична: проста і зважена; для дискретного та інтервального рядів (із закритими та відкритими і інтервалами). Властивості середньої арифметичної.
4. Інші види ступеневих середніх показників: середня гармонічна, середня геометрична, середня квадратична.
5. Мода: визначення поняття, методика обрахунку для дискретних та інтервальних рядів. Переваги і недоліки моди.
6. Медіана: визначення поняття, методика обрахунку для дискретних (з парним та непарним числом варіант, з різною частотою варіант) та інтервальних рядів. Переваги і недоліки медіани.

Методичні рекомендації до практичної роботи:

Завдання 1. Користуючись статистичними даними табл. 1, обрахуйте просту середню арифметичну, моду та медіану кількості міст (за 27-а регіонами) та територіальних громад (за 24-а регіонами) України на 1 січня 2022 р.

Таблиця 1

**Кількість міст та територіальних громад за регіонами України
на 1. 01. 2022 р. (за даними Державної служби статистики України)**

Регіони України	Кількість міст	Кількість територіальних громад
Україна	461	1469*
АР Крим	16	
Вінницька область	18	63
Волинська область	11	54
Дніпропетровська область	20	86
Донецька область	52	66
Житомирська область	12	66

Закарпатська область	11	64
Запорізька область	14	67
Івано-Франківська область	15	62
Київська область	26	69
Кіровоградська область	12	49
Луганська область	37	37
Львівська область	44	73
Миколаївська область	9	52
Одеська область	19	91
Полтавська область	16	60
Рівненська область	11	64
Сумська область	15	51
Тернопільська область	18	55
Харківська область	17	56
Херсонська область	9	49
Хмельницька область	13	60
Черкаська область	16	66
Чернівецька область	11	52
Чернігівська область	16	57
м. Київ	1	
м. Севастополь	2	

* Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя

Середня арифметична являє собою відношення сумарної величини всіх варіант ознаки до їх чисельності. Вона застосовується тоді, коли відомі індивідуальні значення усереднюваної ознаки та їх кількість у сукупності.

Проста середня арифметична застосовується в тих випадках, коли кожне індивідуальне значення ознаки (варіанта) зустрічається один раз або однакове число разів. Проста середня арифметична розраховується шляхом підсумовування окремих одиниць і ділення одержаної суми на кількість величин.

$$\bar{x} = \frac{\text{Сума значень ознак}}{\text{Обсяг сукупності}} [\text{ознака}]$$

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_n)}{n} = \frac{\sum x}{n}$$

— де \bar{x} — проста середня арифметична;

x — окремі значення усереднюваної ознаки (варіанти);

n – число одиниць досліджуваної сукупності (число варіант).

Рішення:

a)

The screenshot shows the Microsoft Excel 2010 interface. The 'Аргументы функции' (Function Arguments) dialog box is open for the 'CRP3H4' function. The dialog box contains the following information:

- Функция:** CRP3H4
- Число1:** 81829
- Число2:** (empty)
- Результат:** 17,07407407
- Описание:** Возвращает среднее арифметическое своих аргументов, которые могут быть числами, именами, ссылками или ссылками на ячейки с числами.
- Справка по этой функции:** (link)
- Buttons:** OK, Отмена (Cancel)

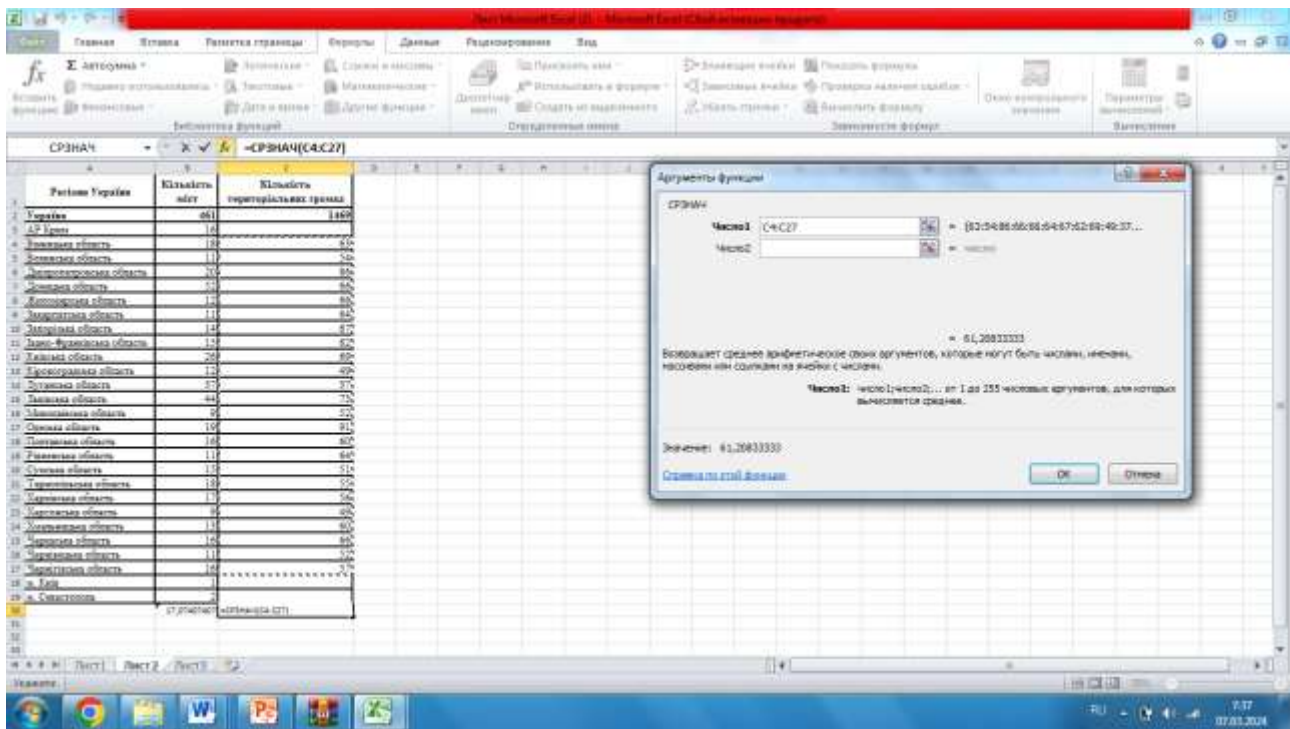
The background spreadsheet shows a table with columns: Регионы Украины, Численность, and Численность территории (всего). The data is organized by region, with 'Д. Киев' and 'Д. Севастополь' at the bottom.

або

$$X_{\text{ср}} = (16+18+11+20+\dots+2)/27 = 461/27 = \mathbf{17,07}$$

(середня кількість міст по 27-и регіонах України)

6)



або

$$X_{cep} = (63+54+86+66+...+57)/24 = 1469/24 = \mathbf{61,21}$$

(середня кількість територіальних громад по 24-х регіонах України)

Мода – це значення варіанти, що найчастіше повторюється в ряду розподілу; найтипівіше, характерне для даної сукупності значення варіанту. Для *дискретних рядів* розподілу моду визначають візуально без будь-яких розрахунків за значенням варіанти з найбільшою частотою (часткою). Якщо в ряду розподілу всі варіанти зустрічаються однаково кількість разів, моди немає.

Рішення:

а) розподіл регіонів України за кількістю міст є дво модальним, оскільки 2 варіанти – 11 і 16 міст – зустрічаються з найбільшою частотою (4 рази). Тобто модами є **11 і 16 міст**;

б) у розподілі регіонів України за кількістю територіальних громад модою є **66 територіальних громад**, оскільки ця варіанта зустрічається найчастіше (3 рази).

Медіана – це варіанта, що розташована в середині ранжируваної сукупності і ділить ранжирований (впорядкований за мірою зростання або

зменшення) ряд на дві рівні за обсягом частини – із значеннями ознаки менше і більше медіани.

Для обчислення медіани для *дискретного ряду з непарним числом варіант* треба:

1) на основі сукупності побудувати ранжируваний ряд і пронумерувати його члени;

2) знайти порядковий номер члена ряду із значенням ознаки рівним медіані:

$$M_e = x_{m/2+0,5}$$

або

$$M_e = x_{(m+1)/2}$$

де x – окремі значення усереднюваної ознаки (варіанти);

m – кількість членів ранжируваного ряду.

Рішення:

Проранжируємо розподіл регіонів України за кількістю міст і пронумеруємо його членів (в нашому ряді є варіанти, які зустрічаються по декілька разів, але зараз не будемо звертати на це увагу):

Регіони України	Кількість міст	№з/п
Україна	461	
м. Київ	1	1
м. Севастополь	2	2
Миколаївська область	9	3
Херсонська область	9	4
Волинська область	11	5
Закарпатська область	11	6
Рівненська область	11	7
Чернівецька область	11	8
Житомирська область	12	9
Кіровоградська область	12	10
Хмельницька область	13	11
Запорізька область	14	12
Івано-Франківська область	15	13
Сумська область	15	14

АР Крим	16	15
Полтавська область	16	16
Черкаська область	16	17
Чернігівська область	16	18
Харківська область	17	19
Вінницька область	18	20
Тернопільська область	18	21
Одеська область	19	22
Дніпропетровська область	20	23
Київська область	26	24
Луганська область	37	25
Львівська область	44	26
Донецька область	52	27

Знайдемо порядковий номер члена ряду із значенням ознаки рівним медіані:

$$M_e = x_{(27+1)/2} = x_{14} = \mathbf{15 \text{ міст}}$$

Для обчислення медіани для *дискретного ряду з парним числом варіант* треба:

- 1) на основі сукупності побудувати ранжируваний ряд і пронумерувати його члени;
- 2) знайти порядкові номери двох центральних елементів ряду:

$$x_{m/2}$$

та

$$x_{m/2+1}$$

- 3) обрахувати середню арифметичну з 2-х центральних елементів ряду:

$$M_e = (x_{m/2} + x_{m/2+1}) / 2$$

Рішення:

Проранжируємо розподіл регіонів України за кількістю територіальних громад і пронумеруємо його членів (*в нашому ряді є варіанти, які зустрічаються по декілька разів, але зараз не будемо звертати на це увагу*):

Регіони України	Кількість територіальних громад	№з/п
-----------------	---------------------------------	------

Україна	1469	
Луганська область	37	1
Кіровоградська область	49	2
Херсонська область	49	3
Сумська область	51	4
Миколаївська область	52	5
Чернівецька область	52	6
Волинська область	54	7
Тернопільська область	55	8
Харківська область	56	9
Чернігівська область	57	10
Полтавська область	60	11
Хмельницька область	60	12
Івано-Франківська область	62	13
Вінницька область	63	14
Закарпатська область	64	15
Рівненська область	64	16
Донецька область	66	17
Житомирська область	66	18
Черкаська область	66	19
Запорізька область	67	20
Київська область	69	21
Львівська область	73	22
Дніпропетровська область	86	23
Одеська область	91	24

Знайдемо порядкові номери двох центральних елементів ряду:

$$X_{24/2} = X_{12} = 60$$

та

$$X_{24/2+1} = X_{13} = 62$$

Обрахуємо середню арифметичну з 2-х центральних елементів ряду:

$$(60+62) / 2 = \mathbf{61 \text{ територіальна громада}}$$

Завдання 2. Користуючись статистичними даними табл. 2, обрахуйте зважену середню арифметичну (середню кількість осіб у домогосподарствах), моду та медіану кількості осіб у домогосподарствах України в 2022 р. Останню категорію (4 і більше осіб) у розрахунках взяти умовно як 5 осіб. Розпишіть всі проміжні етапи розрахунків.

Таблиця 2

Розподіл домогосподарств України за кількістю осіб у їх складі в 2022 році
(за даними Державної служби статистики України)

Кількість осіб	Кількість домогосподарств ¹ , тис.
1	2604,3
2	5252,3
3	3826,4
4 і більше	2866,2

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Зважена середня арифметична застосовується в тих випадках, коли індивідуальні значення ознаки (варіанти) в сукупності зустрічаються багато разів і неоднакове число разів. Зважування в даному випадку здійснюється за частотами, які показують скільки разів повторюється та або інша варіанта. Формула середньої арифметичної зваженої має вигляд:

$$\bar{x} = \frac{(x_1 f_1 + x_2 f_2 + \dots + x_n f_n)}{f_1 + f_2 + \dots + f_n} = \frac{\sum x f}{\sum f},$$

де \bar{x} – зважена середня арифметична;

x – окремі значення усереднюваної ознаки (варіанти);

n – число одиниць досліджуваної сукупності (число варіант);

f – частота, яка показує скільки разів зустрічається варіанта (вага).

При обчисленні зваженої середньої арифметичної, якщо *ряд дискретний*, необхідно:

- помножити кожний варіант на його частоту;
- знайти суму одержаних добутків варіантів на частоти;
- знайти суму частот (якщо вона невідома);
- суму добутків варіантів на частоти розділити на суму частот.

В загальному вигляді підрахунку зваженої середньої арифметичної дискретного ряду можна представити таким чином:

Варіанти, (x)	Частоти, (f)	Добуток варіантів на частоти, (x f)
x ₁	f ₁	x ₁ f ₁
x ₂	f ₂	x ₂ f ₂
...
x _n	f _n	x _n f _n
	$\sum f$ сума частот	$\sum x f$ обсяг явища

$$\bar{X} = \frac{\sum x f}{\sum f}$$

Рішення:

Кількість осіб	Кількість домогосподарств, тис.	Добуток
1	2604,3	2604,3
2	5252,3	10504,6
3	3826,4	11479,2
5	2866,2	14331
сума	14549,2	38919,1
середня зважена арифметична	2,67 (середня кількість осіб у домогосподарствах)	

Як уже зазначалося, для дискретних рядів розподілу **моду** визначають візуально без будь-яких розрахунків за значенням варіанти з найбільшою частотою (часткою). У розподілі домогосподарств України за кількістю осіб у їх складі в 2022 році **моду** є **2 особи**, оскільки ця варіанта зустрічається найчастіше (5252,3тис. разів).

При обчисленні **медіани** для дискретного ряду з різною частотою варіант, необхідно обчислити накопичені частоти. Під **накопиченими частотами** розуміють наростаючий підсумок послідовного підсумовування

частот, починаючи з першого варіанту або першого інтервалу.

Рішення:

Ряд уже проранжирований за кількістю осіб у домогосподарствах, тому ранжирувати додатково не треба. Обрахуємо накопичені частоти:

Кількість осіб	Кількість домогосподарств, тис.	Накопичені частоти
1	2604,3	2604,3
2	5252,3	7856,6 (2604,3 + 5252,3)
3	3826,4	11683 (7856,6 + 3826,4)
5	2866,2	14549,2 (11683 + 2866,2)

Знайдемо накопичену частоту члена ряду із значенням ознаки рівним медіані:

$$M_e = x_{14549,2/2} = x_{7274,6}$$

$x_{7274,6}$ знаходиться в межах накопиченої частоти $x_{7856,6}$. Накопиченій частоті 7856,6 відповідає значення варіанти рівне 2. Отже, **$M_e = 2$ особи.**

Завдання 3. Користуючись статистичними даними табл. 3, обрахуйте зважену середню арифметичну (середню житлову площу домогосподарств), моду та медіану моду та медіану житлової площі домогосподарств України в 2022 р. Розпишіть всі проміжні етапи розрахунків.

Таблиця 3

Розподіл домогосподарств України за розміром житлової площі на одну особу в 2022 році (за даними Державної служби статистики України)

Розмір житлової площі, м ² на одну особу	Кількість домогосподарств ¹ , тис.
до 7,5	843,9
7,5 – 9,0	683,8
9,01 – 13,65	3462,7

13,66 – 20,0	4190,2
20,01 – 25,0	1789,6
понад 25,01	3579,1

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

При обчисленні *зваженої середньої арифметичної*, якщо ряд *інтервальний* (варіанти значення ознак представлені не одним числом, а певним інтервалом – нижньою і верхньою межами), необхідно:

- визначити середню величину кожного інтервалу (його середину) як напівсуму верхньої і нижньої меж;
- визначити середню для всього ряду в тій послідовності, що і для дискретного варіаційного ряду.

Якщо інтервальний ряд *має відкриті інтервали в першій і останній групі*, то спочатку треба визначити невідомі межі інтервалів. Вони визначаються:

а) або експертним шляхом виходячи з сутності ознаки в сукупності чи особливостей шкали;

б) або величина інтервалу в першій групі приймається рівною інтервалу наступної групи, а величина інтервалу в останній групі рівна інтервалу попередньої групи.

Рішення:

Визначимо невідомі межі інтервалів за таким правилом: величина інтервалу в першій групі приймається рівною інтервалу наступної групи (1,5), а величина інтервалу в останній групі рівна інтервалу попередньої групи (4,99).

Розмір житлової площі, м ² на одну особу	Кількість домогосподарств, тис.
6,0 – 7,5	843,9
7,5 – 9,0	683,8
9,01 – 13,65	3462,7
13,66 – 20,0	4190,2
20,01 – 25,0	1789,6
25,01 – 30,0	3579,1

Визначимо середню величину кожного інтервалу (його середину) як напівсуму верхньої і нижньої меж.

Розмір житлової площі, м ² на одну особу	Кількість домогосподарств, тис.	Середина інтервалу
6,0 – 7,5	843,9	6,75
7,5 – 9,0	683,8	8,25
9,01 – 13,65	3462,7	11,33
13,66 – 20,0	4190,2	16,83
20,01 – 25,0	1789,6	22,505
25,01 – 30,0	3579,1	27,505

Визначимо зважену середню арифметичну для всього ряду в тій послідовності, що і для дискретного варіаційного ряду (помножимо кожний варіант на його частоту; знайдемо суму одержаних добутків варіантів на частоти; знайдемо суму частот; суму добутків варіантів на частоти розділимо на суму частот).

Розмір житлової площі, м ² на одну особу	Кількість домогосподарств, тис.	Середина інтервалу	Добуток
6,0 – 7,5	843,9	6,75	5696,325
7,5 – 9,0	683,8	8,25	5641,35
9,01 – 13,65	3462,7	11,33	39232,39
13,66 – 20,0	4190,2	16,83	70521,07
20,01 – 25,0	1789,6	22,505	40274,95
25,01 – 30,0	3579,1	27,505	98443,15
сума	14549,3		259809,2
зважена середня арифметична	17,86 (середній розмір житлової площі в домогосподарствах, м ² на одну особу)		

При обчисленні *моди* в *інтервальному ряду* спочатку визначається модальний інтервал (в рядах з рівними інтервалами – це інтервал з найбільшою частотою, в рядах з нерівними інтервалами – це інтервал з найбільшою щільністю розподілу) і значення моди в середині інтервалу розраховується за формулою:

$$M_o = x_0 + i \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}$$

де M_o – мода;

x_0 – нижня межа модального інтервалу;

i – величина модального інтервалу;

f_{M_0} , f_{M_0-1} , f_{M_0+1} – частота відповідно передмодального, модального та післямодального інтервалів.

Рішення:

Спочатку визначимо модальний інтервал. Модальним буде інтервал 13,66-20,0 м², тому що він має найбільшу частоту $f_{M_0} = 4190,2$ тис.

Підставимо значення у формулу:

$$M_0 = 13,66 + 6,34 * (4190,2 - 3462,7) / ((4190,2 - 3462,7) + (4190,2 - 1789,6)) =$$
$$\mathbf{15,13 \text{ м}^2}$$

Отже, найпоширеніший розмір житлової площі в домогосподарствах – 15,13 м² на особу.

При обчисленні **медіани** для *інтервального ряду*, медіана обчислюється для середини медіанного інтервалу, за який приймається такий, де сума накопичених частот перевищує половину значень частот ряду розподілу.

Порядок обчислення:

1) по сумі частот визначається порядковий номер M_e ;

2) по накопичених частотах визначається медіанний інтервал, тобто інтервал, в якому знаходиться M_e ;

3) по формулі визначається медіана:

В даному випадку формула для розрахунку медіани має вигляд:

$$M_e = x_0 + h (0,5 \sum f - Sx_0) / f_m$$

де x_0 – нижня межа медіанного інтервалу;

h – величина медіанного інтервалу;

$0,5 \sum f$ – половина суми накопичених частот інтервального ряду;

Sx_0 – сума накопичених частот перед медіанним інтервалом;

f_m – частота медіанного інтервалу.

Рішення:

Ряд уже проранжирований за розміром житлової площі у домогосподарствах, тому ранжирувати додатково не треба. Обрахуємо накопичені частоти:

Розмір житлової площі, м ² на одну особу	Кількість домогосподарств, тис.	Накопичені частоти
до 7,5	843,9	843,9
7,5 – 9,0	683,8	1527,7
9,01 – 13,65	3462,7	4990,4
13,66 – 20,0	4190,2	9180,6
20,01 – 25,0	1789,6	10970,2
понад 25,01	3579,1	14549,3

Визначаємо порядковий номер $M_e = 14549,3 : 2 = 7274,7$. За накопиченими частотами бачимо, що 7274,7-й номер елемента ряду знаходиться в інтервалі 13,66-20,0. Це медіанний інтервал.

Рахуємо медіану за формулою:

$$M_e = 13,66 + 6,34 (0,5 * 14549,3 - 4990,4) / 4190,2 = \mathbf{17,12 \text{ м}^2}$$

Отже, можна зробити висновок: в половині домогосподарств України розмір житлової площі складає більше 17,12 м² на особу, а в половині домогосподарств – менше 17,12 м² на особу.

Завдання до практичної роботи:

Завдання 1. Користуючись статистичними даними табл. 1-2, обрахуйте просту середню арифметичну, моду та медіану за двома рядами статистичних показників (на вибір студента, але щоб були статистичні ряди з парним та непарним числом варіант). Розпишіть всі проміжні етапи розрахунків.

Таблиця 1

Демографічні показники за регіонами України в 2021 р. (за даними Державної служби статистики України)

Регіони України	Загальний коефіцієнт народжуваності, ‰	Загальний коефіцієнт смертності, ‰	Коефіцієнт природного приросту, ‰	Частка населення у віці 0-14 років у загальній чисельності населення, ‰	Частка населення у віці 15-64 років у загальній чисельності населення, ‰	Частка населення у віці 65 років і старше у загальній чисельності населення, ‰	Середній вік, років	На 1000 чоловіків припадає жінок	Рівень урбанізації, ‰
Вінницька	6,9	18,2	-11,3	15,2	66,8	18	42,2	1156	52,3
Волинська	9,6	15,1	-5,5	19,3	67,2	13,5	38,5	1118	52,2
Дніпропетровська	6,3	20,8	-14,5	15,1	67,3	17,6	42,3	1190	84,2
Житомирська	7,2	19,9	-12,7	16,1	67	16,9	41,3	1144	59,6
Закарпатська	10,1	14,8	-4,7	19,5	68,2	12,3	37,8	1080	37,2
Запорізька	5,8	21,6	-15,8	14,2	67,3	18,5	43,1	1190	77,5
Івано-Франківська	7,8	15,2	-7,4	16,6	68,8	14,6	40,1	1115	44,6
Київська	7	20,2	-13,2	17,7	67,2	15,1	40,3	1184	61,8
Кіровоградська	6,1	20,8	-14,7	14,7	66,6	18,7	42,8	1166	63,7
Львівська	7,8	16	-8,2	16,1	68,6	15,3	40,6	1111	61,2
Миколаївська	6,4	19,4	-13	15,1	67,6	17,3	42	1156	68,8
Одеська	8,2	18	-9,8	16,8	67,1	16,1	40,6	1122	67,3
Полтавська	5,9	20,3	-14,4	14	67,9	18,1	43	1162	62,7
Рівненська	10,2	14,9	-4,7	20,1	67	12,9	37,9	1104	47,5
Сумська	5,2	20	-14,8	12,8	68	19,2	44	1178	69,8
Тернопільська	7,1	16,4	-9,3	15,4	68,7	15,9	41,2	1134	46,1
Харківська	5,9	21	-15,1	13,5	69	17,5	42,6	1155	81,4
Херсонська	7	20	-13	15,9	67,3	16,8	41,5	1152	61,4
Хмельницька	7	19	-12	15,5	66,9	17,6	42	1150	58
Черкаська	5,8	19,6	-13,8	13,6	67,1	19,3	43,6	1172	57,3
Чернівецька	8,8	16,1	-7,3	17,1	68,3	14,6	39,7	1120	43,3

Чернігівська	5,5	22	-16,5	13,2	66,5	20,3	44,2	1199	66
м. Київ	10	15,3	-5,3	16,5	67,8	15,7	40,4	1169	100
Україна¹	7,3	18,5	-11,2	14,9	67,4	17,7	42,2	1157	69,7

¹ Відносні показники розраховано без урахування відповідних даних по Донецькій та Луганській областях. У зв'язку з відсутністю інформації по частині тимчасово окупованих територій у цих областях здійснення розрахунків відносних показників по ним є некоректним

Таблиця 2

Споживання населенням України продуктів харчування у 2020 р.
(за даними Державної служби статистики України)

Регіони	Споживання населенням, 2020 р. ¹									
	м'яса та м'ясо-продуктів, кг	молока та молоко-продуктів, кг	яєць, шт.	хлібних продуктів, кг	картоплі, кг	овоче-баштанних культур, кг	плодів, ягід, винограду, кг	риби та рибо-продуктів, кг	цукру, кг	олії, кг
Україна	53,8	201,9	228	96,6	134	164	56,5	12,4	27,8	12,3
Вінницька	57,5	200,9	299	111,5	168	174,5	63,6	14,7	30,1	13,3
Волинська	55,1	209,2	285	110,1	178,5	164,5	52,3	13,5	31,4	12,5
Дніпропетровська	65,6	196,4	303	88,8	115,7	168,1	63,7	13	27,3	13,1
Донецька	48,7	171,2	251	93,3	92,5	140,4	42,3	12,5	26,8	10,8
Житомирська	54,2	202,8	318	103,7	183,5	169,2	49,8	14,4	26,6	11,5
Закарпатська	52,2	232,8	281	116,7	153,5	170,1	52,2	8	28,4	12,6
Запорізька	56,1	180,4	275	93,2	96,2	172,6	52,5	13,2	27,4	12,2
Івано-Франківська	49,8	300,6	276	112,8	187,3	154,7	54,4	9,4	31,5	12,5
Київська	60,6	209,4	301	75,9	118,9	173,3	77,5	16,1	23,5	11,9
Кіровоградська	57,8	229,2	327	100,4	144,8	186,6	49,4	13,4	32,6	11,9
Луганська	43,5	150,6	217	90,1	110,9	106,5	43	9,2	27,9	11,6
Львівська	53,1	225,9	278	93,6	180,1	182,8	58,5	9,6	29,2	13,6
Миколаївська	57,2	205,4	266	105,8	123,5	172,4	60	13,4	28,3	13,6
Одеська	50,1	180,2	263	98,2	96,7	165,3	59,5	15,1	26,3	14
Полтавська	50,9	195,5	284	97,2	137,7	179	58,3	12,2	27,9	11,2
Рівненська	48,5	190,5	278	90,5	158,8	145,5	44,7	11	27,6	10,6
Сумська	51,4	180,4	275	98,5	168	163,4	43,3	10,6	29,3	10,7
Тернопільська	47,7	237,8	280	93,6	156,5	162,8	52,1	9,4	24,8	13,1
Харківська	53,5	202,6	272	91,9	103,5	166,4	50,7	9,5	23,6	11,2
Херсонська	55	195,8	255	99,6	135,6	181,5	50,6	14,5	30,6	12,9

Хмельницька	53,8	204,2	289	115,4	169,2	162,6	62,8	11,1	28,9	12,7
Черкаська	52,1	222,2	275	113,2	164,4	168,3	60,9	14,2	33,8	12,5
Чернівецька	44,4	220,1	281	109,2	143,7	175,7	65,7	10,3	29,9	12,9
Чернігівська	53	207,5	269	114,1	164,1	177,5	55,7	12,8	38,2	13,4

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Завдання 2. Користуючись статистичними даними табл. 3, обрахуйте зважену середню арифметичну (середню кількість осіб у домогосподарствах), моду та медіану за одним дискретним рядом статистичних показників (*тобто за одним регіоном, на вибір студента*). Останню категорію (4 і більше осіб) у розрахунках взяти умовно як 5 осіб. Розпишіть всі проміжні етапи розрахунків.

Завдання 3. Користуючись статистичними даними табл. 4, обрахуйте зважену середню арифметичну (середню житлову площу домогосподарств), моду та медіану за одним інтервальним рядом статистичних показників (*тобто за одним регіоном, на вибір студента*). Розпишіть всі проміжні етапи розрахунків.

Таблиця 3

Розподіл домогосподарств за кількістю осіб у їх складі за регіонами України в 2022 році
(за даними Державної служби статистики України)

Кількість осіб	Кількість домогосподарств за регіонами України, тис.																									
	Україна всього ¹	Вінницька область	Волинська область	Дніпропетровська область	Донецька область	Житомирська область	Закарпатська область	Запорізька область	Івано-Франківська область	Київська область	Кіровоградська область	Луганська область	Львівська область	Миколаївська область	Одеська область	Полтавська область	Рівненська область	Сумська область	Тернопільська область	Харківська область	Херсонська область	Хмельницька область	Черкаська область	Чернівецька область	Чернігівська область	м. Київ
1	2604,3	143,2	46,7	275,2	118,8	94,7	31,2	143,7	31,0	107,9	110,1	69,7	159,1	76,2	101,8	130,0	46,0	105,5	64,2	214,1	54,7	110,9	99,4	52,9	105,5	117,4
2	5252,3	214,5	91,4	503,1	460,0	162,5	84,2	261,7	144,3	267,0	137,3	146,2	176,0	138,1	373,8	213,5	110,4	122,3	78,7	368,0	154,9	140,7	204,8	62,9	159,7	476,2
3	3826,4	131,6	54,3	378,0	248,3	150,2	69,1	158,7	119,1	146,3	94,3	87,7	183,3	145,3	223,7	130,6	91,9	107,1	89,3	402,5	113,3	93,5	116,7	93,1	94,8	299,0
4 і більше	2866,2	120,1	138,7	160,7	66,1	63,6	166,2	113,9	148,3	153,1	52,9	36,4	289,0	64,0	178,1	93,7	122,3	86,9	111,3	91,5	62,4	113,2	71,4	93,4	53,8	214,9

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Таблиця 4

Розподіл домогосподарств за розміром житлової площі на одну особу за регіонами України в 2022 році

(за даними Державної служби статистики України)

Розмір житлової площі, м ² на одну особу	Кількість домогосподарств за регіонами України, тис.																										
	Україна всього ¹	Вінницька область	Волинська область	Дніпропетровська область	Донецька область	Житомирська область	Закарпатська область	Запорізька область	Івано-Франківська область	Київська область	Кіровоградська область	Луганська область	Львівська область	Миколаївська область	Одеська область	Полтавська область	Рівненська область	Сумська область	Тернопільська область	Харківська область	Херсонська область	Хмельницька область	Черкаська область	Чернівецька область	Чернігівська область	м. Київ	
до 7,5	843,9	39,0	21,5	75,1	33,9	28,3	22,1	28,5	29,7	23,6	15,8	16,7	69,4	16,1	44,7	18,2	36,7	16,9	14,4	78,6	30,8	23,4	31,5	4,8	21,1	100,8	
7,5 – 9,0	683,8	20,1	11,9	42,1	33,0	2,4	4,2	40,0	30,5	20,9	24,1	10,9	33,1	8,9	43,0	27,8	24,5	30,8	18,9	64,6	16,6	38,9	3,0	15,1	14,5	111,9	
9,01 – 13,65	3462,7	129,8	102,0	334,5	192,9	94,7	104,2	141,7	79,2	134,2	84,8	70,7	221,2	113,1	201,8	124,9	90,8	119,8	78,7	287,3	87,1	113,2	84,2	90,1	61,7	323,4	
13,66 – 20,0	4190,2	151,7	103,6	324,0	350,1	162,5	78,2	204,8	114,2	238,0	98,7	95,5	188,1	129,2	269,4	169,2	91,5	96,6	117,8	350,8	87,5	97,1	128,0	82,2	109,7	344,4	
20,01 – 25,0	1789,6	76,8	28,5	219,9	127,7	68,8	56,1	87,5	57,1	82,3	41,8	36,0	98,5	52,5	76,3	64,7	38,5	42,2	24,4	100,1	85,6	32,1	67,9	42,6	62,1	122,9	
понад 25,01	3579,1	192,0	63,6	321,3	155,4	114,4	85,9	175,6	131,9	175,3	129,4	110,2	197,0	103,8	242,2	162,9	88,6	115,6	89,3	194,8	77,9	153,5	177,7	67,4	144,8	104,1	

¹ Без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Практичне заняття 9-10

СТАТИСТИЧНА БАЗА ДАНИХ. МАТРИЦІ ДАНИХ

Питання для обговорення:

1. Статистична інформація, статистичні дані, статистичні показники, статистична сукупність: визначення і сутність понять.
2. Статистична сукупність: проста статистична сукупність або статистичний ряд, генеральна і вибіркова сукупність (вибірка).
3. Статистичні ознаки та їх класифікація.
4. Статистичні ряди розподілу та їх види. Атрибутивні (якісні) та варіаційні (кількісні) ряди розподілу (дискретні, інтервальні). Ранжировані та неранжировані ряди розподілу.
5. Статистичні шкали та їх види.
6. Бази даних.
7. Матриці даних та вимоги до них.
8. Теорія ймовірності та закон великих чисел.

