

Міністерство освіти і науки України
Харківська державна академія фізичної культури
Кафедра інформатики та біомеханіки

Я. В. Жерновнікова, С. С. Пятисоцька, О. М. Ольховий

Статистична обробка та аналіз наукових даних

Навчальний посібник

для здобувачів другого (магістерського) рівня вищої освіти за
спеціальністю 17 «Терапія та реабілітація»

Харків
ХДАФК
2025

УДК 004/519.25(075)
Ж 59

*Затверджено Вченою радою ХДАФК
(Протокол № 6 від 19 травня 2025 року)*

Рецензенти **І. В. Рубан** – Перший проректор Харківського національного університету радіоелектроніки, доктор технічних наук, професор.
Л. В. Подрігало – Завідувач кафедри спортивної медицини та гігієни Харківської державної академії фізичної культури, доктор медичних наук, професор.

Жерновнікова Я.В.

Ж 59 Статистична обробка та аналіз наукових даних : навчальний посібник для здобувачів другого (магістерського) рівня вищої освіти за спеціальністю І7 «Терапія та реабілітація» / Жерновнікова Я. В., Пятисоцька С. С., Ольховий О. М. – Харків : ХДАФК, 2025. 124 с.

Навчальний посібник з курсу «Статистична обробка та аналіз наукових даних» розроблено відповідно до навчального плану і складається з теоретичного матеріалу, методичних рекомендацій для проведення практичних занять, питань до самоконтролю, завдань для самостійної роботи, тестових завдань, додатків, списку рекомендованої літератури. Вивчення дисципліни допоможе здобувачам вищої освіти орієнтуватися в методах математичної статистики, особливостях їх застосування в практичних ситуаціях, проводити кількісну обробку емпіричних даних наукових досліджень, інтерпретувати отримані результати.

УДК 004/519.25(075)

© Жерновнікова Я. В., Пятисоцька С. С.,
Ольховий О. М.
©ХДАФК (ФЦВС), 2025

Вступ	4
Практична робота №1 Використання Google-таблиць для аналізу даних.....	6
Практична робота №2. Використання можливостей онлайн-сервісу RawGraphs для візуалізації даних.....	17
Практична робота №3. Сортуння та фільтрація статистичних даних.....	30
Практична робота №4. Робота зі зведеними таблицями.....	35
Практична робота №5. Створення таблиць даних у програмі SPSS.....	50
Практична робота №6. Описова статистика у програмах MS Excel та SPSS.....	60
Практична робота №7. Параметричні методи перевірки статистичних гіпотез у програмах MS Excel та SPSS.....	69
Практична робота №8. Непараметричні методи перевірки статистичних гіпотез.....	79
Практична робота №9. Кореляційний аналіз із використанням параметричних методів у програмах MS Excel та SPSS.....	84
Практична робота №10. Регресійний аналіз у програмах MS Excel та SPSS.....	92
Тестові завдання	104
Список рекомендованої літератури	119
Додатки	121

Вступ

Навчальна дисципліна «Статистична обробка та аналіз наукових даних» є нормативною дисципліною освітньо-професійної програми підготовки здобувачів другого магістерського рівня вищої освіти за спеціальністю І7 «Терапія та реабілітація». Зміст дисципліни забезпечує спеціальну підготовку майбутніх фахівців з фізичної терапії та реабілітації відповідно сучасних вимог.

Мета вивчення дисципліни: теоретичне і практичне оволодіння основами математико-статистичного аналізу числових даних у галузі охорони здоров'я з використанням комп'ютерних технологій.

Основні завдання навчальної дисципліни:

- засвоїти основний понятійний апарат теорії ймовірностей як теоретичну основи статистики;
- ознайомити здобувачів вищої освіти з основними методами статистичного спостереження та первинної обробки експериментальних даних з використанням електронних таблиць;
- оволодіти практичними навиками використання комп'ютерних технологій для розв'язання основних завдань статистичного аналізу.

Компетентності:

- здатність оцінювати, критично обговорювати та застосовувати результати наукових досліджень у практичній діяльності;
- здатність проводити наукові дослідження відповідного рівня, публікувати результати в українських та міжнародних наукових виданнях;
- здатність генерувати нові ідеї, брати участь у інноваційних проектах для вирішення клінічних, наукових та освітніх завдань;
- здатність впроваджувати сучасні наукові дані у практичну діяльність.

В результаті вивчення дисципліни здобувач вищої освіти повинен демонструвати уміння здійснювати науково-дослідну діяльність:

- знаходити, аналізувати та використовувати необхідну фахову інформацію;
- застосовувати пошуковий процес;
- розробляти і модифікувати пошукові стратегії;
- застосовувати сучасні цифрові технології та спеціалізоване програмне забезпечення, методи статистичного аналізу даних для розв'язання складних задач;
- демонструвати розуміння методів дослідження та уміння доповідати у науковому форматі;
- описувати відповідні методи дослідження, які можуть використовуватися для вивчення різноманітних дослідницьких питань;

- описувати різні методи збору даних;
 - використовувати стандартизований формат ведення записів;
 - інтегрувати сучасні наукові дані у практичну діяльність;
 - презентувати результати власних досліджень усно / письмово для фахівців і нефахівців;
 - аргументувати доцільність науково обґрунтованого навчання.
- Відповідно до робочої програми навчальної дисципліни на практичні заняття відводиться 28 академічних годин, які розподілені на 10 практичних робіт.
- Метою виконання практичних робіт є закріплення матеріалу, вивченого під час лекційних занять, набуття навичок математико-статистичної обробки емпіричних даних з використанням електронних таблиць Microsoft Excel та програми SPSS, демонстрація зв'язку набутих знань з практичними завданнями досліджень в галузі охорони здоров'я.

Практичне заняття №1

Тема «Використання Google-таблиць для аналізу даних»

Мета заняття: ознайомитись із основними можливостями роботи з електронними таблицями онлайн-офісу. Закріпити навички створення, редагування, та форматування таблиць, проводити базові обчислення, надавати спільний доступ для інших користувачів.

Зміст і хід заняття

1. Загальні теоретичні відомості.

Виконання та захист практичного завдання.

1. Загальні теоретичні відомості.

Для роботи з електронними таблицями використовується редактор Google Sheets, що дозволяє заносити дані до рядків і стовпчиків електронної таблиці, а також проводити базові обчислення.

Програма дозволяє перетворення даних у формат Microsoft Excel та Open Document.

Передбачена можливість доступу інших користувачів до створеної таблиці як з правом редагування, так і без (щонайбільше десяти користувачам одночасно).

За допомогою електронних таблиць Google можна:

- імпортувати й експортувати файли типів *.xls, *.pdf або *.html-файли;
- формувати клітинки і змінювати формули, обчислюючи їх результати, і відображати дані в потрібному вигляді;
- спілкуватися в режимі реального часу з іншими користувачами, що редагують цю ж таблицю;
- вставляти таблицю або її частину у свій блог або на Web-сайт.

Щоб розпочати роботу з електронними таблицями, необхідно виконати наступні дії, а саме на диску Google натиснути кнопку Створити таблицю.

Отримаємо вікно вигляду (рис. 1), яке містить такі вкладки: Файл, Редагувати, Вигляд, Вставити, Формат, Дані, Інструменти, Доповнення.

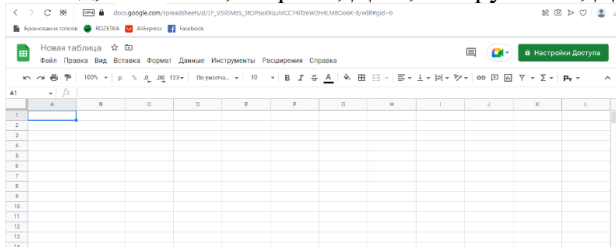


Рис. 1. Вікно електронних таблиць Google

Під час роботи з документом, якщо ви не назвали його на початку роботи, він буде автоматично збережений під назвою, яка складається з декількох перших слів набраного тексту. Щоб уникнути цього, необхідно зберегти документ на початку роботи (ще до набору тексту) або перейменувати його пізніше, після завершення роботи.

На панелі інструментів головного вікна електронних таблиць знаходяться кнопки: роздрукувати, повернення, назад, формат заливки, грошовий та процентний формат, зменшити (збільшити) кількість десяткових цифр, інші формати, назва та розмір шрифту, жирний, курсив, закреслення, колір тексту, колір заповнення, межі клітинки, об'єднання клітинок, вирівнювання по горизонталі та вертикалі, перенесення тексту в декілька рядків у клітинці, гіперпосилання, коментар, діаграма, фільтр, функції, транслітерація.

Інтерфейс вікна електронних таблиць дуже схожий на інтерфейс MS Excel 2010. Наявні усі засоби редагування і форматування клітинок та роботи з діаграмами, як і у офісному редакторі, що встановлений на персональному комп'ютері.

Для надання спільного доступу редагування або перегляду електронних таблиць використовується кнопка Спільний доступ.

Практичне завдання

Завдання 1. Створення та редагування Google-таблиці.

1. Перейдіть на офіційний сайт Google Sheets за посиланням https://workspace.google.com/intl/ru_ru/products/sheets/ та натисніть кнопку для переходу безпосередньо до аркушів.

2. Якщо у вас ще немає особистого профілю в Google, розробники запропонують створити його. Цей крок є обов'язковим для початку користування усіма інструментами та можливості зберігати зміни у хмарі. Для цього натисніть на спеціальну кнопку на сторінці та дотримуйтесь інструкцій від розробників.

3. Після авторизації з'явиться основне вікно керування документами, в якому показані раніше створені таблиці, шаблони та кнопки для додавання нового файлу. У цьому вікні можна скористатись засобом сортування за часом створення та редагування таблиць в акаунті.

4. Створіть нову таблицю за прикладом (табл. 1).

Таблиця 1

Вхідні дані

Ім'я	Стать	Вік	Довжина тіла	Маса тіла	ОГК у спокої	ОГК на видиху	Індекс Пінье	Оцінка індексу Пінье	Індекс міщності статури	Оцінка індексу міщності статури
1	2	3	4	5	6	7	8	9	10	11
Артем		9	146	37,6	69	67				
Альона		11	140	50,1	84	82				
Марина		11	144	30	70	69				
Валерія		10	137	32,9	62	61				
Олександр		10	147	33,3	66	65				
Олена		10	142	43,7	73	71				
Кирило		10	140	28,6	59	58				
Богдан		9	142	34	66	64				
Олександр		10	146	29,8	61	60				
Сергій		10	138	39,5	69	67				
Кирило		10	141	25,9	63	61				
Богдан		10	140	36,2	72	70				
Олександр		10	148	44,9	74	72				
Сергій		10	143	42,9	78	76				
Віталій		10	145	35,2	69	66				
Світлана		10	153	61,9	85	83				
Микола		10	151	35,5	67	65				

1	2	3	4	5	6	7	8	9	10	11
Вікторія		10	146	45,5	75	75				
Дмитро		10	141	39,4	67	65				
Вікторія		10	143	32,2	60	59				
Марк		11	149	37,5	67	67				
Сергій		10	150	35,9	66	65				
Данило		10	154	58,4	86	83				
Єлизавета		11	143	31,9	63	62				
Богдан		11	135	33,9	65	64				
Данило		11	139	44,2	81	80				

5. Відредагуйте таблицю. Оберіть будь яку клітинку з заголовків та натисніть на кутову кнопку в лівому верхньому куті аркуша таким чином виділивши всі клітинки аркуша.

6. Подвійним кліком на межі між колонками автоматично налаштуйте ширину кожної колонки відповідно до тексту який в ній міститься.

7. Закріпіть верхній рядок таблиці. Перетягніть горизонтальну сіру лінію в лівому верхньому куті на один рядок вниз.

Завдання 2. Розрахунок формул.

1. Зробіть розрахунки індексів фізичного розвитку в 8-му та 10-му стовбцях за наведеними формулами:

$$\text{Індекс Піньє} = \frac{\text{ОГК}_{(\text{в спокої})} * 100\%}{\text{довжина тіла}};$$

$$\text{Індекс міцності статури} = \text{довжина тіла} - (\text{маса тіла} + \text{ОГК}_{(\text{на видиху})}).$$

2. Використовуючи функцію ЕСЛИ розрахуйте «Оцінку індексу Піньє» та «Оцінку індексу міцності статури».

Результати оцінки індексу Піньє оцініть за наступною шкалою:

- величина індексу менше 50% – вузька грудна клітина;
- більше 55% – широка грудна клітина;
- 51 – 54% – нормальна грудна клітина.

Результати оцінки індексу міцності статури оцініть за наступною шкалою:

- 10-15 – міцна статура;
- 16-20 – добра статура;
- 21-25 – середня статура;
- 26-30 – слабка статура;
- 31 і більше – дуже слабка статура.

Завдання 3. Робота зі статистикою по стовпцям.

Статистика стовпчиків дозволяє отримувати відомості про значення в тому чи іншому стовпці. Завдяки цьому можна виявляти проблеми, отримувати уявлення про наявні дані тощо. Доступні засоби візуалізації даних (кількість і розподіл), таблиці частотності (визначення найпоширеніших і рідкісних значень), а також зведена статистика на рівні стовпця.

1. Виділіть колонку «Індекс Піньє» → виберіть вкладку Дані → **Статистика по стовпцях.**

2. Праворуч з'явиться вікно «Статистика по стовпцях», де представлено розгорнуту описову статистику за даними, що містяться в цій колонці, та розподіл значень у вигляді діаграми. Це корисний

інструмент для швидкого представлення кількісних та якісних характеристик емпіричних даних.

3. Перейдіть до колонки «Оцінка індексу Піньє», що містить текстові значення, оберіть **Кількісне відображення розподілу**. В цій колонці записано три категорії і більшість значень стосується категорії «Вузька грудна клітина» (рис. 2).

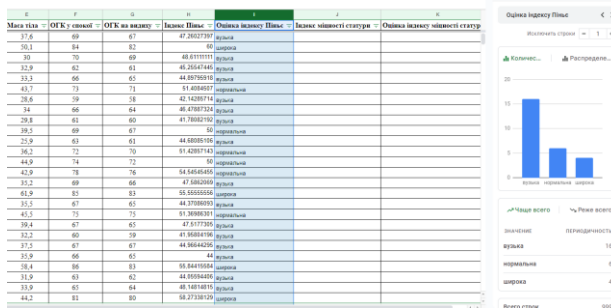


Рис. 2. Статистика по стовпцях

Завдання 4. Переклад тексту

Ще один засіб роботи з текстовими даними – переклад іншими мовами. Для цього Google пропонує використовувати власну функцію, інтегровану з фірмовим перекладачем. Потрібно лише обрати, якою мовою здійснювати переклад, використовуючи для цього кодові позначення.

1. Додайте ще одну колонку поруч з колонкою «Оцінка індексу Піньє».

2. Виділіть клітинку **J2** → виберіть вкладку **Вставлення** → **Σ Функція** → **Google** → **GOOGLETRANSLATE** та оберіть функцію **=GOOGLETRANSLATE(I2;"UK";"EN")**, де **I2** - клітинка з текстом для перекладу, **UK** – мова оригіналу, а **EN** – мова перекладу.

3. Натисніть **Enter** та ознайомтеся з результатом. Скопіюйте (розтягніть) формулу, щоб перекласти всі інші слова (номери клітинок будуть замінені автоматично).

Завдання 5. Створення списку значень.

Списки, що випадають, дозволяють вибрати значення для таблиці із заздалегідь підготовлених переліків.

Функція корисна, коли в таблицю багато разів потрібно вносити ті самі дані – наприклад, види спорту, спортивні розряди, стать спортсменів. Крім того списки, що випадають, прискорюють роботу, завдяки їм також зменшується ймовірність внесення помилкових даних.

1. Виділіть діапазон клітинок B2:B11, потім клацніть на них правою кнопкою миші і виберіть з контекстного меню **Список, що розкривається**. Праворуч з'явилася панель «Правила перевірки даних».

2. У розділі «Критерії» залиште варіант «Список, що розкривається» і перейдіть до розділу нижче. Замість «Варіант 1» і «Варіант 2» введіть значення для списку – «Чоловіча», «Жіноча».

3. У невеликих віконцях ліворуч від значень списку задайте їх колір. «Чоловіча» – синій колір. «Жіноча» – червоний колір та натисніть «Готово».

Тепер до всіх клітинок виділеного діапазону прикріплений список, що розкривається. Щоб його викликати, натисніть на стрілку вниз у правій частині клітинки.

4. Заповніть стовбець «Стать» відповідно до даних стовбця «Ім'я» (рис. 3).

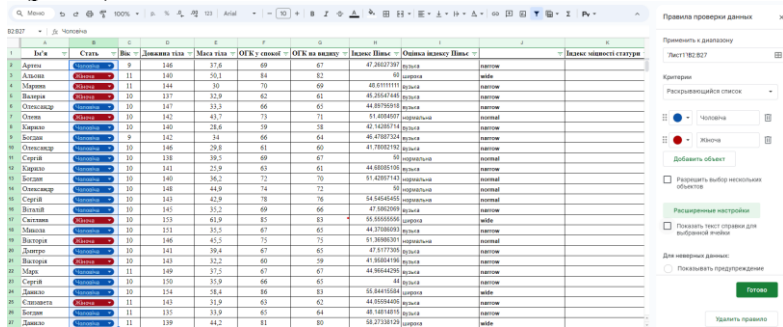


Рис. 3. Приклад списку значень

Щоб змінити зовнішній вигляд клітинок списку, що випадає, в панелі налаштувань натисніть кнопку «Розширені налаштування» і прокрутіть панель вниз.

У категорії «Стиль показу» можна вибрати, як виглядатимуть значення клітинок списку, що випадає. За замовчуванням стоїть стиль «Чіп». Крім нього, можна вибрати стиль «Стрілка» або «Звичайний».

Для того щоб видалити список, що випадає, необхідно його виділити та вибрати вкладку Дані → Налаштувати перевірку даних та натиснути кнопку **Видалити правило**.

Завдання 6. Створення діаграми.

1. Виділіть стовбець «Оцінка індексу Піньє».
2. Виберіть вкладку **Вставлення** → **Діаграма**.
3. У вікні «Редактор діаграм» відкрийте вкладку **Налаштування** та змініть тип діаграми на «Об'ємна кругова діаграма».

4. Відкрийте вкладку **Додаткові**, щоб змінити зовнішній вигляд діаграми. Встановіть для неї фон, колір межі та шрифт на ваш розсуд (рис. 4).



Рис. 4. Приклад кругової діаграми

Також можна візуалізувати дані у вигляді таблиці.

1. Виділіть весь діапазон даних.
2. Виберіть вкладку **Вставлення** → **Діаграма**.
3. У вікні «Редактор діаграм» відкрийте вкладку **Налаштування** та змініть тип діаграми «Таблиця».

Цю таблицю можна скролити або перенести на окремий аркуш.

4. В правому верхньому куті таблиці натисніть лівою кнопкою миші на три крапки та виберіть з контекстного меню «Перемістити на окремий аркуш».

Данні в такій таблиці можна сортувати і це не вплине на джерело даних. Готову візуалізацію можна завантажити, опублікувати або використовувати для дашбордів (рис. 5).

Ім'я	Статьа	Вік	Довжина тіла	Маса тіла	ІІТ в у стані	ІІТ в на вантажі	Часовий пояс	Оцінка індексу Пінье	Часовий пояс	Оцінка індексу Пінье	Часовий пояс	Оцінка індексу Пінье
Анна	Жінка	8	140	25.0	48	57	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Анна	Жінка	11	140	30.1	54	62	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Марина	Жінка	11	140	30	70	60	48:17:11	нормальна	нормальна	нормальна	нормальна	нормальна
Вікторія	Жінка	10	137	20.0	40	47	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Світлана	Жінка	10	147	10.0	48	55	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Олена	Жінка	10	140	40.0	70	71	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Гарет	Чоловік	10	140	20.0	50	59	47:14:07	нормальна	нормальна	нормальна	нормальна	нормальна
Вікторія	Жінка	10	140	30.0	50	54	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Світлана	Чоловік	10	140	20.0	61	60	47:10:00	нормальна	нормальна	нормальна	нормальна	нормальна
Олеся	Чоловік	10	140	30.0	50	57	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Гарет	Чоловік	10	141	20.0	60	61	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Вікторія	Чоловік	10	140	20.0	70	72	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Світлана	Чоловік	10	140	40.0	74	70	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Олеся	Чоловік	10	140	40.0	70	70	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Вікторія	Чоловік	10	140	20.0	40	48	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Олена	Жінка	10	130	10.0	40	40	50:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Марина	Чоловік	10	130	20.0	47	48	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Вікторія	Жінка	10	140	40.0	70	70	51:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Дарина	Чоловік	10	140	20.0	67	67	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Вікторія	Жінка	10	140	20.0	60	60	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Олеся	Чоловік	10	140	20.0	60	61	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Марина	Чоловік	10	140	20.0	60	61	47:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Дарина	Чоловік	10	140	20.0	60	61	52:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Світлана	Жінка	11	140	20.0	40	42	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна
Вікторія	Чоловік	11	130	10.0	40	40	48:14:07	нормальна	нормальна	нормальна	нормальна	нормальна
Дарина	Чоловік	11	130	40.0	61	60	48:00:00	нормальна	нормальна	нормальна	нормальна	нормальна

Рис. 5. Приклад діаграми «Таблиця»

Завдання 7. Створення зведеної таблиці.

1. Помістіть табличний курсор в будь-яку клітинку діапазону даних.
2. Виберіть вкладку **Вставлення** → **Створити зведену таблицю**

3. У діалоговому вікні що відкрилося виберіть розташування зведеної таблиці «на існуючий лист», в полі «Діапазон» виберіть клітинку «N2».

Після того, як створили зведену таблицю в правій частині екрана відкрився **Редактор зведеної таблиці**, під полем рекомендацій знаходяться кнопки додавання даних до зведеної таблиці.

4. Натисніть кнопку **Рядки** і виберіть значення **Стать** та **Вік**. Це означає, що в зведеній таблиці значення **Стать** та **Вік** будуть розміщуватися в стовпцях. Також тут через випадають списки можна вибрати порядок сортування і відобразити додаткове поле «**підсумок**», призначене для виведення суми всіх даних з таблиці (рис. 6).

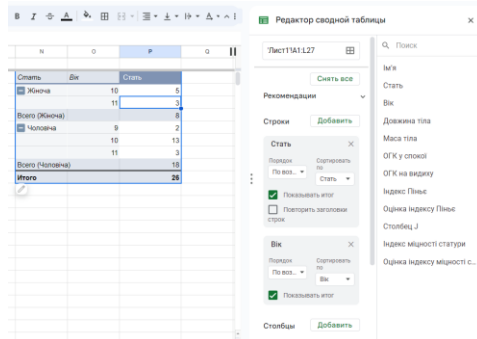


Рис. 6. Редагування зведеної таблиці

5. Виберіть дані, що підраховуються, натисніть кнопку **Значення**, і виберіть значення **Стать**. Змінити принцип обчислення підсумкової суми для кожної клітинки можна через список, що випадає, обрати «**підсумувати за**», де на вибір доступно багато функцій (рис. 7).

Стать	Вік	Стать
Жіноча	10	5
	11	3
Всього (Жіноча)		8
Чоловіча	9	2
	10	13
	11	3
Всього (Чоловіча)		18
Итого		26

Рис. 7. Приклад зведеної таблиці

Отже, за допомогою зведеної таблиці впорядкували дані із загального списку та розраховували кількість спортсменів відповідно до віку та статі.

Для того перегляду вхідних даних для клітинки зі зведеної таблиці виконайте наступні кроки.

6. Клацніть правою кнопкою миші на клітинку P3. Виберіть з контекстного меню **Показати відомості про клітинку**.

7. Відкриється новий аркуш із вихідними даними (рис. 8).

	A	B	C	D	E	F	G	H	I	J	K	L
1	Ім'я	Стать	Вік	Довжина тіла	Маса тіла	ОГК у слові	ОГК на висоту	Індекс Пінч	Оцінка Індeksu Пінч	Індекс міцності статури	Оцінка Індeksu міцності статури	
2	Валерія	Жінка	10	137	32.9	62	61	45.26	вузька	laplow	43.1 дуже слабка статура	
3	Олена	Жінка	10	142	43.7	73	71	51.41	нормальна	normal	27.3 слабка статура	
4	Світлана	Жінка	10	153	61.9	85	83	55.56	широка	wide	8.1 міцна статура	
5	Вікторія	Жінка	10	146	45.5	75	75	51.37	нормальна	normal	25.5 слабка статура	
6	Вікторія	Жінка	10	143	32.2	60	59	41.96	вузька	laplow	51.8 дуже слабка статура	

Рис. 8. Показати відомості про клітинку

Примітка. Зведена таблиця оновлюється кожного разу після зміни клітинок з вхідними даними.

Завдання 8. Сорткування даних в таблиці.

1. Виконайте сорткування за зростанням по стовпцю «Довжина тіла». Виберіть вкладку **Дані** → **Сортувати лист** → **Сортувати лист по стовпцю A (A > Я)**.

При використанні такого методу сорткування не потрібно виділяти заголовки таблиць, оскільки їх значення також потрапляють до параметру і будуть перемішані з усіма іншими. Можна використати закріплення першого рядка, щоб заголовки завжди залишались у вихідному положенні. Здійснюється це через вкладку **Вид**.

Завдання 9. Створення фільтра.

Режим фільтрації – метод відбору даних, який динамічно налаштовується для кожного стовпця таблиці. Стане в нагоді при роботі з великим діапазоном даних для зміни їх розташування, враховуючи наявні параметри пошуку. Такий фільтр допоможе не тільки розташувати значення в заданому порядку, але й відібрати певний діапазон даних за встановленим правилом.

1. Створіть копію Листа 1.

2. На другому листі виділіть перший рядок таблиці.

3. Виберіть вкладку **Дані** → **Створити фільтр**. Після виконання команди біля заголовків стовбців з'являться кнопки **Фільтра** . Фільтрація доступна у трьох різних варіантах. Можна видалити або показати клітинки з конкретним кольором, умовою чи значенням.

4. Відфільтруйте спортсменів, які мають вузьку грудну клітину та масу тіла від 30 до 40 кг.

Відстежувати параметри, що діють, досить просто: якщо відповідний значок фільтра  знаходиться праворуч від назви стовпця, то зараз діють певні умови і окрема частина вмісту прихована.

Таким чином можна налаштувати розташування даних у потрібному порядку або зовсім приховати їх на певний час. При цьому розрахунки у функціях теж зміняться, якщо якісь задіяні клітинки зараз неактивні. Це дуже зручно при розрахунках чи перегляді загальної інформації.

Завдання 10. Налаштування загального доступу та захист аркушів.

1. Створіть новий аркуш та дайте йому назву «Самостійна робота».
2. Натисніть кнопку «Налаштування доступу».
3. Дайте назву документу «Практична робота 4 Прізвище Ім'я та група студента (написати особисті дані)».
4. Введіть електронні адреси користувачів, котрим потрібно надати доступ до файлу.
5. Вкажіть рівень доступу для цих користувачів – **Редактор**.
6. Натисніть **Надіслати**.

Для того, щоб заборонити користувачам вносити зміни до таблиці, зробіть наступні кроки:

7. **Виберіть вкладку Дані → Захистити аркуші та діапазони.**
8. Натисніть **Додати аркуш або діапазон**.
9. Виберіть **Аркуш 1**.
10. Натисніть **Задати дозволи**.
11. Натисніть кнопку **Готово**.

Для того, щоб переглянути зміни в таблиці, натисніть «Остання зміна»  у верхньому правому куті.

Самостійна робота.

1. Створіть власну базу даних, яка буде містити мінімум 7 ознак та 15 об'єктів.
2. Розрахуйте декілька формул та функцій.
3. Створіть декілька зведених таблиці за показниками вхідних даних.
4. Прорангуйте емпіричні дані за зростанням або за спаданням (на окремому аркуші).
5. Побудуйте одну діаграму на ваш вибір.
6. Надайте спільний доступ для 3-х користувачів, щоб вони відредагували ваш документ, і перегляньте зміни, які вніс кожен користувач.

Контрольні питання

1. Які дані можна вводити в клітинки електронної таблиці?
2. Як змінити назву електронної таблиці?
3. Як надати спільний доступ до електронної таблиці?

Практичне заняття № 2

Тема «Використання можливостей онлайн-сервісу RawGraphs для візуалізації даних»

Мета заняття: навчити створювати різні типи діаграм в онлайн-сервісі RawGraphs.

Зміст і хід заняття

1. Загальні теоретичні відомості.

Виконання та захист практичного завдання.

1. Загальні теоретичні відомості.

Візуалізація емпіричних даних – це невід’ємна частина математичної статистики. Правильне відображення інформації за допомогою графічного зображення емпіричних даних найчастіше є не менш важливим, ніж етапи збору, обробки й аналізу даних, оскільки отримані дані мають бути представлені в наочній і зрозумілій формі. Різноманітні діаграми, графіки та схеми дають змогу описувати складні системи, процеси та виявляти можливі взаємозв’язки між ознаками.

RawGraphs – це онлайн-сервіс з відкритим кодом для створення візуалізацій статистичних даних, призначених для подальшої модифікації. Цей сервіс можна використовувати без встановлення та завантаження будь-якого програмного забезпечення. Він представляє підхід до візуалізації даних на основі діаграм: кожна візуальна модель є незалежним модулем, що надає різні візуальні змінні, які можна використовувати для зіставлення вимірювань даних. Отже, користувачі можуть створювати складні візуалізації даних. Зображення, отримані після парсингу даних у RawGraphs, можна експортувати та редагувати.

Завантаження та форматування даних для RawGraphs.



Дані можна завантажити такими способами:

- вставити з таблиці;
- завантажити файл у табличному форматі;
- із URL-адреси;
- із запиту SPARQL;
- з одного з представлених зразків даних.

Завантаження файлу.

Для того щоб завантажити дані з файлу, натисніть кнопку «Завантажити дані» та виберіть файл на своєму комп'ютері.

Приймаються такі формати:

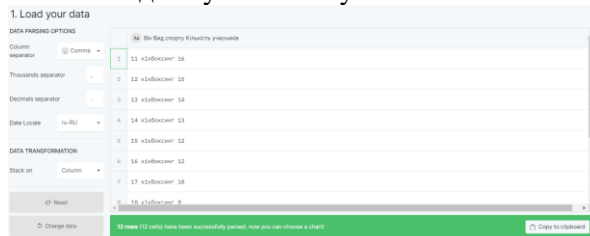
.tsv
.csv
.dsv
.json

Параметри аналізу даних

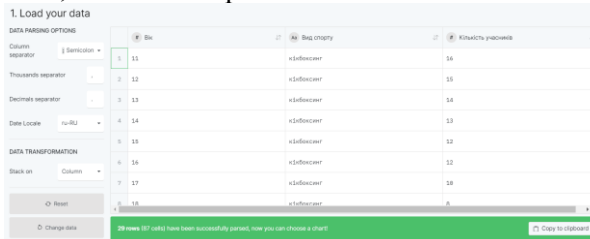
При завантаженні даних до RAWGraphs, є кілька варіантів того, як програмне забезпечення має аналізувати набір даних.

Розділювач стовпців

RAWGraphs може вгадати символ, який використовується для визначення стовпців таблиці. Якщо цього не станеться, всі значення будуть представлені в одному стовпчику.



Якщо це сталося, можна визначити символ і вибрати між табуляцією, крапкою з комою, комою та вертикальною межею.



Розділювачі тисяч та десяткових дробів.