Завдання до практичної роботи:

Завдання 1. У зв'язку з неможливістю, а також часто і з недоцільністю або трудністю поглибленого аналізу всіх одиниць спостереження, які складають генеральну сукупність, у більшості випадків статистичний аналіз проводять тільки за певною частиною одиниць – вибірковою сукупністю. Теоретичне обґрунтування вибіркового методу дає математична теорія вірогідності і обґрунтований цією теорією закон великих чисел. Сформулюйте теорію вірогідності та закон великих чисел.

Завдання 2. Побудуйте схему, на якій відобразить, як співвідносяться поняття «база даних» і «матриця даних». На схемі (або нижче як висновок) зробіть відповідні пояснення, які б відображали сутність понять та чим відрізняються і чим схожі ці поняття.

Завдання 3. Користуючись офіційним сайтом Державної служби статистики України (<https://www.ukrstat.gov.ua/>), проведіть вибірковий збір статистичних показників для своєї інформаційної бази даних. Необхідно зібрати 20-25 показників по регіонах України, віддаючи перевагу відносними і середнім показникам. Для цього потрібно:

а) визначитись з тематикою інформаційної бази (наприклад, демографічна ситуація, статеві-вікова структура населення, працересурсний потенціал, міграційна ситуація, економіка, сільське господарство, доходи і витрати населення, споживання населення, туризм і рекреація, охорона здоров'я, освіта, культура, торгівля тощо);

б) відібрати 10-15 статистичних показників по регіонах України з обраної теми;

в) відібрати додатково ще близько 10 загальних статистичних показників, які відображають фактори, що впливають на ті процеси, об'єкти чи явища, які ви обрали темою вашої інформаційної бази.

Наприклад, ви обрали темою вашої інформаційної бази демографічну ситуацію в Україні. Ваша інформаційна база може включати наступні демографічні показники по регіонах України:

*загальний коефіцієнт народжуваності,
загальний коефіцієнт смертності,
коефіцієнт природного приросту,
сумарний коефіцієнт народжуваності (коефіцієнт фертильності),
коефіцієнт дитячої смертності,
середній вік матері при народженні дитини,
брутто-показник відтворення,
нетто-показник відтворення,
довжина жіночого покоління,
вікові коефіцієнти народжуваності,
вікові коефіцієнти смертності,
коефіцієнти смертності за статтю,
коефіцієнти смертності за європейським стандартом,
коефіцієнти смертності за європейським стандартом за окремими причинами смерті,*

середня очікувана тривалість життя при народженні (обидві статті, чоловіки, жінки),

середня очікувана тривалість життя при досягненні віку 15 років, 45 років, 60 років, 65 років тощо.

На демографічну ситуацію впливає багато факторів, зокрема статеві-вікова структура населення, сімейний стан, тип місцевості (сільська чи міська), рівень захворюваності, рівень освіти, економічні фактори, екологічні фактори, міграційна ситуація в регіоні тощо. Тому серед загальних статистичних показників, які відображають фактори впливу на демографічну ситуацію, можна взяти такі:

*частка населення у віці 0-14 років у загальній чисельності населення,
частка населення у віці 15-64 років у загальній чисельності населення,
частка населення у віці 65 років і старше у загальній чисельності населення,
середній вік населення,
рівень урбанізації,
кількість прибулих на 1000 осіб наявного населення,
кількість вибулих на 1000 осіб наявного населення,
міграційний приріст на 1000 осіб наявного населення,
ВРП на 1000 осіб наявного населення,
наявний дохід у розрахунку на одну особу,
середньомісячна номінальна заробітна плата штатних працівників,
частка жінок у загальній чисельності населення,
частка чоловіків у загальній чисельності населення,
на 1000 жінок припадає чоловіків,
коефіцієнт шлюбності,
коефіцієнт розлучуваності тощо.*

PS: всі дані мають бути на один і той же рік (не важливо який) та охоплювати всі регіони України (якщо якогось показника немає по всім регіонам, то краще його не брати, оскільки не можна, щоб у базі даних були пропуски). З 2014 року немає статистичних даних по тимчасово окупованим територіям Автономної Республіки Крим та місту Севастополь, тому ці регіони можна не включати в інформаційну базу даних (якщо ви обрали рік, починаючи з 2014-го). Деякі відносні показники не рахуються по Луганській і Донецькій областях. У зв'язку з відсутністю інформації по частині тимчасово окупованих територій у цих областях здійснення розрахунків деяких відносних показників по ним є некоректним. Тому можна або не брати ці показники, або виключити ці області з інформаційної бази даних.


Практичне заняття 11-12

ПОБУДОВА ОДНОВИМІРНИХ СТАТИСТИЧНИХ МОДЕЛЕЙ ВЕКТОРІВ ЗМІННИХ

Питання для обговорення

1. Статистичний аналіз: поняття, методи, види. Дескриптивний (описовий) та аналітичний (індуктивний) аналіз.
2. Статистики одновимірного аналізу
 - значення процентилей (Percentile): проценти́лі, кварта́лі (Quartile), міжквартильний розмах (Interquartile range);
 - міри центральної тенденції (Central Tendency): мода (Mode), медіана (Median), середнє (Mean);
 - міри варіації (Dispersion): стандартне відхилення (Std. deviation), дисперсія (Variance), розмах (Range), мінімум (Minimum), максимум (Maximum);
 - характеристики форми розподілу (Distribution): асиметрія (Skewness), ексцес (Kurtosis);
 - підсумок з п'яти значень: мінімум (Minimum), перший (або нижній) квартиль (Q25 або Q1 або lower half), медіана (Median, Q50 або Q2), третій (або верхній) квартиль (Q75 або Q3 або upper half), максимум (Maximum).
3. Нормальний розподіл випадкових величин (Normal distribution).
4. Візуалізація одновимірного аналізу. Ящиківі (коробчасті) діаграми (boxplot, «ящик з вусами», біржова діаграма).
5. Центральний момент першого, другого, третього порядків.

Методичні рекомендації до практичної роботи:

1. Відкриваємо програму Statistica 8.0. Натискаємо ОК. Закриваємо всі вікна.
2. Натискаємо кнопку File → New () або Ctrl+N. Натискаємо ОК.
3. З'являється таблиця, в яку необхідно додати дані з бази даних (табл 1).

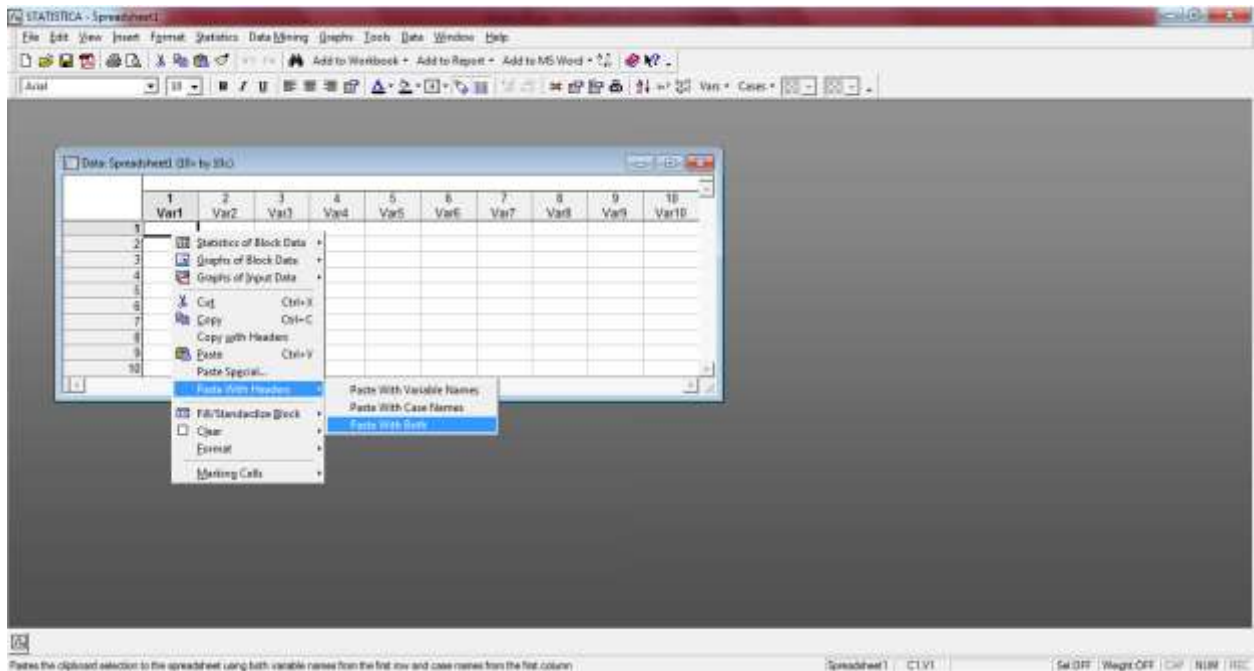
Таблиця 1

Показники демографічної ситуації населення по регіонах України у 2021 році

	живонароджен их на 1000 осіб наявного населення	померлих на 1000 осіб наявного населення	природний приріст (скорочення) на 1000 осіб наявного населення	шлюбів на 1000 осіб наявного населення	розірвань шлюбів на 1000 осіб наявного населення
Вінницька область	6,9	18,2	-11,3	5,3	3,3
Волинська область	9,6	15,1	-5,5	5,4	2,6
Дніпропетровська область	6,3	20,8	-14,5	6,0	3,7
Житомирська область	7,2	19,9	-12,7	5,1	3,2
Закарпатська область	10,1	14,8	-4,7	5,3	2,2
Запорізька область	5,8	21,6	-15,8	5,4	3,4
Івано-Франківська область	7,8	15,2	-7,4	5,2	2,9
Київська область	7,0	20,2	-13,2	6,2	3,8
Кіровоградська область	6,1	20,8	-14,7	4,5	3,2
Львівська область	7,8	16,0	-8,2	5,5	2,4
Миколаївська область	6,4	19,4	-13,0	5,4	3,4
Одеська область	8,2	18,0	-9,8	6,7	3,6
Полтавська область	5,9	20,3	-14,4	5,1	3,3
Рівненська область	10,2	14,9	-4,7	5,5	2,6
Сумська область	5,2	20,0	-14,8	4,7	3,0
Тернопільська область	7,1	16,4	-9,3	5,1	2,7
Харківська область	5,9	21,0	-15,1	5,9	3,6
Херсонська область	7,0	20,0	-13,0	5,3	3,2
Хмельницька область	7,0	19,0	-12,0	5,0	3,3
Черкаська область	5,8	19,6	-13,8	5,3	3,2
Чернівецька область	8,8	16,1	-7,3	5,5	2,8
Чернігівська область	5,5	22,0	-16,5	4,7	3,1
м.Київ	10,0	15,3	-5,3	8,6	3,6
Україна¹	7,3	18,5	-11,2	5,8	3,2

¹ Відносні показники розраховано без урахування відповідних даних по Донецькій та Луганській областях. У зв'язку з відсутністю інформації по частині тимчасово окупованих територій у цих областях здійснення розрахунків відносних показників по ним є некоректним

4. Для цього копіюємо табл. 1 і вставляємо її у таблицю програми Statistica. Дана операція робиться наступним чином. У першій ячійці правою кнопкою миші у вікні, що відкривається, натискаємо «Paste With Headers → Paste With Both». Таким чином, ми копіюємо у таблицю програми Statistica наші показники, а також назви показників та назви регіонів України (рис. 1).



STATISTICA - [Data: Spreadsheet2* (10v by 24c)]											
File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help											
Add to Workbook Add to Report Add to MS Word											
Arial 10 B I U											
	живога роджен ини	1000 осіб на 1000	мн приріст (на 1000)	осіб наяног	шлюбів на 1000	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10	
Вінницька область	6,9	18,2	-11,3	5,3	3,3						
Волинська область	9,6	15,1	-5,5	5,4	2,6						
Дніпропетровська область	6,3	20,8	-14,5	6	3,7						
Житомирська область	7,2	19,9	-12,7	5,1	3,2						
Закарпатська область	10,1	14,8	-4,7	5,3	2,2						
Запорізька область	5,8	21,6	-15,8	5,4	3,4						
Івано-Франківська область	7,8	15,2	-7,4	5,2	2,9						
Київська область	7	20,2	-13,2	6,2	3,8						
Кіровоградська область	6,1	20,8	-14,7	4,5	3,2						
Львівська область	7,8	16	-8,2	5,5	2,4						
Миколаївська область	6,4	19,4	-13	5,4	3,4						
Одеська область	8,2	18	-9,8	6,7	3,6						
Полтавська область	5,9	20,3	-14,4	5,1	3,3						
Рівненська область	10,2	14,9	-4,7	5,5	2,6						
Сумська область	5,2	20	-14,8	4,7	3						
Тернопільська область	7,1	16,4	-9,3	5,1	2,7						
Харківська область	5,9	21	-15,1	5,9	3,6						
Херсонська область	7	20	-13	5,3	3,2						
Хмельницька область	7	19	-12	5	3,3						
Черкаська область	5,8	19,6	-13,8	5,3	3,2						
Чернівецька область	8,8	16,1	-7,3	5,5	2,8						
Чернігівська область	5,5	22	-16,5	4,7	3,1						
м. Київ	10	15,3	-5,3	8,6	3,6						
24											



	1 живонароджених на 1000 осіб наявного населення	2 померлих на 1000 осіб наявного населення	3 природний приріст (скорочення) на 1000 осіб наявного населення	4 шлюбів на 1000 осіб наявного населення	5 розірвань шлюбів на 1000 осіб наявного населення	
Вінницька область	6,9	18,2	-11,3	5,3	3,3	
Волинська область	9,6	15,1	-5,5	5,4	2,6	
Дніпропетровська область	6,3	20,8	-14,5	6	3,7	
Житомирська область	7,2	19,9	-12,7	5,1	3,2	
Закарпатська область	10,1	14,8	-4,7	5,3	2,2	
Запорізька область	5,8	21,6	-15,8	5,4	3,4	
Івано-Франківська область	7,8	15,2	-7,4	5,2	2,9	
Київська область	7	20,2	-13,2	6,2	3,8	
Кіровоградська область	6,1	20,8	-14,7	4,5	3,2	
Львівська область	7,8	16	-8,2	5,5	2,4	
Миколаївська область	6,4	19,4	-13	5,4	3,4	
Одеська область	8,2	18	-9,8	6,7	3,6	
Полтавська область	5,9	20,3	-14,4	5,1	3,3	
Рівненська область	10,2	14,9	-4,7	5,5	2,6	
Сумська область	5,2	20	-14,8	4,7	3	
Тернопільська область	7,1	16,4	-9,3	5,1	2,7	
Харківська область	5,9	21	-15,1	5,9	3,6	
Херсонська область	7	20	-13	5,3	3,2	
Хмельницька область	7	19	-12	5	3,3	
Черкаська область	5,8	19,6	-13,8	5,3	3,2	
Чернівецька область	8,8	16,1	-7,3	5,5	2,8	
Чернігівська область	5,5	22	-16,5	4,7	3,1	
м.Київ	10	15,3	-5,3	8,6	3,6	

Рис. 1. Додавання бази даних у таблицю програми Statistica

5. Після цього на панелі задач натискаємо «Statistics → Basic Statistics/Tables → Descriptive Statistics → OK».
6. Відкривається діалогове вікно (рис. 2).

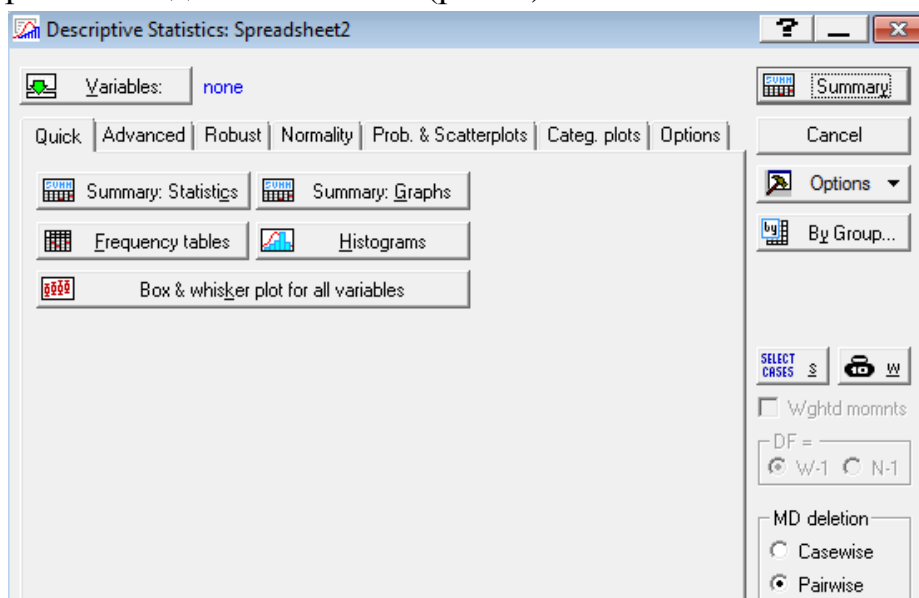


Рис. 2. Діалогове вікно описової статистики

7. Натискаємо кнопку «Variables» та обираємо показники для аналізу (рис. 3).

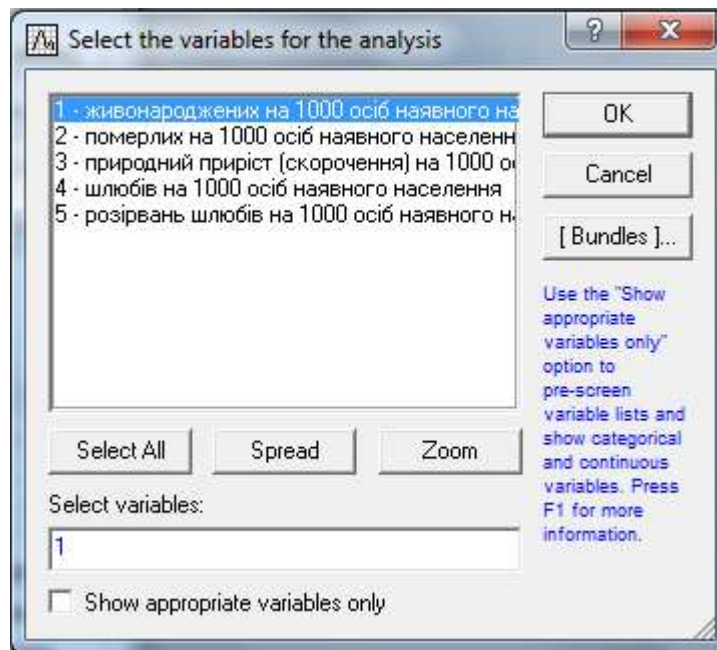


Рис. 3. Вибір показника для аналізу

8. Обираємо опцію «Advanced». Галочки ставимо навпроти «Valid N» (кількість значень), «Mean» (середня арифметична), «Sum» (сума), «Median» (медіана), «Mode» (мода), «Standard Deviation» (стандартне відхилення), «CI for Sample SD» (довірчий інтервал), «Coefficient of variation» (коефіцієнт варіації), «Variance» (дисперсія), «Skewness» (асиметрія), «Kurtosis» (ексцес), «Minimum & maximum» (мінімальне та максимальне значення), «Lower & upper quartiles» (нижній та верхній квартилі), «Range» (розмах), «Quartile range» (квартильний розмах) (рис. 4).

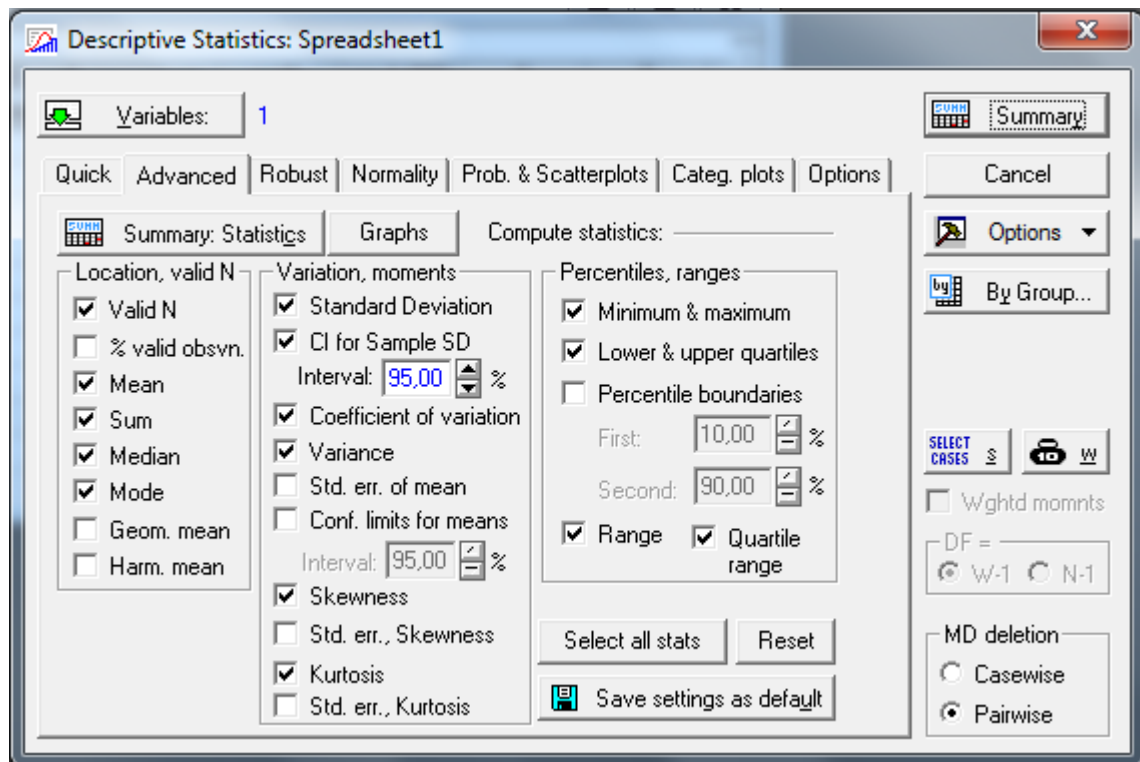


Рис. 4. Опції для описової статистики

9. В опції «Options» галочку ставимо навпроти «Median/Quartiles/Range». Тобто в графіку Box and Whiskers Plot буде відображатися медіана, квартилі, максимальне та мінімальне значення вибірки.
10. Далі ми повертаємося у вкладку «Quick» та натискаємо кнопку «Summary: Graph» (рис. 5).

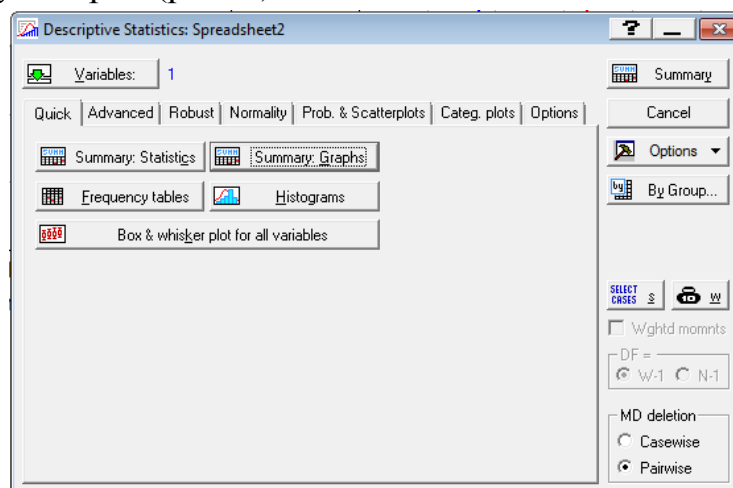


Рис. 5. Натискання кнопки «Summary: Graph»

11. Як результат, з'являється наступний результуючий графік (рис. 6). Для того, щоб його скопіювати, треба натиснути праву кнопку миші та у вікні, що відкривається, обрати «Copy Graph». За цим графіком можна

зрозуміти, чи відповідає даний розподіл нормальному.

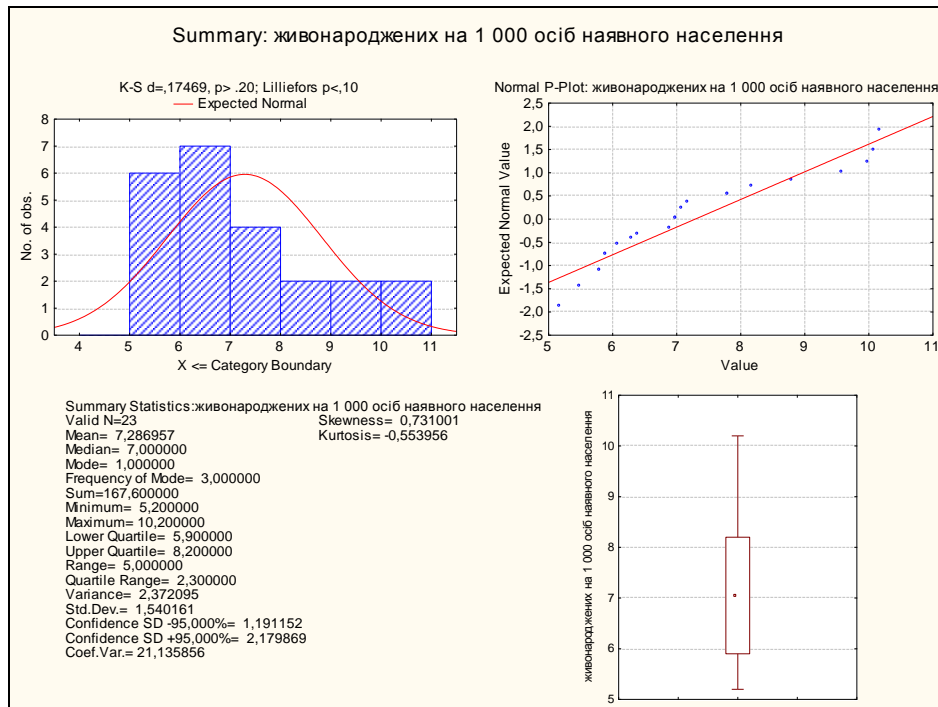


Рис. 6. Результуючий графік одновимірного статистичного аналізу

Аналізуючи отримані результати дескриптивного аналізу, можна зробити наступні висновки:

- дескриптивний аналіз проводився за 23-а значеннями (варіантами) (Valid N = 23);
- середня арифметична значень загального коефіцієнту народжуваності за регіонами України в 2021 р. дорівнює 7,3 ‰ (Mean = 7,286957);
- медіана значень загального коефіцієнту народжуваності за регіонами України в 2021 р. дорівнює 7,0 ‰ (Median = 7,000000);
- всі значення (варіанти) загального коефіцієнту народжуваності за регіонами України в 2021 р. зустрічаються по одному разу, тому моди немає (Mode = 1,000000);
- мінімальне значення загального коефіцієнту народжуваності за регіонами України в 2021 р. складає 5,2 ‰ (Minimum = 5,200000);
- максимальне значення загального коефіцієнту народжуваності за регіонами України в 2021 р. складає 10,2 ‰ (Maximum = 10,200000);
- різниця між максимальним і мінімальним значеннями (розмах) загального коефіцієнту народжуваності за регіонами України в 2021 р. складає 5,0 ‰ (Range = 5,000000);
- нижній або перший квантиль (Q25 або Q1 – 25 % процентиль) розподілу значень загального коефіцієнту народжуваності за регіонами України в 2021 р. дорівнює 5,9 ‰ (Lower Quartile = 5,900000). Це

означає, що нижче 5,9 ‰ розташовується 25 ‰ значень загального коефіцієнту народжуваності за регіонами України;

- верхній або третій кuartиль (Q75 або Q3 – 75 ‰ процентиль) розподілу значень загального коефіцієнту народжуваності за регіонами України в 2021 р. дорівнює 8,2 ‰ (Upper Quartile = 8,200000). Це означає, що нижче 8,2 ‰ розташовується 75 ‰ значень загального коефіцієнту народжуваності за регіонами України;
- різниця між верхнім і нижнім кuartилями (міжкuartильний розмах) розподілу значень загального коефіцієнту народжуваності за регіонами України в 2021 р. складає 2,3 ‰ (Quartile Range = 2,300000);
- коефіцієнт варіації розподілу значень загального коефіцієнту народжуваності за регіонами України в 2021 р. складає 21,1 ‰ (Coefficient of variation = 21,125856), тобто менше 33 ‰, що свідчить про однорідність вибірки та коректність обрахування мір центральної тенденції (середньої арифметичної, моди, медіани);
- дисперсія, тобто відхилення значень загального коефіцієнту народжуваності від центру розподілу, складає 2,4 ‰ (Variance = 2,372095);
- стандартне відхилення, тобто відхилення значень загального коефіцієнту народжуваності від середнього значення, складає 1,5 ‰ (Std. Deviation = 1,540161);
- асиметрія значень загального коефіцієнту народжуваності за регіонами України в 2021 р., дорівнює 0,7 (Skewness = 0,731001). Асиметрія свідчить про наявність або відсутність симетрії даних, тобто характеризує форму розподілу, його відмінність від нормального розподілу. В нашому випадку вона більше 0, тобто є позитивною. Це означає, що крива розподілу зміщена вліво або в бік менших значень (розподіл з «довгим правим хвостом»);
- ексцес значень загального коефіцієнту народжуваності за регіонами України в 2021 р., дорівнює -0,6 (Kurtosis = -0,553956). Ексцес також характеризує форму розподілу, його відмінність від нормального розподілу, а саме – крутість кривої розподілу. В нашому випадку він менше 0, тобто є негативним. Це означає, що крива розподілу більш полого, ніж нормальна крива.

12. Можна окремо виводити

гістограму розподілу – внизу натиснути «Descriptive Statistics», повернутись у вкладку «Quick» та натиснути кнопку «Histograms» (рис. 7);

«ящик з вусами» – внизу натиснути «Descriptive Statistics», повернутись у вкладку «Quick» та натиснути кнопку «Box and Whiskers Plot all variables» (рис. 8);

діаграму розсіювання – внизу натиснути «Descriptive Statistics», вибрати вкладку «Prob. Scatterplots» та натиснути кнопку «Normal probability plot» (рис. 9);

дескриптивні статистики – внизу натиснути «Descriptive Statistics», повернутись у вкладку «Quick» та натиснути кнопку «Summary: Statistics» (табл. 2);

показники ряду розподілу – внизу натиснути «Descriptive Statistics», повернутись у вкладку «Quick» та натиснути кнопку «Frequency tables» (табл. 3).

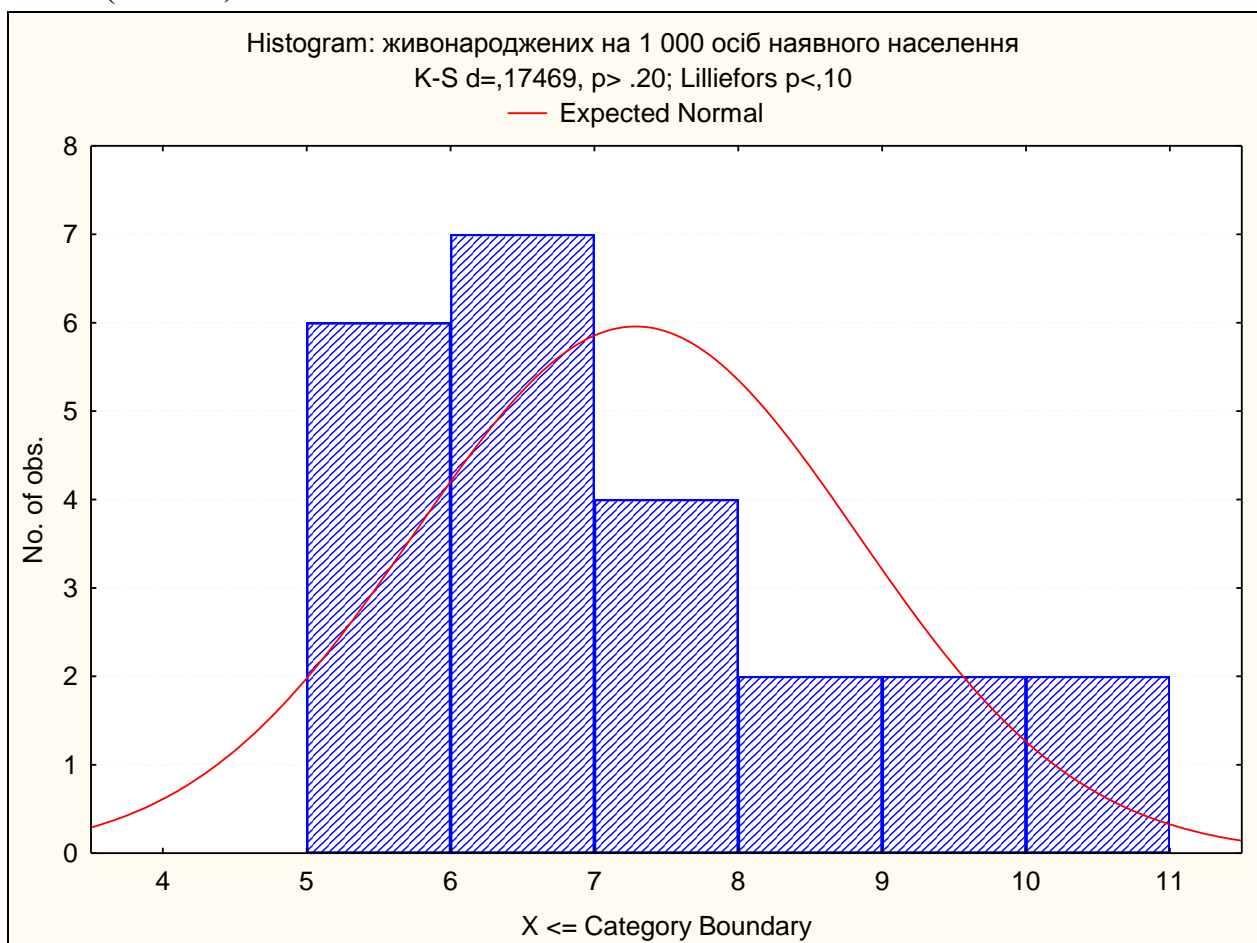


Рис. 7. Гістограма розподілу народжуваності за регіонами України в 2021 р.