Файли даних, які містять формати, часто неправильно інтерпретуються. Наприклад, файл, що містить числа французькою мовою, використовує коми для визначення десяткових дробів, тоді як англійською мовою десяткові дробі визначаються точками/періодами. У RAWGraphs користувач може самостійно задавати необхідні символи для правильного розпізнавання даних, зокрема дат.

Якщо символ неправильний, може виникнути проблема із встановленням правильного типу даних. Наприклад, визначити «,» для

роздільників тисяч, але таблиця використовує «.», число 1.000.000 (один мільйон) буде проаналізовано як рядок, а не як число.

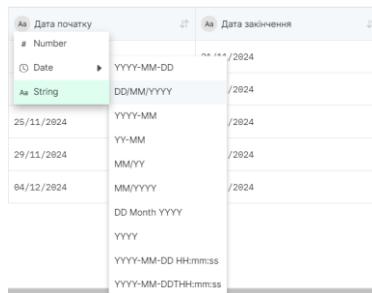
Формат Дата.

Дати в RAWGraphs можуть бути відформатовані так: «ДД Місяць РРРР». За допомогою цього списку, що розкривається, можна вибрати мову для позначення «Місяця». Якщо вибрано неправильну мову, програма інтерпретуватиме значення в стовпці як текстові рядки, а не як дати.

Типи даних.

Після перевірки правильності всіх роздільників можна перевірити, як проаналізовано стовпці, і за необхідності змінити тип даних. Цей крок важливий, оскільки деякі змінні, які можуть бути використані для побудови діаграми, приймають лише певні типи даних. Піктограма поруч з ім'ям стовпця визначає тип даних стовпців.

Для зміни типу даних потрібно натиснути на піктограму поруч з ім'ям стовпця і вибрати потрібний тип. Для встановлення типу даних «Дата» для стовпця потрібно вказати обраний формат, вибравши один з варіантів зі списку.



Якщо стовпець може бути проаналізовано за допомогою нового типу даних, під таблицею з'явиться повідомлення про те, що рядки таблиці були успішно проаналізовані.

Якщо стовпець не може бути проаналізований з новим типом даних, всі клітинки, що містять помилку, будуть виділені, а під таблицею з'явиться повідомлення з попередженням.

	Написання курсиві статті	Дата початку	Дата закінчення	Статус
1	Вибір теми	19/11/2024	21/11/2024	Перший
2	Огляд літератури	22/11/2024	24/11/2024	Другий
3	Збір та аналіз даних	25/11/2024	28/11/2024	Третій
4	Написання статті	29/11/2024	02/12/2024	Четвертий
5	Рефлексія в мистецтві	04/12/2024	06/12/2024	П'ятий

Yes, these fields are 1 of 5 fields. Написання курсиві статті. There are 5 fields in 5 rows total.

Copy to clipboard

Зміна значення клітинку.

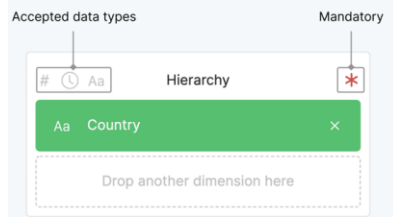
Для зміни значення клітинку необхідно двічі клацнути по клітинці та написати нове значення. Якщо нове значення не відповідає типу даних, клітинка буде виділено жовтим кольором.

Зіставлення розмірів даних зі змінними діаграми.

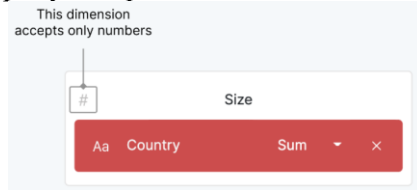
Етап відображення є найважливішим етапом створення візуалізації в RAWGraphs. На цьому кроці необхідно зіставити дані з візуальними змінними вибраної діаграми. Для цього потрібно перетягнути зелені прямокутники (розміри ваших даних) і перемістити їх у вікно під назвою змінних.

Пояснення графічних елементів, які можна побачити у змінних діаграми.

Сірі піктограми ліворуч визначають, які типи даних приймає ця візуальна змінна. У цьому випадку змінна «Ієрархія» приймає всі типи даних і є обов'язковою. Якщо праворуч є червоний *, це означає, що змінна є обов'язковою, і діаграму неможливо створити без цієї змінної.



Якщо змінна не приймає тип даних вибраного розміру, прямокутник стає червоним і діаграму створити неможливо.



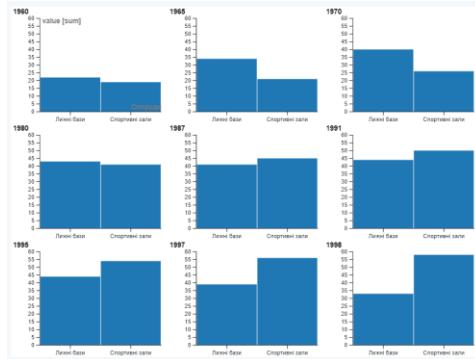
Деякі змінні, наприклад «Ієрархія», приймають більше ніж один вимір, тоді як інші, наприклад «Розмір», приймають лише одну змінну. Важливим є порядок розміщення розмірів у змінній, яка приймає більше одного. Для зміни порядку їх розташування треба перетягнути їх усередину.

Серія.

Невелика кратна діаграма (іноді її називають гратчастою, сітчастою або панельною діаграмою) – це серія подібних графіків або діаграм із однаковим масштабом і осями, що дозволяє легко їх порівнювати. В ній

використовується кілька переглядів для відображення різних розділів набору даних.

У RAWGraphs для деяких видів діаграм (гістограма, багатонабірна гістограма, гістограма з накопиченням, діаграма Veeswarrт, рельєфна діаграма, лінійна діаграма, потоковий графік) можна знайти змінну під назвою «Series». Змінна «Серія» дозволяє створити діаграму для кожного значення виміру.



Експорт візуалізації.

Візуалізацію можна експортувати в такі формати:

- .svg
- .png
- .jpg
- .rawgraphs

Щоб експортувати візуалізацію, просто введіть назву файлу, виберіть формат і натисніть «Завантажити».

5. Export

viz

.svg

Download

.png і .jpg

Це два найпоширеніші формати зображень для оцінювання, і вони добре працюють, якщо необхідно включити візуалізацію до публікації або до слайд-шоу у вихідному вигляді. Оскільки це растрові формати, то не можна редагувати окремі елементи візуалізації, такі як тексти, фігури, осі тощо.

.svg і .svg

Це відкритий формат векторних зображень, сумісний із більшістю програм графічного дизайну, а також є дружнім до Інтернету.

Практичне завдання

Завдання 1. Побудова матричної діаграми.

Розглянемо приклад побудови матричної діаграми (кореляційна матриця) за даними наповненості спортивних секцій спортсменами різного віку.

Для того, щоб отримати потрібний вигляд діаграми в RawGraphs необхідно виконати наступні дії.

1. Перейти за посиланням <https://rawgraphs.io/>
2. Натиснути кнопку **Use it now!**.
3. Завантажити дані (табл. 2).

Таблиця 2

Вхідні дані

Вік	Вид спорту	Кількість учасників
11	кікбоксинг	16
12	кікбоксинг	15
13	кікбоксинг	14
14	кікбоксинг	13
15	кікбоксинг	12
16	кікбоксинг	12
17	кікбоксинг	10
18	кікбоксинг	8
11	футзал	14
12	футзал	14
13	футзал	13
14	футзал	12
15	футзал	12
16	футзал	11
17	футзал	10
18	футзал	10
11	тхеквондо	13
12	тхеквондо	13
13	тхеквондо	15
14	тхеквондо	15
11	скелелазіння	12
12	скелелазіння	13
13	скелелазіння	14
14	скелелазіння	5
15	скелелазіння	4
16	скелелазіння	3
11	худ. гімнастика	6
12	худ. гімнастика	5
13	худ. гімнастика	5

Дані можна скопіювати із MS Excel або GoogleSheets. Коли онлайн сервіс успішно розпізнає отримані дані, вони відобразяться у вигляді таблиці (рис. 4). На рис.9 зображено розподіл даних по стовпцях «Вік», «Вид спорту», «Кількість учасників».

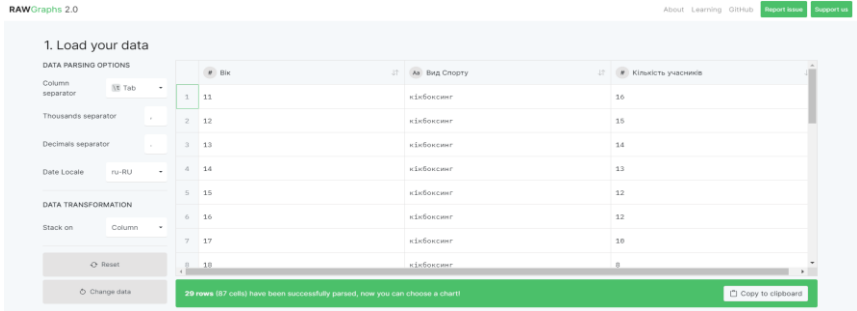


Рис. 9. Вхідні дані

4. Вибрати 1 із 31 можливих видів графіків, який треба побудувати за вхідними даними (рис. 10). Для прикладу обираємо **matrix plot** – **матричний графік**, котрий дозволяє порівнювати два категоріальні вимірювання, розташовуючи їх по горизонтальній та вертикальній осях. Кожен квадрат представляє собою можливу кореляцію між двома змінними. Кожен квадрат в матриці зафарбований кольором, який відповідає силі та напрямку кореляції між двома змінними. Більш темні кольори зазвичай вказують на щільнішу кореляцію, а більш світлі – на слабку.

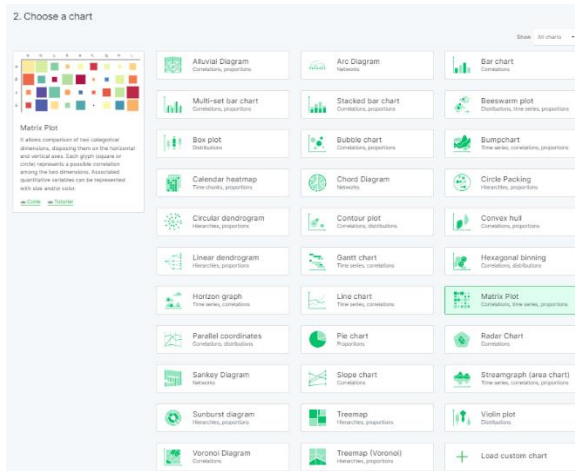


Рис. 10. Види діаграм в RawGraphs.

5. Для зіставлення завантажених даних із вибраним видом діаграми оберіть стовпці для змінних та відображених вимірювань (рис. 11).

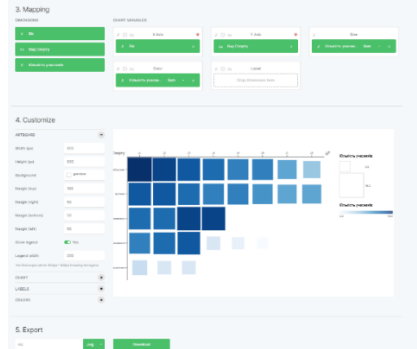


Рис. 11. Налаштування графіку та візуалізація даних у RawGraphs

Отримана кореляційна матриця допомагає провести аналіз зв'язків між видом спорту та віком вихованців спортивної школи. Спостерігається зацікавленість кікбоксингом у дітей 11 років, а також підвищений інтерес до занять тхеквондо у дітей 13-14 років. Таким чином, можна виявити сильні та слабкі кореляції, які можуть бути не очевидні при використанні інших типів візуалізацій.

Сервіс не зберігає завантажені користувачем дані, вони обробляються лише веб-браузером. Інструмент дозволяє експортувати отримані візуалізації в популярних форматах для подальшого редагування.

Завдання 2. Побудова кругової діаграми.

Побудуйте кругову діаграму за показниками загальної фізичної підготовленості борців вільного стилю. Проаналізуйте отримані дані.

1. Завантажити дані (табл. 3).

Таблиця 3

Вхідні дані

Спортсмени	Згинання та розгинання рук в упорі лежачи (кількість разів)	Підтягування на перекладині (кількість разів)	Піднімання ніг у висі на шведській стінці (кількість разів)
спортсмен 1	17	11	20
спортсмен 2	20	16	25
спортсмен 3	22	12	24
спортсмен 4	20	10	26
спортсмен 5	15	17	27
спортсмен 6	18	22	22
спортсмен 7	17	20	27
спортсмен 8	16	13	28
спортсмен 9	18	15	29
спортсмен 10	25	18	31

2. Вибрати **Pie chart** – **кругова діаграма**. Вона дозволяє побачити пропорції між величинами, що складають ціле, за допомогою секторів кола.

3. Виконайте зіставлення завантажених даних із вибраним видом діаграми. Оберіть відповідні стовпці для змінних і параметрів, що будуть відображені (рис. 12).

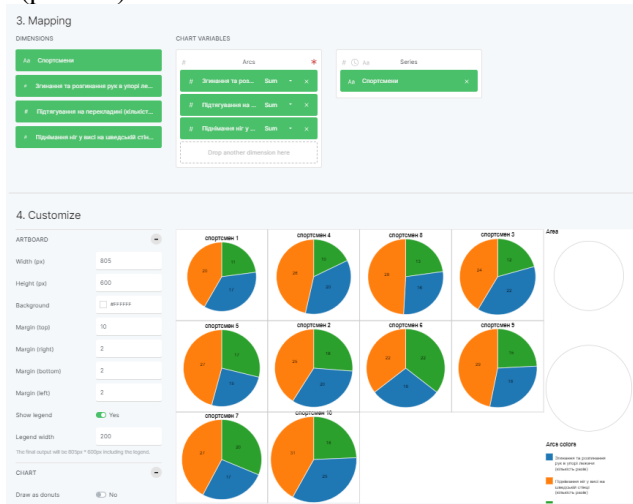


Рис. 12. Налаштування графіка та візуалізація даних у RawGraphs

Завдання 3. Побудова рельєфної діаграми та гістограми.

Побудуйте рельєфну діаграму та гістограму за даними кількості спортивних залів та лижних баз за досліджуваний проміжок часу.

1. Завантажити дані (табл. 4).

Таблиця 4

Вхідні дані

Споруди	1960	1965	1970	1980	1987	1991	1995	1997	1998
Спортивні зали	19	21	26	41	45	50	54	56	58
Лижні бази	22	34	40	43	41	44	44	39	33

2. Вибір **Bumpchart**– **рельєфна діаграма (часові ряди, кореляції, пропорції)** дозволяє порівнювати кілька категорій у безперервному вимірі та вдосконалювати їх сортування. За замовчуванням сортування ґрунтується на розмірі потоку.

3. Ліворуч у розділі **Data transformation / Перетворення даних** знайдіть команду для розміщення набору даних у певному стовпці. Зі списку, що випадає, оберіть «Споруди» (рис. 13).

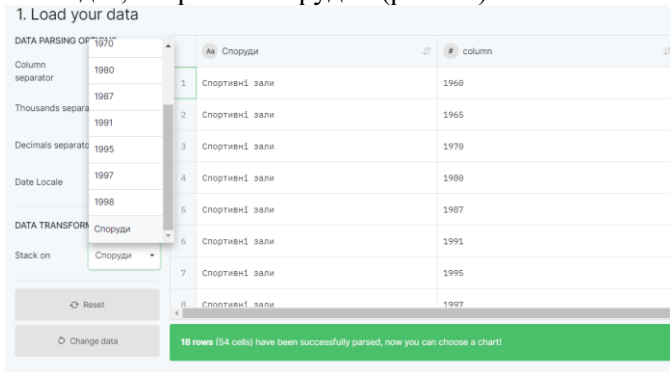


Рис. 13. Розташування даних

У першому стовпці вказані назви спортивних споруд, у другому – роки, а в третьому – значення.

За допомогою цієї операції визначено основний вимір (у даному випадку Споруди), який став першим стовпцем нового набору даних. У другому стовпці з назвою **column / стовбець** показані всі заголовки попередньої таблиці, що залишилися. У третьому стовпці з назвою **value / значення** показані значення попередньої таблиці.

4. Зіставте завантажені дані із обраним видом діаграми. Оираємо стовпці для змінних та відображених вимірювань (рис. 14).

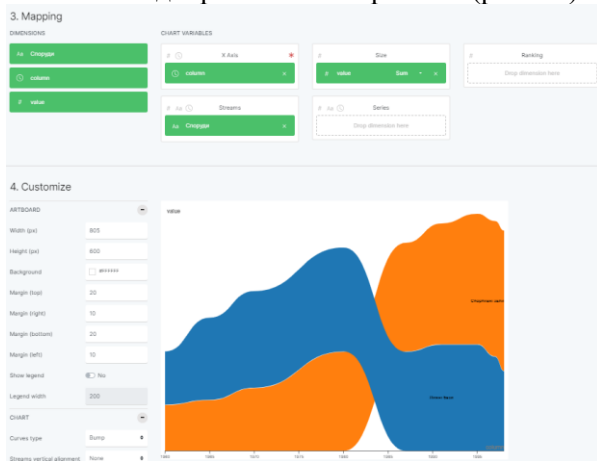


Рис. 14. Налаштування графіка та візуалізація даних у RawGraphs

5. Таким самим способом побудуйте **Bar chart** /гістограму (рис. 15).

Гістограма являє собою набір вертикальних стовпчиків, висота яких визначається значеннями даних. Гістограма є одним з найбільш поширених типів діаграм, цей тип корисний для порівняння значень числових даних в одному або декількох рядах даних, а також для подання зміни даних у часі або за категоріями.

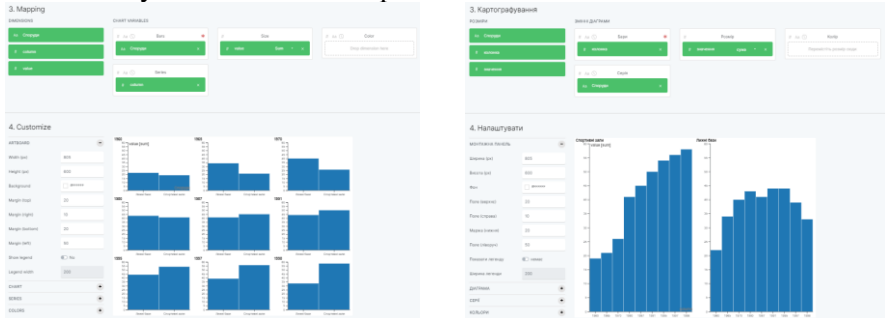


Рис. 15. Налаштування графіка та візуалізація даних у RawGraphs

Завдання 4. Побудова діаграми Ганта.

Діаграма Ганта – це інструмент для візуалізації робочого процесу. Він допомагає планувати проекти, керувати ними та структурувати робочі процеси. Графік названо на честь Генрі Ганта – американського інженера, завдяки якому цей метод планування став всесвітньо відомим. Розглянемо класичний приклад діаграми Ганта, яка представлена у вигляді стовпчастого графіка.

1. Завантажте дані (табл. 5).

Таблиця 5

Вхідні дані

Написання наукової статті	Дата початку	Дата закінчення	Етапи
Вибір теми	19/11/2024	21/11/2024	Перший
Огляд літератури	22/11/2024	24/11/2024	Другий
Збір та аналіз даних	25/11/2024	28/11/2024	Третій
Написання статті	29/11/2024	03/12/2024	Четвертий
Публікація в журналі	04/12/2024	05/12/2024	П'яти

2. Оберіть **Canntt chart** – діаграми Ганта. Діаграми Ганта ілюструють дати початку та закінчення кінцевих елементів та підсумкових елементів проекту.

3. При завантаженні таблиці з вхідними даними в стовбцях «Дата початку» та «Дата закінчення» оберіть формат «Дата» (рис. 16).

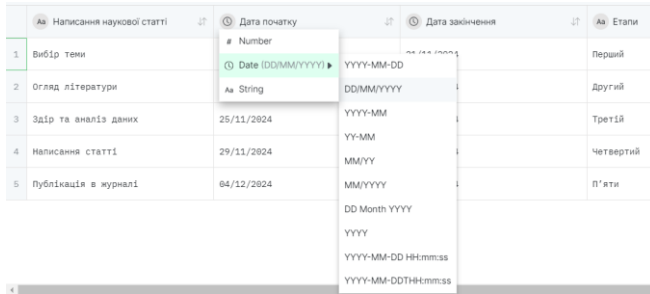


Рис. 16. Налаштування формату «Дата»

4. Зіставте завантажені дані з обраним видом діаграми. Оберіть стовпці для змінних та відображених вимірювань (рис. 17).

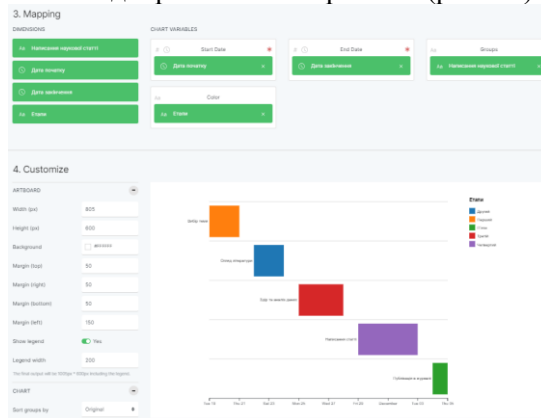


Рис. 17. Налаштування графіка та візуалізація даних у RawGraphs

Самостійна робота

1. Побудуйте точковий графік залежності маси та довжини тіла.

Спортсмен	Маса тіла, кг	Довжина тіла, см
Спортсмен 1	67	178
Спортсмен 2	71	182
Спортсмен 3	65	175
Спортсмен 4	72	185
Спортсмен 5	69	180
Спортсмен 6	66	177
Спортсмен 7	71	183
Спортсмен 8	70	181
Спортсмен 9	66	176
Спортсмен 10	73	187

2. Побудуйте пелюсткову та бульбашкову діаграми за показникам рухової активності футболістів протягом одного матчу. Проаналізуйте отримані дані.

Обсяг рухової активності (ОРА)			
Ігрове амплуа	хвилини	км	ОРА, %
Крайній захисник	70.9	8.7	20
Центральний захисник	69.0	9.0	21
Півзахисник	79.9	9.7	35
Крайній нападуючий	70.0	8.9	5
Центральний нападник	70.6	9.7	15

3. Побудуйте багатонабірну гістограму розподілу за даними групи учнів підліткового віку за видами порушення постави.

Стать	Сутулість	Круглоувігнута спина	Плоска спина	Сколіотична постава
Дівчата	4	2	2	4
Хлопці	5	3	3	4

Контрольні питання

1. Що таке RAWGraphs і які його основні переваги у візуалізації даних?
2. Чому візуалізація емпіричних даних вважається важливою складовою математичної статистики?
3. Які формати файлів підтримує RAWGraphs під час імпорту даних?
4. Як завантажити та формувати дані для RawGraphs?
5. Що відбувається, якщо тип даних у клітинці не відповідає обраному формату?
6. Як зіставити розміри даних із змінними діаграми?
7. У чому особливість змінної «Серія» та для яких типів діаграм вона доступна?
8. Які формати файлів доступні для експорту візуалізацій у RAWGraphs?
9. Як система повідомляє про успішний або неуспішний аналіз даних у RAWGraphs?
10. Чому важливо правильно визначити типи даних перед створенням візуалізації в RAWGraphs?

Практичне заняття № 3

Тема «Сортування та фільтрація статистичних даних»

Мета заняття: навчитися сортувати та форматувати дані в таблиці Microsoft Excel.

Зміст і хід заняття

1. Сортування даних в таблицях.
 2. Фільтрація даних.
 3. Обчислення проміжних підсумків.
- Виконання та захист практичного завдання.

1. Сортування даних в таблицях.

Сортування – це зміна відносного положення даних у списку відповідно до значень або типу даних.

Дані звичайно сортуються за алфавітом, за числовим значенням або за датою. За порядком сортування даних в Excel числа сортуються у порядку зростання: від найменшого від’ємного до найбільшого додатного числа. При сортуванні тексту порівнюється зміст клітинок за символами зліва направо. Регістр символів при сортуванні не враховується.

У тексті, окрім літер, можуть використовуватися цифри та інші символи. Упорядкована послідовність усіх можливих символів в Excel при сортуванні така:

(пробіл)!«#\$\$%&()*+,./:;? @ [\] л _ " { |) ~ + < = > 0123456789ABCDEFGHIJKLMN0PQRSTUVWXYZABBrflEE?;K3MI (укр.) ї Й КЛМНОПРСТУФХІЧШЦ'ЬНЬЗЮЯ

Сортування за одним параметром. Для сортування таблиці клацніть по будь-якій заповненій клітинці та натисніть одну з кнопок на панелі інструментів:  **Сортировка от А до Я** ,  **Сортировка от Я до А**.

При цьому буде проведено сортування суцільної частини таблиці, яка не має порожніх рядків або стовпців. Рядки переставлятимуться повністю. Заголовки (підписи) стовпців до процесу сортування за замовченням не потрапляють.

Можливе сортування не всього списку, а лише його окремої частини. Для цього потрібно виділити діапазон даних і застосувати команду **Дані** → **Сортування**. Якщо сортування не задовольняє запиту, його можна скасувати. Для цього треба клацнути по кнопці Відмінити на панелі інструментів або натиснути комбінацію клавіш Ctrl+Z.

2. Фільтрація даних.

Однією з найпоширеніших операцій зі списками є вибір записів, або фільтрація.

Фільтрація – це процес пошуку і вибору записів відповідно до встановлених критеріїв.

Фільтри також спрощують процес введення та видалення записів зі списку. При фільтрації записи, які не відповідають зазначеним критеріям, приховуються, але їх порядок розміщення в таблиці залишається незмінним і вони не вилучаються з таблиці.

В Excel є такі типи фільтрів: автофільтр та розширений фільтр.

1. **Автофільтр** – це фільтр, який дозволяє задати прості критерії пошуку записів. У результаті відображаються ті записи, що задовольняють умові пошуку, і приховуються всі інші.

2. **Розширений фільтр** – це фільтр для задання складних критеріїв пошуку, а також відображення результатів фільтрації в окремій області таблиці.

3. Обчислення проміжних підсумків.

Проміжні підсумки – це узагальнюючі значення (суми, середнього, кількості тощо). Вони обчислюються для груп об'єктів, представлених у таблиці, а також є засобом обчислення цих значень. Проміжні підсумки доцільно обчислювати лише для таблиць, що впорядковані за значеннями певним параметром.

Перш ніж встановлювати проміжні підсумки, слід відсортувати список. Це дозволить згрупувати рядки, за якими потрібно обчислити підсумки. Після цього можна обчислити проміжні підсумки для кожного стовпця із числовими даними.

Якщо дані не мають формату списку або потрібно вивести лише один підсумок, можна скористатись функцією Автосума замість автоматичних обчислень.

Для обчислення значень проміжних підсумків використовують підсумкову функцію, наприклад *Сума (SUM)* або *Середнє арифметичне (AVERAGE)*.

Проміжні підсумки можна вивести у списку з використанням декількох типів обчислення.

Загальні підсумки обчислюють за докладними відомостями, а не за значеннями в рядках проміжних підсумків. Наприклад, у разі використання підсумкової функції **AVERAGE** загальний підсумок повертає середнє значення для всіх рядків списку, а не для проміжних підсумків.

Значення проміжних і загальних підсумків автоматично повторно обчислюються після кожної зміни вхідних даних.

Не всі таблиці та набори даних підходять для того, щоб застосувати до них функцію проміжних підсумків. До головних умов відносяться такі:

- таблиця повинна мати формат звичайної області клітинок;
- шапка таблиці повинна складатися з одного рядка, і розміщуватися на першому рядку аркуша;
- в таблиці не повинно бути порожніх рядків.

Практичне завдання

Завдання 1. Сорткування даних в таблиці.

1. Створіть таблицю за прикладом (табл. 6).

Таблиця 6

Вхідні дані

№	ППП	Вид спорту	Дата народження	Довжина тіла (см)	Маса тіла (кг)	ОГК в спокої (см)	ОГК видих (см)	Показник міцності статури
1	Черкашин А.В.	Плавання	12.12.2010	146	37,6	69	67	41,4
2	Зайко О.С.	Футбол	05.08.2010	140	50,1	84	82	7,9
3	Кіряшко В.П.	Баскетбол	03.04.2011	144	30,0	70	69	45,0
4	Мазур М.Г.	Баскетбол	23.04.2009	140	32,9	62	61	46,1
5	Сухно І.В.	Бокс	03.07.2012	147	33,3	66	65	48,7
6	Самойлов М.Ю.	Футбол	01.03.2012	142	43,7	73	71	27,3
7	Рогов С.О.	Гімнастика	03.06.2011	140	28,6	59	58	53,4
8	Бойко А.Ю.	Плавання	18.12.2010	142	34,0	66	64	44,0
9	Шарапов І.В.	Бокс	30.07.2010	146	29,8	61	60	56,2
10	Арестов В.С.	Бокс	02.09.2011	138	39,5	69	67	31,5

2. Відсортуйте дані в таблиці за стовпцем «ППП» за зростанням.

Для цього:

- виділіть діапазон клітинок A1:I11;
- оберіть команду **Данні** → панель інструментів **Сорткування** і **фільтр** → кнопка **Сорткування**;
- оберіть сортувати по стовпцю **ППП** → **Порядок сортування Від А до Я**.

3. Клацніть правою кнопкою миші на вкладці **Аркуш 1**. Виберіть в контекстному меню команду **Перейменувати**. Введіть нове ім'я «Сорткування».

4. Клацніть правою кнопкою миші на вкладці **Сорткування** і оберіть в контекстному меню команду **Перемістити або скопіювати**.

5. У вікні, що відкрилося, встановіть прапорець **Створити копію**, щоб вибраний лист копіювався, а не переміщувався.

6. У списку перед аркушем клацніть на пункті **Перемістити в кінець**, щоб скопіювати перший лист в кінець книги та клацніть на кнопці **ОК**.

7. Створіть три копії цього аркуша.

8. На другому аркуші виконайте сортування даних по третьому стовпцю у такій послідовності: Плавання, Бокс, Футбол, Гімнастика, Баскетбол.

Для цього зробіть такі кроки:

- виділіть діапазон клітинок A1:P11;
- оберіть команду **Данні** → панель інструментів **Сортування і фільтр** → кнопка **Сортування**;
- оберіть стовбець → Вид спорту, **Сортування** → **Значення, порядок** → **Настроюваний список**;
- у наступному вікні запишіть необхідну послідовність видів спорту;
- натисніть **ОК**.

9. На третьому листі підведіть проміжні підсумки:

- виконайте сортування за зростанням по стовпцю Вид спорту;
- виділіть діапазон A1: P11;
- оберіть команду **Данні** → панель інструментів **Структура** → кнопка **Проміжний підсумок**;
- в діалоговому вікні в рядку «При кожній зміні в» виберіть стовпець «Вид спорту», в рядку «Операція» виберіть «Середнє», у вікні «Додати підсумки по» виберіть «Маса тіла (кг)». Натисніть кнопку **ОК**.

10. Перейменуйте Аркуш 3 в «Підсумки».

11. Аркуш 4 перейменувати в «Фільтр».

12. Для включення фільтра виділіть діапазон даних і оберіть команду **Данні** → панель інструментів **Сортування і Фільтр** → кнопка **Фільтр**



Фільтр

. Після виконання команди біля заголовків стовпців з'являться кнопки **Фільтра** .