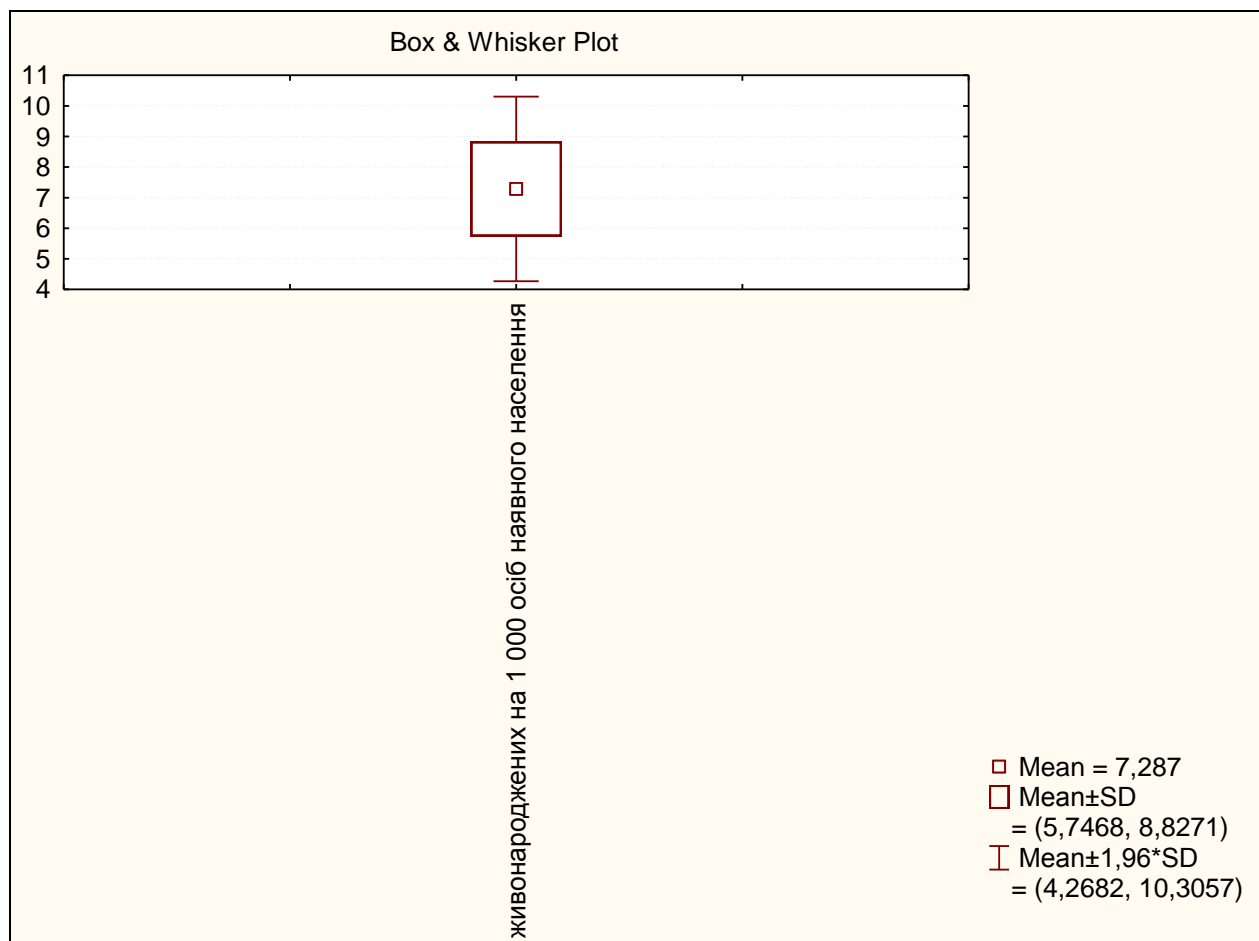


Рис. 8. Коробчаста діаграма («ящик з вусами») розподілу народжуваності за регіонами України в 2021 р.

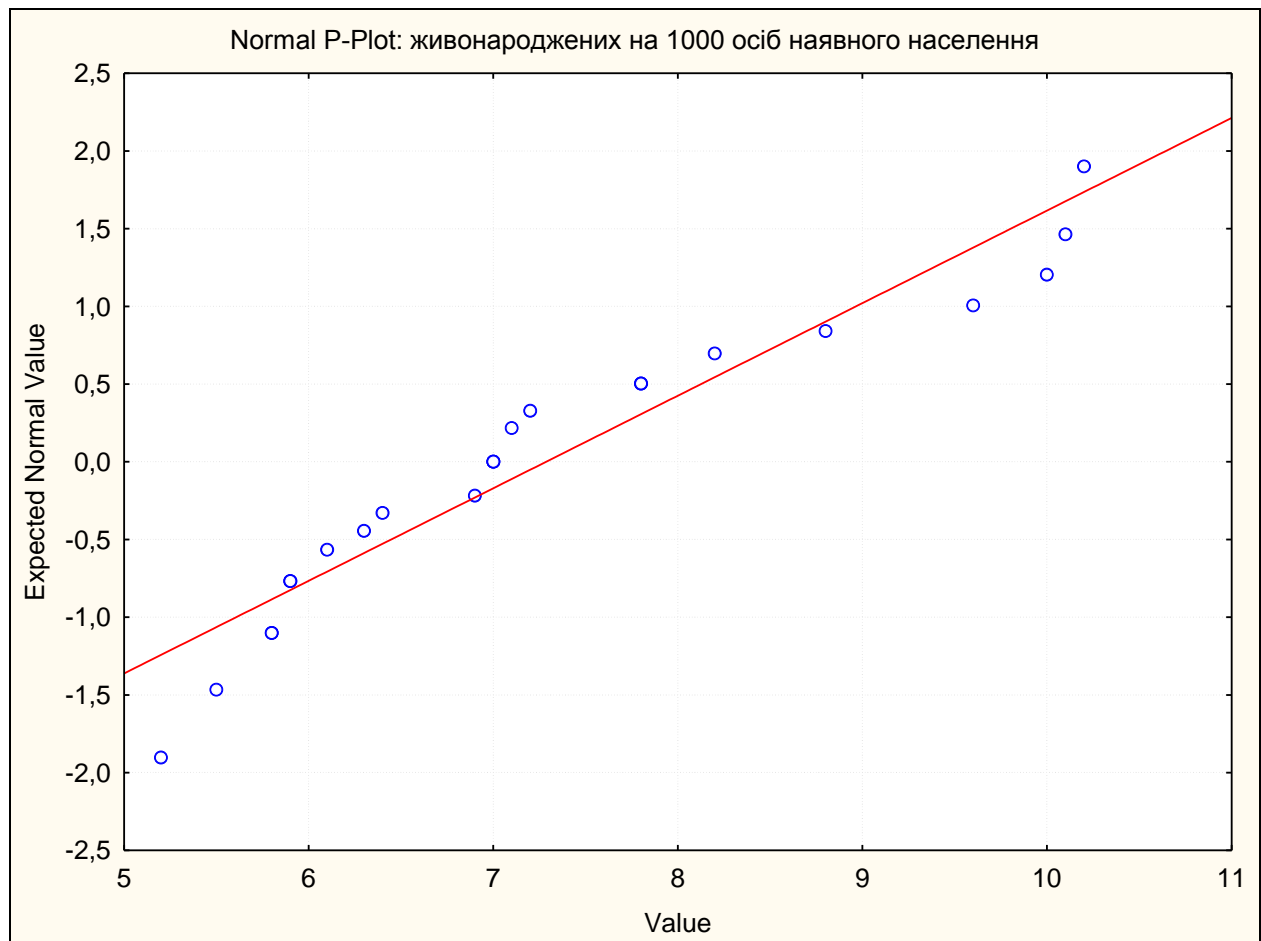


Рис. 9. Розсіяна діаграма розподілу народжуваності за регіонами України в 2021 р.

Таблиця 2

Дескриптивні статистики народжуваності за регіонами України в 2021 р.

Descriptive Statistics (Spreadsheet1)																			
	Valid N	Mean	Median	Mode	Frequency of Mode	Sum	Minimum	Maximum	Lower - Quartile	Upper - Quartile	Range	Quartile - Range	Variance	Std. Dev.	Confidence SD - 95,000 %	Confidence SD - +95,000 %	Coe f. Var .	Sk ewness	K urtosis
живонароджених на 1 000 осіб наявного населення	23	7,286957	7,000000	7,000000	3	167,000000	5,200000	10,200000	5,900000	8,200000	5,000000	2,300000	2,372095	1,540161	1,191152	2,179869	21,13586	0,731001	-0,553956

Таблиця 3

Показники ряду розподілу народжуваності

за регіонами України в 2021 р.

Frequency table: живонароджених на 1 000 осіб наявного населення (Spreadsheet1) K-S d=,17469, p> .20; Lilliefors p<,10

	Cou nt	Cumulative - Count	Percent - of Valid	Cumul % - of Valid	% of all - Cases	Cumulative % - of All
4,000000<x<=5,0 00000	0	0	0,00000	0,0000	0,00000	0,0000
5,000000<x<=6,0 00000	6	6	26,08696	26,0870	25,00000	25,0000
6,000000<x<=7,0 00000	7	13	30,43478	56,5217	29,16667	54,1667
7,000000<x<=8,0 00000	4	17	17,39130	73,9130	16,66667	70,8333
8,000000<x<=9,0 00000	2	19	8,69565	82,6087	8,33333	79,1667
9,000000<x<=10, 00000	2	21	8,69565	91,3043	8,33333	87,5000
10,00000<x<=11, 00000	2	23	8,69565	100,0000	8,33333	95,8333
Missing	1	24	4,34783		4,16667	100,0000

13. Для того щоб зробити одновимірний статистичний аналіз за іншим показником, треба внизу натиснути «Descriptive Statistics», у діалоговому вікні натискаємо кнопку «Variables» та обираємо новий показник для аналізу (рис. 10).

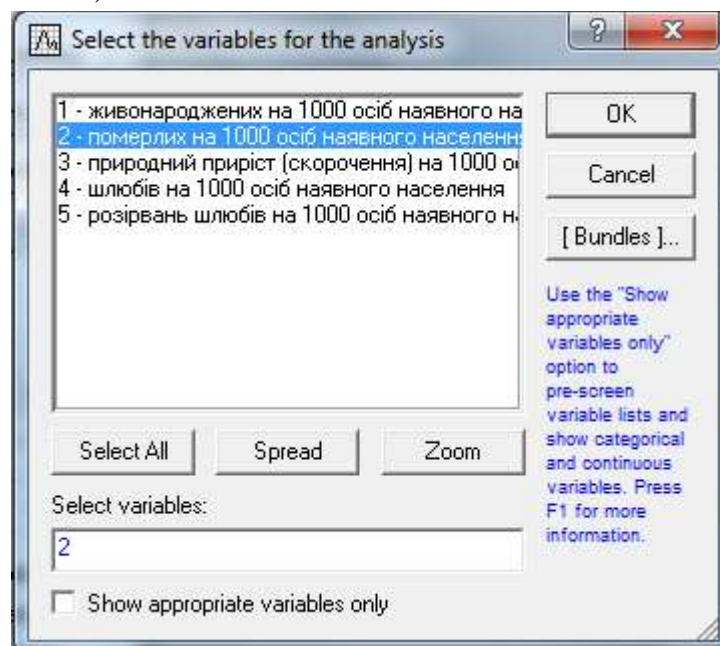


Рис. 10. Вибір показника для аналізу

Завдання до практичної роботи:


Завдання 1. Проведіть дескриптивний одновимірний статистичний аналіз якогось статистичного показника зі своєї бази даних, взятого в динаміці (за 2-3 роки). Проаналізуйте отримані параметри описової статистики за групами дескриптивних статистик: а) значення процентилей (Percentile); б) міри центральної тенденції (Central Tendency); в) міри варіації (Dispersion); г) характеристики форми розподілу (Distribution); д) підсумок з п'яти значень. Порівняйте їх за різні роки. Зробіть висновки. Візуалізуйте різними способами (окрім тих варіантів візуалізації, які надає програма «Statistica»).

Практичне заняття 13-14
ПОБУДОВА ДВОВИМІРНИХ СТАТИСТИЧНИХ МОДЕЛЕЙ НА
МАТРИЦІ ВИХІДНИХ ДАНИХ.
КОРЕЛЯЦІЙНИЙ СТАТИСТИЧНИЙ АНАЛІЗ

Питання для обговорення

1. Основні поняття кореляційного аналізу: кореляція, кореляційний зв'язок, кореляційна залежність, кореляційний аналіз.
2. Кореляційний аналіз: поняття, мета, завдання, види.
3. Кореляційні залежності та їх види за формою, напрямом, силою, значущістю.
4. Коефіцієнт кореляції: різноманіття і правила вибору.
5. Лінійний коефіцієнт кореляції Пірсона r (коефіцієнт парної кореляції) та коефіцієнт рангової кореляції Спірмена r_s .
6. Псевдокореляція.
7. Візуалізація кореляційного аналізу. Діаграма розсіювання.

Методичні рекомендації до практичної роботи:

1. Відкриваємо програму Statistica 8.0. Натискаємо ОК. Закриваємо всі вікна.
2. Натискаємо кнопку File → New () або Ctrl+N. Натискаємо ОК.
3. З'являється таблицю, в яку необхідно додати дані з бази даних (табл 1).

Таблиця 1

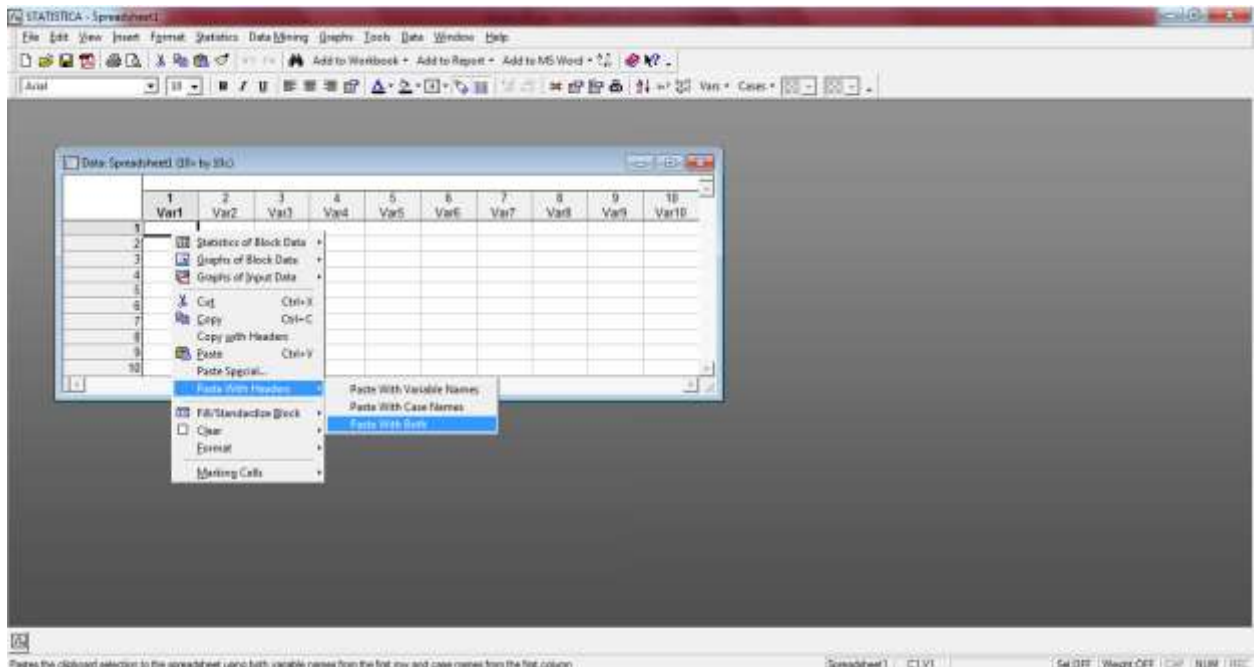
Показники демографічної ситуації населення по регіонах України у 2021 році

	Загальний коефіцієнт народжуван ості, ‰	Загальний коефіцієнт смертності, ‰	Коефіцієнт природного приросту, ‰	Коефіцієнт шлюбності, ‰	Коефіцієнт розлучуван ості, ‰	Частка населення у віці 0-14 років у загальній чисельності населення, %	Частка населення у віці 15-64 років у загальній чисельності населення, %	Частка населення у віці 65 років і старше у загальній чисельності населення, %	Середній вік, років	На 1000 чоловіків припадає жінок
Вінницька	6,9	18,2	-11,3	5,3	3,3	15,2	66,8	18,0	42,2	1156
Волинська	9,6	15,1	-5,5	5,4	2,6	19,3	67,2	13,5	38,5	1118
Дніпропетровська	6,3	20,8	-14,5	6,0	3,7	15,1	67,3	17,6	42,3	1190
Житомирська	7,2	19,9	-12,7	5,1	3,2	16,1	67,0	16,9	41,3	1144
Закарпатська	10,1	14,8	-4,7	5,3	2,2	19,5	68,2	12,3	37,8	1080
Запорізька	5,8	21,6	-15,8	5,4	3,4	14,2	67,3	18,5	43,1	1190
Івано-Франківська	7,8	15,2	-7,4	5,2	2,9	16,6	68,8	14,6	40,1	1115
Київська	7,0	20,2	-13,2	6,2	3,8	17,7	67,2	15,1	40,3	1184
Кіровоградська	6,1	20,8	-14,7	4,5	3,2	14,7	66,6	18,7	42,8	1166
Львівська	7,8	16,0	-8,2	5,5	2,4	16,1	68,6	15,3	40,6	1111
Миколаївська	6,4	19,4	-13,0	5,4	3,4	15,1	67,6	17,3	42,0	1156
Одеська	8,2	18,0	-9,8	6,7	3,6	16,8	67,1	16,1	40,6	1122
Полтавська	5,9	20,3	-14,4	5,1	3,3	14,0	67,9	18,1	43,0	1162
Рівненська	10,2	14,9	-4,7	5,5	2,6	20,1	67,0	12,9	37,9	1104
Сумська	5,2	20,0	-14,8	4,7	3,0	12,8	68,0	19,2	44,0	1178
Тернопільська	7,1	16,4	-9,3	5,1	2,7	15,4	68,7	15,9	41,2	1134
Харківська	5,9	21,0	-15,1	5,9	3,6	13,5	69,0	17,5	42,6	1155
Херсонська	7,0	20,0	-13,0	5,3	3,2	15,9	67,3	16,8	41,5	1152
Хмельницька	7,0	19,0	-12,0	5,0	3,3	15,5	66,9	17,6	42,0	1150

Черкаська	5,8	19,6	-13,8	5,3	3,2	13,6	67,1	19,3	43,6	1172
Чернівецька	8,8	16,1	-7,3	5,5	2,8	17,1	68,3	14,6	39,7	1120
Чернігівська	5,5	22,0	-16,5	4,7	3,1	13,2	66,5	20,3	44,2	1199
м.Київ	10,0	15,3	-5,3	8,6	3,6	16,5	67,8	15,7	40,4	1169
Україна¹	7,3	18,5	-11,2	5,8	3,2	14,9	67,4	17,7	42,2	1157

¹ Відносні показники розраховано без урахування відповідних даних по Донецькій та Луганській областях. У зв'язку з відсутністю інформації по частині тимчасово окупованих територій у цих областях здійснення розрахунків відносних показників по ним є некоректним

4. Для цього копіюємо табл. 1 і вставляємо її у таблицю програми Statistica. Дана операція робиться наступним чином. У першій ячійці правою кнопкою миші у вікні, що відкривається, натискаємо «Paste With Headers → Paste With Both». Таким чином, ми копіюємо у таблицю програми Statistica наші показники, а також назви показників та назви регіонів України (рис. 1).



STATISTICA - [Data: Spreadsheet2* (11v by 24c)]

	Загальний коефіцієнт	коєфіцієнт природн	ент	коєфіцієнт	коєфіцієнт	населен	років у загальні	і старше у	Середній вік,	На 1000 чоловік	11 NewVar
Вінницька	6,9	18,2	-11,3	5,3	3,3	15,2	66,8	18	42,2	1156	
Волинська	9,6	15,1	-5,5	5,4	2,6	19,3	67,2	13,5	38,5	1118	
Дніпропетровська	6,3	20,8	-14,5	6	3,7	15,1	67,3	17,6	42,3	1190	
Житомирська	7,2	19,9	-12,7	5,1	3,2	16,1	67	16,9	41,3	1144	
Закарпатська	10,1	14,8	-4,7	5,3	2,2	19,5	68,2	12,3	37,8	1080	
Запорізька	5,8	21,6	-15,8	5,4	3,4	14,2	67,3	18,5	43,1	1190	
Івано-Франківська	7,8	15,2	-7,4	5,2	2,9	16,6	68,8	14,6	40,1	1115	
Київська	7	20,2	-13,2	6,2	3,8	17,7	67,2	15,1	40,3	1184	
Кіровоградська	6,1	20,8	-14,7	4,5	3,2	14,7	66,6	18,7	42,8	1166	
Львівська	7,8	16	-8,2	5,5	2,4	16,1	68,6	15,3	40,6	1111	
Миколаївська	6,4	19,4	-13	5,4	3,4	15,1	67,6	17,3	42	1156	
Одеська	8,2	18	-9,8	6,7	3,6	16,8	67,1	16,1	40,6	1122	
Полтавська	5,9	20,3	-14,4	5,1	3,3	14	67,9	18,1	43	1162	
Рівненська	10,2	14,9	-4,7	5,5	2,6	20,1	67	12,9	37,9	1104	
Сумська	5,2	20	-14,8	4,7	3	12,8	68	19,2	44	1178	
Тернопільська	7,1	16,4	-9,3	5,1	2,7	15,4	68,7	15,9	41,2	1134	
Харківська	5,9	21	-15,1	5,9	3,6	13,5	69	17,5	42,6	1155	
Херсонська	7	20	-13	5,3	3,2	15,9	67,3	16,8	41,5	1152	
Хмельницька	7	19	-12	5	3,3	15,5	66,9	17,6	42	1150	
Черкаська	5,8	19,6	-13,8	5,3	3,2	13,6	67,1	19,3	43,6	1172	
Чернівецька	8,8	16,1	-7,3	5,5	2,8	17,1	68,3	14,6	39,7	1120	
Чернігівська	5,5	22	-16,5	4,7	3,1	13,2	66,5	20,3	44,2	1199	
м. Київ	10	15,3	-5,3	8,6	3,6	16,5	67,8	15,7	40,4	1169	
24											



STATISTICA - [Data: Spreadsheet2* (10v by 23c)]

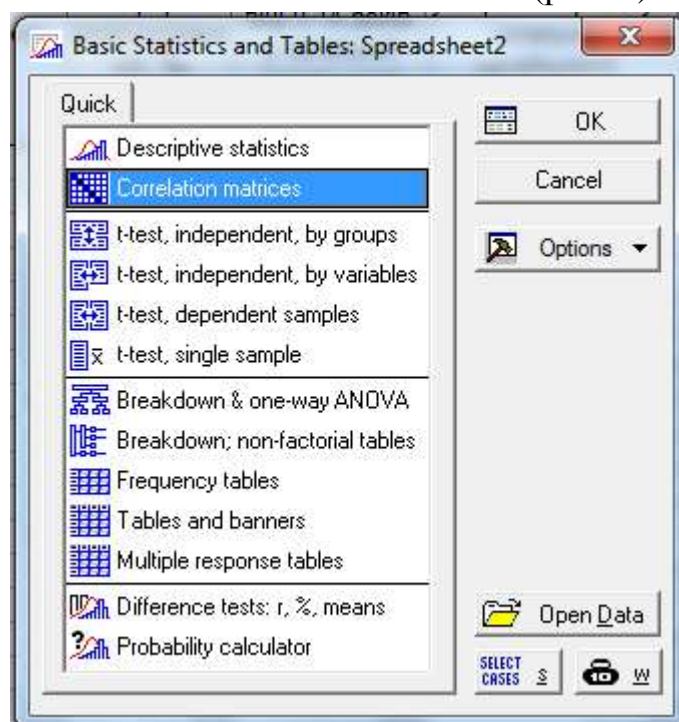
File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

Anal 10 B I U

	1	2	3	4	5	6	7	8	9	10
	Загальний коефіцієнт народжуван ості, %	Загальний коефіцієнт смертності, %	Коефіцієнт природного приросту, %	Коефіцієнт щільності, %	Коефіцієнт розпучуван ості, %	Частка населення у віці 0-14 років у загальній чисельності населення, %	Частка населення у віці 15-64 років у загальній чисельності населення, %	Частка населення у віці 65 років і старше у загальній чисельності населення, %	Середній вік, років	На 1000 чоловіків припадає жінок
Вінницька	6,9	18,2	-11,3	5,3	3,3	15,2	66,8	18	42,2	1156
Волинська	9,6	15,1	-5,5	5,4	2,6	19,3	67,2	13,5	38,5	1118
Дніпропетровська	6,3	20,8	-14,5	6	3,7	15,1	67,3	17,6	42,3	1190
Житомирська	7,2	19,9	-12,7	5,1	3,2	16,1	67	16,9	41,3	1144
Закарпатська	10,1	14,8	-4,7	5,3	2,2	19,5	68,2	12,3	37,8	1080
Запорізька	5,8	21,6	-15,8	5,4	3,4	14,2	67,3	18,5	43,1	1190
Івано-Франківська	7,8	15,2	-7,4	5,2	2,9	16,6	68,8	14,6	40,1	1115
Київська	7	20,2	-13,2	6,2	3,8	17,7	67,2	15,1	40,3	1184
Кіровоградська	6,1	20,8	-14,7	4,5	3,2	14,7	66,6	18,7	42,8	1166
Львівська	7,8	16	-8,2	5,5	2,4	16,1	68,6	15,3	40,6	1111
Миколаївська	6,4	19,4	-13	5,4	3,4	15,1	67,6	17,3	42	1156
Одеська	8,2	18	-9,8	6,7	3,6	16,8	67,1	16,1	40,6	1122
Полтавська	5,9	20,3	-14,4	5,1	3,3	14	67,9	18,1	43	1162
Рівненська	10,2	14,9	-4,7	5,5	2,6	20,1	67	12,9	37,9	1104
Сумська	5,2	20	-14,8	4,7	3	12,8	68	19,2	44	1178
Тернопільська	7,1	16,4	-9,3	5,1	2,7	15,4	68,7	15,9	41,2	1134
Харківська	5,9	21	-15,1	5,9	3,6	13,5	69	17,5	42,6	1155
Херсонська	7	20	-13	5,3	3,2	15,9	67,3	16,8	41,5	1152
Хмельницька	7	19	-12	5	3,3	15,5	66,9	17,6	42	1150
Черкаська	5,8	19,6	-13,8	5,3	3,2	13,6	67,1	19,3	43,6	1172
Чернівецька	8,8	16,1	-7,3	5,5	2,8	17,1	68,3	14,6	39,7	1120
Чернівецька	5,5	22	-16,5	4,7	3,1	13,2	66,5	20,3	44,2	1199
м. Київ	10	15,3	-5,3	8,6	3,6	16,5	67,8	15,7	40,4	1169

Рис. 1. Додавання бази даних у таблицю програми Statistica

5. Після цього на панелі задач натискаємо «Statistics → Basic Statistics/Tables → Correlation Matrix → OK» (рис. 2).



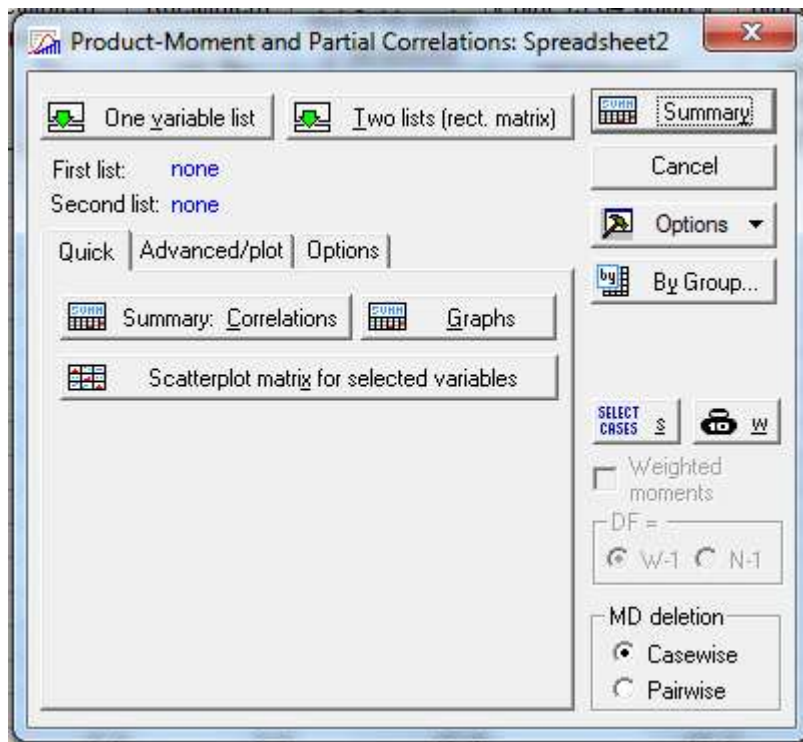


Рис. 2. Діалогове вікно виконання кореляційного аналізу

6. Натискаємо кнопку «Two lists (rect. matrix)». У вікні, що відкрилося, обираємо необхідні показники («Select All»). Натискаємо «OK» (рис. 3).

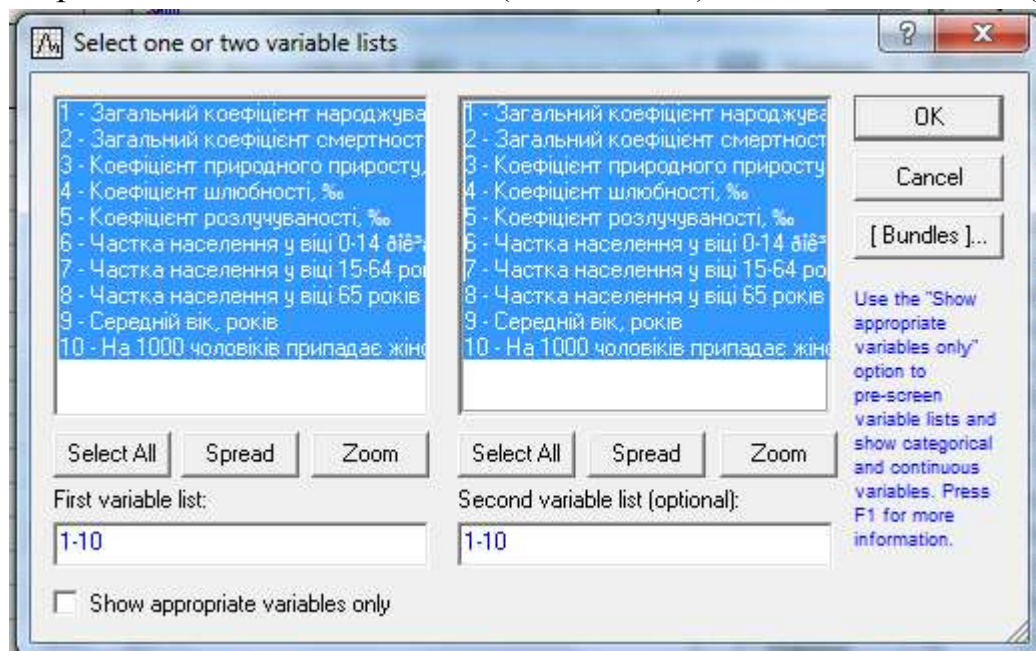


Рис. 3. Вибір показників для аналізу

7. Обираємо вкладку «Option» та перевіряємо, щоб у рядку «p-level for highlighting» було «,05». Це і є р-рівень – статистичний рівень значущості. При такому р-рівні результати нашого статистичного аналізу будуть статистично значущими (рис. 4).

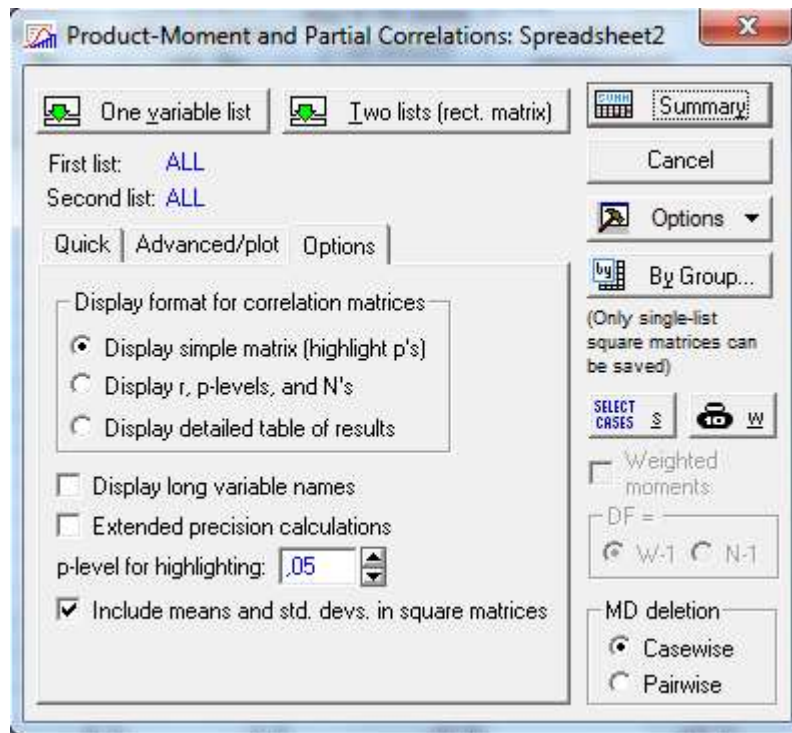


Рис. 4. Перевірка статистичного рівня значимості (p-рівень)

Після цього переходимо на вкладку «Quick» та натискаємо на кнопку «Summary: Correlations». В результаті отримуємо кореляційну матрицю. Червоним кольором підсвічуються ті коефіцієнти кореляції, для яких $p < 0,05$. Ці коефіцієнти кореляції значимі та можуть враховуватися при аналізі (рис. 5).