12. Відфільтруйте дані спортсменів, які народились в 2010 році з Показником міцності статури від 40 до 45:

- для відбору спортсменів, які народились в 2010 році натисніть на кнопку  і виберіть відповідний рік;
- для відбору «Показника міцності статури» від 40 до 45 натисніть на кнопку  і виконайте команду: **Числові фільтри** → **Між**.

13. Перейменуйте Аркуш 5 в «Розширений фільтр».

14. На п'ятому аркуші поруч з вихідною таблицею створіть таблицю умов. Особливості таблиці такі: рядок заголовків повинен повністю збігатися з шапкою фільтрованої таблиці. Щоб уникнути помилок, скопіюйте рядок заголовків у вихідній таблиці та вставте до таблиці умов критерії відбору Прізвище спортсмена за літерою «С».

Діапазон умов

№	ІПШ	Вид спорту	Дата народження	Довжина тіла (см)	Маса тіла (кг)	ОГК в спокої (см)	ОГК видих (см)	Показник міцності статури
	С							

15. Оберіть команду **Данні** → панель інструментів **Сортування і фільтр** → кнопка **Додатково**.

16. У вікні **Розширеного фільтру** оберіть спосіб обробки інформації (на цьому ж аркуші або на іншому), задайте вихідний діапазон (табл. 6) і діапазон умов (табл. 7). Рядки заголовків повинні бути включені до діапазонів. Натисніть **ОК**.

Контрольні питання

1. Що таке сортування даних у контексті роботи з таблицями Microsoft Excel?

2. Якими способами можна сортувати дані в Microsoft Excel?

3. У якому порядку Microsoft Excel сортує числові значення?

4. Яким чином можна відсортувати лише частину списку, а не всю таблицю?

5. Що таке фільтрація в Microsoft Excel?

6. Яка головна відмінність між сортуванням і фільтрацією даних?

7. Чи змінюється порядок записів у таблиці після фільтрації?

8. Які типи фільтрів підтримує Microsoft Excel?

9. Для яких завдань доцільно використовувати автофільтр?

10. Коли варто використовувати розширений фільтр замість автофільтра?

16. Що таке проміжні підсумки в Microsoft Excel?

17. Для чого потрібно попередньо відсортувати таблицю перед обчисленням проміжних підсумків?

18. Які типи обчислень можуть використовуватись для підсумків? Назвіть приклади.

19. Яка різниця між загальними та проміжними підсумками?

20. Які умови необхідні для коректної роботи функції проміжних підсумків у Excel?

Практична робота № 4

Тема «Робота зі зведеними таблицями»

Мета заняття: поглибити знання по темі «Зведені таблиці»; навчити обчисленню в зведеній таблиці, створювати зведені діаграми, проводити фільтрацію зведених таблиць за допомогою зрізів та часової шкали.

Зміст і хід заняття

1. Поняття зведеної таблиці.
2. Обчислення в зведеній таблиці.
3. Створення обчислюваного поля та елемента, що обчислює, в зведеній таблиці.
4. Створення зведених діаграм.
5. Фільтрація зведених таблиць за допомогою роздільників та часової шкали.

Виконання та захист практичного завдання.

1. Поняття зведеної таблиці.

Зведені таблиці в Excel застосовують для узагальнення інформації з баз даних, які можуть зберігатися як в робочих книгах Excel, так і в зовнішніх файлах.

Зведена таблиця – це динамічна таблиця підсумкових даних, вилучених або розрахованих на основі інформації, що міститься в базі даних. Базами даних для створення зведених таблиць можуть бути дані на робочому аркуші, організовані у вигляді таблиці, або зовнішні бази даних. За допомогою засобів зведених таблиць можна перетворити велику кількість малозрозумілих чисел в одну невелику і зручну для аналізу таблицю.

Зведені таблиці дозволяють створювати динамічні перехресні таблиці з структурою, що легко змінювати. Дані в таких таблицях узагальнюються за кількома параметрами. Крім того, за допомогою зведеної таблиці можна обчислити проміжні підсумки з будь-яким рівнем деталізації. Однією з найбільш вражаючих властивостей зведеної таблиці є її інтерактивність. Після створення зведеної таблиці можна змінювати наявні дані та додавати спеціальні формули для виконання додаткових розрахунків. Більш того, після створення зведеної таблиці можна групувати її елементи. Зі зведеної таблиці можна зробити в необхідний звіт, застосувавши функцію форматування.

Єдиним недоліком зведених таблиць є те, що на відміну від підсумкових звітів, створених з використанням формул, вони не оновлюються автоматично при зміні вхідних даних. Але з оновленням

зведеної таблиці не виникає особливих проблем, досить виконати команду **Оновити**.

Зведені таблиці вперше з'явилися в Excel 97. Засіб створення зведених таблиць в Excel 2007 значно вдосконалено, а Excel 2010-2019 додано нові можливості, і тепер створення і застосування зведених таблиць стало набагато простіше.

2. Обчислення в зведеній таблиці.

В зведених таблицях для обчислення підсумкових значень найчастіше використовується функція підсумовування. Але при необхідності завжди можна змінити функцію, яка обчислює, задану в діалоговому вікні **Параметри значення поля** (рис. 18).

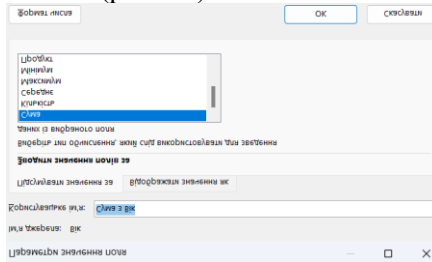


Рис. 18. Діалогове вікно **Параметри значення поля**

Вкладка *Підсумувати значення за* використовується для вибору функції, що обчислює. Можна обрати функції **Сума**, **Кількість**, **Середнє**, **Максимум**, **Мінімум**, **Добуток**, **Кількість чисел**, **Незсунуте відхилення**, **Зміщена дисперсія**, **Незсунута дисперсія**.

Попередньо обчислені значення у зведеній таблиці можна відображати у різній формі з застосуванням додаткових обчислень. Форма представлення значень обирається в списку **Відображати значення як** однойменної вкладки діалогового вікна **Параметри значення поля** (табл. 8).

Таблиця 8

Функції списку **Відображати значення як**

Функція	Опис
Без обчислення	Відключити додаткові обчислення
Відмінність	Показує дані як різниці між значеннями і значенням <i>елемент</i> поля, обраного в списку <i>Поле</i>
Частка	Показує дані як процентні відносини значень до значення <i>елемент</i> поля, обраного в списку <i>Поле</i>
Зведена відмінність	Відображає дані як процентні відношення значень до різниці значень і значення <i>елемент</i> поля, обраного в списку <i>Поле</i>

Зі зростаючим підсумком у	Відображає дані у вигляді суми, що накопичена, в поле, обраному в списку <i>Поле</i>
% від зростаючого підсумка у	Відображає дані у вигляді % від суми, що накопичена, в поле, обраному в списку <i>Поле</i>
Частка суми рядків	Відображає дані в кожному рядку (категорії) в вигляді відсотку від підсумкового значення рядка
Частка суми стовпців	Відображає дані в кожному стовпці у вигляді відсотку від підсумкового значення стовпця
Частка загальної суми	Показує всі дані у вигляді відсотку від загального підсумкового значення
Сортування від найменшого до найбільшого	Ранг вибраних значень в певному полі з урахуванням того, що найменшому з них присвоюється значення 1, а іншим – значення більш високого рангу відповідно.
Сортування від найбільшого до найменшого	Ранг вибраних значень в певному полі з урахуванням того, що найбільшим значенням в полі присвоюється значення 1, а кожному меншому значенню – більш високий ранг.
Індекс	Дані обчислюються за формулою $((\text{значення в клітинці}) \times (\text{головний загальний підсумок})) / ((\text{загальний підсумок по строкам}) \times (\text{загальний підсумок по стовпцям}))$

3. Створення обчислюваного поля і елемента, що обчислюється, в зведених таблиці.

Оскільки зведена таблиця є особливим типом діапазону даних, в неї не можна вставляти нові рядки або стовпці. Це означає, що не можна вставити до зведеної таблиці формули, які будуть виконувати операції над даними. Але можна створювати нові обчислювані поля для зведеної таблиці. В полі, що обчислює, виконуються розрахунки з використанням значення з інших полів.

Як правило, для відображення нової інформації в зведеній таблиці використовуються поля, що обчислюються. Цей спосіб дозволяє уникнути створення нового поля даних у вихідній базі даних. У більшості випадків значно простіше вставити новий стовпець до вихідного діапазону з формулою, яка виконує необхідне обчислення. Застосування полів, що обчислюються, зручніше в тому випадку, коли дані надходять з джерела, яким важко маніпулювати, наприклад, з зовнішніх баз даних.

Поле, що обчислюється. Нове поле створюється на основі існуючих полів зведеної таблиці. Якщо зведена таблиця створена на основі таблиці робочого аркуша, то поле, що обчислюється, еквівалентно новому стовпцю в цій таблиці. При цьому в новому стовпці виконуються певні обчислення. Поле, що обчислюється, повинно знаходитися в області

«Значення» зведеної таблиці. Не можна створювати поля, що обчислюються, на основі полів з області «Рядки», «Стовпці» або «Фільтри».

Крім полів, що обчислюються, Excel дозволяє створити елементи, що обчислюються, для будь-якого поля зведеної таблиці. Значимо, що створення полів, що обчислюються, дозволяє уникнути додавання нових полів в вихідну таблицю даних, а створення елементів, що обчислюються, - нових рядків.

Елемент, що обчислюються. Елемент, що обчислюються, розраховується з використанням вмісту інших елементів, які знаходяться в тому ж полі зведеної таблиці. Якщо зведена таблиця створена на основі таблиці робочого аркуша, то створення елемента, що обчислюються, еквівалентно вставці однієї чи декількох рядків в цю таблицю. При цьому в нові рядки записуються формули, що використовують значення з інших рядків. Елемент, що обчислюються, повинен знаходитися в області «Рядки», «Стовпці» або «Фільтри», але не може перебувати в області «Значення».

Для створення полів, що обчислюються, або елементів не можна використовувати стандартні функції Excel, тобто не можна вводити формули безпосередньо в клітинці. Замість цього формули вводять в спеціальному діалоговому вікні і зберігають разом з даними зведеної таблиці.

4. Створення зведених діаграм.

Зведені діаграми – це графічне представлення даних, узагальнених в зведених таблицях.

Excel пропонує кілька способів побудови зведених діаграм:

- Помістити табличний курсор в будь-яку клітинку зведеної таблиці і виконати команду **Аналіз зведених таблиць** → **Знаряддя** → **Зведена діаграма**.
- Вибрати будь-яку клітинку в існуючій зведеній таблиці та виконати команду **Вставлення** → **Діаграми** → **Зведена діаграма**.
- Виконати команду **Вставлення** → **Діаграми** → **Зведена діаграма**. Якщо курсор знаходиться поза межами зведеної таблиці, програма запросить джерело даних і створить зведену діаграму.
- Виконати команду **Вставлення** → **Діаграми** → **Зведена діаграма** → **Зведена діаграма і зведена таблиця**. Програма запросить джерело даних, створить зведену діаграму і зведену таблицю. Ця команда доступна тільки в тому випадку, коли покажчик знаходиться поза межами зведеної таблиці.

Зауваження щодо зведених діаграм.

Між зведеною таблицею і зведеною діаграмою програма Excel автоматично встановлює двосторонній зв'язок. Тому, якщо внесені будь-які структурні зміни в одному з цих об'єктів, ці зміни відразу відображаються у іншому об'єкті.

При активізації зведеної діаграми панель **Аналіз зведених таблиць** замінюється на панель **Аналіз зведених діаграм**. На цій панелі область **Легенда (Ряди)** замінює область **Колони**, а область **Вісь (Категорії)** – область **Рядки**.

Поля-кнопки зведеної діаграми містять ті ж самі елементи управління, що і заголовки полів зведеної таблиці. Ці елементи управління дозволяють фільтрувати дані, відображені в зведеній таблиці та зведеній діаграмі. Тому, якщо були внесені будь-які зміни у зведеній діаграмі, то ці зміни будуть відображені і в зведеній таблиці.

Якщо зведена діаграма пов'язана зі зведеною таблицею, і видаляється вихідна зведена таблиця, то зведена діаграма зберігається. При цьому до формули «Ряд діаграми» будуть підставлені вихідні дані зі зведеної таблиці у вигляді масиву.

За замовчуванням зведена діаграма розташовується на тому самому робочому аркуші, на якому знаходиться вихідна зведена таблиця. Для переміщення зведеної діаграми на інший робочий аркуш або на аркуш діаграм, треба виконати команду **Аналіз зведених діаграм** → **Дії** → **Перемістити діаграму**.

На основі однієї зведеної таблиці можна створити будь-яку кількість зведених діаграм. При цьому кожна з цих діаграм можна налаштовувати і форматовувати незалежно від іншої. Але на всіх цих діаграмах будуть відображені одні й ті ж самі дані.

При виборі звичайної діаграми піктограми відображаються зліва направо: **Елементи діаграми**, **Стилі діаграми** і **Фільтри діаграми**. У зведених діаграмах піктограма **Фільтри діаграми** не відображається.

5. Фільтрація зведених таблиць за допомогою роздільників та часової шкали

До зведених таблиць можна також застосовувати зрізи і часову шкалу.

Роздільник являє собою інтерактивний елемент керування, що дозволяє здійснювати фільтрацію даних у зведеній таблиці. Використання зрізів можливо тільки в Excel 2010 і пізніших версіях. Нижче наводиться зведена таблиця з двома зрізами. Кожен зріз являє собою певне поле. В цьому випадку зведена таблиця відображає дані для досліджуваних, ці показники наведені за два місяці (рис. 19).

Сумма по полю	Показники	Названия столбцов	
Названия строк	Иванов	Фомин	Общий итог
Березень	1574	1252	2826
Січень	1345	1118	2463
Общий итог	2919	2370	5289

Прізвище
 Иванов
 Попов
 Фомін
 Шарапов

Місяць
 Березень
 Лютий
 Січень

Рис. 19. Приклад зведеної таблиці з двома роздільниками

Фільтрацію аналогічного типу можна виконати, скориставшись для цього мітками зведеної таблиці. Роздільник можна використовувати для створення зручної у використанні інтерактивної «приладової панелі».

Для того, щоб додати один або кілька роздільників на робочий аркуш, необхідно вибрати будь-яку клітинку у зведеній таблиці і виконати команду **Вставлення** → **Фільтри** → **Роздільник**. У діалоговому вікні **Вставлення роздільників** встановити прапорці для потрібних полів з повного списку всіх полів зведеної таблиці і натиснути кнопку **ОК**.

Роздільники можна переміщати по робочому аркушу, змінювати їх розміри і зовнішній вигляд. Для видалення результатів фільтрації треба клацнути на піктограмі, розташованій у верхньому правому куті роздільника.

Для того, щоб за допомогою роздільника відфільтрувати дані зі зведеної таблиці, досить клацнути на одній кнопці. Для відображення набору даних треба натиснути на кнопках відповідних роздільників, утримуючи клавішу «Ctrl». Для вибору ряду послідовно розташованих кнопок треба клацнути на першій кнопці і, утримуючи клавішу «Shift», клацнути на останній кнопці набору.

Часові шкали (timeline) вперше з'явилися в Excel 2013. Концептуально вони подібні до роздільників, але цей елемент управління призначений для спрощення фільтрації даних у зведених таблицях за часовими показниками.

Часові шкали застосовуються тільки в тому випадку – коли в зведеній таблиці є поле у форматі *Дата*. Цей засіб не працює з даними у форматі *Час*. Для додавання часової шкали треба вибрати клітинку в зведеній таблиці і виконати команду **Вставлення** → **Фільтри** → **Часова шкала**. У діалоговому вікні з'явиться перелік всіх полів у форматі *Дата*. Якщо зведена таблиця не містить поле, відформатоване як дата, програма виведе повідомлення про помилку. Часова шкала має ті ж можливостями для форматування, що і роздільник. Вона дозволяє створити зручну інтерактивну панель, яка істотно спрощує фільтрацію зведених таблиць.