Correlations (Spreadsheet2)										
Marked correlations are significant at $p < .05000$										
N=23 (Casewise deletion of missing data)										
Variable	Загальний коефіцієнт народжуваності, ‰	Загальний коефіцієнт смертності, ‰	Коефіцієнт природного приросту, ‰	Коефіцієнт шлюбності, ‰	Коефіцієнт розлучуваності, ‰	Частка населення у віці 0-14 років у загальній чисельності населення, ‰	Частка населення у віці 15-64 років у загальній чисельності населення, ‰	Частка населення у віці 65 років і старші у загальній чисельності населення, ‰	Середній вік, років	На 1000 чоловіків припадає жінок
Загальний коефіцієнт народжуваності, ‰	1.00	-0.88	0.96	0.46	-0.48	0.91	0.13	-0.90	-0.94	-0.75
Загальний коефіцієнт смертності, ‰	-0.88	1.00	-0.98	-0.29	0.65	-0.76	-0.39	0.85	0.84	0.83
Коефіцієнт природного приросту, ‰	0.96	-0.98	1.00	0.37	-0.60	0.64	0.30	-0.89	-0.90	-0.82
Коефіцієнт шлюбності, ‰	0.46	-0.29	0.37	1.00	0.41	0.25	0.12	-0.28	-0.30	-0.03
Коефіцієнт розлучуваності, ‰	-0.48	0.65	-0.60	0.41	1.00	-0.45	-0.31	0.53	0.51	0.73
Частка населення у віці 0-14 років у загальній чисельності населення, ‰	0.91	-0.76	0.64	0.25	-0.45	1.00	0.00	-0.94	-0.97	-0.73
Частка населення у віці 15-64 років у загальній чисельності населення, ‰	0.13	-0.39	0.30	0.12	-0.31	0.00	1.00	0.36	-0.21	-0.40
Частка населення у віці 65 років і старші у загальній чисельності населення, ‰	-0.90	0.85	-0.89	-0.28	0.53	-0.94	-0.35	1.00	0.99	0.82
Середній вік, років	-0.94	0.84	-0.90	-0.30	0.51	-0.97	-0.21	0.89	1.00	0.82
На 1000 чоловіків припадає жінок	-0.75	0.83	-0.82	0.03	0.73	-0.73	-0.40	0.82	0.82	1.00

Рис. 5. Кореляційна матриця

8. Аналіз кореляційної матриці дозволяє зробити наступні висновки:

1) дуже сильна (коефіцієнт кореляції $> \pm 0,90-0,99$) позитивна кореляційна залежність характерна для таких показників, як

- коефіцієнти народжуваності та природного приросту ($r = 0,96$),
- частка населення у віці 0-14 років у загальній чисельності населення та загальний коефіцієнт народжуваності ($r = 0,91$),

- середній вік та частка населення у віці 65 років і старше у загальній чисельності населення ($r = 0,99$).

Це означає, що зі збільшенням/зменшенням значення одного показника, збільшується/зменшується значення іншого показника. Обидва показника міняються у тому ж напрямі;

2) дуже сильна (коефіцієнт кореляції $> \pm 0,90-0,99$) негативна кореляційна залежність характерна для таких показників, як

- загальний коефіцієнт народжуваності та частка населення у віці 65 років і старше у загальній чисельності населення ($r = -0,90$),
- загальний коефіцієнт народжуваності та середній вік населення ($r = -0,94$),
- коефіцієнти смертності та природного приросту ($r = -0,98$),
- коефіцієнт природного приросту та середній вік населення ($r = -0,90$),
- частки населення у віці 0-14 років та 65 років і старше у загальній чисельності населення ($r = -0,94$),
- частка населення у віці 0-14 років у загальній чисельності населення та середній вік населення ($r = -0,97$).

Це означає, що зі збільшенням/зменшенням значення одного показника, зменшується/збільшується значення іншого показника. Показники рухаються у протилежних напрямках;

3) сильна (коефіцієнт кореляції $\pm 0,70-0,89$) позитивна кореляційна залежність характерна для таких показників, як

- загальний коефіцієнт смертності та частка населення у віці 65 років і старше у загальній чисельності населення ($r = 0,85$),
- загальний коефіцієнт смертності та середній вік населення ($r = 0,84$),
- загальний коефіцієнт смертності та на 1000 чоловіків припадає жінок ($r = 0,83$),
- частка населення у віці 0-14 років у загальній чисельності населення та коефіцієнт природного приросту ($r = 0,84$),
- коефіцієнт розлучуваності та на 1000 чоловіків припадає жінок ($r = 0,73$),
- частка населення у віці 65 років і старше у загальній чисельності населення та на 1000 чоловіків припадає жінок ($r = 0,82$),
- середній вік населення та на 1000 чоловіків припадає жінок ($r = 0,82$);

4) сильна (коефіцієнт кореляції $\pm 0,70-0,89$) негативна кореляційна залежність характерна для таких показників, як

- загальні коефіцієнти народжуваності та смертності ($r = -0,88$),
- загальний коефіцієнт народжуваності та на 1000 чоловіків припадає жінок ($r = -0,75$),

- загальний коефіцієнт смертності та частка населення у віці 0-14 років у загальній чисельності населення ($r = -0,76$),
- коефіцієнт природного приросту та частка населення у віці 65 років і старше у загальній чисельності населення ($r = -0,89$),
- коефіцієнт природного приросту та на 1000 чоловіків припадає жінок ($r = -0,82$),
- частка населення у віці 0-14 років у загальній чисельності населення та на 1000 чоловіків припадає жінок ($r = -0,73$);

5) значна (коефіцієнт кореляції $\pm 0,50-0,69$) позитивна кореляційна залежність характерна для таких показників, як

- загальний коефіцієнт смертності та коефіцієнт розлучуваності ($r = 0,65$),
- коефіцієнт розлучуваності та частка населення у віці 65 років і старше у загальній чисельності населення ($r = 0,53$),
- коефіцієнт розлучуваності та середній вік населення ($r = 0,51$);

6) значна (коефіцієнт кореляції $\pm 0,50-0,69$) негативна кореляційна залежність характерна для таких показників, як

- коефіцієнти природного приросту та розлучуваності ($r = -0,60$);

7) помірна (коефіцієнт кореляції $\pm 0,30-0,49$) позитивна кореляційна залежність характерна для таких показників, як

- коефіцієнти народжуваності та шлюбності ($r = 0,46$);

8) помірна (коефіцієнт кореляції $\pm 0,30-0,49$) негативна кореляційна залежність характерна для таких показників, як

- коефіцієнти народжуваності та розлучуваності ($r = -0,48$),
- коефіцієнт розлучуваності та частка населення у віці 0-14 років у загальній чисельності населення ($r = -0,45$).

З 40-а пар показників (10 показників можуть між собою утворити 40 пар, якщо не враховувати парою показник із самим собою) статистично значущі кореляції показали 29 пар показників. Найбільше кореляційних зв'язків (8 з 9-и можливих) показав такий показник як загальний коефіцієнт народжуваності (корелював зі всіма показниками, окрім частки населення у віці 15-64 років у загальній чисельності населення). Жодного статистично значущого кореляційного зв'язку не показав такий показник як частка населення у віці 15-64 років у загальній чисельності населення; всього один статистично значущий кореляційний зв'язок показав такий показник як коефіцієнт шлюбності (корелював тільки із загальним коефіцієнтом народжуваності).

9. Одним з найкращих варіантів візуалізації кореляційного аналізу, є діаграма розсіювання. Її можна побудувати як у програмі Statistica, так і в програмі Excel. В першому випадку слід повернутися у діалогове вікно виконання кореляційного аналізу (рис. 2). Для цього в лівому нижньому куті натиснути «Product-Moment and Partial Correlations» та у вкладці «Advanced/plot» обрати «2D scatterplot». Програма побудує всі діаграми розсіювання для всіх можливих пар показників. Наприклад, на рис. 6 показана діаграма розсіювання для таких показників як загальні коефіцієнти народжуваності та смертності. Кожна точка – це регіон України, який характеризується двома показниками – загальним коефіцієнтом народжуваності (відкладений по горизонтальній осі x) та загальним коефіцієнтом народжуваності смертності (відкладений по вертикальній осі y).

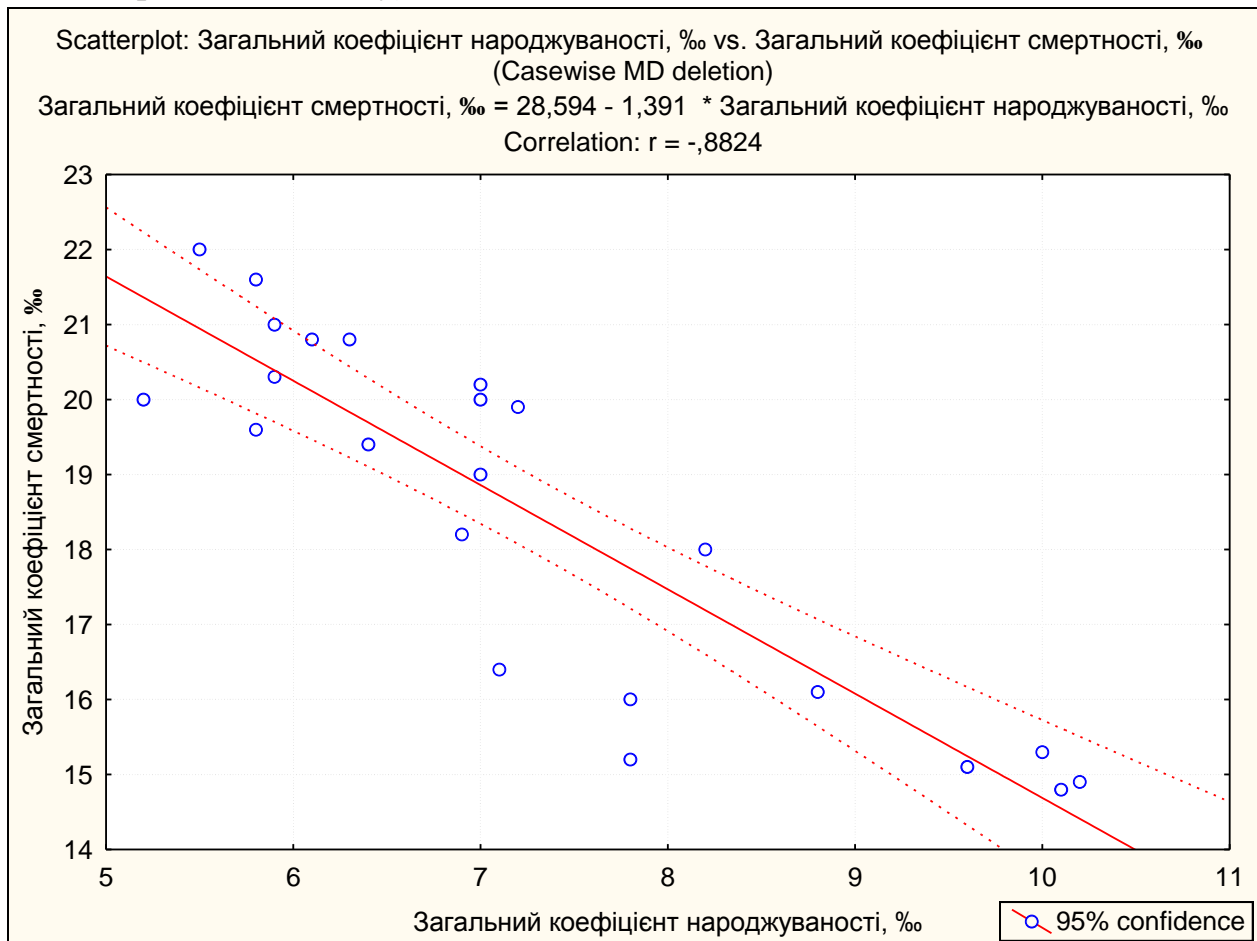


Рис. 6. Діаграма розсіювання для показників народжуваності та смертності за регіонами України в 2021 р.

В програмі Excel діаграму розсіювання для кожної пари показників слід робити окремо. Для цього обираємо такий вид діаграми як точкова діаграма (рис. 7).

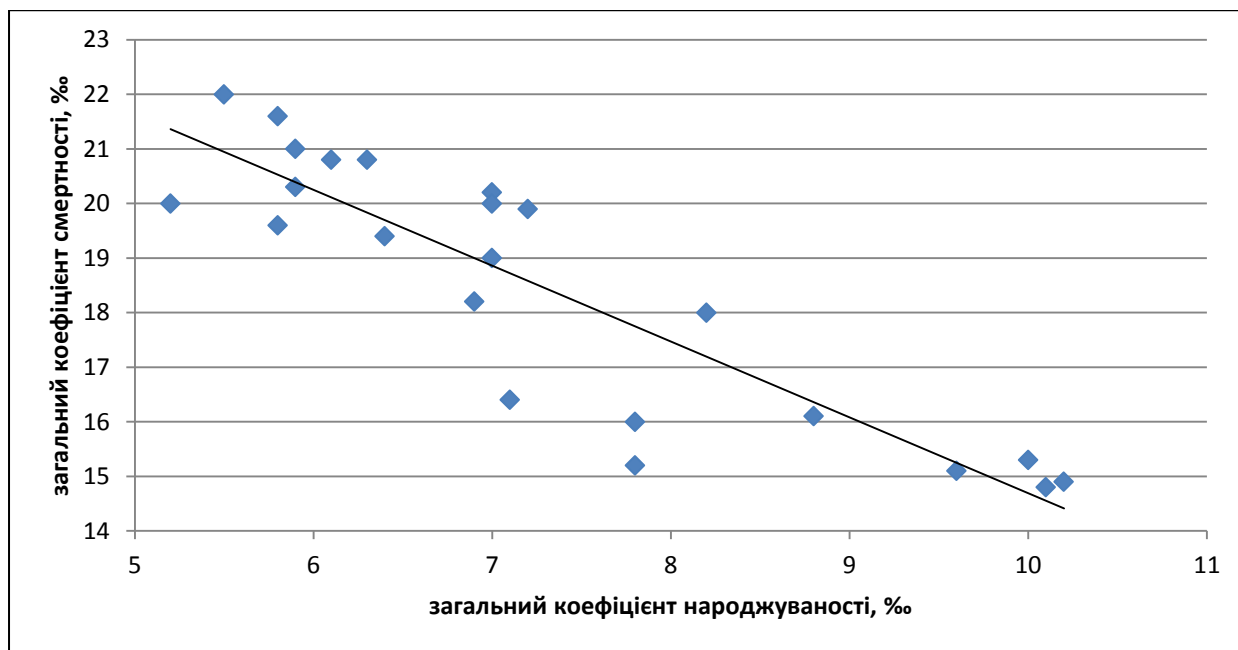


Рис. 7. Діаграма розсіювання для показників народжуваності та смертності за регіонами України в 2021 р.

Можна диференціювати пуансони регіонів на діаграмі. Наприклад, проранжирувати регіони України за загальним коефіцієнтом народжуваності, поділити їх на групи (наприклад, регіони із загальним коефіцієнтом народжуваності 5,0-5,9 ‰; 6,0-6,9 ‰; 7,0-7,9 ‰; 8,0-8,9 ‰; 9,0-9,9 ‰; 10,0-10,9 ‰). Побудувати в Excel наступну робочу таблицю (табл. 1). Побудувати по ній діаграму розсіювання (рис. 8). В легенді діаграми за потреби можна перерахувати назви регіонів. Можна ранжирувати регіони не за народжуваністю, а за смертністю (все це залежить від суб'єктивного рішення дослідника).

Таблиця 1

Регіони України	Загальний коефіцієнт народжуваності, ‰	Загальний коефіцієнт смертності, ‰					
Сумська	5,2	20					
Чернігівська	5,5	22					
Запорізька	5,8	21,6					
Черкаська	5,8	19,6					
Полтавська	5,9	20,3					
Харківська	5,9	21					
Кіровоградська	6,1	20,8					
Дніпропетровська	6,3	20,8					
Миколаївська	6,4	19,4					
Вінницька	6,9	18,2					

Київська	7			20,2			
Херсонська	7			20			
Хмельницька	7			19			
Тернопільська	7,1			16,4			
Житомирська	7,2			19,9			
Івано-Франківська	7,8			15,2			
Львівська	7,8			16			
Одеська	8,2				18		
Чернівецька	8,8				16,1		
Волинська	9,6					15,1	
м.Київ	10						15,3
Закарпатська	10,1						14,8
Рівненська	10,2						14,9

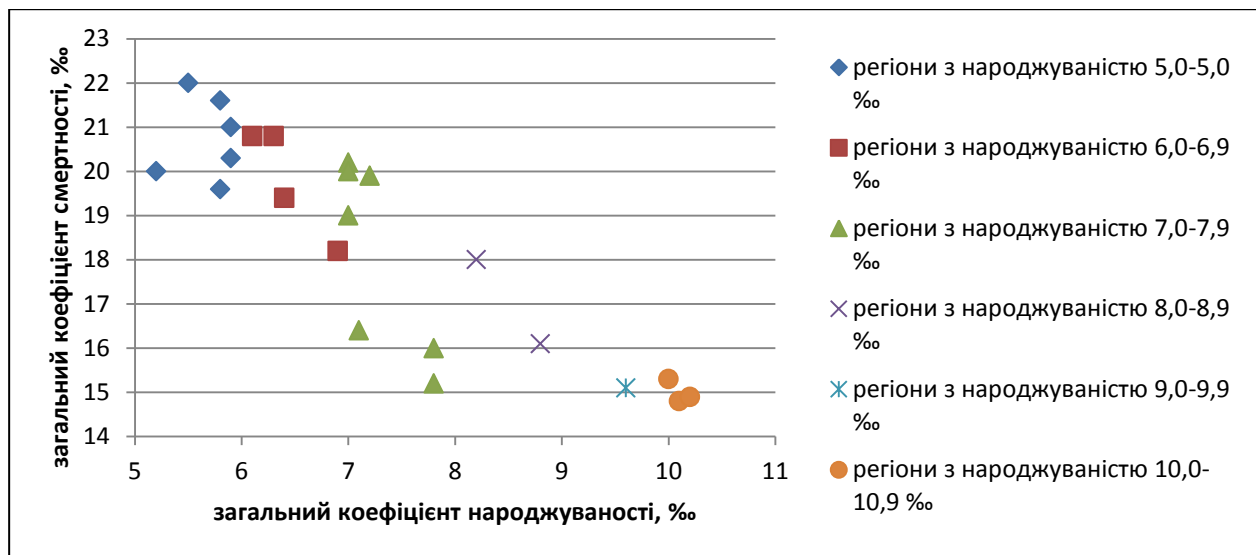


Рис. 8. Діаграма розсіювання для показників народжуваності та смертності за регіонами України в 2021 р.

Завдання до практичної роботи:

Завдання 1. Проведіть кореляційний аналіз показників зі своєї бази даних (5-10 показників). Бажано для аналізу обирати показники, які, на ваш погляд, взаємозалежать між собою. Результатом кореляційного аналізу є кореляційна матриця. Наведіть її. Користуючись табл. 2, опишіть результати кореляційного аналізу за прикладом (як це зроблено в методичних рекомендаціях до практичної роботи).

Таблиця 2

Відповідність значення коефіцієнту кореляції та тісноти зв'язку ознак

Коефіцієнт кореляції	Тіснота зв'язку
1,00	Зв'язок функціональний
$> \pm 0,90-0,99$	Дуже сильний
$\pm 0.70-0.89$	Сильний
$\pm 0.50-0.69$	Значний
0.30-0,49	Помірний
$\pm 0.20-0.29$	Слабкий
$< \pm 0.20$	Дуже слабкий або відсутній

Завдання 2. В Excel та Statistica побудуйте діаграми розсіювання за якоюсь парою показників серед тих, які ви обрали для кореляційного аналізу (з попереднього завдання). Зробіть в Excel диференціацію пуансонів.


Практичне заняття 15-16

ПОБУДОВА ДВОВИМІРНИХ СТАТИСТИЧНИХ МОДЕЛЕЙ НА МАТРИЦІ ВИХІДНИХ ДАНИХ. РЕГРЕСІЙНИЙ СТАТИСТИЧНИЙ АНАЛІЗ

Питання для обговорення

1. Кореляційний і регресійний аналіз: риси схожості і відмінності.
2. Регресійний аналіз: поняття, мета, завдання.
3. Види регресійного аналізу. Лінійний регресійний аналіз.
4. Етапи регресійного аналізу.
5. Рівняння регресії і лінія регресії.
6. Метод найменших квадратів.
7. Показники регресійного аналізу.
8. Аналіз залишків.
9. Припущення та обмеження регресійного аналізу.
10. Вимоги до регресійних моделей.

Методичні рекомендації до практичної роботи:

1. Відкриваємо програму Statistica 8.0. Натискаємо ОК. Закриваємо всі вікна.
2. Натискаємо кнопку File → New () або Ctrl+N. Натискаємо ОК.
3. З'являється таблиця, в яку необхідно додати дані з бази даних (табл 1).

Таблиця 1

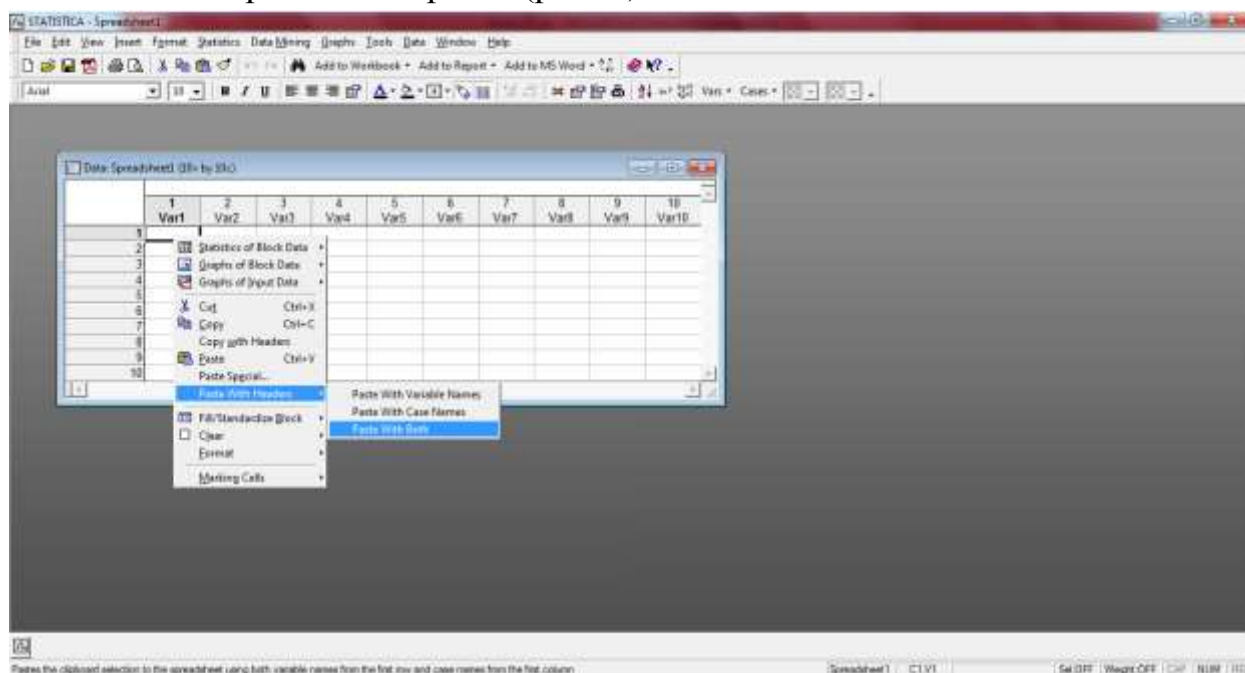
Показники демографічної ситуації населення по регіонах України у 2021 році

	Середній вік, років	Загальний коефіцієнт народжуван ості, ‰	Частка населення у віці 0-14 років у загальній чисельності населення, %	Частка населення у віці 15-64 років у загальній чисельності населення, %	Частка населення у віці 65 років і старше у загальній чисельності населення, %	Середня очікувана тривалість життя при народженні, років
Вінницька	42,2	6,9	15,2	66,8	18,0	71,18
Волинська	38,5	9,6	19,3	67,2	13,5	70,70
Дніпропетровська	42,3	6,3	15,1	67,3	17,6	68,42
Житомирська	41,3	7,2	16,1	67,0	16,9	68,21
Закарпатська	37,8	10,1	19,5	68,2	12,3	69,45
Запорізька	43,1	5,8	14,2	67,3	18,5	68,64
Івано-Франківська	40,1	7,8	16,6	68,8	14,6	72,03
Київська	40,3	7,0	17,7	67,2	15,1	68,10

Кіровоградська	42,8	6,1	14,7	66,6	18,7	68,63
Львівська	40,6	7,8	16,1	68,6	15,3	71,51
Миколаївська	42,0	6,4	15,1	67,6	17,3	69,09
Одеська	40,6	8,2	16,8	67,1	16,1	69,21
Полтавська	43,0	5,9	14,0	67,9	18,1	70,01
Рівненська	37,9	10,2	20,1	67,0	12,9	69,92
Сумська	44,0	5,2	12,8	68,0	19,2	70,34
Тернопільська	41,2	7,1	15,4	68,7	15,9	72,11
Харківська	42,6	5,9	13,5	69,0	17,5	68,73
Херсонська	41,5	7,0	15,9	67,3	16,8	68,16
Хмельницька	42,0	7,0	15,5	66,9	17,6	70,27
Черкаська	43,6	5,8	13,6	67,1	19,3	70,57
Чернівецька	39,7	8,8	17,1	68,3	14,6	71,27
Чернігівська	44,2	5,5	13,2	66,5	20,3	68,94
м.Київ	40,4	10,0	16,5	67,8	15,7	71,07
Україна¹	42,2	7,3	14,9	67,4	17,7	69,77

¹ Відносні показники розраховано без урахування відповідних даних по Донецькій та Луганській областях. У зв'язку з відсутністю інформації по частині тимчасово окупованих територій у цих областях здійснення розрахунків відносних показників по ним є некоректним

- Для цього копіюємо табл. 1 і вставляємо її у таблицю програми Statistica. Дана операція робиться наступним чином. У першій ячійці правою кнопкою миші у вікні, що відкривається, натискаємо «Paste With Headers → Paste With Both». Таким чином, ми копіюємо у таблицю програми Statistica наші показники, а також назви показників та назви регіонів України (рис. 1).





STATISTICA - [Data: Spreadsheet2* (10v by 24c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

Add to Workbook Add to Report Add to MS Word

Arial 10 B I U

	Середній вік,	коефіцієнт	років у загальні	років у загальні	старше у середній	тривалість	7 Var7	8 Var8	9 Var9	10 Var10	
Вінницька	42,2	6,9	15,2	66,8	18	71,18					
Волинська	38,5	9,6	19,3	67,2	13,5	70,7					
Дніпропетровська	42,3	6,3	15,1	67,3	17,6	68,42					
Житомирська	41,3	7,2	16,1	67	16,9	68,21					
Закарпатська	37,8	10,1	19,5	68,2	12,3	69,45					
Запорізька	43,1	5,8	14,2	67,3	18,5	68,64					
Івано-Франківська	40,1	7,8	16,6	68,8	14,6	72,03					
Київська	40,3	7	17,7	67,2	15,1	68,1					
Кіровоградська	42,8	6,1	14,7	66,6	18,7	68,63					
Львівська	40,6	7,8	16,1	68,6	15,3	71,51					
Миколаївська	42	6,4	15,1	67,6	17,3	69,09					
Одеська	40,6	8,2	16,8	67,1	16,1	69,21					
Полтавська	43	5,9	14	67,9	18,1	70,01					
Рівненська	37,9	10,2	20,1	67	12,9	69,92					
Сумська	44	5,2	12,8	68	19,2	70,34					
Тернопільська	41,2	7,1	15,4	68,7	15,9	72,11					
Харківська	42,6	5,9	13,5	69	17,5	68,73					
Херсонська	41,5	7	15,9	67,3	16,8	68,16					
Хмельницька	42	7	15,5	66,9	17,6	70,27					
Черкаська	43,6	5,8	13,6	67,1	19,3	70,57					
Чернівецька	39,7	8,8	17,1	68,3	14,6	71,27					
Чернігівська	44,2	5,5	13,2	66,5	20,3	68,94					
м.Київ	40,4	10	16,5	67,8	15,7	71,07					
24											



	1 Середній вік, років	2 Загальний коефіцієнт народжуваності, ‰	3 Частка населення у віці 0-14 років у загальній чисельності населення, ‰	4 Частка населення у віці 15-64 років у загальній чисельності населення, ‰	5 Частка населення у віці 65 років і старше у загальній чисельності населення, ‰	6 Середня очікувана тривалість життя при народженні, років
Вінницька	42,2	6,9	15,2	66,8	18	71,18
Волинська	38,5	9,6	19,3	67,2	13,5	70,7
Дніпропетровська	42,3	6,3	15,1	67,3	17,6	68,42
Житомирська	41,3	7,2	16,1	67	16,9	68,21
Закарпатська	37,8	10,1	19,5	68,2	12,3	69,45
Запорізька	43,1	5,8	14,2	67,3	18,5	68,64
Івано-Франківська	40,1	7,8	16,6	68,8	14,6	72,03
Київська	40,3	7	17,7	67,2	15,1	68,1
Кіровоградська	42,8	6,1	14,7	66,6	18,7	68,63
Львівська	40,6	7,8	16,1	68,6	15,3	71,51
Миколаївська	42	6,4	15,1	67,6	17,3	69,09
Одеська	40,6	8,2	16,8	67,1	16,1	69,21
Полтавська	43	5,9	14	67,9	18,1	70,01
Рівненська	37,9	10,2	20,1	67	12,9	69,92
Сумська	44	5,2	12,8	68	19,2	70,34
Тернопільська	41,2	7,1	15,4	68,7	15,9	72,11
Харківська	42,6	5,9	13,5	69	17,5	68,73
Херсонська	41,5	7	15,9	67,3	16,8	68,16
Хмельницька	42	7	15,5	66,9	17,6	70,27
Черкаська	43,6	5,8	13,6	67,1	19,3	70,57
Чернівецька	39,7	8,8	17,1	68,3	14,6	71,27
Чернігівська	44,2	5,5	13,2	66,5	20,3	68,94
м. Київ	40,4	10	16,5	67,8	15,7	71,07

Рис. 1. Додавання бази даних у таблицю програми Statistica

5. Після цього на панелі задач натискаємо «Statistics → Multiple Regression». В закладці «Advanced» виставляємо галочку навпроти «Review Descriptive statistics, correlation matrix» (рис. 2).