Практичне завдання

Завдання 1. Створення зведеної таблиці.

1. Створіть таблицю за прикладом (табл. 9).
2. Помістити табличний курсор в будь яку клітинку діапазону даних.
3. Виберіть вкладку **Вставлення** → **блок Таблиці** → **Зведена**



таблиця .

4. У діалоговому вікні, що відкрилося, виберіть розташування звіту зведеної таблиці «Новий аркуш».

5. Натисніть ОК.

6. Клацніть правою кнопкою миші на імені поля «Стать» в панелі **Поля зведеної таблиці** і виберіть в меню область його місця розташування **Додати в назви рядків**.

Другий спосіб заповнення макета зведеної таблиці.

Перетягніть ім'я поля з верхнього списку в одне з чотирьох полів панелі **Поля зведеної таблиці**.

7. Поле «Рівень біологічного розвитку» перемістіть в область «Значення». Зведена таблиця буде обчислювати суму всіх значень в полі «Рівень біологічного розвитку».

8. Поле «Рівень біологічного розвитку» перемістіть в область «Стовпці». Значення поля «Рівень біологічного розвитку» будуть заголовками стовпців зведеної таблиці.

Зведена таблиця відображає кількість респондентів, розподілених за рівнями біологічного розвитку та статтю (рис. 20). Зведена таблиця оновлюється автоматично при кожній зміні на панелі *Поля зведеної таблиці*.

Кількість з Рівень біологічного розвитку	Позначки стов	успівшений	принищений	Загальний підсумок
Позначка рядка	середній			
жін		8	5	13
чол		2	3	5
Загальний підсумок		10	8	18

Рис. 20. Приклад зведеної таблиці «Завдання 1»

Таблиця 9

Результати опитування респондентів на питання анкети «Здоровий спосіб життя»

№	Стать	Дата заповнення анкети	Вік	Рівень біологічного розвитку	Маса тіла	Довжина тіла	Як часто Ви приймаєте лікарські препарати?	Чи дотримуєтеся Ви здорового раціону і режиму харчування?	Чи займаєтесь ви спортом?	Відзначте причини, які викликають у Вас стрес
1	чол	01.04.2025	18	пришвидшений	60	183	Рідко	Так	Займаюся регулярно	Проблеми в сім'ї
2	жін	01.04.2025	19	пришвидшений	57	175	Рідко	Так	Займаюся регулярно	Інше
3	жін	01.04.2025	19	уповільнений	51	161	Рідко	Інколи	Займаюся регулярно	Фінансові труднощі
4	жін	02.04.2025	20	уповільнений	49	156	Рідко	Інколи	Інколи	Проблеми в колективі
5	чол	03.04.2025	18	пришвидшений	82	176	Рідко	Інколи	Інколи	Самотність
6	чол	03.04.2025	19	пришвидшений	100	180	Рідко	Ні	Не займаюся	Інше
7	чол	04.04.2025	19	середній	60	185	Постійно	Інколи	Займаюся регулярно	Самотність
8	жін	04.04.2025	18	середній	59	169	Доволі часто	Інколи	Не займаюся	Інше
9	чол	04.04.2025	19	уповільнений	73	175	Рідко	Так	Займаюся регулярно	Проблеми в навчанні
10	жін	05.04.2025	18	пришвидшений	80	170	Постійно	Інколи	Інколи	Самотність
11	жін	05.04.2025	21	середній	65	168	Рідко	Інколи	Інколи	Фінансові труднощі
12	жін	07.04.2025	18	уповільнений	48	161	Рідко	Так	Займаюся регулярно	Фінансові труднощі
13	чол	07.04.2025	18	пришвидшений	86	186	Рідко	Так	Займаюся регулярно	Фінансові труднощі
14	жін	07.04.2025	19	пришвидшений	59	172	Доволі часто	Так	Займаюся регулярно	Фінансові труднощі
15	чол	07.04.2025	21	уповільнений	60	175	Рідко	Так	Інколи	Інше
16	чол	07.04.2025	20	уповільнений	67	178	Рідко	Так	Інколи	Проблеми в сім'ї
17	жін	07.04.2025	21	середній	65	165	Рідко	Так	Займаюся регулярно	Інше
18	жін	09.04.2025	18	середній	54	170	Доволі часто	Інколи	Займаюся регулярно	Самотність
19	жін	10.04.2025	19	уповільнений	56	160	Рідко	Так	Займаюся регулярно	Проблеми в сім'ї
20	жін	12.04.2025	19	середній	53	167	Рідко	Так	Займаюся регулярно	Інше
21	чол	12.04.2025	21	пришвидшений	90	183	Рідко	Так	Займаюся регулярно	Інше
22	жін	12.04.2025	18	середній	59	173	Рідко	Інколи	Займаюся регулярно	Проблеми в навчанні
23	жін	12.04.2025	20	пришвидшений	60	180	Рідко	Ні	Не займаюся	Фінансові труднощі
24	жін	14.04.2025	18	уповільнений	53	163	Рідко	Ні	Не займаюся	Фінансові труднощі
25	жін	14.04.2025	20	середній	60	172	Доволі часто	Інколи	Займаюся регулярно	Проблеми в навчанні
26	чол	14.04.2025	19	середній	67	179	Рідко	Ні	Займаюся регулярно	Фінансові труднощі
27	жін	14.04.2025	18	середній	58	167	Рідко	Інколи	Займаюся регулярно	Проблеми в колективі
28	чол	14.04.2025	18	пришвидшений	90	178	Рідко	Інколи	Займаюся регулярно	Фінансові труднощі
29	жін	14.04.2025	20	пришвидшений	50	175	Постійно	Інколи	Інколи	Фінансові труднощі

Завдання 2. Обчислення в зведеній таблиці.

1. Створіть дві копії зведеної таблиці окремо на кожному аркуші.
2. На другому аркуші клацніть правою кнопкою миші на підсумковому значенні зведеної таблиці і виберіть з контекстного меню **Параметри значення поля**.
3. У діалоговому вікні **Параметри значення поля** на вкладці **Відображати значення як** виберіть функцію **Частка загальної суми** (рис. 21).

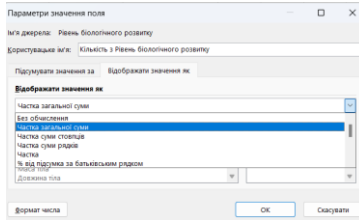


Рис. 21. Діалогове вікно **Параметри значення поля**

4. На четвертому аркуші розрахуйте функцію **Частка суми стовбців**.
 5. Побудуйте зведену таблицю на окремому аркуші.
У цій зведеній таблиці:
 - поле «Маса тіла» розміщується в області «Значення»;
 - поле «Рівень біологічного розвитку» розміщується в області «Стовпці»;
 - поле «Стать» розміщується в області «Рядки».
 6. На четвертому аркуші клацніть правою кнопкою миші на підсумковому значенні зведеної таблиці і виберіть з контекстного меню **Параметри значення поля**.
 7. У діалоговому вікні **Параметри значення поля** на вкладці **Підсумувати значення за** виберіть функцію **Середнє**.
- Зведена таблиця відображає середні показники маси тіла для чоловіків і жінок за кожною групою (рис. 22).

Середнє з Маса тіла	Позначки стовпців				
Позначки рядків	середній	уповільнений	пришвидшений	Загальний підсумок	
жін	59	51	61	58	
чол	63	67	85	76	
Загальний підсумок	60	57	74	64	

Рис. 22. Приклад зведеної таблиці «Завдання 2»

Завдання 3. Фільтрація зведеної таблиці.

У цій зведеній таблиці:

- поле «Відзначте причини, які викликають у Вас стрес» розміщується в області «Фільтри»;

- поле «Вік» розміщується в області «Фільтри»;
- поле «Стать» розміщується в області «Рядки»;
- поле «Відзначте причини, які викликають у Вас стрес» розміщується в області «Значення», підраховуються кількість елементів цього поля.

У цій зведеної таблиці в область «Фільтри» розміщено два поля. Наприклад, це дозволяє відобразити тільки ті дані, для яких значення в поле «Вік» дорівнює «18», значення в поле «Відзначте причини, які викликають у Вас стрес» дорівнює «Фінансові труднощі» (рис. 23).

Відзначте причини, які викликають у Вас стрес: Фінансові труднощі		↑
Вік 18		↑
Позначки рядків	Кількість з Відзначте причини, які викликають у Вас стрес:	
жін		2
чол		3
Загальний підсумок		5

Рис. 23. Приклад зведеної таблиці «Завдання 3»

Завдання 4. Фільтрація зведених таблиць за допомогою часової шкали.

У цій зведеної таблиці:

- поле «Вік» розміщується в області «Значення»;
- поле «Вік» розміщується в області «Стовпці»;
- поле «Стать» розміщується в області «Рядки».

1. Встановіть табличний курсор на одному з елементів зведеної таблиці.
2. Виберіть команду **Вставлення** → **Фільтри** → **Часова шкала**.
3. У діалоговому вікні **Вставка часових шкал** встановіть прапорець поруч із полем «Дата заповнення анкети».
4. Натисніть **ОК**.
5. Натисніть на стрілку поруч з відображеним часовим рівнем і виберіть варіант «Дні».
6. Перетягніть смугу прокрутки часової шкали до періоду часу 01–04 квітня 2025 р. (рис. 24).

Кількість з Вік	Позначки стовпців								
Позначки рядків	18	19	20	Загальний підсумок					
жін	1	2	1	4					
чол	2	3		5					
Загальний підсумок	3	5	1	9					

Дата заповнення

1–4 Кві 2025 р.

БЕР 2025 Р. КВІ 2025 Р.

29 30 31 1 2 3 4 5 6 7 8 9

Дні

Рис. 24. Приклад зведеної таблиці «Завдання 4»

Завдання 5. Фільтрація зведених таблиць за допомогою роздільників.

1. Створіть таблицю за прикладом (табл. 10).

Таблиця 10

Вхідні дані

Прізвище	Місяць	Показники
Фомін	Січень	1118
Фомін	Лютий	1960
Фомін	Березень	1252
Попов	Січень	1271
Попов	Лютий	1557
Попов	Березень	1679
Шарапов	Січень	1247
Шарапов	Лютий	1238
Шарапов	Березень	1028
Іванов	Січень	1345
Іванов	Лютий	1784
Іванов	Березень	1574

2. За цими даними побудуйте зведену таблицю.
У цій зведеній таблиці:
 - поле «Показники» розміщується в області «Значення»;
 - поле «Прізвище» розміщується в області «Стовпці»;
 - поле «Місяць» розміщується в області «Рядки».
 3. Встановіть табличний курсор на одному з елементів зведеної таблиці.
 4. Виберіть команду **Вставлення** → **Фільтри** → **Роздільник**.
 5. У діалоговому вікні **Вставлення роздільника** встановіть прапорці для полів «Прізвище» та «Місяць».
 6. Натисніть ОК.
 7. Утримуючи клавішу «Ctrl» клацніть на кнопках зрізів: «Фомін», «Лютий».
 8. За даними зведеної таблиці побудуйте зведену діаграму.
 9. Виберіть команду **Аналіз зведених таблиць** → **Зведена діаграма**.
 10. У діалоговому вікні **Вставлення діаграми** виберіть тип діаграми **Стовпчаста**.
 11. Натисніть ОК.
- На зведеній діаграмі знаходяться кнопки вибору полів, що дозволяють застосувати фільтри до цих діаграм.
12. Натисніть лівою кнопкою миші на фільтр кнопки «Прізвище» та залиште лише прізвища «Іванов» та «Попов».
 13. Для видалення кнопки вибору полів, виберіть команду **Аналіз зведених діаграм** → **Кнопки полів** → **Приховати всі** (рис. 25).

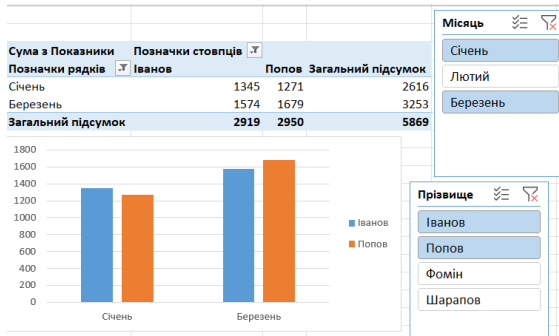


Рис. 25. Приклад зведеної таблиці «Завдання 5»

При створенні та використанні зведеної діаграми майте на увазі наступне:

1) Між зведеною таблицею і зведеною діаграмою Excel автоматично встановлюється двосторонній зв'язок. Тому при внесенні будь-яких структурних змін в одному з цих об'єктів, вони відразу будуть відображені на іншому об'єкті.

2) На панелі **Область фільтра зведеної таблиці**, яка з'являється після виділення зведеної діаграми, заголовками осей вказані назви полів, поміщені в області **Назви рядків** і **Назви стовпців** зведеної таблиці. Тут також представлені засоби фільтрації значень цих полів. Тому при внесенні будь-яких змін до зведеної діаграми за допомогою засобів фільтрації панелі **Область фільтра зведеної таблиці** вони відобразяться і в зведеній таблиці.

Завдання 6. Створення обчислюваного поля.

1. Створіть таблицю за прикладом (табл. 11).
2. За цими даними побудуйте зведену таблицю на окремому аркуші. У цій зведеної таблиці:

- поле «Тривалість процедур» розміщується в області «Значення»;
- поле «Фізичний терапевт» розміщується в області «Стовпці»;
- поле «Місяць» розміщується в області «Рядки».

1. Клацніть лівою кнопкою миші на одному з елементів таблиці.

2. Виберіть команду **Аналіз зведених таблиць** → **Обчислення** → **Поля, елементи та набори** → **Обчислюване поле**.

3. У діалоговому вікні **Вставка обчислюваного поля** введіть ім'я поля «Середня тривалість» і формулу «= Тривалість процедур (хв) / 'Проведено сеансів'».

Формулу можна вводити вручну або двічі клацнути на іменах відповідних полів в списку **Поля**. Скопіюйте його в поле **Формула**

подвійним натисканням на елементі. Оскільки в назві поля «Проведено сеансів» є пробіл, програма автоматично виокремить цю назву в одинарні лапки.

Таблиця 11

Вхідні дані

№	Фізичний терапевт	Місяць	Тривалість процедур (хв)	Проведено сеансів	№	Фізичний терапевт	Місяць	Тривалість процедур (хв)	Проведено сеансів
1	Ганна	Січень	1230	41	13	Володимир	Січень	1125	37
2	Ганна	Лютий	1100	38	14	Володимир	Лютий	1150	39
3	Ганна	Березень	1190	36	15	Володимир	Березень	1205	40
4	Ганна	Квітень	1325	44	16	Володимир	Квітень	1235	41
5	Ганна	Травень	1380	46	17	Володимир	Травень	1290	43
6	Ганна	Червень	1260	42	18	Володимир	Червень	1255	42
7	Ганна	Липень	1400	47	19	Володимир	Липень	1310	44
8	Ганна	Серпень	1350	45	20	Володимир	Серпень	1345	45
9	Ганна	Вересень	1415	48	21	Володимир	Вересень	1390	47
10	Ганна	Жовтень	1385	46	22	Володимир	Жовтень	1440	48
11	Ганна	Листопад	1470	49	23	Володимир	Листопад	1460	49
12	Ганна	Грудень	1500	50	24	Володимир	Грудень	1510	51

Створене поле буде додано до області «Значення» зведеної таблиці. З цим полем можна працювати так само, як і з будь-яким іншим полем, за одним лише винятком: його не можна переміщати до областей «Рядки», «Колони» і «Фільтри» (воно має розташовуватися тільки в області «Значення»).

4. Замініть ім'я поля «Сума по полю Середня тривалість» на «Середня тривалість сеансів».

5. Виберіть команду **Конструктор** → **Макет** в блоці **Параметри стилю зведеної таблиці** виберіть **Строкати стовпці** (рис. 26).

Позиція стовпця		Ганна		Володимир		Разом Сума з Тривалість процедур (хв)		Разом Середня тривалість сеансів	
Позначки рядків	Сума з Тривалість процедур (хв)	Середня тривалість сеансів	Сума з Тривалість процедур (хв)	Середня тривалість сеансів	Сума з Тривалість процедур (хв)	Середня тривалість сеансів	Сума з Тривалість процедур (хв)	Середня тривалість сеансів	Сума з Тривалість процедур (хв)
Січень	1125	30,40540541	1230	30	2355	30,19230789	1125	30,19230789	2355
Лютий	1150	29,48717949	1100	30	2250	29,20727273	1150	29,20727273	2250
Березень	1205	30,125	1190	30	2395	30,11311378	1205	30,11311378	2395
Квітень	1235	30,12195122	1325	30	2560	30,11363636	1235	30,11363636	2560
Травень	1290	30	1380	30	2670	30	1290	30	2670
Червень	1255	29,88095238	1260	30	2515	29,84047619	1255	29,84047619	2515
Липень	1310	29,77272727	1400	30	2710	29,78021578	1310	29,78021578	2710
Серпень	1345	29,48680809	1350	30	2695	29,94444444	1345	29,94444444	2695
Вересень	1390	29,57446809	1415	30	2805	29,52631579	1390	29,52631579	2805
Жовтень	1440	30	1385	30	2825	30,05315149	1440	30,05315149	2825
Листопад	1460	29,79591837	1470	30	2930	29,89795918	1460	29,89795918	2930
Грудень	1510	29,60784314	1500	30	3010	29,8019002	1510	29,8019002	3010
Загальний підсумок	15715	29,87642586	16005	30,08458647	31720	29,98109641	15715	29,98109641	31720

Рис. 26. Приклад зведеної таблиці «Завдання б»

Завдання 7. Вставка елемента, що обчислює.

1. Встановіть табличний курсор на одному з елементів стовпця А.
2. Виберіть команду **Аналіз зведених таблиць** → **Обчислення** → **Поля, елементи та набори** → **Елемент, що обчислюється**.
3. У діалоговому вікні **Вставлення елемента, що обчислюється** введіть ім'я поля «Комісія за кв.1» і формулу «=10% * (Січень + Лютий + Березень)».
4. Натисніть **ОК**.
5. Повторіть п. 2–3 для створення обчислюваних елементів: «Комісія за кв.2» формула «= 11% * (Квітень + Травень + Червень)», «Комісія за кв.3» формула «= 12% * (Липень + Серпень + Вересень)», «Комісія за кв.4» формула «= 12,5% * (Жовтень + Листопад + Грудень)».

6. Натисніть **ОК**.

Комісія тут – умовне число, що може відображати розрахунок бонусу фізичного терапевта за квартал.

7. Виберіть команду **Конструктор** → **Макет** → **Загальна сума** → **Вимкнути для рядків і стовпців**.
8. Виділіть діапазон клітинок А6:А17.
9. Виберіть команду **Аналіз зведених таблиць** → **Група** → **Групувати виділення**.
10. Виділіть діапазон клітинок А19:А26.
11. Виберіть команду **Аналіз зведених таблиць** → **Група** → **Групувати виділення**.
12. Перейменуйте назву груп: «Група1» – «Тривалість процедур», «Група2» – «Комісійні за квартал».
13. Встановіть табличний курсор на одному з елементів таблиці.
14. Виберіть команду **Конструктор** → **Проміжні підсумки** → **Відображати всі проміжні підсумки в нижній частині групи** (рис. 26).

Позначи стовпці		Ганна			
Володимир					
Позначи рядків	Сума з Тривалість процедур (кв)	Середня тривалість сеансів	Сума з Тривалість процедур (кв)	Середня тривалість сеансів	
Тривалість процедур					
Січень	1125	30,40540541	1230	30	
Лютий	1150	29,48717949	1100	28,94736842	
Березень	1205	30,125	1190	33,05555556	
Квітень	1235	30,12195122	1325	30,11363636	
Травень	1290	30	1380	30	
Червень	1255	29,88095238	1260	30	
Липень	1310	29,77272727	1400	29,78723404	
Серпень	1345	29,88888889	1350	30	
Вересень	1390	29,57446809	1415	29,47916667	
Жовтень	1440	30	1385	30,10869565	
Листопад	1460	29,79591837	1470	30	
Грудень	1510	29,60784114	1500	30	
Тривалість процедур П	15715	29,87642586	16005	30,08458647	
Комісійні за квартал					
Комісія за кв.1	348	30	352	30,60869565	
Комісія за кв.2	415,8	30	436,15	30,03787879	
Комісія за кв.3	485,4	29,74264706	499,8	29,75	
Комісія за кв.4	551,25	29,7972973	544,375	30,03448276	
Комісійні за квартал П	1800,45	29,86811546	1832,325	30,06522274	

Рис. 27. Приклад зведеної таблиці «Завдання 7»

Самостійна робота

Побудуйте декілька зведених таблиць за даними плану реабілітації лікування пацієнтів.

№	Вік	Діагноз	Вид терапії	Тривалість (хв)	Результат (оцінка 1-5)
1	45	Інсульт	Фізична терапія	60	4
2	62	Заміна колінного суглоба	Фізична терапія	45	5
3	38	Травма спини	Масаж	30	3
4	55	Розсіянийсклероз	Ерготерапія	50	4
5	70	Перелом стегна	Фізична терапія	60	4
6	48	Синдром зап'ястного каналу	Ерготерапія	40	3
7	58	Біль у попереку	Масаж	45	2
8	31	Спортивна травма	Фізична терапія	55	5
9	68	Інсульт	Логопедія	40	3
10	42	Травма плеча	Фізична терапія	50	4

- Зведена таблиця 1: Середня тривалість терапії за видом терапії.

Вид терапії	Середня тривалість (хв)
Ерготерапія	45
Логопедія	40
Масаж	37,5
Фізична терапія	54

- Зведена таблиця 2: Середній результат терапії за діагнозом

Діагноз	Середній результат
Біль у попереку	2
Заміна колінного суглоба	5
Інсульт	3,5
Перелом стегна	4
Розсіянийсклероз	4
Синдром зап'ястного каналу	3
Спортивна травма	5
Травма плеча	4
Травма спини	3

- Зведена таблиця 3: Кількість пацієнтів за віковими групами та видом терапії.

Кількість з	Вид терапії	Позначки стовпців				
Вікова група	Ерготерапія	Логопедія	Масаж	Фізична терапія	Загальний підсумок	
30-49	1		1	3	5	
50-70	1	1	1	2	5	
Загальний підсумок	2	1	2	5	10	

Контрольні питання

- Розкрийте поняття «зведена таблиця».
- Для чого призначена зведена таблиця?
- Опишіть покроково порядок дій для створення зведеної таблиці.
- Чи можна оновлювати дані у зведеній таблиці?

Практична робота № 5

Тема «Підготовка даних анкетного опитування до аналізу за допомогою SPSS»

Мета заняття: на навчальному прикладі опитування респондентів вивчити процес підготовки даних для обробки в програмі SPSS.

Зміст і хід заняття

1. Загальні теоретичні відомості.

Виконання та захист практичного завдання.

1. Загальні теоретичні відомості.

Визначення змінних.

У редакторі даних двічі клацніть на клітинці з написом **var** або клацніть на ярлику **Variable view/Перегляд змінних** у нижній межі таблиці. В обох випадках виконання цих команд приведе до переходу у режим перегляду змінних, який забезпечує редактор даних (рис. 28).

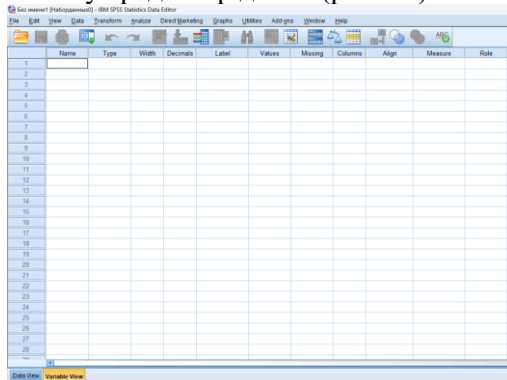


Рис. 28. Редактор даних: перегляд змінних

Ім'я змінної.

Задання імені змінної:

- Введіть у текстовому полі **Name/Ім'я** вибране ім'я змінної.

При виборі імені змінної слід дотримуватися певних правил:

- ✓ Імена змінних можуть містити літери латинського алфавіту та цифри. Крім того, допускаються спеціальні символи (підкреслення), . (крапка), а також символи @ та #. Не допускається використання пробілів, знаків інших алфавітів та спеціальних символів, таких як ! ? " *.

- ✓ Ім'я змінної має починатися з літери.

- ✓ Останній символ імені не може бути точкою або символом підкреслення (_).

✓ Імена змінних нечутливі до регістру, тобто великі та малі літери не розрізняються.

Приклади допустимих імен змінних:

- ✓ Статя
- ✓ Група_1
- ✓ Маса_тіла
- ✓ Група11

Приклади неприпустимих імен змінних:

1та1	Ім'я починається не з літери
Маса тіла (кг)	Ім'я містить пробіл та дужки
Ім'я!	Символ "!" не дозволяється

Тип змінної

Як видно з електронної таблиці, новостворені в SPSS змінні за замовчуванням є чисельними з максимальною довжиною вісім знаків, причому дрібна частина складається з двох знаків.

Для зміни типу змінної треба клацнути у клітинці на кнопці з трьома точками: Відкриється діалогове вікно **Define Variable Type/Визначення типу змінної** (рис. 29).

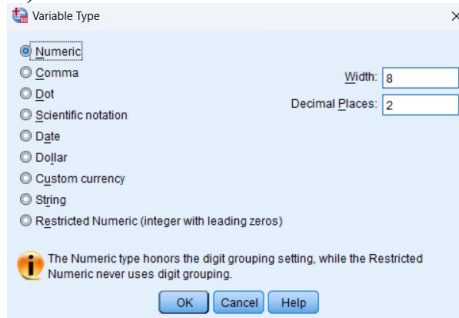


Рис. 29. Діалогове вікно Define Variable Type (для чисельної змінної)

У SPSS існують такі типи змінних:

Numeric (Чисельний)	До допустимих значень належать цифри, перед якими стоїть знак плюс або мінус та десятковий роздільник. Знак плюс перед числом, на відміну мінуса, не відображається. У текстовому полі Length (Довжина) визначається максимальна кількість знаків, включаючи позицію для десяткового роздільника. У текстовому полі Decimals (Десяткові розряди) вводиться кількість знаків дробової частини, що відображаються.
------------------------	--

Comma (Кома)	До допустимих значень відносяться цифри, перед якими стоїть знак плюс або мінус, крапка, як десятковий роздільник і одна або кілька ком в якості роздільників груп розрядів. Якщо коми опускаються під час введення, вони вставляються автоматично. Довжина такої змінної дорівнює максимальній кількості знаків, включаючи десятковий роздільник і коми між групами розрядів.
Dot (крапка)	До допустимих значень відносяться цифри, перед якими стоїть знак плюс або мінус, кома, як десятковий роздільник і одна або кілька крапок як роздільники груп розрядів. Якщо крапки опускаються під час введення, вони вставляються автоматично.
Scientific notation (Експоненційне подання)	При введенні даних дозволяються всі допустимі чисельні значення, включаючи експоненціальне представлення, про яке свідчить буква E або D, що міститься в числі, а також знак плюс або мінус.
Date (Дата)	Допустимі значення — дата та/або час.
Dollar (Долар)	До допустимих значень відносяться: знак долара, крапка, як десятковий роздільник і коми, як роздільники груп розрядів. Якщо знак долара або коми опускаються під час введення, вони вставляються автоматично.
Special currency (Спеціальна валюта)	Користувач може вказувати власні формати валюти. У полі Length у цьому випадку задається максимальна кількість знаків, включаючи всі знаки, які користувач ставить. Позначення валюти під час введення не вказується, воно вставляється автоматично.
String (Рядок)	Рядок символів. До допустимих значень належать: літери, цифри та спеціальні символи. Розрізняються короткі та довгі рядкові змінні. Короткі строкові змінні можуть містити не більше восьми знаків. У більшості процедур SPSS застосування довгих змінних обмежується або взагалі не допускається.

При введенні та виведенні даних треба враховувати такі особливості:

- **Чисельні формати.** У чисельних форматах десятковим роздільником може бути крапка, або кома. Тип десяткового роздільника залежить від параметрів діалогового вікна Мова та стандарти (Regional Settings) на панелі керування Windows. Точне значення змінної зберігається всередині програми, а *Редактор даних* відображає на екрані лише задану кількість десяткових розрядів. Значення, які мають більше десяткових розрядів, округляються. Для обчислень застосовується точне значення.

- **Формати дати та часу.** У форматах дати в якості роздільника між значеннями дня, місяця та числа можуть застосовуватися навскісна риса, дефіс, пробіл, кома або крапка. Можна вибрати один з декількох форматів дати (dd-mm-уууу, dd-mmm-уу, mm/dd/уууу тощо). Дата у форматі dd-mmm-уу відображається з роздільником-дефісом та скороченням назви місяця з трьох літер. Дата у форматах dd/mm/уу та mm/dd/уу відображається з роздільником – навскісною рисою та номером місяця замість назви.

- Всього доступно 27 різних форматів дати і часу, які відображаються в списку, що розгортається. У форматах часу в якості роздільників між значеннями годин, хвилин і секунд можуть використовуватися двокрапка, крапка або пробіл.

- **Спеціальна валюта.** Формати відображення валюти ССА, ССВ, ССС, ССD та СЕ задаються за допомогою вкладки Currency (Валюта), яка відкривається за допомогою команди у меню Edit (Правка) / Options... (Параметри...).

Мітка змінної/Label

Мітка змінної – це назва, яка дозволяє описати змінну докладніше. Мітка змінної може містити до 256 символів. У мітках змінних розрізняються великі та малі літери. Вони відображаються у тому вигляді, в якому були введені.

Мітки значень/Values

Мітки значень – це назва, яка дозволяє детальніше описати можливі значення змінної. Так, наприклад, для змінної «Стать» можна задати мітку «жіноча» для значення «1» та мітку «чоловіча» для значення «2».

Пропущені значення/Missing values

У SPSS допускаються два види пропущених значень:

- Пропущені значення, що визначаються системою (System-defined missing values). Якщо в матриці даних є незаповнені чисельні клітинки, система SPSS самостійно ідентифікує їх як пропущені значення. Цей факт відображається у матриці даних за допомогою коми «,».

- Пропущені значення, що задаються користувачем (User-defined missing values). У певних випадках у змінних відсутні значення. Наприклад, якщо на запитання не було відповіді, відповідь невідома, або існують інші причини, користувач може за допомогою кнопки **Missing** оголосити ці значення як пропущені.

Стовпці/Columns

Поле **Columns** визначає ширину, яку матиме в таблиці цей стовпець, при відображенні значень. Ширину стовпця можна також змінити безпосередньо у вікні редактора даних. Для цього треба помістити

вказівник миші на роздільник між двома заголовками стовпців з іменами змінних. Вигляд покажчика зміниться. Подвійна стрілка, що з'явилася, вказує, що відповідний стовпець можна розширити або звужити шляхом перетягування.

Вирівнювання/Alignment.

Тут можна встановити вид вирівнювання значень, тобто визначити, як вони відображатимуться в таблиці. Можливі види вирівнювання – **Right/праворуч**, **Left/ліворуч** та **Center/по центру**. Для вибору виду вирівнювання треба клацнути на кнопці .