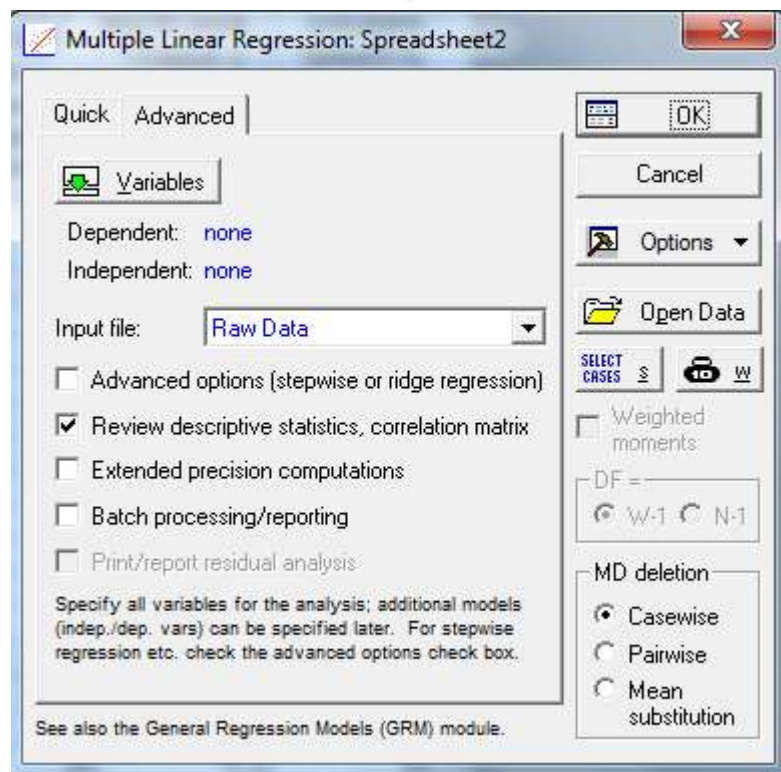
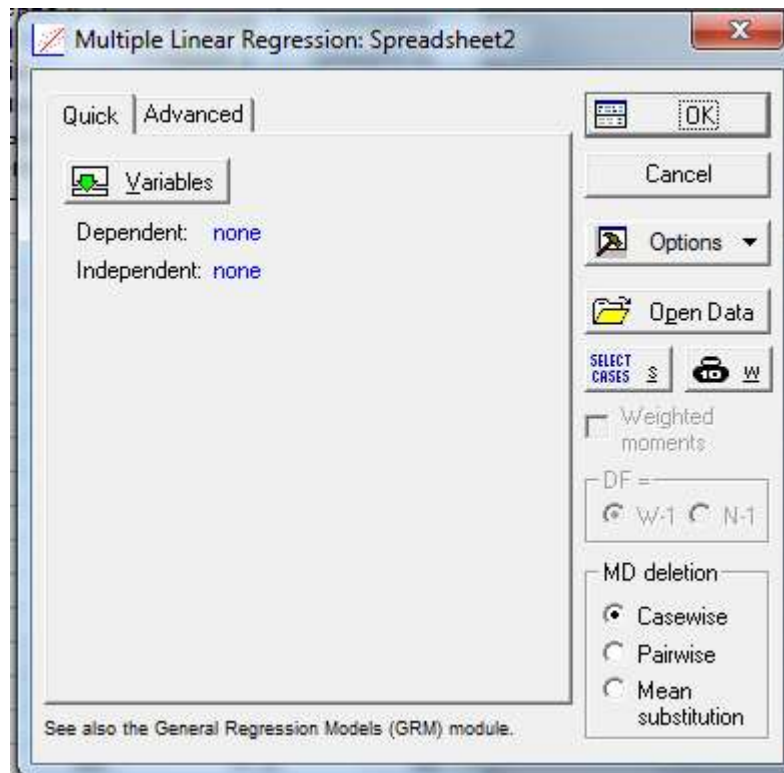
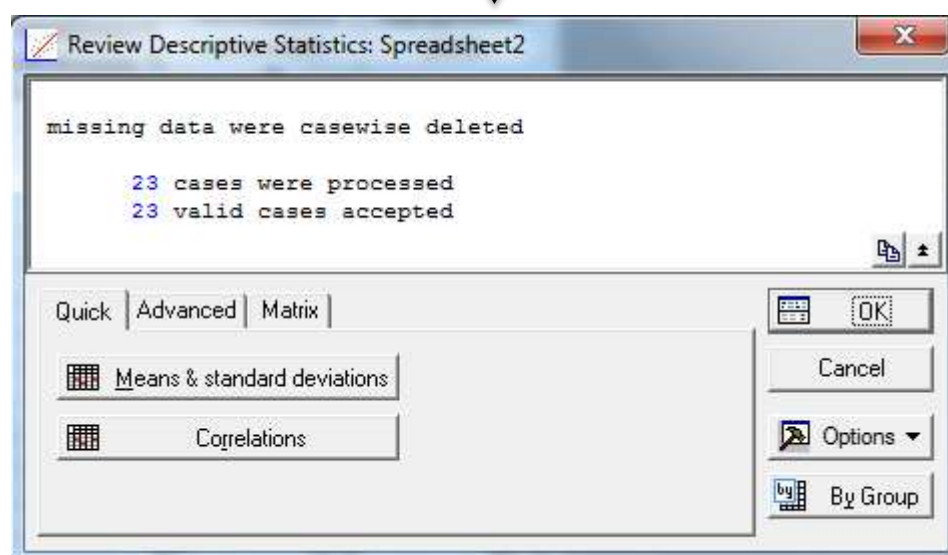
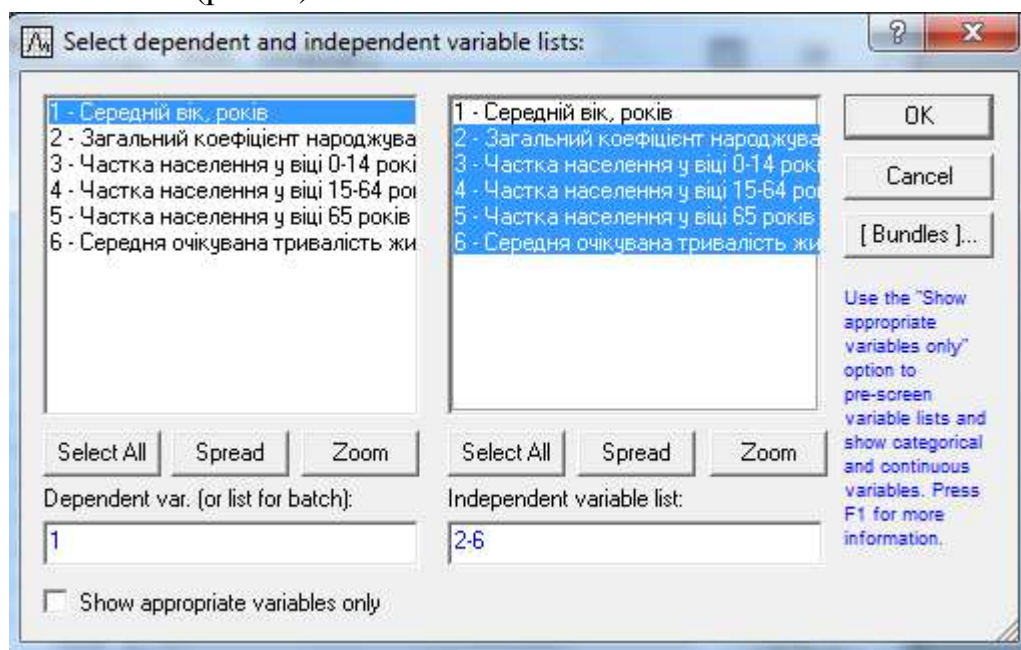


Рис. 2. Діалогове вікно регресійного аналізу

6. В закладці «Variables» в залежних змінних (зліва) обираємо Var1 (середній вік), в незалежних (справа) Var2-Var6 (незалежні змінні) → OK → OK (два рази!). У вікні, що з'явилося, натискаємо на кнопку

«Correlation (рис. 3).



STATISTICA - [Workbook1] - Correlations (Spreadsheet2)

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Workbook Window Help

10 Add to Workbook Add to Report Add to MS Word

Correlations (Spreadsheet2)

Variable	Загальний коефіцієнт народжуваності, %	Частка населення у віці 0-14 років у загальній чисельності населення, %	Частка населення у віці 15-64 років у загальній чисельності населення, %	Частка населення у віці 65 років і старше у загальній чисельності населення, %	Середня очікувана тривалість життя при народженні, років	Середній вік, років
Загальний коефіцієнт народжуваності, %	1.000000	0.908200	0.132196	-0.896301	0.324791	0.936206
Частка населення у віці 0-14 років у загальній чисельності населення, %	0.908200	1.000000	0.804230	-0.937853	0.119272	-0.974937
Частка населення у віці 15-64 років у загальній чисельності населення, %	0.132196	0.804230	1.000000	-0.350996	0.473997	-0.212788
Частка населення у віці 65 років і старше у загальній чисельності населення, %	-0.896301	-0.937853	-0.350996	1.000000	0.276178	0.986762
Середня очікувана тривалість життя при народженні, років	0.324791	0.119272	0.473997	0.276178	1.000000	-0.235128
Середній вік, років	-0.936206	-0.974937	-0.212788	0.986762	-0.235128	1.000000

Рис. 3. Виконання кореляційного аналізу

7. Отримана кореляційна матриця (останній стовпчик) показує, які показники мають високу кореляцію з середнім віком. Саме їх треба використовувати для регресійного аналізу. Отже, частку населення у віці 15-64 років у загальній чисельності населення та середню очікувану тривалість життя при народженні слід убрати.
8. В лівому нижньому куті нажимаємо «Multiple Linear Regression» → «Cancel». В закладці «Advanced» виставляємо галочку навпроти «Advanced option (stepwise or ridge regression)» або просто убираємо всі галочки (рис. 4).

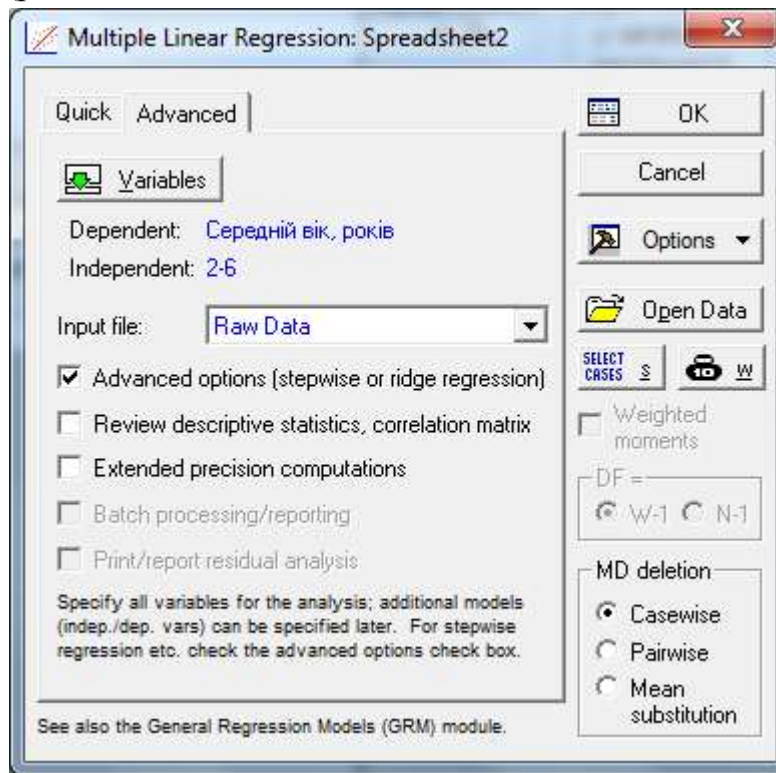


Рис. 4. Діалогове вікно регресійного аналізу

9. В закладці «Variables» в залежних змінних (зліва) обираємо Var1 (середній), в незалежних (справа) Var2, Var3, Var5 → OK → OK → OK (три рази!). З'являється вікно з показниками регресійного аналізу (рис. 5).

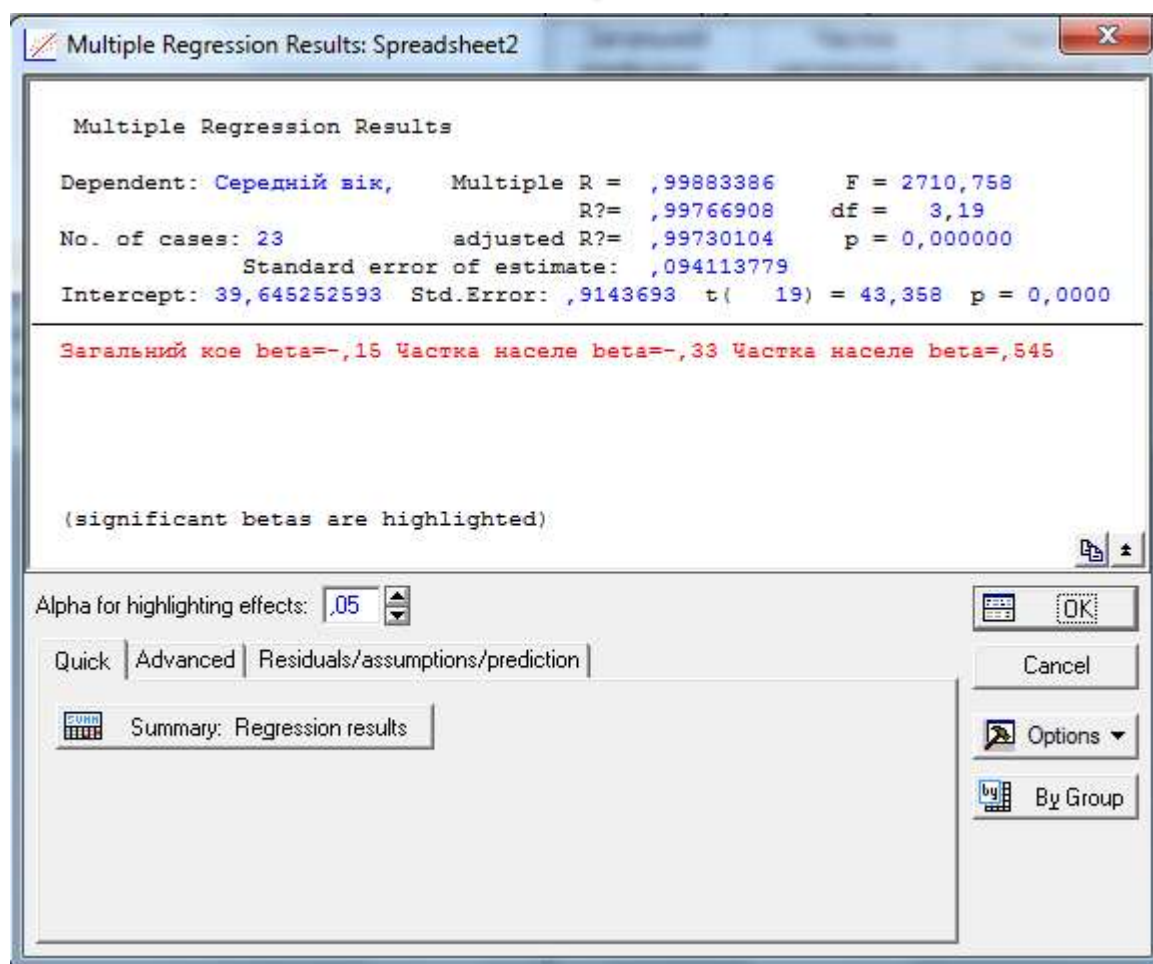
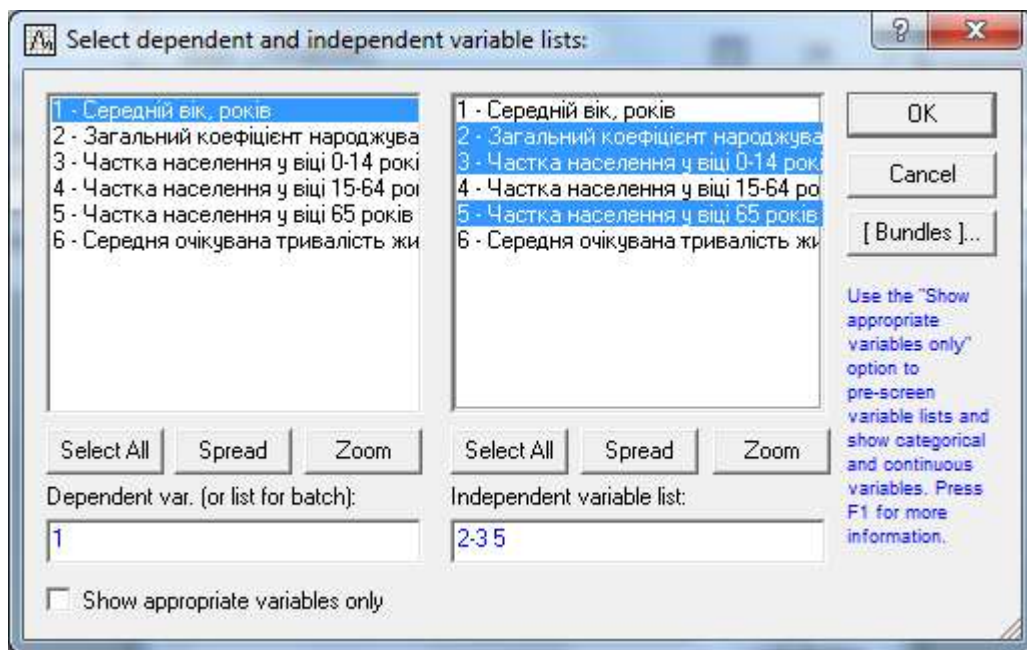


Рис. 5. Показники регресійного аналізу

В інформаційній частині містяться короткі відомості про результати регресійного аналізу, а саме:

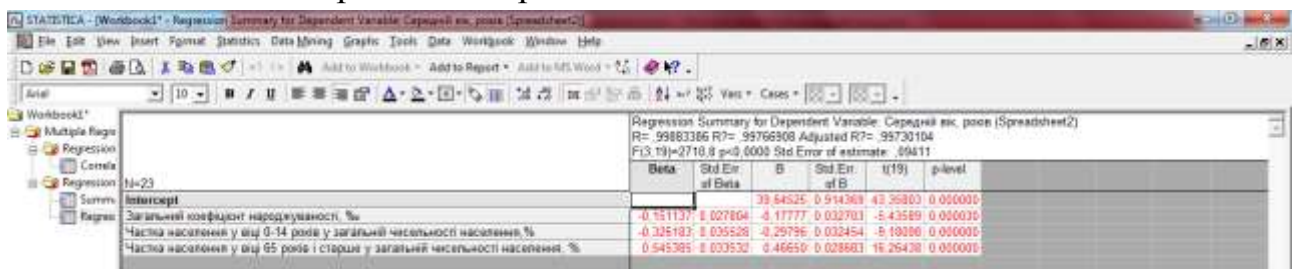
- **Dependent** (ім'я залежної змінної – середній вік);
- **No. of cases** (число спостережень – 23 регіони України);
- **Multiple R** (коефіцієнт множинної кореляції $R = 0,99883386$);
- **R?** (коефіцієнт детермінації $R^2 = 0,99766908$);
- **adjusted R?** (скоригований коефіцієнт детермінації $\text{adjusted } R^2 = 0,99730104$);
- **Standard error of estimate** (стандартна помилка оцінки $= 0,094113779$);
- **Intercept** (оцінка вільного члена регресії $= 39,645252593$);
- **Std. Error** (стандартна помилка оцінки вільного члена $= 0,9143693$);
- **t** (значення t-критерію $t(19) = 43,358$);
- **p** (рівень значущості $p = 0,0000$);
- **F** (значення F-критерію $F = 2710,758$);
- **df** (число ступенів вільності $df = 3,19$).

Під лінією подані стандартизовані оцінки коефіцієнтів регресії.

Серед наведених показників слід звернути увагу на коефіцієнт детермінації R^2 та рівень статистичної значущості p-рівень. R^2 дорівнює 0,99766908 або 99,8 %. Це означає, що 99,8 % варіації залежної ознаки (середнього віку) пояснюється варіацією незалежних ознак (загального коефіцієнту народжуваності, частки населення у віці 0-14 років та 65+ у загальній чисельності населення). **Регресійна модель повинна пояснити не менше 80 % варіації залежної ознаки, тобто $R^2 > 0,8$ (80 %).**

P-рівень дорівнює 0,0000, що означає дуже високу статистичну значущість регресійної моделі. **Коефіцієнти рівняння регресії і його вільний член повинні бути значущими на p-рівні 0,05.**

10. Натискаємо на кнопку «Summary: Regression Results». З'являється матриця (рис. 6). Матриця містить стандартизовані (Beta) і звичайні (B) оцінки коефіцієнтів регресії, їх стандартні помилки, t-критерії (в дужках указано ступені вільності) і рівні значущості. Коефіцієнти Beta оцінюються за стандартизованими даними, мають вибіркове середнє значення 0 і середньоквадратичне відхилення 1.



Regression Summary for Dependent Variable: Середній вік, роки (Spreadsheet2)

R = .99883386 R² = .99766908 Adjusted R² = .99730104
F(3, 19) = 2710.8 p < .0000 Std. Error of estimate: .09411

	Beta	Std. Err. of Beta	B	Std. Err. of B	t(19)	p-level
Intercept			39.64525	0.914369	43.35803	0.000000
Загальний коефіцієнт народжуваності, ‰	-0.151137	0.027804	-4.17777	0.032793	-5.43589	0.000030
Частка населення у віці 0-14 років у загальній чисельності населення, ‰	-0.335183	0.035528	-0.25796	0.032454	-5.18096	0.000005
Частка населення у віці 65 років і старші у загальній чисельності населення, ‰	0.545395	0.033532	0.46650	0.028683	16.26438	0.000001

Рис. 6. Матриця регресійного аналізу

11.Бачимо, що p-level для всіх показників менше 0,05 (якщо по якомусь показнику p-level був би більше 0,05, то його треба було б видалити з незалежних змінних у закладці «Variables» в листі, що справа). Для розрахування рівняння регресії нам потрібна колонка В. Вона відображає вільний член b_0 та коефіцієнти регресії. Таким чином, рядок Intercept – це b_0 , рядок загальний коефіцієнт народжуваності – b_2 , частка населення у віці 0-14 років – b_3 , частка населення у віці 65 років і старше – b_5 (по порядковим номерам незалежних змінних у таблиці на рис. 1). Формуємо рівняння регресії.

$$y=b_0+b_1x_1+b_2x_2+\dots+b_nx_n$$



$y = 39,64525 + (-0,17777x_2) + (-0,29796x_3) + (0,46650x_5)$

де x_2 , x_3 , x_5 – це значення незалежних змінних, тобто загального коефіцієнту народжуваності, частки населення у віці 0-14 років, частки населення у віці 65 років і старше.

12.Додаткові можливості здійснення регресійного аналізу в програмі Statistica. У діалоговому вікні «Multiple Regression Results» на вкладці «Advanced» знаходяться кнопки аналізу, що дозволяють побудувати (рис. 7):

- Summary: Regression results – таблицю результатів регресійного аналізу (рис. 6);
- ANOVA (Overall goodness of fit) – таблицю дисперсійного аналізу;
- Covariance of coefficients (Коефіцієнти коваріації) – таблицю коефіцієнтів коваріації;
- Current sweep matrix – матрицю статистичного зв'язку між коефіцієнтами кореляції;
- Partial correlations – таблицю частинних коефіцієнтів кореляції;
- Redundancy – аналіз надлишковості;
- Stepwise regression summary – поетапний регресійний аналіз;
- ANOVA adjusted for mean – ANOVA з поправкою на середнє.

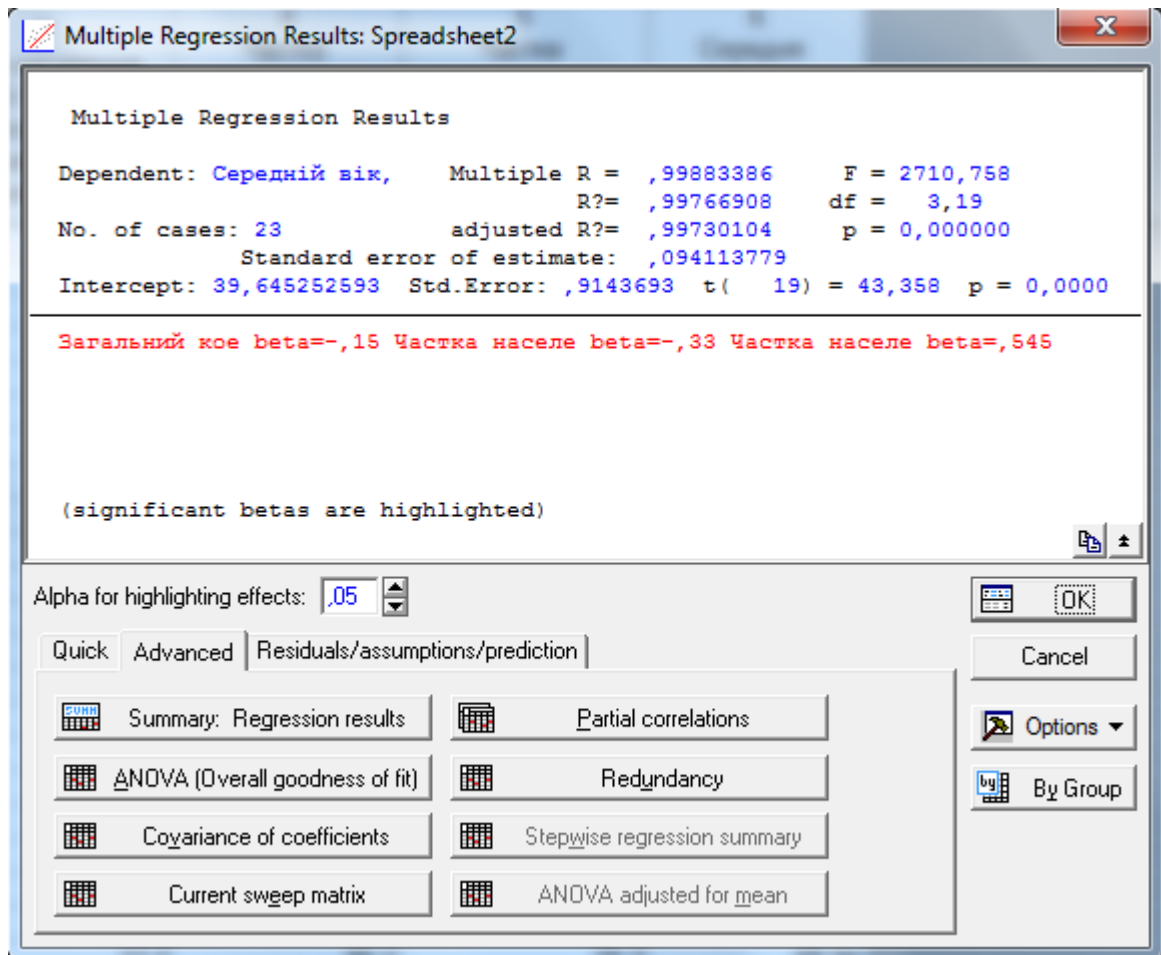


Рис. 7. Додаткові можливості здійснення регресійного аналізу

13. Важливим етапом регресійного аналізу є аналіз залишків (випадкових відхилень). Щоб провести аналіз залишків необхідно перейти на вкладку «Residuals/assumptions/prediction» (Залишки/припущення/прогноз) діалогового вікна «Multiple Linear Regression». Далі потрібно натиснути кнопку «Perform residual analysis» (Виконати аналіз залишків), у результаті чого відкриється робоче вікно «Residuals Analysis» (Аналіз залишків) (рис. 8).

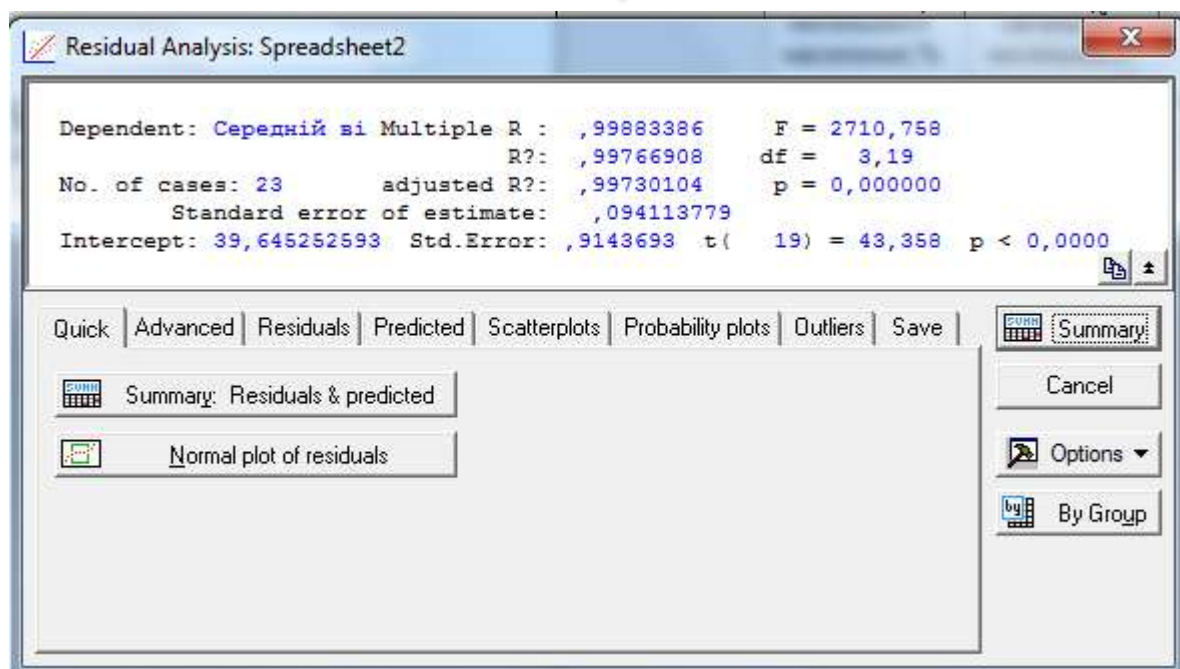
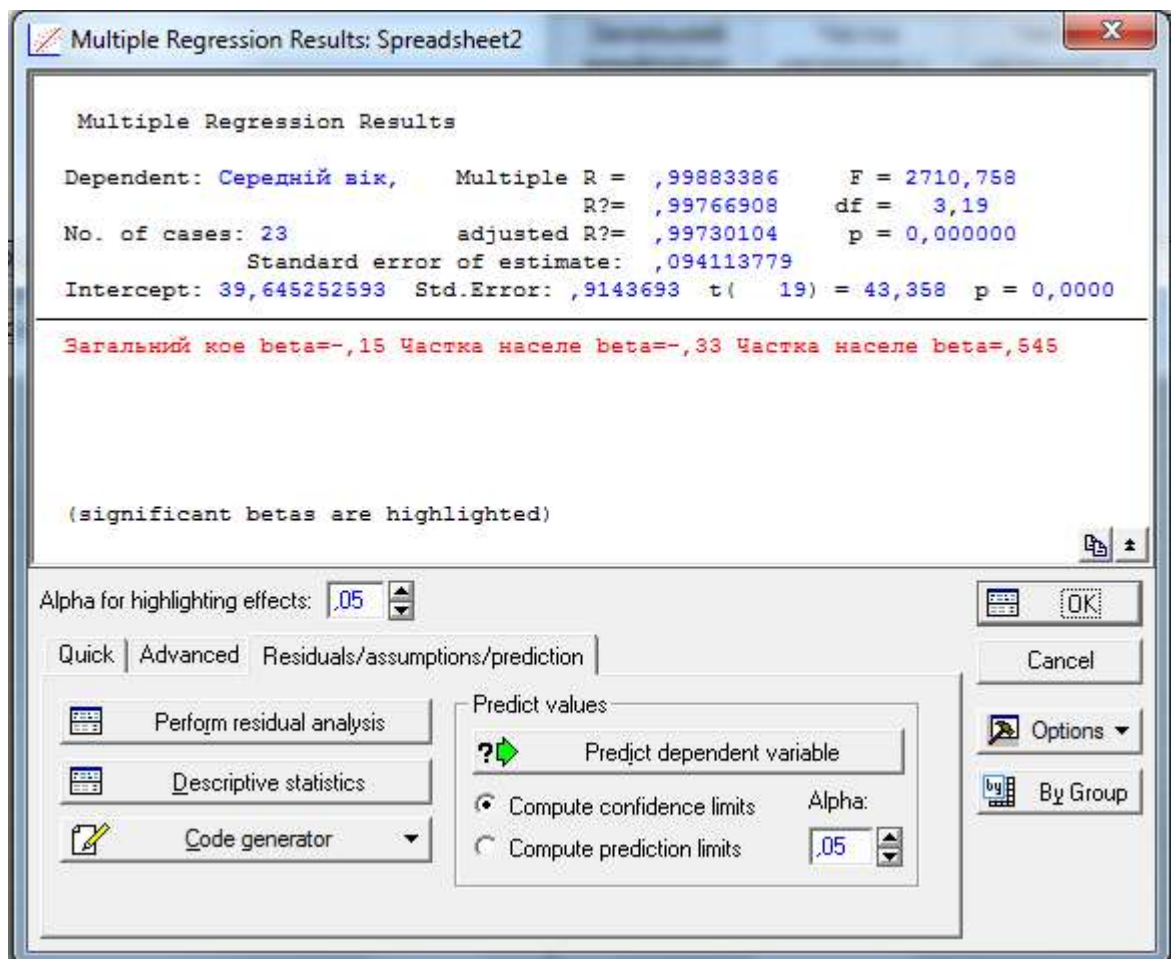


Рис. 8. Робоче вікно «Residuals Analysis»

14. У діалоговому вікні можна перевірити залишки на нормальність

розподілу, натискуючи на кнопку «Histogram of residuals» (Гістограма залишків) на вкладці «Residuals» (рис. 9) або на кнопку «Normal plot of residuals» (Графік залишків) на вкладці «Quick» (рис. 10). Якщо залишки розподілені нормально, то вони будуть знаходитися близько коло червоної, апроксимуючої лінії.

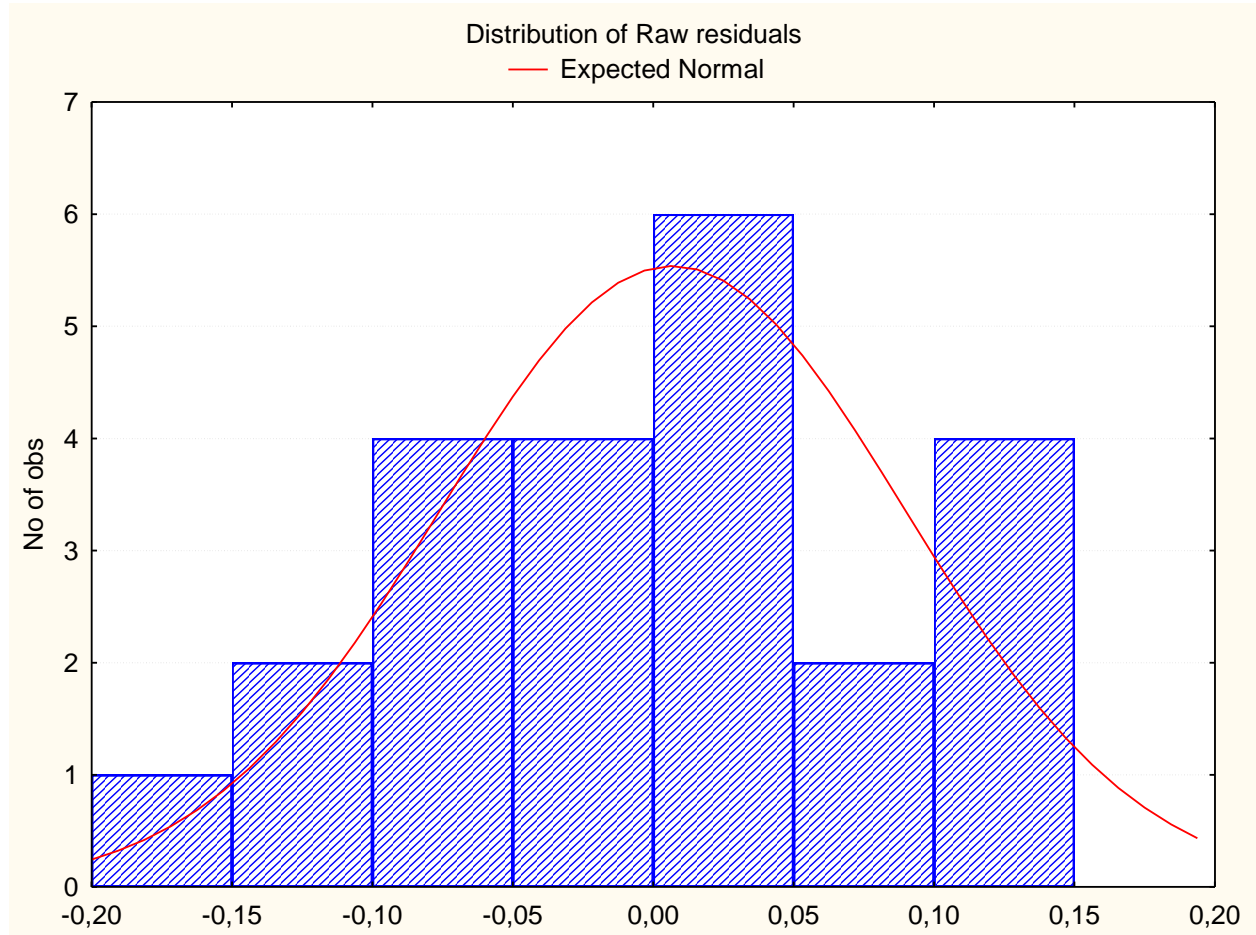


Рис. 9. Гістограма розподілу залишків

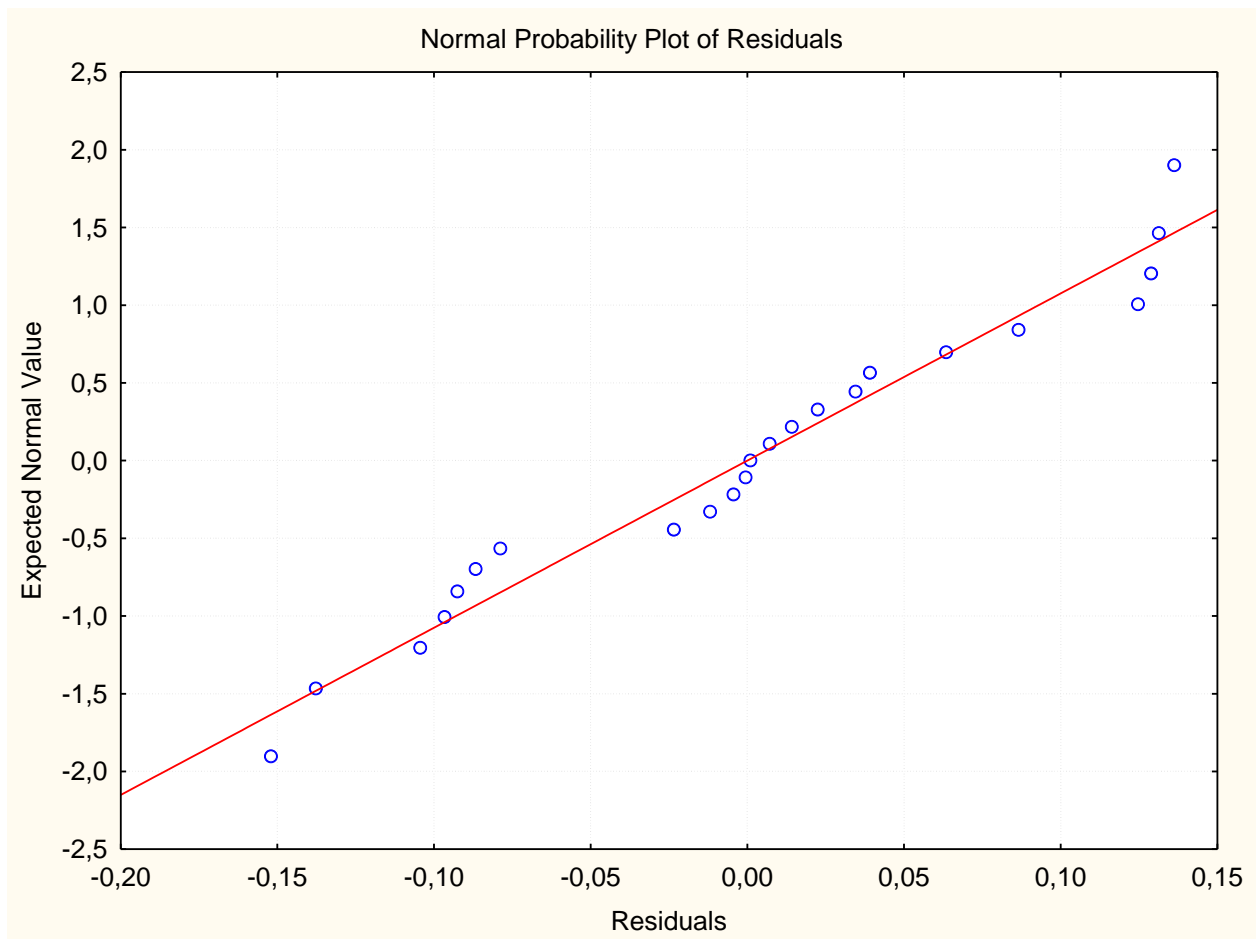


Рис. 10. Графік розподілу залишків

15. Випадковий характер залишків можна перевірити, натиснувши кнопку «Predicted vs. Residual» у вкладці «Scatterplot» (рис. 11). З'являється графік залежності залишків від теоретичних значень результативної ознаки. Якщо на графіку отримана горизонтальна смуга, то залишки є випадковими величинами.

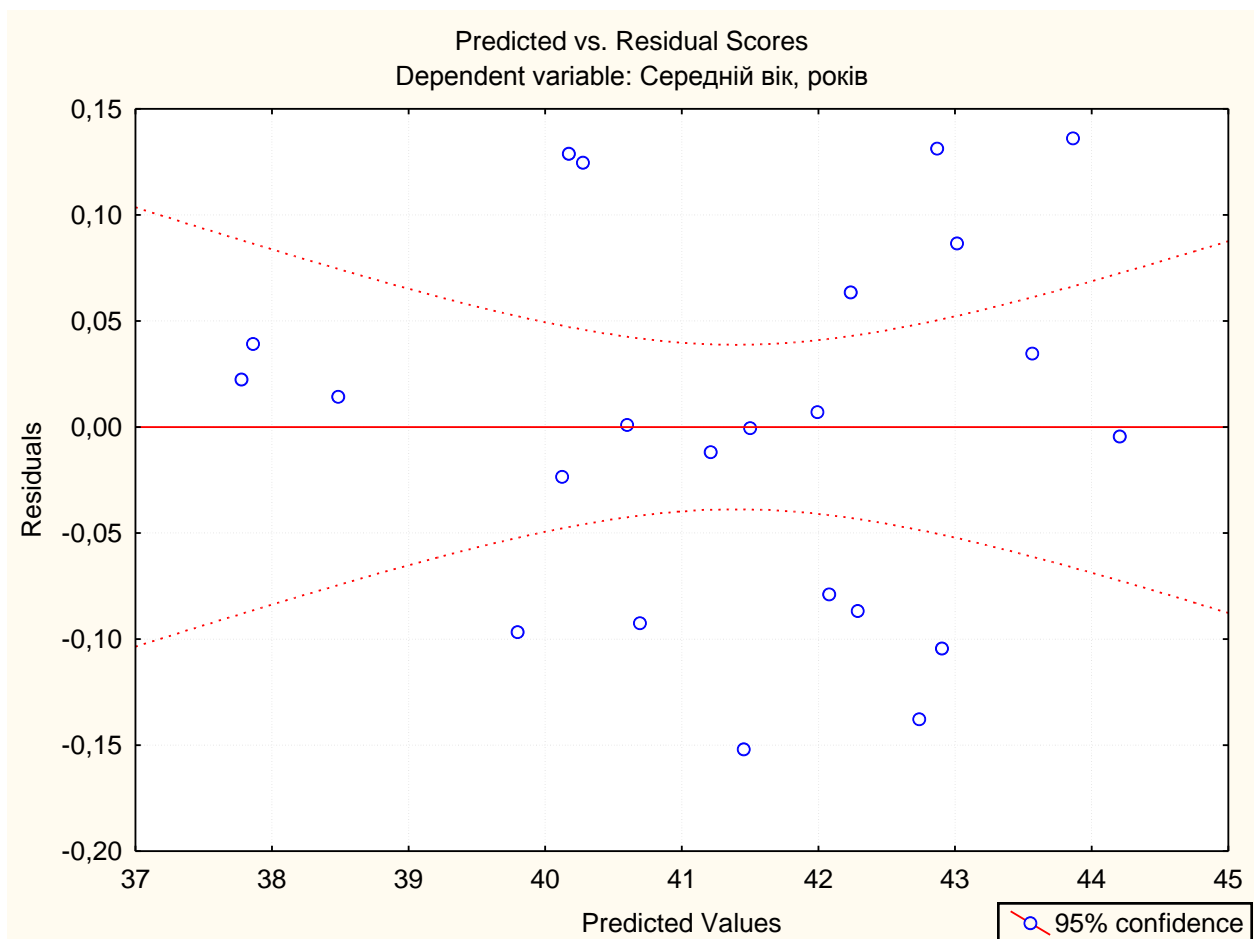


Рис. 11. Графік залежності залишків
від теоретичних значень результативної ознаки

16. Відсутність автокореляції залишків, коли значення залишків розподілені незалежно один від одного, можна за допомогою статистики Дарбіна-Уотсона. Для цього нажимаємо кнопку «Durbin-Watson statistic» у вкладці «Advanced» (рис. 12). **Автокореляція відсутня, якщо виконується наступна умова: $1.5 < DW < 2.5$.** В нашому випадку $DW = 2,191739$, тобто більше 1,5, але менше 2,5.

STATISTICA - [Workbook1* - Durbin-Watson d (Spreadsheet2)]

File Edit View Insert Format Statistics Data Mining Graphs Tools

Workbook1*

- Multiple Regre
- Regression
- Correla
- Regression
- Summi
- Regres:
- Redunc

Durbin-Watson d (Spreadsheet2) and serial correlation of residuals	
	Durbin-Watson d
Estimate	2,191739

Рис. 12. Статистика Дарбіна-Уотсона

Завдання до практичної роботи:

Завдання 1. Проведіть регресійний аналіз показників зі своєї бази даних (5-10 показників). Для цього оберіть будь-який показник (він буде залежною змінною, тобто відгуком) та декілька інших, від яких, на Ваш погляд, залежить відгук. Це будуть незалежні змінні (предиктори). Наведіть показники регресійного аналізу та матрицю регресійного аналізу. Розрахуйте рівняння регресії. Зробіть аналіз залишків будь-яким одним способом (випадковий характер залишків, відсутність автокореляції залишків, нормальний розподіл залишків, гомоскедастичність, нульова середня величина залишків, незалежна від x). Зробіть висновки.

Практичне заняття 17-18
ПОБУДОВА БАГАТОВИМІРНИХ СТАТИСТИЧНИХ МОДЕЛЕЙ НА
МАТРИЦІ ВИХІДНИХ ДАНИХ.
КЛАСТЕР-АНАЛІЗ ВИХІДНИХ ДАНИХ.
ІНТЕРПРЕТАЦІЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Питання для обговорення

1. Багатовимірний аналіз.
2. Кластерний аналіз: поняття, задачі, види.
3. Кластер та його характеристики.
4. Стандартизація (standardization) або нормування (normalization) бази даних. Лінійне шкалювання.
5. Методи кластерного аналізу:
 - 5.1. Ієрархічні методи кластерного аналізу (агломераційні і дивизимні).
 - 5.2. Неієрархічні методи кластерного аналізу (ітеративні, факторні, методи згущень, методи, які використовують теорію графів).
6. Міри подібності (евклідова відстань, манхеттенська відстань, відстань Чебишева, відсоток незгоди). Сутність евклідової відстані.
7. Методи об'єднання або зв'язки (метод Варда, метод одиничного зв'язку, метод повного зв'язку, метод середнього зв'язку, центроїдний метод). Сутність методу Варда.
8. Дендрограма кластеризації.

Методичні рекомендації до практичної роботи:

1. Проводимо лінійне шкалювання бази даних, по якій будемо робити кластерний аналіз (табл. 1). Суть лінійного шкалювання полягає у нормуванні статистичних параметрів, тобто приведення їх у індекси в інтервалі від 0 до 1. Для позитивних параметрів (загальний коефіцієнт народжуваності, коефіцієнт природного приросту, коефіцієнт шлюбності, частка населення у віці 0-14 років, 15-64 років та 65+ років у загальній чисельності населення, середній вік, на 1000 чоловіків припадає жінок) розрахунки проводяться за формулою:

$$I_j = \frac{X_{i,j} - X_{\min,j}}{X_{\max,j} - X_{\min,j}}$$

де I_j – індекс j -того індикатора (індекс має амплітуду від 0 до 1);

$X_{i,j}$ – поточне значення j -того індикатора;

$X_{\max, j}$ – найбільше значення j -того індикатора в ряду спостереження;

$X_{\min, j}$ – найнижче значення j -того індикатора в ряду спостереження.

Для негативних показників (загальний коефіцієнт смертності, коефіцієнт розлучуваності) розрахунки проводяться за зворотною формулою:

$$I_j = 1 - \frac{X_{i,j} - X_{\min,j}}{X_{\max,j} - X_{\min,j}}$$

Таблиця 1

Показники демографічної ситуації населення по регіонах України у 2021 році

Регіони України	Загальний коефіцієнт народжуваності, ‰	Загальний коефіцієнт смертності, ‰	Коефіцієнт природного приросту, ‰	Коефіцієнт шлюбності, ‰	Коефіцієнт розлучуваності, ‰	Частка населення у віці 0-14 років у загальній чисельності населення, %	Частка населення у віці 15-64 років у загальній чисельності населення, %	Частка населення у віці 65 років і старше у загальній чисельності населення, %	Середній вік, років	На 1000 чоловіків припадає жінок
Вінницька	6,9	18,2	-11,3	5,3	3,3	15,2	66,8	18,0	42,2	1156
Волинська	9,6	15,1	-5,5	5,4	2,6	19,3	67,2	13,5	38,5	1118
Дніпропетровська	6,3	20,8	-14,5	6,0	3,7	15,1	67,3	17,6	42,3	1190
Житомирська	7,2	19,9	-12,7	5,1	3,2	16,1	67,0	16,9	41,3	1144
Закарпатська	10,1	14,8	-4,7	5,3	2,2	19,5	68,2	12,3	37,8	1080
Запорізька	5,8	21,6	-15,8	5,4	3,4	14,2	67,3	18,5	43,1	1190
Івано-Франківська	7,8	15,2	-7,4	5,2	2,9	16,6	68,8	14,6	40,1	1115
Київська	7,0	20,2	-13,2	6,2	3,8	17,7	67,2	15,1	40,3	1184
Кіровоградська	6,1	20,8	-14,7	4,5	3,2	14,7	66,6	18,7	42,8	1166
Львівська	7,8	16,0	-8,2	5,5	2,4	16,1	68,6	15,3	40,6	1111
Миколаївська	6,4	19,4	-13,0	5,4	3,4	15,1	67,6	17,3	42,0	1156
Одеська	8,2	18,0	-9,8	6,7	3,6	16,8	67,1	16,1	40,6	1122
Полтавська	5,9	20,3	-14,4	5,1	3,3	14,0	67,9	18,1	43,0	1162
Рівненська	10,2	14,9	-4,7	5,5	2,6	20,1	67,0	12,9	37,9	1104
Сумська	5,2	20,0	-14,8	4,7	3,0	12,8	68,0	19,2	44,0	1178
Тернопільська	7,1	16,4	-9,3	5,1	2,7	15,4	68,7	15,9	41,2	1134
Харківська	5,9	21,0	-15,1	5,9	3,6	13,5	69,0	17,5	42,6	1155
Херсонська	7,0	20,0	-13,0	5,3	3,2	15,9	67,3	16,8	41,5	1152
Хмельницька	7,0	19,0	-12,0	5,0	3,3	15,5	66,9	17,6	42,0	1150

Черкаська	5,8	19,6	-13,8	5,3	3,2	13,6	67,1	19,3	43,6	1172
Чернівецька	8,8	16,1	-7,3	5,5	2,8	17,1	68,3	14,6	39,7	1120
Чернігівська	5,5	22,0	-16,5	4,7	3,1	13,2	66,5	20,3	44,2	1199
м.Київ	10,0	15,3	-5,3	8,6	3,6	16,5	67,8	15,7	40,4	1169
Україна¹	7,3	18,5	-11,2	5,8	3,2	14,9	67,4	17,7	42,2	1157

¹ Відносні показники розраховано без урахування відповідних даних по Донецькій та Луганській областях. У зв'язку з відсутністю інформації по частині тимчасово окупованих територій у цих областях здійснення розрахунків відносних показників по ним є некоректним



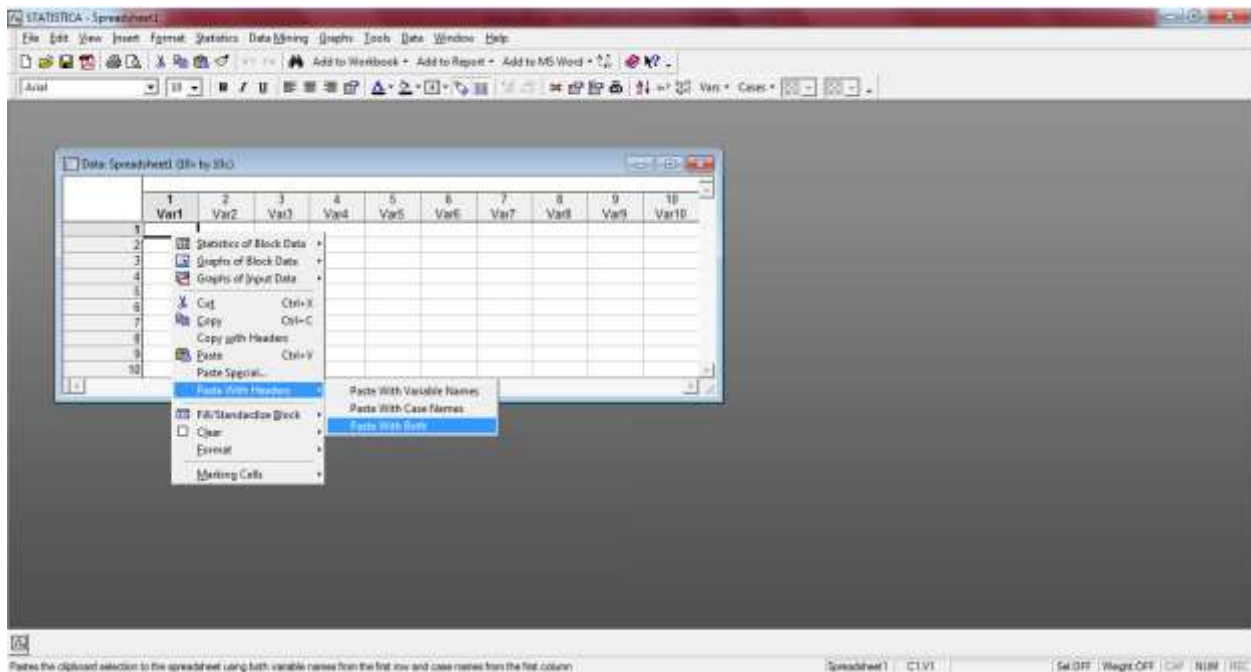
Таблиця 2

Індекси показників демографічної ситуації населення по регіонах України у 2021 році

	Загальний коефіцієнт народжуваності , ‰	Загальний коефіцієнт смертності, ‰	Коефіцієнт природного приросту, ‰	Коефіцієнт шлюбності, ‰	Коефіцієнт розлучуваності , ‰	Частка населення у віці 0-14 років у загальній чисельності населення, %	Частка населення у віці 15-64 років у загальній чисельності населення, %	Частка населення у віці 65 років і старше у загальній чисельності населення, %	Середній вік, років	На 1000 чоловіків припадає жінок
Вінницька	0,3400	0,5278	0,4407	0,1951	0,3125	0,3288	0,1200	0,2875	0,6875	0,6387
Волинська	0,8800	0,9583	0,9322	0,2195	0,7500	0,8904	0,2800	0,8500	0,1094	0,3193
Дніпропетровська	0,2200	0,1667	0,1695	0,3659	0,0625	0,3151	0,3200	0,3375	0,7031	0,9244
Житомирська	0,4000	0,2917	0,3220	0,1463	0,3750	0,4521	0,2000	0,4250	0,5469	0,5378
Закарпатська	0,9800	1,0000	1,0000	0,1951	1,0000	0,9178	0,6800	1,0000	0,0000	0,0000
Запорізька	0,1200	0,0556	0,0593	0,2195	0,2500	0,1918	0,3200	0,2250	0,8281	0,9244
Івано-Франківська	0,5200	0,9444	0,7712	0,1707	0,5625	0,5205	0,9200	0,7125	0,3594	0,2941

Київська	0,3600	0,2500	0,2797	0,4146	0,0000	0,6712	0,2800	0,6500	0,3906	0,8739
Кіровоградська	0,1800	0,1667	0,1525	0,0000	0,3750	0,2603	0,0400	0,2000	0,7812	0,7227
Львівська	0,5200	0,8333	0,7034	0,2439	0,8750	0,4521	0,8400	0,6250	0,4375	0,2605
Миколаївська	0,2400	0,3611	0,2966	0,2195	0,2500	0,3151	0,4400	0,3750	0,6563	0,6387
Одеська	0,6000	0,5556	0,5678	0,5366	0,1250	0,5479	0,2400	0,5250	0,4375	0,3529
Полтавська	0,1400	0,2361	0,1780	0,1463	0,3125	0,1644	0,5600	0,2750	0,8125	0,6891
Рівненська	1,0000	0,9861	1,0000	0,2439	0,7500	1,0000	0,2000	0,9250	0,0156	0,2017
Сумська	0,0000	0,2778	0,1441	0,0488	0,5000	0,0000	0,6000	0,1375	0,9688	0,8235
Тернопільська	0,3800	0,7778	0,6102	0,1463	0,6875	0,3562	0,8800	0,5500	0,5313	0,4538
Харківська	0,1400	0,1389	0,1186	0,3415	0,1250	0,0959	1,0000	0,3500	0,7500	0,6303
Херсонська	0,3600	0,2778	0,2966	0,1951	0,3750	0,4247	0,3200	0,4375	0,5781	0,6050
Хмельницька	0,3600	0,4167	0,3814	0,1220	0,3125	0,3699	0,1600	0,3375	0,6563	0,5882
Черкаська	0,1200	0,3333	0,2288	0,1951	0,3750	0,1096	0,2400	0,1250	0,9063	0,7731
Чернівецька	0,7200	0,8194	0,7797	0,2439	0,6250	0,5890	0,7200	0,7125	0,2969	0,3361
Чернігівська	0,0600	0,0000	0,0000	0,0488	0,4375	0,0548	0,0000	0,0000	1,0000	1,0000
м.Київ	0,9600	0,9306	0,9492	1,0000	0,1250	0,5068	0,5200	0,5750	0,4063	0,7479

2. Відкриваємо програму Statistica 8.0. Натискаємо ОК. Закриваємо всі вікна.
3. Натискаємо кнопку File → New (📄) або Ctrl+N. Натискаємо ОК.
4. З'являється таблиця, в яку необхідно додати пронормовану базу даних (табл. 2).
5. Для цього копіюємо в таблиці Excel пронормовану базу даних і вставляємо її у таблицю програми Statistica. Дана операція робиться наступним чином. У першій ячійці правою кнопкою миші у вікні, що відкривається, натискаємо «Paste With Headers → Paste With Both». Таким чином, ми копіюємо у таблицю програми Statistica наші індекси показників, а також назви показників та назви регіонів України (рис. 1).



STATISTICA - [Data: Spreadsheet2* (11v by 24c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

Anal 10

	коєфіцієнт	н	сп	поєфіцієнт	ент	років у	років у	старше	Середній	чоловіки	11
	ент	коєфіцієнт	природн	ент	розлучу	загалні	загалні	у	й вік	в	NewVar
Вінницька	0,34	0,5277778	0,440678	0,195122	0,3125	0,328767	0,12	0,2875	0,6875	0,638655	
Волинська	0,88	0,958333	0,932203	0,219512	0,75	0,890411	0,28	0,85	0,109375	0,319328	
Дніпропетровська	0,22	0,166667	0,169492	0,365854	0,0625	0,315068	0,32	0,3375	0,703125	0,92437	
Житомирська	0,4	0,291667	0,322034	0,146341	0,375	0,452055	0,2	0,425	0,546875	0,537815	
Закарпатська	0,98	1	1	0,195122	1	0,917808	0,68	1	0	0	
Запорізька	0,12	0,055556	0,059322	0,219512	0,25	0,191781	0,32	0,225	0,828125	0,92437	
Івано-Франківська	0,52	0,944444	0,771186	0,170732	0,5625	0,520548	0,92	0,7125	0,359375	0,294118	
Київська	0,36	0,25	0,279661	0,414634	0	0,671233	0,28	0,65	0,390625	0,87395	
Кіровоградська	0,18	0,166667	0,152542	0	0,375	0,260274	0,04	0,2	0,78125	0,722689	
Львівська	0,52	0,833333	0,70339	0,243902	0,875	0,452055	0,84	0,625	0,4375	0,260504	
Миколаївська	0,24	0,361111	0,29661	0,219512	0,25	0,315068	0,44	0,375	0,65625	0,638655	
Одеська	0,6	0,555556	0,567797	0,536585	0,125	0,547945	0,24	0,525	0,4375	0,352941	
Полтавська	0,14	0,236111	0,177966	0,146341	0,3125	0,164384	0,56	0,275	0,8125	0,689076	
Рівненська	1	0,986111	1	0,243902	0,75	1	0,2	0,925	0,015625	0,201681	
Сумська	0	0,277778	0,144068	0,04878	0,5	0	0,6	0,1375	0,96875	0,823529	
Тернопільська	0,38	0,777778	0,610169	0,146341	0,6875	0,356164	0,88	0,55	0,53125	0,453782	
Харківська	0,14	0,138889	0,118644	0,341463	0,125	0,09589	1	0,35	0,75	0,630252	
Херсонська	0,36	0,277778	0,29661	0,195122	0,375	0,424658	0,32	0,4375	0,578125	0,605042	
Хмельницька	0,36	0,416667	0,381356	0,121951	0,3125	0,369063	0,16	0,3375	0,65625	0,588235	
Черкаська	0,12	0,333333	0,228814	0,195122	0,375	0,109589	0,24	0,125	0,90625	0,773109	
Чернівецька	0,72	0,819444	0,779661	0,243902	0,625	0,589041	0,72	0,7125	0,296875	0,336134	
Чернігівська	0,06	0	0	0,04878	0,4375	0,054795	0	0	1	1	
м. Київ	0,96	0,930556	0,949153	1	0,125	0,506849	0,52	0,575	0,40625	0,747899	

24



STATISTICA - [Data: Spreadsheet2* (10v by 23c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

Anal 10

	1	2	3	4	5	6	7	8	9	10
	Загальний	Загальний	Коефіцієнт	Коефіцієнт	Коефіцієнт	Частка населення	Частка населення	Частка населення	Середній	На 1000
	коєфіцієнт	коєфіцієнт	природного	коєфіцієнт	розлучуваності	у віці 0-14 років у загальній чисельності населення, %	у віці 15-64 років у загальній чисельності населення, %	у віці 65 років і старших у загальній чисельності населення, %	й вік, років	чоловіків на жінку
Вінницька	0,34	0,52777778	0,440677966	0,195121951	0,3125	0,328767123	0,12	0,2875	0,6875	0,63865462
Волинська	0,88	0,95833333	0,93220339	0,219512195	0,75	0,890410959	0,28	0,85	0,109375	0,319327731
Дніпропетровська	0,22	0,16666667	0,169491525	0,365853659	0,0625	0,315068493	0,32	0,3375	0,703125	0,924369748
Житомирська	0,4	0,29166667	0,322033898	0,146341463	0,375	0,452054795	0,2	0,425	0,546875	0,537815126
Закарпатська	0,98	1	1	0,195121951	1	0,917808219	0,68	1	0	0
Запорізька	0,12	0,05555556	0,059322039	0,219512195	0,25	0,191780822	0,32	0,225	0,828125	0,924369748
Івано-Франківська	0,52	0,944444444	0,771186441	0,170731707	0,5625	0,520547945	0,92	0,7125	0,359375	0,294117647
Київська	0,36	0,25	0,279661017	0,414634146	0	0,671232877	0,28	0,65	0,390625	0,87394958
Кіровоградська	0,18	0,16666667	0,152542373	0	0,375	0,260273973	0,04	0,2	0,78125	0,722689076
Львівська	0,52	0,83333333	0,703389831	0,243902439	0,875	0,452054795	0,84	0,625	0,4375	0,260504202
Миколаївська	0,24	0,361111111	0,296610169	0,219512195	0,25	0,315068493	0,44	0,375	0,65625	0,638655462
Одеська	0,6	0,55555556	0,56779661	0,536585366	0,125	0,547945205	0,24	0,525	0,4375	0,352941176
Полтавська	0,14	0,236111111	0,177966102	0,146341463	0,3125	0,164383562	0,56	0,275	0,8125	0,68907563
Рівненська	1	0,986111111	1	0,243902439	0,75	1	0,2	0,925	0,015625	0,201680672
Сумська	0	0,27777778	0,144067797	0,0487804878	0,5	0	0,6	0,1375	0,96875	0,823529412
Тернопільська	0,38	0,77777778	0,610169492	0,146341463	0,6875	0,356164384	0,88	0,55	0,53125	0,453781513
Харківська	0,14	0,13888889	0,118644068	0,341463415	0,125	0,095890411	1	0,35	0,75	0,630252101
Херсонська	0,36	0,27777778	0,296610169	0,195121951	0,375	0,424657534	0,32	0,4375	0,578125	0,605042917
Хмельницька	0,36	0,41666667	0,381355932	0,12195122	0,3125	0,369863014	0,16	0,3375	0,65625	0,588235294
Черкаська	0,12	0,33333333	0,228813559	0,195121951	0,375	0,109589041	0,24	0,125	0,90625	0,773109244
Чернівецька	0,72	0,81944444	0,779661017	0,243902439	0,625	0,589041096	0,72	0,7125	0,296875	0,336134454
Чернігівська	0,06	0	0	0,0487804878	0,4375	0,0547945205	0	0	1	1
м. Київ	0,96	0,93055556	0,949152542	1	0,125	0,506849315	0,52	0,575	0,40625	0,74789916

Рис. 1. Додавання бази даних у таблицю програми Statistica

9. Після цього на панелі задач натискаємо «Statistics» → «Multivariate Exploratory Techniques» → «Cluster Analysis» (рис. 2).

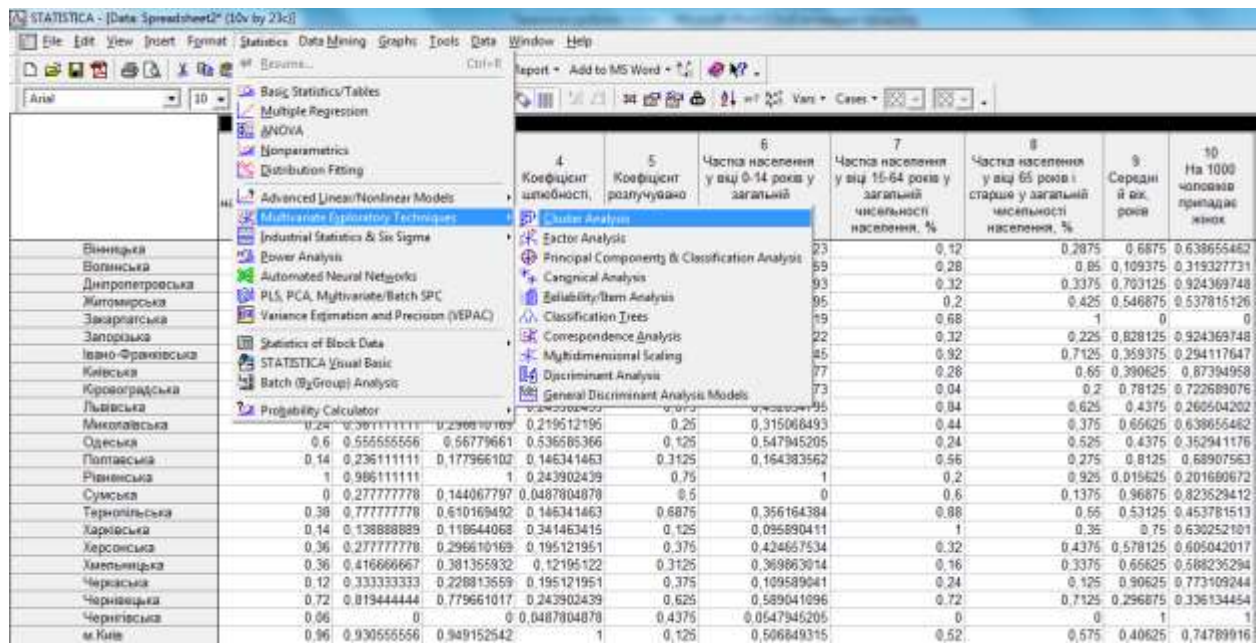


Рис. 2. Алгоритм вибору кластерного аналізу в програмі Statistica 8.0

10. Після цього з'являється діалогове вікно «Clustering method» («Метод кластеризації»). Ми обираємо метод «Joining (tree clustering)» та натискаємо «OK» (рис. 3).

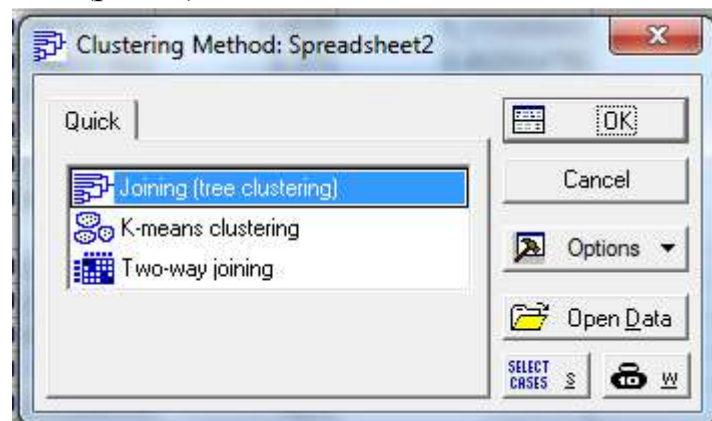


Рис. 3. Вибір методу кластеризації

11. Відкривається діалогове вікно «Cluster Analysis: Joining (tree clustering)». Натискаємо вкладку «Advanced», натискаємо кнопку «Variables». Обираємо показники для аналізу (рис. 4). Натискаємо «OK».

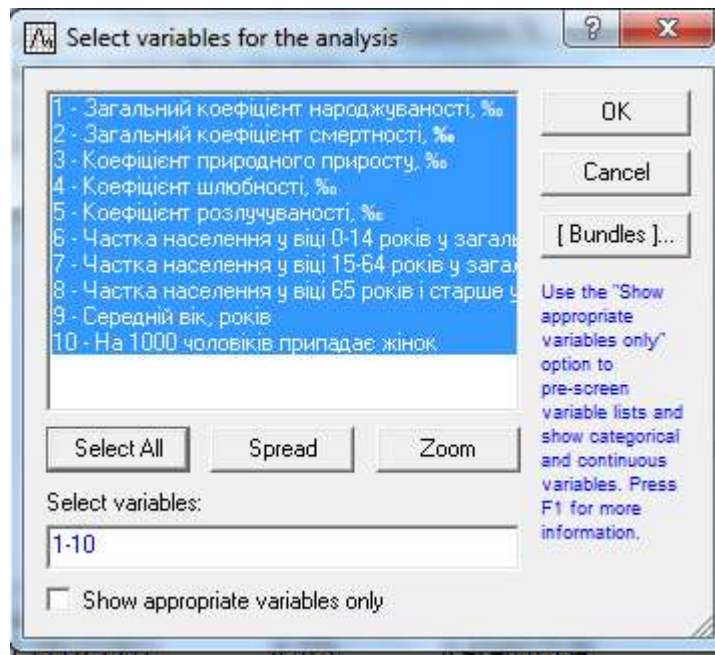


Рис. 4. Вибір показників для аналізу

12. Наступним кроком є вибір методу об'єднання кластерів та міри подібності. У полі «Input file» має бути «Raw data», «Cluster» – «Cases (rows)» (кластерний аналіз буде виконуватися за рядками, тобто групуватися будуть адміністративно-територіальні одиниці, а не показники), «Agglomeration (linkage) rule» – «Ward's method» (метод Варда), «Distance measure» – «Euclidean distances» (евклідова відстань). Після цього натискаємо «OK» (рис. 5).



Рис. 5. Вибір методу об'єднання кластерів та міри подібності (відстані)

13. Далі відкривається діалогове вікно «Joining results». Галочка має стояти навпроти поля «Rectangular branches». Натискаємо на кнопку «Vertical

icle plot» (рис. 6).



Рис. 6. Вибір напрямку дендрограми кластеризації

14. Таким чином, ми отримуємо вертикальну діаграму (рис. 7).

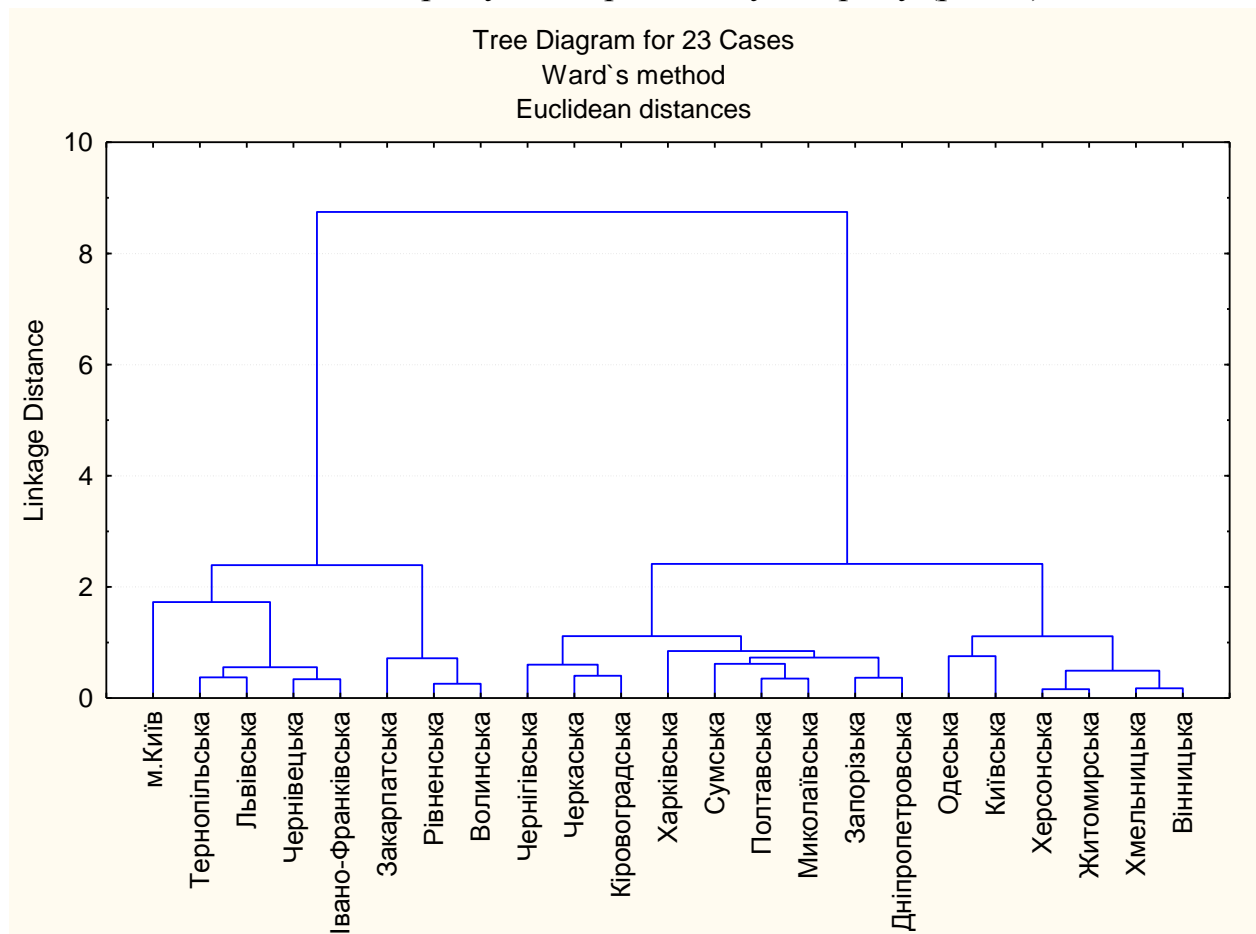


Рис. 7. Дендрограма кластеризації

15. Далі потрібно повернутися (у нижньому лівому куті шукаємо вкладку «Joining results») та у вкладці «Advanced» натиснути кнопку «Amalgamation schedule». В результаті отримуємо таблицю з результатами об'єднань регіонів за ступенем міри відстані (рис. 8). Обираємо певний рівень. Наприклад, ми обрали рівень об'єднання 1,0. Візуально на дендрограмі ми бачимо виділення 5-и кластерів. В таблиці Amalgamation Schedule ця відстань лежить між відстанями об'єднання 0,8466054 та 1,113385. Ми маємо брати до уваги відстані зв'язку, що є більшими, оскільки класифікація об'єктів відбувається з найбільшої відстані. В даному випадку, такою відстанню є 2,412549 (відстань 8,745249 є останньою, на ній всі наші регіони об'єднуються в один кластер). Таким чином, для виділення груп регіонів ми обираємо п'ять відстаней зв'язку (рис. 8).

Linkage distance	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. No. 7	Obj. No. 8	Obj. No. 9	Obj. No. 10	Obj. No. 11	Obj. No. 12	Obj. No. 13
0.8466054	Житомирська	Хмельницька											
1.113385	Вінницька	Хмельницька											
1.115810	Дніпропетровська	Закарпатська	Львівська	Тернопільська									
1.728193	Івано-Франківська	Чернівецька	Полтавська	Сумська									
2.392927	Волинська	Рівненська	Закарпатська	Миколаївська	Полтавська	Сумська							
2.412549	Вінницька	Хмельницька	Житомирська	Харківська	Київська	Одеська	Дніпропетровська	Закарпатська	Львівська	Тернопільська	Чернівецька	Чернівецька	Хмельницька

Рис. 8. Ступінь об'єднання регіонів у кластери за мірою подібності (відстані)

16. На обраних відстанях зв'язку виділяємо групи регіонів, які об'єдналися на кожному з 5-и рівнів (зверху вниз, тобто від відстані 1,113385 до відстані 2,412549). Враховуючи ієрахічність класифікації, на кожному рівні перелік регіонів може бути подібний. Нашою задачею є виокремлення регіонів та вилучення тих, що вже повторювалися на попередніх рівнях об'єднання. Найкраще таблицю Amalgamation Schedule скопіювати у Word (табл. 3) і там провести виокремлення. В результаті ми отримаємо табл. 4. Після цього, для наочності, виділені групи на дендрограмі можна позначити, як це зроблено на рис. 9.

Таблиця 3

Ступінь об'єднання регіонів у кластери за мірою подібності

та виділення груп районів

Amalgamation Schedule (Spreadsheet2) Ward's method Euclidean distances																							
	Obj. No. - 1	Ob j. No. - 2	Ob j. No. - 3	Obj. No. - 4	Ob j. No. - 5	Ob j. No. - 6	Obj. No. - 7	Ob j. No. - 8	Ob j. No. - 9	Ob j. No. - 10	Ob j. No. - 11	Ob j. No. - 12	Obj. No. - 13	Ob j. No. - 14	Ob j. No. - 15	Ob j. No. - 16	Ob j. No. - 17	Ob j. No. - 18	Obj. No. - 19	Ob j. No. - 20	Ob j. No. - 21	Ob j. No. - 22	Ob j. No. - 23
1, 1 1 3 3 8 5	Вінницька	Хмельницька	Житомирська	Херсонська	Київська	Одеська																	
1, 1 5 8 1 0	Дніпропетровська	Закарпатська	Миколаївська	Полтавська	Сумська	Харківська	Кіровоградська	Черкаська	Чернігівська														
1, 7 2 8 1 9 3	Івано-Франківська	Чернівецька	Львівська	Тернопільська	м. Київ																		
2, 3 9 2 9 2 7	Волинська	Рівненська	Закарпатська	Івано-Франківська	Чернівецька	Львівська	Тернопільська	м. Київ															
2, 4 1 2 5 4 9	Вінницька	Хмельницька	Житомирська	Херсонська	Київська	Одеська	Дніпропетровська	Закарпатська	Миколаївська	Полтавська	Сумська	Харківська	Кіровоградська	Черкаська	Чернігівська								
8, 7 4 5 2 4 9	Вінницька	Хмельницька	Житомирська	Херсонська	Київська	Одеська	Дніпропетровська	Закарпатська	Миколаївська	Полтавська	Сумська	Харківська	Кіровоградська	Черкаська	Чернігівська	Волинська	Рівненська	Закарпатська	Івано-Франківська	Чернівецька	Львівська	Тернопільська	м. Київ



Таблиця 4

Ступінь об'єднання регіонів у кластери за мірою подібності та виділення груп районів (повторювані райони були видалені)

Amalgamation Schedule (Spreadsheet2) Ward's method Euclidean distances

[illegible]

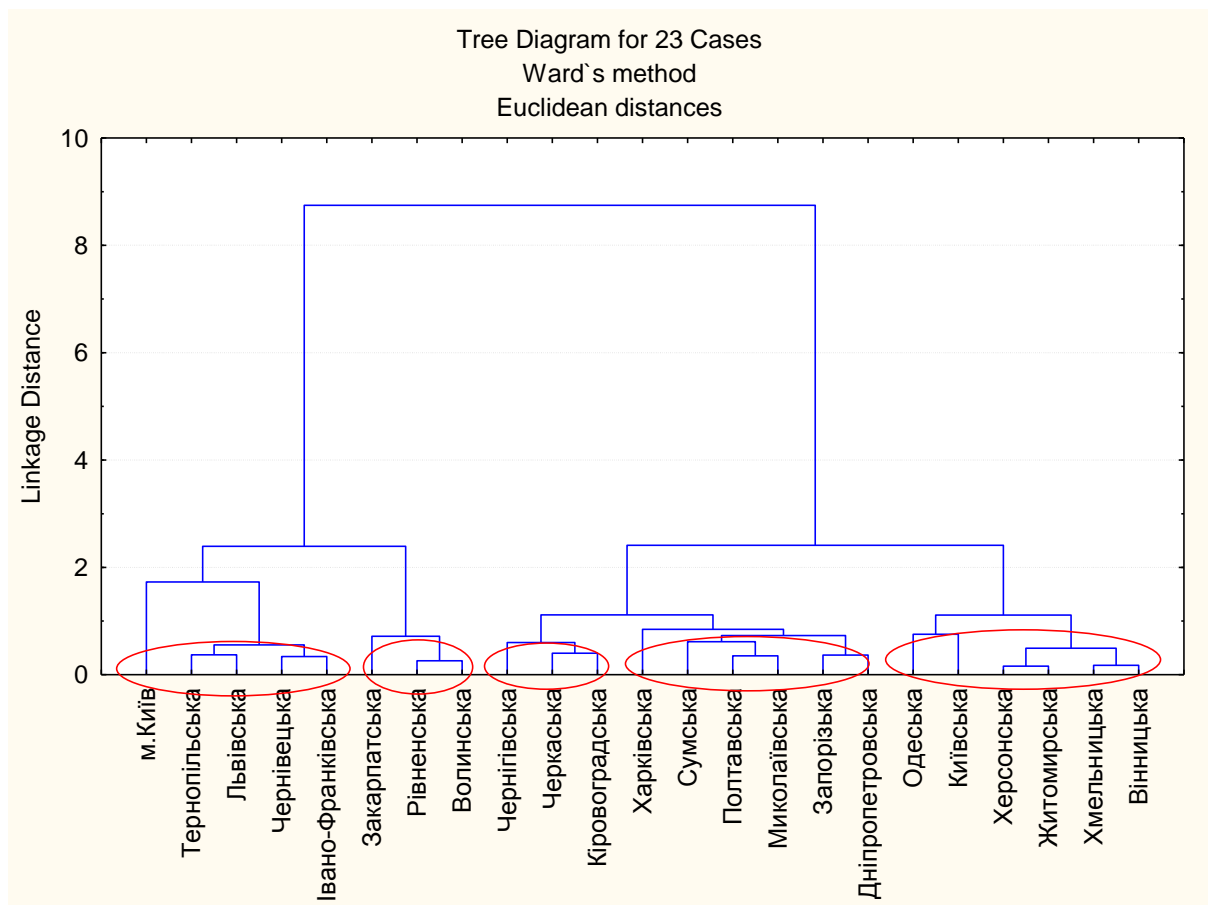


Рис. 9. Кластери регіонів України
за особливостями демографічної ситуації в 2021 р.

Отже, в залежності від отриманої дендрограми, можливе виділення кластерів та субкластерів. При групуванні регіонів необхідно дивитися на таблицю Amalgamation Schedule та дендрограму.

Завдання до практичної роботи:


Завдання 1. Проведіть кластерний аналіз показників зі своєї бази даних (5-10 показників), попередньо нормалізувавши їх (методом лінійного шкалювання). Наведіть таблицю з індексами, яку отримали після лінійного шкалювання. Побудуйте дендрограму кластеризації. Проаналізуйте таблицю з результатами об'єднань адміністративно-територіальних одиниць за ступенем міри відстані. Які регіони України об'єдналися найпершими, тобто мають найбільшу міру подібності? За скільки кроків всі регіони України об'єдналися в один кластер? На якій відстані всі регіони України об'єдналися в один кластер? Виділіть кластери адміністративно-територіальних одиниць за допомогою таблиці Amalgamation Schedule. Позначте їх на дендрограмі кластеризації або картосхемі.

Практичне заняття 19-20
ФАКТОРНИЙ АНАЛІЗ ВИХІДНИХ ДАНИХ.
ІНТЕРПРЕТАЦІЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ.

Питання для обговорення

1. Факторний аналіз: визначення поняття, мета і завдання.
2. Метод головних компонент (principal component analysis, PCA).
3. Етапи факторного аналізу:
 - 3.1. Підготовка даних до факторного аналізу.
 - 3.2. Визначення кількості факторів (за критеріями Кайзера, кам'янистого осипу Кеттела, інтерпретації та інваріантності, кумулятивним відсотком). Власні значення (eigenvalues) факторів.
 - 3.3. Обертання факторів – ортогональне і косокутне. Метод Varimax (варімакс).
 - 3.4. Ідентифікація факторів. Факторні навантаження (factor loadings).
 - 3.5. Обрахунок факторних ваг (factor scores).
 - 3.6. Інтерпретація факторів.

Методичні рекомендації до практичної роботи:

1. Відкриваємо програму Statistica 8.0. Натискаємо ОК. Закриваємо всі вікна.
2. Натискаємо кнопку File → New () або Ctrl+N. Натискаємо ОК.
3. З'являється таблиця, в яку необхідно додати базу даних (табл. 1).

Таблиця 1

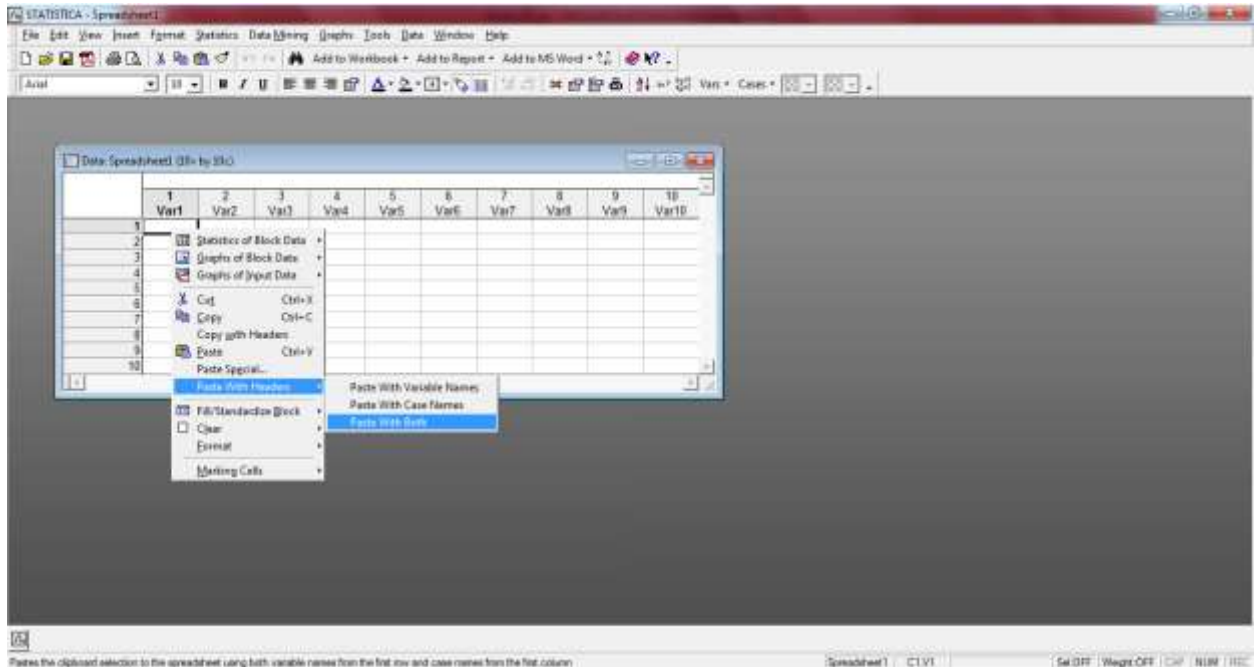
Показники, що впливають на демографічну ситуацію в Україні, 2021 р.

Регіони України	Загальний коефіцієнт народжуваності, ‰	Загальний коефіцієнт смертності, ‰	Коефіцієнт природного приросту, ‰	Частка населення у віці 0-14 років у загальній чисельності населення, %	Частка населення у віці 15-64 років у загальній чисельності населення, %	Частка населення у віці 65 років і старше у загальній чисельності населення, %	Середній вік, років	На 1000 чоловіків припадає жінок	Рівень урбанізації, %	Кількість міст	Кількість селищ міського типу	Кількість сіл	кількість прибулих на 1000 осіб наявного населення, ‰	кількість вибулих на 1000 осіб наявного населення, ‰	міграційний приріст на 1000 осіб наявного населення, ‰	ВРП, млн. грн. на 1000 осіб наявного населення	Наявний дохід у розрахунку на одну особу, грн	Середньомісячна номінальна заробітна плата штатних працівників, грн.
Вінницька	6,9	18,2	-11,3	15,2	66,8	18	42,2	1156	52,3	18	29	1456	10,78	12,41	-1,63	115,0	86274	12488
Волинська	9,6	15,1	-5,5	19,3	67,2	13,5	38,5	1118	52,2	11	22	1054	11,25	11,64	-0,39	90,6	70061	11445
Дніпропетровська	6,3	20,8	-14,5	15,1	67,3	17,6	42,3	1190	84,2	20	45	1435	9,72	9,81	-0,09	188,1	113085	13669
Житомирська	7,2	19,9	-12,7	16,1	67	16,9	41,3	1144	59,6	12	43	1613	14,10	15,22	-1,13	96,6	79328	11625
Закарпатська	10,1	14,8	-4,7	19,5	68,2	12,3	37,8	1080	37,2	11	19	578	5,38	5,25	0,13	60,8	60386	12235
Запорізька	5,8	21,6	-15,8	14,2	67,3	18,5	43,1	1190	77,5	14	22	914	8,12	9,28	-1,16	139,7	97924	13782
Івано-Франківська	7,8	15,2	-7,4	16,6	68,8	14,6	40,1	1115	44,6	15	24	766	10,73	10,16	0,56	88,5	71944	11988
Київська	7	20,2	-13,2	17,7	67,2	15,1	40,3	1184	61,8	26	30	1126	31,94	15,18	16,76	162,4	98771	14290
Кіровоградська	6,1	20,8	-14,7	14,7	66,6	18,7	42,8	1166	63,7	12	27	990	10,22	13,48	-3,26	110,2	76623	11180
Львівська	7,8	16	-8,2	16,1	68,6	15,3	40,6	1111	61,2	44	34	1850	11,27	10,94	0,34	119,5	89441	12530
Миколаївська	6,4	19,4	-13	15,1	67,6	17,3	42	1156	68,8	9	17	885	8,98	11,00	-2,03	113,7	85575	13807
Одеська	8,2	18	-9,8	16,8	67,1	16,1	40,6	1122	67,3	19	33	1122	13,11	10,37	2,73	115,5	96851	12326
Полтавська	5,9	20,3	-14,4	14	67,9	18,1	43	1162	62,7	16	20	1803	15,01	14,79	0,21	197,2	95770	12988
Рівненська	10,2	14,9	-4,7	20,1	67	12,9	37,9	1104	47,5	11	16	999	13,71	14,88	-1,17	77,8	70826	12762
Сумська	5,2	20	-14,8	12,8	68	19,2	44	1178	69,8	15	20	1454	12,37	14,59	-2,22	101,6	87410	11844
Тернопільська	7,1	16,4	-9,3	15,4	68,7	15,9	41,2	1134	46,1	18	17	1023	12,50	11,80	0,70	79,8	67467	11595

Харківська	5,9	21	-15,1	13,5	69	17,5	42,6	1155	81,4	17	60	1674	16,36	14,55	1,81	123,0	92746	12051
Херсонська	7	20	-13	15,9	67,3	16,8	41,5	1152	61,4	9	31	656	8,72	10,78	-2,07	88,0	76532	11398
Хмельницька	7	19	-12	15,5	66,9	17,6	42	1150	58	13	24	1414	13,60	13,71	-0,11	97,6	78500	12326
Черкаська	5,8	19,6	-13,8	13,6	67,1	19,3	43,6	1172	57,3	16	14	824	12,52	13,72	-1,20	113,0	79621	11965
Чернівецька	8,8	16,1	-7,3	17,1	68,3	14,6	39,7	1120	43,3	11	8	398	8,66	8,18	0,48	61,3	64130	11183
Чернігівська	5,5	22	-16,5	13,2	66,5	20,3	44,2	1199	66	16	29	1464	11,54	13,05	-1,51	118,3	76777	11363
м.Київ	10	15,3	-5,3	16,5	67,8	15,7	40,4	1169	100	1	0	0	16,65	14,71	1,94	432,3	225321	20558
Україна¹	7,3	18,5	-11,2	14,9	67,4	17,7	42,2	1157	69,7	461	881	28369	11,59	11,07	0,52	132,4	90036	14014

¹ Відносні показники розраховано без урахування відповідних даних по Донецькій та Луганській областях. У зв'язку з відсутністю інформації по частині тимчасово окупованих територій у цих областях здійснення розрахунків відносних показників по ним є некоректним

4. Дана операція робиться наступним чином. Ставимо курсор у першу ячейку, натискаємо праву кнопку мишки. У вікні, що відкривається, натискаємо «Paste With Headers → Paste With Both». Таким чином, ми копіюємо у таблицю програми Statistica наші показники, а також назви показників та назви регіонів (рис. 1).



	кoeffi- сип	кoeffициєнт природн	роєк у заставля	роєк у заставля	роєк у заставля	Середн й коеф	коєф и	Роль у заставля	Кількост ь селуц	Кількост ь селуц	Кількост ь селуц	на 1000 осіб	на 1000 осіб	на 1000 осіб	на 1000 осіб	розросту	заробіт	Нова
Волинська	6,9	18,2	-11,3	15,2	68,8	18	42,2	1156	52,3	18	29	1456	10,77637	12,40531	-1,62834	114,9581	86274	12498
Рівненська	9,6	15,1	-6,5	15,3	67,2	13,5	38,5	1118	52,2	11	22	1854	11,24611	11,63846	-0,39164	90,6014	70051	11445
Житомирська	6,3	20,8	-14,5	15,1	67,3	17,6	42,3	1190	84,2	20	46	1436	9,723606	9,870	-0,09139	108,0723	113986	13669
Київська	7,2	19,9	-12,7	16,1	67	16,9	41,3	1144	53,6	12	43	1613	14,09631	15,22266	-1,12635	96,62879	79328	11625
Закарпатська	10,1	14,8	-4,7	15,5	68,2	12,3	37,8	1080	37,2	11	19	578	5,383792	5,25201	0,131782	60,76936	80386	12235
Закарпатська	5,8	21,6	-15,6	14,2	67,3	18,5	43,1	1190	77,5	14	22	914	8,118589	8,282486	-1,1639	139,7878	97324	13782
Чернівецька	7,8	15,2	-7,4	16,6	68,8	14,6	40,1	1116	44,6	16	24	786	10,72632	10,1668	0,560725	88,53237	71944	11998
Київська	7	20,2	-13,2	17,7	67,2	15,1	40,3	1184	61,8	26	36	1126	31,94066	35,17593	16,78472	142,359	98771	14299
Київська	6,1	20,8	-14,7	14,7	66,6	18,7	42,8	1166	63,7	12	27	950	10,22228	13,47996	-3,25768	110,1723	76623	11588
Львівська	7,9	16	-6,2	16,1	68,6	15,3	40,6	1111	61,2	44	34	1850	11,27462	10,93048	0,33614	119,5182	89441	12530
Львівська	6,4	19,4	-13	15,1	67,6	17,3	42	1156	68,8	9	17	885	8,91583	11,60364	-2,6278	113,7281	86575	13807
Севська	8,2	18	-9,8	16,8	67,1	16,1	40,6	1122	67,3	19	33	1122	13,1067	10,37343	2,733275	115,5354	96851	12326
Полтавська	5,9	20,3	-14,4	14	67,9	18,1	43	1162	62,7	16	20	1883	15,00969	14,79498	0,211713	197,2176	95770	12988
Рівненська	10,2	14,9	-4,7	26,1	67	12,9	37,9	1184	47,5	11	16	999	13,71188	14,8011	-1,16822	77,8247	70826	12762
Сумська	5,2	20	-14,8	12,8	68	19,2	44	1178	69,8	15	20	1454	12,38862	14,58815	-2,22163	101,6189	87416	11684
Тернопільська	7,1	16,4	-9,3	15,4	68,7	15,9	41,2	1134	46,1	18	17	1023	12,50155	11,79879	0,701763	79,75332	67467	11595
Львівська	5,9	21	-15,1	13,5	69	17,5	42,6	1155	81,4	17	60	1674	16,36077	14,56197	1,8088	123,3476	92746	12051
Львівська	7	20	-13	15,8	67,3	16,8	41,5	1152	61,4	9	31	656	8,715073	10,78477	-2,06969	88,04131	76532	11398
Львівська	7	19	-12	15,5	68,9	17,6	42	1150	58	13	24	1414	13,68401	13,71224	-0,16823	97,55304	78580	12326
Черкаська	5,8	19,6	-13,8	13,6	67,1	19,3	43,6	1172	57,3	16	14	824	12,51865	13,72137	-1,20268	112,9813	79621	11965
Черкаська	8,8	16,1	-7,3	17,1	68,3	14,6	38,7	1120	43,3	11	8	398	8,661844	8,177823	0,484021	61,29662	64136	11583
Черкаська	6,5	22	-16,5	13,2	66,5	20,3	44,2	1189	66	18	29	1484	11,54261	13,81306	-1,51045	118,2865	76777	11363
м. Київ	10	15,3	-5,3	16,5	67,8	15,7	40,4	1169	508	1	0	0	16,64735	14,70548	1,941875	432,3326	225321	20558



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Загальний коефіцієнт народжуваності, ‰	Загальний коефіцієнт смертності, ‰	Коефіцієнт природного приросту, ‰	Частка населення у віці 0-14 років у загальній чисельності населення, ‰	Частка населення у віці 15-64 років у загальній чисельності населення, ‰	Частка населення у віці 65 років і старше у загальній чисельності населення, ‰	Середній вік, років	На 1000 чоловік в припадає жінок	Рівень урбанізації, ‰	Кількість міст	Кількість населення, тис. осіб	Кількість міст, тис. осіб	Кількість населення, тис. осіб	Кількість міст, тис. осіб	Кількість населення, тис. осіб	ВВП, млн грн на 1000 осіб населення
Вінницька	6,9	18,2	-11,3	15,2	66,9	18	42,2	1156	52,3	18	29	1456	10,7769714	12,485309	-1,62823758	114,9691
Волинська	9,6	15,1	-5,5	19,3	67,2	13,1	38,5	1118	52,2	11	22	1054	11,3468131	11,628443	-0,29163622	95,60014
Дніпропетровська	6,3	20,8	-14,5	15,1	67,3	17,6	42,1	1190	84,2	20	45	1435	9,72360596	9,9149999	-0,09138395	188,0722
Житомирська	7,2	15,9	-12,7	16,1	67	16,9	41,3	1144	59,6	12	43	1615	14,0963095	15,222657	-1,12634772	96,50778
Закарпатська	10,1	14,6	-4,7	19,5	68,2	12,3	37,8	1080	37,2	11	19	576	5,38179205	5,2520097	0,12978273	66,76935
Запорізька	5,8	21,6	-15,8	14,2	67,3	18,3	43,1	1190	77,5	14	22	914	8,11868926	9,2824856	-1,16384639	139,1078
Івано-Франківська	7,8	15,2	-7,4	18,6	68,8	14,8	40,1	1115	44,6	15	24	786	10,7255245	10,1648	0,56072471	86,53236
Київська	7	20,2	-13,2	17,7	67,2	15,1	40,3	1184	61,8	26	30	1126	31,9406555	15,175934	16,7647218	162,2989
Кіровоградська	6,1	20,8	-14,7	14,7	66,6	18,7	42,8	1166	63,7	12	27	990	10,2222832	13,479958	-3,25767501	110,1722
Львівська	7,8	16	-8,2	16,1	68,6	15,3	40,6	1111	61,2	44	34	1816	11,2746768	10,930477	0,356148151	119,5182
Миколаївська	6,4	19,4	-13	15,1	67,6	17,3	42	1156	68,8	9	17	885	8,97583029	11,863635	-2,02780492	113,7201
Одеська	8,2	18	-9,8	16,8	67,1	16,1	40,6	1122	67,3	19	33	1122	13,1667045	10,37343	2,73327459	115,5363
Полтавська	6,9	20,3	-14,4	14	67,9	18,1	43	1162	62,7	16	20	1083	15,0086927	14,79498	0,213712662	197,2175
Рівненська	10,2	14,9	-4,7	20,1	67	12,9	37,9	1104	47,5	11	16	989	13,7118755	14,881098	-1,1692229	77,62478
Сумська	5,2	20	-14,8	12,8	68	19,2	44	1178	69,8	15	20	1454	12,3666212	14,588153	-2,22153138	101,6188
Тернопільська	7,1	16,4	-9,3	15,4	68,7	15,9	41,2	1134	46,1	18	17	1023	12,5015538	11,798791	0,701762628	79,75331
Харківська	5,9	21	-15,1	13,5	69	17,5	42,6	1155	61,4	17	60	1074	16,3607688	14,551909	1,80879975	123,0476
Хмельницька	7	20	-13	15,9	67,3	16,8	41,5	1152	61,4	9	31	656	8,71507331	10,784768	-2,06985063	88,041
Хмельницька	7	19	-12	15,5	66,9	17,6	42	1150	58	13	24	1414	13,6640084	13,712241	-0,10823312	97,55303
Черкаська	6,8	19,6	-13,8	13,6	67,1	19,3	43,8	1172	57,3	16	14	824	12,5186949	13,721372	-1,2026769	112,9913
Чернівецька	8,8	16,1	-7,3	17,1	68,3	14,6	39,7	1120	43,3	11	8	398	9,69184442	9,1778233	0,486021126	61,29661
Чернівецька	5,5	22	-16,5	13,2	66,5	20,3	44,2	1199	66	16	29	1464	11,5426111	13,853084	-1,51045277	118,2864
м. Київ	10	15,3	-5,3	16,5	67,8	15,7	40,4	1169	100	1	0	0	16,547354	14,705479	1,94187517	432,3326

Рис. 1. Додавання бази даних у таблицю програми Statistica 8.0

- Після цього на панелі задач натискаємо «Statistics» → «Multivariate Exploratory Techniques» → «Factor Analysis» (рис. 2).

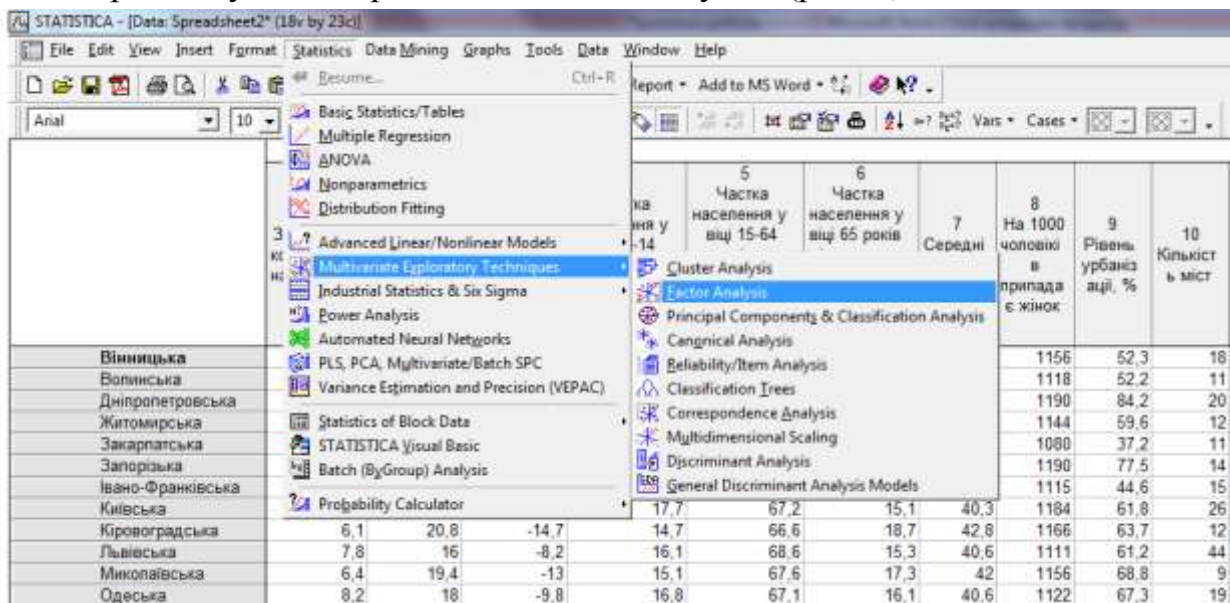


Рис. 2. Алгоритм вибору факторного аналізу в програмі Statistica 8.0

- Відкривається вікно факторного аналізу. Натискаємо на кнопку «Variables» та обираємо показники для аналізу (рис. 3). Натискаємо «ОК» і потім ще «ОК».

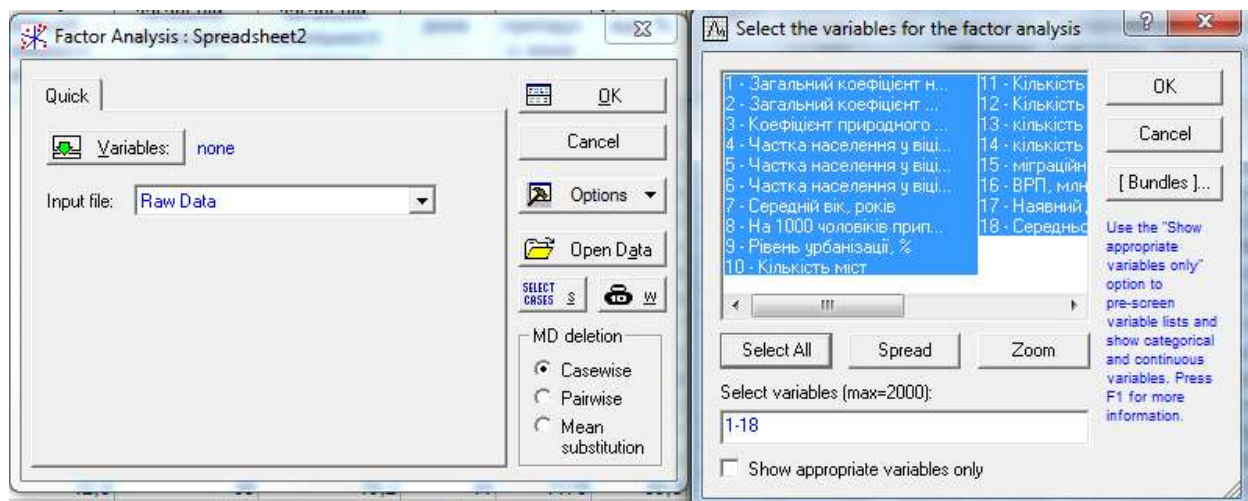


Рис. 3. Вибір вихідних даних і елементів виконання факторного аналізу

7. В результаті відкривається діалогове вікно «Define method of factor extraction» (рис. 4). Обираємо кількість факторів, яка в 2 рази менша за кількість показників (в нашому випадку 9), в полі «Minimum eigenvalue» має бути 1,000. Це є власне значення фактору, яке має бути більше 1 (за критерієм Кайзера). Натискаємо «ОК».

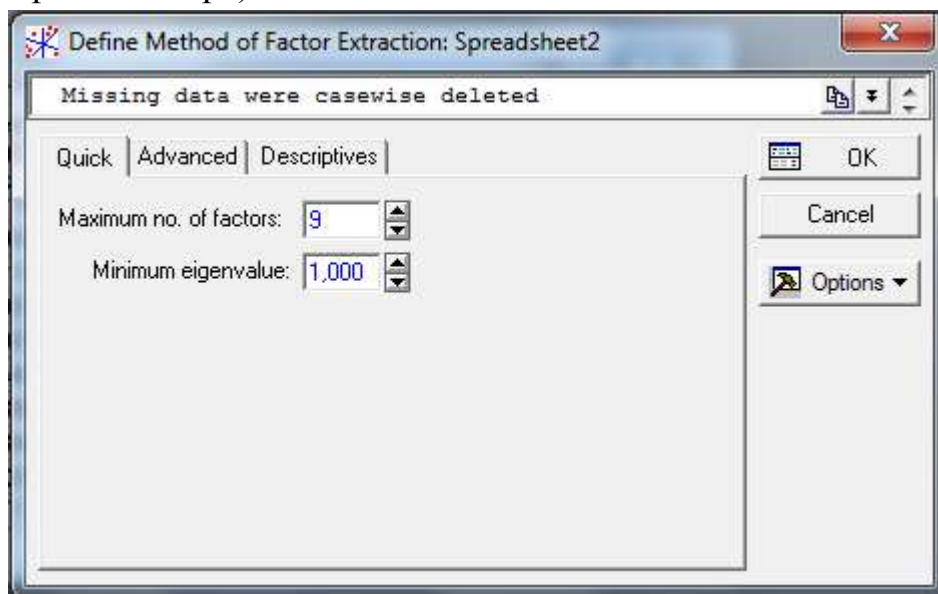


Рис. 4. Вибір кількості факторів

8. У вкладці «Advanced» у полі «Extraction Method» обираємо «Principal components» (метод головних компонент) (рис. 5). Натискаємо «ОК».

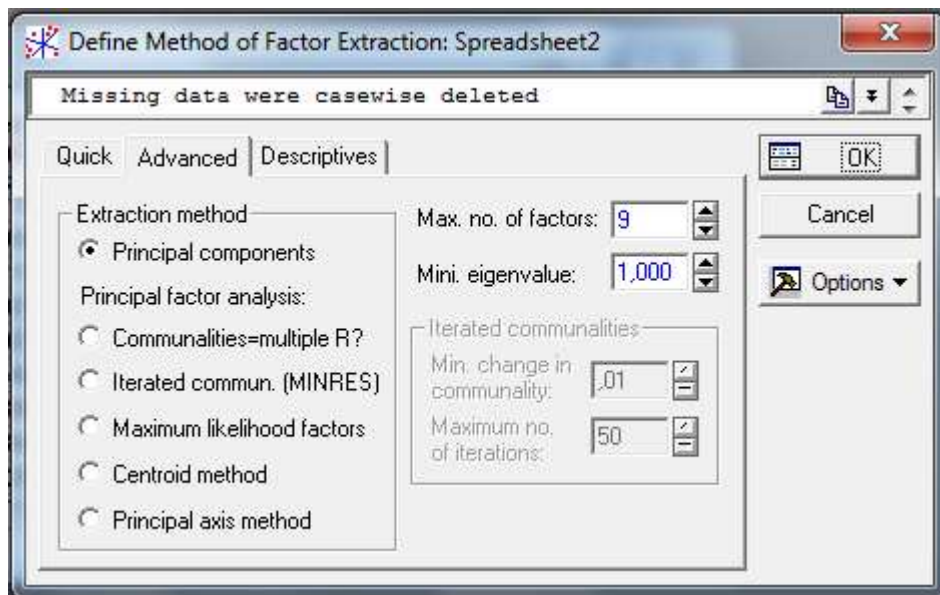


Рис. 5. Вибір методу виділення факторів

9. Відкривається діалогове вікно «Factor Analysis Results». На цьому етапі можна подивитися факторні навантаження (Summary: factor loadings або Summary), факторні ваги (Factor Scores), графіки (Plot of Eigenvalues, Plot of loadings 2D і 3D), власні значення факторів (Eigenvalues) без обертання осей координат (Unrotated) (рис. 6).

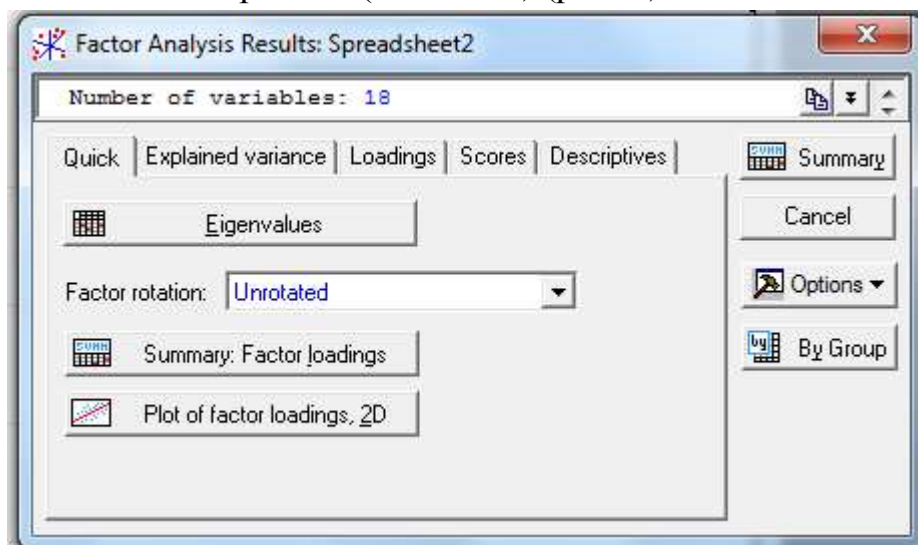


Рис. 6. Діалогове вікно факторного аналізу без обертання (Unrotated)

10. У полі «Factor rotation» обираємо «Varimax normalized». Тобто ми обрали варімаксий метод обертання факторів (рис. 7).

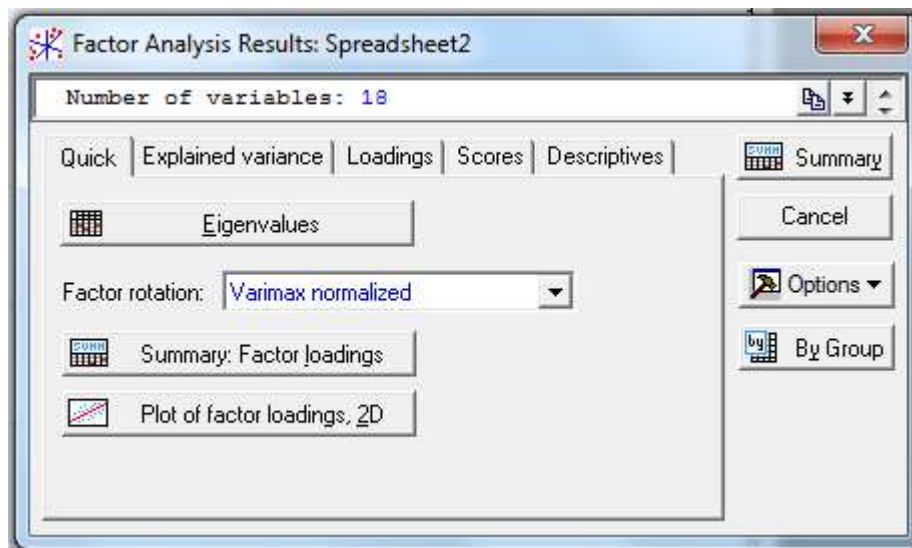


Рис. 7. Вибір методу обертання факторів

11. У вкладці «Loadings» в полі «Highlight factor loadings greater than:» задати порогове значення факторних навантажень, які будуть підсвічуватися, якщо значення навантажень буде більше цього значення. Ці значення потрібно враховувати при ідентифікації факторів (рис. 8).

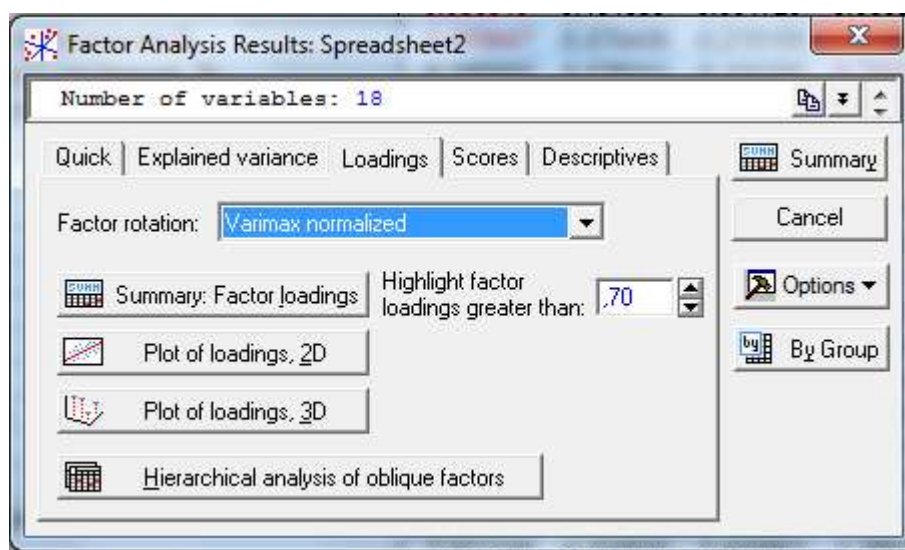
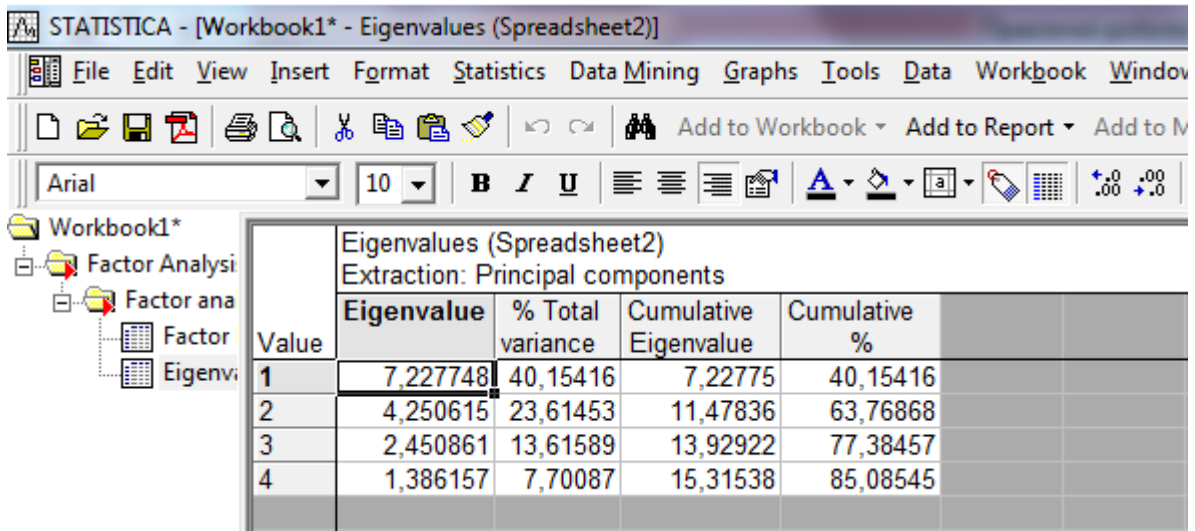


Рис. 8. Задання порогового значення факторних навантажень

12. Повертаємося до вкладки «Quick» та натискаємо кнопку «Eigenvalue». Відкривається таблиця із власними значеннями факторів – абсолютними, відносними та кумулятивними значеннями дисперсії кожного з факторів. У нашому випадку перший фактор пояснює приблизно 40 % дисперсії, другий фактор – 23,6 % дисперсії, третій фактор – 13,6 % дисперсії, а четвертий фактор – 7,7 % дисперсії. Разом

вони описують 85 % дисперсії, що свідчить про високу факторизацію. За допомогою таблиці визначаємо кількість факторів. Eigenvalue має бути більше за 1 (критерій Кайзера). Таким чином, у даному випадку за цим критерієм можна виділити 4 фактори. Cumulative % має бути вище 75 % (критерій кумулятивного відсотку). У даному випадку за цим критерієм можна виділити 3 фактори (рис. 9). Найбільш значимим фактором є перший, який має найбільші власні значення.



Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	7,227748	40,15416	7,22775	40,15416
2	4,250615	23,61453	11,47836	63,76868
3	2,450861	13,61589	13,92922	77,38457
4	1,386157	7,70087	15,31538	85,08545

Рис. 9. Таблиця власних значень факторів

13. Повертаємося до вкладки «Explained variance» та натискаємо кнопку «Scree plot». З'являється графік дисперсій (рис. 10). На ньому знаходимо таке місце, де зменшення дисперсії (власних значень) зліва направо максимально уповільнюється. Це місце відповідає четвертій точці, тобто за критерієм «кам'янистого осипу» Кеттеля можна виділити 4 фактори. У більшості випадків, графік дисперсій оцінюється за кількістю факторів, які лежать вище значення 1. Знову ж таки, цей критерій призводить до виділення 4 факторів, що підтверджує результати розрахунку критерію Кайзера.

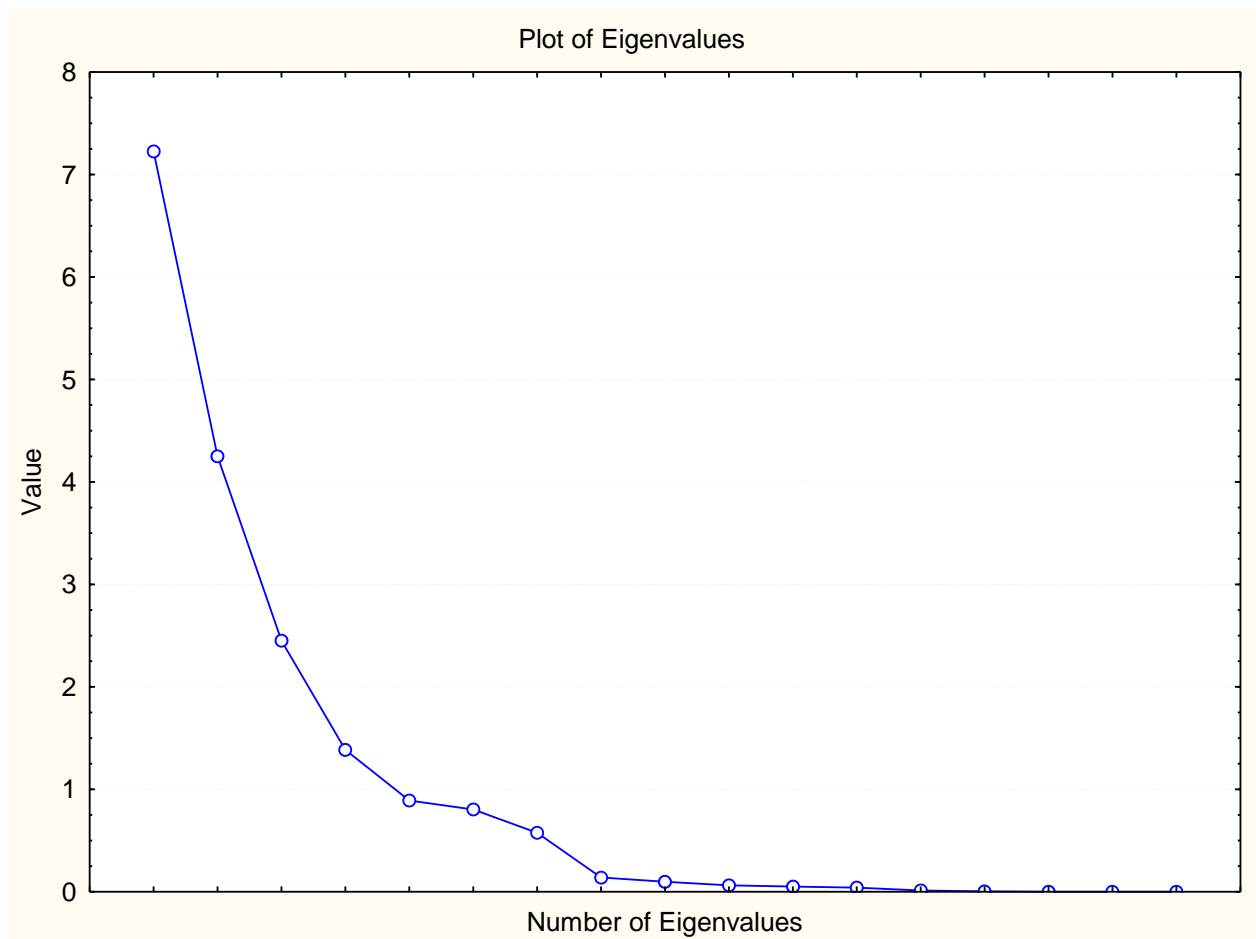


Рис. 10. Графік дисперсії

14. У лівому нижньому куту шукаємо «Factor Analysis Result» та повертаємося до діалогового вікна. Натискаємо кнопку «Summary». Перед нами відкривається результуюча таблиця факторних навантажень (рис. 11). Факторні навантаження (factor loadings) – це коефіцієнти кореляції кожного із аналізованих показників із кожним з виділених факторів. Чим тісніший зв’язок показника із фактором, тим більшим є його факторне навантаження. Факторні навантаження знаходяться в межах від -1 до +1. Позитивний знак факторного навантаження вказує на прямий зв’язок показника з фактором, негативний – на обернений. Таблиця факторних навантажень, отримана в результаті факторного аналізу, містить кількість строк, рівну кількості показників (стовпчиків у вихідній матриці), та кількість стовпчиків, рівну кількості виділених факторів. Програма виділила 4 фактори. Зміст факторів визначають показники (змінні), що мають найбільші факторні навантаження (найближчі по модулю до одиниці). Оскільки в «Highlight factor loadings greater than» ми обрали значення 0,7, то в зміст факторів обрані

показники зі значенням факторного навантаження більше 0,7 (по модулю). Вони виділені червоним кольором.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Загальний коефіцієнт народжуваності, ‰	-0.937528	0.118482	-0.012877	-0.213514
Загальний коефіцієнт смертності, ‰	0.925890	-0.055278	0.193818	0.037922
Коефіцієнт природного приросту, ‰	-0.957538	0.082141	-0.127179	-0.109181
Частка населення у віці 0-14 років у загальній чисельності населення, ‰	-0.914118	-0.151651	0.192639	-0.171336
Частка населення у віці 15-64 років у загальній чисельності населення, ‰	-0.348027	0.215261	-0.360705	0.627172
Частка населення у віці 65 років і старше у загальній чисельності населення, ‰	0.976744	0.067301	-0.055207	-0.057214
Середній вік, років	0.972201	0.088284	-0.080252	0.052271
На 1000 чоловіків припадає жінок	0.842279	0.330610	0.265068	-0.144327
Рівень урбанізації, ‰	0.478178	0.785748	0.087159	0.049920
Кількість міст	0.041551	-0.169507	0.201351	0.806157
Кількість селищ міського типу	0.350260	-0.229856	0.180039	0.616572
Кількість сіл	0.463618	-0.317811	0.224420	0.601400
кількість прибулих на 1000 осіб наявного населення, ‰	0.005096	0.230739	0.930663	0.158412
кількість вибулих на 1000 осіб наявного населення, ‰	0.380503	0.176758	0.633207	-0.085579
міграційний приріст на 1000 осіб наявного населення, ‰	-0.248713	0.181558	0.785796	0.263510
ВРП, млн. грн. на 1000 осіб наявного населення	0.061503	0.947742	0.179929	-0.127497
Найвищий дохід у розрахунку на одну особу, грн	0.017869	0.968784	0.134896	-0.115995
Середньомісячна номінальна заробітна плата штатних працівників, грн	-0.153317	0.928450	0.176517	-0.176834
Expl Var	7.020760	3.830996	2.398080	2.065545
Prp.Totl	0.390042	0.212833	0.133227	0.114752

Рис. 11. Факторні навантаження

В нашому випадку перший фактор в основному обумовлює дисперсію загальних коефіцієнтів народжуваності, смертності, коефіцієнту природного приросту, частки населення у віці 0-14 років та 65+ років, середнього віку та кількості жінок на 1000 чоловіків. Цей фактор умовно можна назвати демографічним або структурно-демографічним, оскільки включає показники демографічних процесів і статеві-вікової структури населення.

Другий фактор в основному обумовлює дисперсію ВРП на душу населення, доходу і середньомісячної зарплати, а також рівня урбанізації. Його можна інтерпретувати як економічний або урбанізаційно-економічний або фактор рівня життя. В принципі, показник рівня урбанізації ускладнює інтерпретацію другого фактору, оскільки всі інші показники цього фактору – суто економічні. Крім того, серед інших показників другого фактору він має найнижче факторне навантаження (0,785748). Дозволяється ігнорувати показники, які ускладнюють інтерпретацію фактору та мають найнижче навантаження серед інших показників цього фактору.

Третій фактор в основному обумовлює дисперсію прибулих і міграційного приросту. Це міграційний фактор.

Четвертий фактор обумовлює дисперсію тільки одного показника – кількості міст. Його можна назвати розселенським або екістичним фактором.

Декілька показників не увійшли до жодного фактору. Але якщо

15. Знову в лівому нижньому куту шукаємо «Factor Analysis Result» та повертаємося до діалогового вікна. Переходимо до вкладки «Scores» та натискаємо кнопку «Factor scores». Відкривається таблиця з факторними вагами (рис. 12). Факторні ваги (factor scores) – це показники, що відіграють роль оцінок вкладів територіальних одиниць у кожний з факторів.

STATISTIKA - [Workbook1* - Factor Scores (Spreadsheet2)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Workbook Windows

Arial 10 B I U

Workbook1*

- Factor Analysis
 - Factor analysis
 - Factor scores
 - Eigenvalues
 - Plot of factor loadings
 - Factor loadings
 - Factor scores

Factor Scores (Spreadsheet2)					
Rotation: Varimax normalized					
Extraction: Principal components					
Case	Factor 1	Factor 2	Factor 3	Factor 4	
Вінницька	0,38277	-0,430525	0,00103	-0,20561	
Волинська	-1,32333	-0,725790	0,24786	-0,63881	
Дніпропетровська	0,79335	0,806085	-0,38609	0,82474	
Житомирська	0,34943	-0,716662	0,73790	-0,10378	
Закарпатська	-1,99810	-0,643338	-1,25262	-0,05026	
Запорізька	1,05292	0,524698	-0,85528	-0,33447	
Івано-Франківська	-1,00459	-0,198757	-0,76795	0,62807	
Київська	-0,40985	0,108665	3,84358	0,54119	
Кіровоградська	1,00344	-0,603805	-0,08407	-1,01760	
Львівська	-0,75589	0,266666	-0,49805	2,90315	
Миколаївська	0,41061	0,236019	-0,77152	-0,58167	
Одеська	-0,42217	-0,072219	0,27015	0,29199	
Полтавська	0,77197	0,336079	0,26285	0,41556	
Рівненська	-1,59580	-0,799832	0,86956	-1,10263	
Сумська	1,20624	0,061046	-0,41367	0,18932	
Тернопільська	-0,44653	-0,328882	-0,45752	0,59348	
Харківська	0,76133	0,468948	0,10311	2,12464	
Херсонська	0,30456	-0,467903	-0,54015	-0,69292	
Хмельницька	0,35527	-0,536847	0,47528	-0,53725	
Черкаська	0,93749	-0,303164	-0,06577	-0,82774	
Чернівецька	-1,08670	-0,540483	-0,94498	-0,47222	
Чернігівська	1,62909	-0,519447	0,13945	-0,58229	
м.Київ	-0,91551	4,079450	0,08688	-1,36488	

Регіону з більшою факторною вагою властивий більший рівень прояву

властивостей виділеного фактора (більший рівень зв'язку з фактором). Позитивні факторні ваги відповідають регіонам, що мають рівень прояву властивостей фактора більше середнього, а негативні – нижче середнього. Таблиця факторних ваг складається з кількості строк, рівній кількості регіонів, та кількості стовпчиків, рівній кількості виділених факторів.

Як бачимо з таблиці, найбільші факторні ваги по першому демографічному фактору мають Чернігівська, Сумська, Запорізька, Кіровоградська, Черкаська області. По другому економічному фактору – м. Київ, Дніпропетровська, Запорізька, Харківська області. По третьому міграційному фактору – Київська, Рівненська, Житомирська, Хмельницька області. По четвертому розселенському фактору – Львівська, Харківська, Дніпропетровська, Івано-Франківська області.

Завдання до практичної роботи:

Завдання 1. Проведіть факторний аналіз показників зі своєї бази даних (10-15 показників). База даних має включати як показники з обраної Вами теми дослідження, так і показники, які відображають фактори, що впливають на ті процеси, об'єкти чи явища, які Ви обрали темою інформаційної бази. За різними критеріями (Кайзера, кумулятивного відсотку, Кеттеля) обґрунтуйте кількість факторів. Наведіть таблицю власних значень факторів, графік дисперсії («кам'янистого осипу»). Виділіть найбільш значимий фактор. Проведіть обертання факторів методом варімакс. Наведіть таблицю факторних навантажень. За аналізом факторних навантажень ідентифікуйте фактори (виясніть, які показники визначають кожний фактор та дайте назву факторам). Наведіть таблицю з факторними вагами. Визначте регіони, які роблять найбільший вклад у кожний фактор.