Шкала виміру/Measure.

Тут можна задати шкалу змінної: номінальна (шкала найменувань), порядкова або метрична. За замовченням приймається метрична шкала виміру. Ця відмінність має значення лише для створення інтерактивних графіків, де номінальна і порядкова шкала вимірів об'єднуються у «категоріальний» тип. При завантаженні файлів, створених в попередніх версіях SPSS, або якщо шкала вимірів не визначається однозначно, програма SPSS автоматично передбачає, що дані представлені у метричній шкалі. Однак, якщо відповідна змінна має мітки значень або має менше 24 різних значень, то задається порядкова шкала.

Практичне завдання

Завдання 1. Підготовка даних до аналізу.

У таблиці 12 наведено вибірку з відповідей респондентів на питання анкети.

Таблиця 12

Вхідні дані

Номер анкети	Стать	Вік	Довжина тіла	Маса тіла	Чи дотримується Ви здорового раціону і режиму харчування?	Чи займаєтесь ви спортом?
1	жіноча	21	1,65	65,0	Так	Займаюсь регулярно
2	жіноча	21	1,68	65,0	Інколи	Інколи
3	жіноча	21	1,72	59,7	Інколи	Займаюсь регулярно
4	жіноча	20	1,67	58,0	Інколи	Займаюсь регулярно
5	жіноча	20	1,67	53,0	Так	Займаюсь регулярно
6	жіноча	19	1,69	59,0	Інколи	Не займаюсь
7	чоловіча	19	1,73	62,0	Інколи	Займаюсь регулярно
8	чоловіча	18	1,70	65,0	Інколи	Не займаюсь
9	чоловіча	19	1,85	60,0	Інколи	Займаюсь регулярно
10	чоловіча	18	1,79	66,6	Ні	Інколи

1. Відкрийте SPSS.
2. Для опису змінних в редакторі даних двічі клацніть на ярлику **Variable View** в лівому нижньому кутку екрану.
3. Визначте змінні так, як показано на рис.30.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Номер_анкети	Numeric	3	0	Порядковий номер	None	None	8	Center	Nominal
2	Стать	Numeric	1	0	Стать хворого	[0, немає д...	None	8	Center	Nominal
3	Вік	Numeric	2	0	Вік хворого	None	None	8	Center	Scale
4	Довжина_тіла	Numeric	3	2	м	None	None	8	Center	Scale
5	Маса_тіла	Numeric	2	1	кг	None	None	8	Center	Scale
6	Харчування	Numeric	1	0	Чи дотримуєтесь Ви здорового раціону і режиму харчування?	[0, немає д...	None	8	Center	Ordinal
7	Заняття_спортом	Numeric	1	0	Чи займаєтесь ви спортом?	[0, немає д...	None	8	Center	Ordinal

Рис. 30. Визначення змінних в редакторі SPSS

Визначення змінної «Номер анкети»

4. Введіть в поле **Name/Ім'я** текст «Номер_анкети» і натисніть на клавішу **Enter** або **Tab** для підтвердження.

5. Для встановлення розміру змінної клацніть в поле **Type/Тип** на кнопці з трьома крапками. Відкриється діалогове вікно **Variable Type/Визначення типу змінної**. Виберіть тип **Numeric/Числовий** і встановіть число символів «3» (ми припускаємо, що у нас не більше 999 рядків, в цьому прикладі – відповідей респондентів). Натисніть кнопку **OK** для підтвердження та перейдіть до наступного поля за допомогою клавіши **Tab**.

6. В полі **Decimals/Знаків після коми** встановіть «0» знаків після коми.

7. Для мітки змінної задайте текст «Порядковий номер».

8. Для міток значень погодьтеся із запропонованим програмою налаштуванням **None**, натиснувши **Enter**.

9. Погодьтеся із запропонованим програмою налаштуваннями в поле **Missing/Пропущені**, встановіть «8» в полі **Columns/Стовпці**, «Center» – в полі **Align/Вирівнювання** та виберіть налаштування «Nominal» в поле **Measure/Міра**.

Визначення змінної «Стать»

10. Внесіть в поле **Name** текст «Стать» і натисніть на клавішу **Enter** або **Tab** для підтвердження.

11. Для встановлення розміру змінної клацніть в полі **Type** на кнопці з трьома крапками. Відкриється діалогове вікно **Variable Type/Визначення типу змінної**. Прийміть запропоноване налаштування **Numeric/числовий**, встановіть довжину «1» і кількість десяткових розрядів «0», оскільки в цій змінній будуть зберігатися тільки значення 1, 2 або 0. Натисніть кнопку **OK** для підтвердження налаштування і перейдіть до наступного поля за допомогою клавіши **Tab**.

12. Для мітки змінної задайте текст «Стать хворого».

13. Клацніть в полі **Values** на кнопці. Відкриється діалогове вікно **Value Labels/Визначення міток значень**. Мітки значень визначаються наступним чином.

Визначення міток значень змінної «Стать» (рис. 31)

14. Спочатку введіть в поле **Value/Значення** число «1».

15. Введіть в поле **label/Мітка** текст «жіноча».

16. Клацніть на кнопці **Add/Додати**. Мітка значення буде додана до списку.

17. Повторіть ці дії для значень «2» — «чоловіча» і «0» — «немає даних». Натисніть кнопку **OK** для підтвердження налаштування.

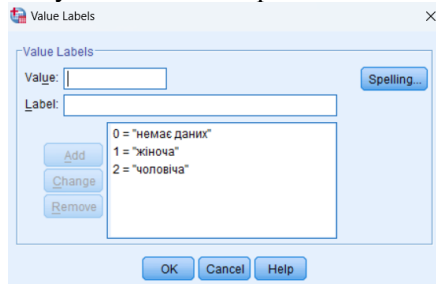


Рис. 31. Визначення міток значень змінної «Стать» у вікні **Value Labels**

Визначення міток значень змінної «Вік»

18. Введіть в поле **Name** текст «Вік» та натисніть клавішу **Enter** або **Tab**.

19. Щоб встановити розмір змінної, клацніть в полі **Type** на кнопці з трьома крапками. Відкриється діалогове вікно **Variable Type**. Прийміть запропоноване налаштування **Numeric**, встановіть довжину «2» (ми припускаємо, що всі респонденти не старше 99 років) і кількість десяткових розрядів «0». Натисніть кнопку **OK** для підтвердження налаштування і перейдіть до наступного поля за допомогою клавіши **Tab**.

20. У полях **Width** і **Decimals** прийміть налаштування, запропоновані за замовченням. Для мітки змінної введіть текст «Вік хворого», а для міток значень прийміть запропоноване налаштування **None**, натиснувши **Tab**.

21. Щоб задати пропущені значення, клацніть в полі **Missing** на кнопці з трьома крапками. Відкриється діалогове вікно **Missing Values**. За замовчуванням обрано варіант **No missing values/Немає пропущених значень**, тобто всі значення розглядаються як допустимі.

22. Прийміть пропонувані налаштування «8» в полі **Columns**, «Center» в полі **Align** і «Scale» в полі **Measure**.

23. Таким же способом запишіть значення змінних «Довжина тіла» та «Маса тіла»

Визначення міток значень змінної «Харчування» (рис. 32)

24. Введіть в поле **Name** текст «Харчування» і натисніть кнопку **Tab** для підтвердження.

25. Для встановлення розміру змінної клацніть в поле **Type** на кнопці з трьома крапками. Відкриється діалогове вікно **Variable Type**. Прийміть запропоноване налаштування **Numeric** і встановіть довжину «1» і кількість десяткових розрядів «0», оскільки в цій змінній будуть зберігатися тільки значення від 1 до 3 і 0 як відсутнє значення. Натисніть кнопку **OK** для підтвердження налаштування і перейдіть до наступного поля за допомогою клавіши **Tab**.

26. Для мітки змінної задайте текст «Чи дотримуєтеся Ви здорового раціону і режиму харчування?».

27. Клацніть в полі **Values** на кнопці. Відкриється діалогове вікно **Value Labels**.

28. Спочатку введіть в поле **Value/Значення** число «1».

29. Введіть в поле **label/Мітка** текст «так».

30. Клацніть на кнопці **Add/Додати**. Мітка значення буде додана в список. Повторіть ці дії для значень «2» — «ні», «3» — «інколи», і «0» — «немає даних».

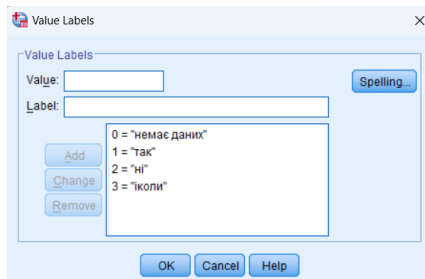


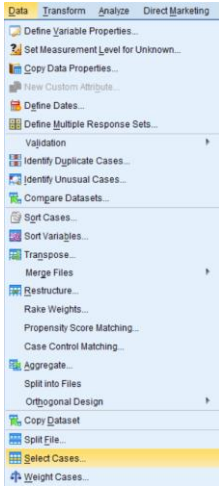
Рис. 32. Визначення міток значень змінної «Харчування» у вікні **Value Labels**

31. Таким самим способом запишіть значення змінної «Заняття спортом».

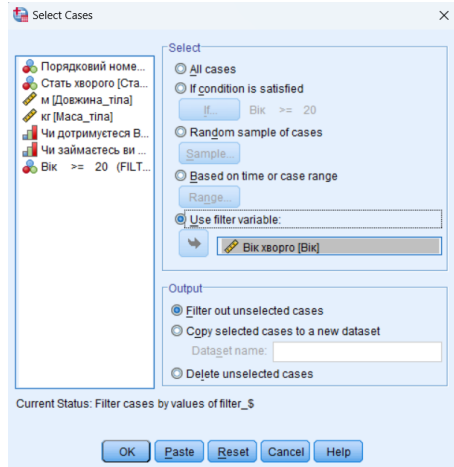
32. Перейдіть на вкладку **Data View** і введіть дані (відповіді 10 респондентів) (згідно таб.7).

Завдання 2. Фільтр даних.

1. Виберіть команду **Data/Дані** → **Select Cases/Вибір спостережень** (рис. 33).



а)



б)

Рис. 33. Виклик діалогового вікна **Select Cases** (а) та діалогове вікно **Select Cases** (б)

2. Для того, щоб в аналізі були задіяні лише ті рядки, в яких вік опитуваних респондентів перевищує 20 років, необхідно поставити крапку в розділі **Use filter variable/Використовувати фільтруючу змінну**. У лівій частині діалогового вікна **Select Cases** мишкою оберіть змінну «Вік» і натисніть на піктограму .

3. Поставте крапку в розділі **If condition is satisfied/Якщо виконана умова** та натисніть кнопку **If**.

4. Додайте обмеження ≥ 20 .

5. Натисніть **Continue**, а потім **OK**.

В результаті отримано перекреслені рядки, які не задіяні в обчисленнях. Також добавлений стовпець **filter_\$** (рис. 34), в якому дані можуть набувати двох значення 0 та 1:

0 = Not Selected/Не вибрано; 1 = Selected/Вибрано.

	Номер_анкет и	Стать	Вік	Довжина_т...	Маса_тіла	Харчування	Заняття_спо ртом	filter_\$
1	1	1	21	1,65	65,0	1	1	1
2	2	1	21	1,68	65,0	3	3	1
3	3	1	21	1,72	59,7	3	1	1
4	4	1	20	1,67	58,0	3	1	1
5	5	1	20	1,67	53,0	1	1	1
6	6	1	19	1,69	59,0	3	2	0
7	7	2	19	1,73	62,0	3	1	0
8	8	2	18	1,70	65,0	3	2	0
9	9	2	19	1,85	60,0	3	1	0
10	10	2	18	1,79	66,6	2	3	0

Рис. 34. Діалогове вікно SPSS з виключеними даними

Завдання 3. Сортування даних.

Відсортуйте дані в таблиці по стовбцю «Маса тіла» за зростанням.

Для цього:

1. У вікні редактора даних виберіть команду **Data/Дані → Sort Cases/Сортування даних**.
2. У викликаному діалоговому вікні вкажіть змінну «Маса тіла», за якою потрібно виконати сортування.
3. Вкажіть напрям сортування **Sort Order/Порядок сортування → Ascending/В порядку зростання** (рис. 35).

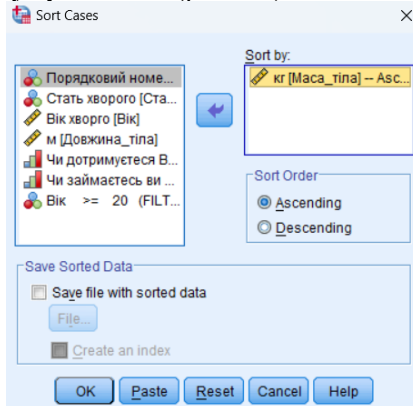


Рис. 35. Діалогове вікно **Sort Cases**

Самостійна робота

Створіть базу даних в програмі SPSS за результатами власного опитування.

Контрольні питання

1. Як перейти до опису змінних в SPSS?
2. Наведіть приклади ситуацій, у яких потрібно використовувати змінні типу «String». Які обмеження існують щодо їх використання в процедурах SPSS?
3. Що встановлюють в колонці «Decimals» при описі змінних?
4. Що таке мітки значень?
5. Які проблеми можуть виникнути, якщо ім'я змінної не відповідає правилам найменування? Наведіть приклади.
6. У яких випадках потрібно змінити шкалу вимірювання змінної в SPSS, і як це може вплинути на результати аналізу?

Практична робота № 6

Тема «Описова статистика у програмах MS Excel та SPSS»

Мета заняття: ознайомитися з основними поняттями описової статистики, сформувані вміння обчислювати значення статистичних параметрів вибірки.

Зміст і хід заняття

1. Загальні теоретичні відомості.

Виконання та захист практичного завдання.

1. Загальні теоретичні відомості.

Описові статистики – це різні показники, що обчислюються, та характеризують розподіл значень змінної. Ці показники умовно можна розбити на кілька груп. Перша група – міри центральної тенденції, навколо яких «групуються» дані: середнє значення, медіана та мода. Друга група характеризує мінливість значень змінної щодо середнього: стандартне відхилення та дисперсія. Діапазон мінливості характеризується мінімумом, максимумом та розмахом. Асиметрія та ексцес є мірою відхилення форми розподілу від нормального. Крім того, існують величини, що виражають похибки деяких статистик: стандартна похибка середнього, стандартна похибка асиметрії та стандартна похибка ексцесу.

Основні статистичні показники

Середнє арифметичне значення – це узагальнююча величина, що характеризує рівень варіюючої ознаки в якісно однорідній сукупності.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}.$$

Медіана – це таке значення x_i , яке знаходиться в середині варіаційного ряду.

Модойо вибірки (M_0) називають значення варіанти, що найчастіше зустрічається у вибірці, тобто значення x_i з найбільшою частотою n_i .

Дисперсія вибірки – величина, за допомогою якої характеризують розсіювання або скупченість навколо центра розсіювання статистичних даних. Це середнє арифметичне квадратів відхилень кожного значення ознаки від середнього значення по вибірці. Цю величину ще називають середнім квадратом відхилень.

$$D = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}$$

Середнє квадратичне (стандартне) відхилення – це величина, за допомогою якої характеризують розсіювання або скупченість навколо центра розсіювання статистичних даних. Ця величина показує, на скільки в середньому кожна варіанта вибірки відрізняється від середньої арифметичної величини цієї вибірки.

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

Коефіцієнт варіації (v) є характеристикою, яка оцінює відношення середнього квадратичного відхилення, тобто фактору розсіювання, до середнього арифметичного, тобто основного фактору ряду:

$$v = \frac{\sigma \cdot 100\%}{\bar{x}}$$

- від 0 до 10% – група однорідна;
- від 10 до 20% – середній ступень однорідності;
- 20 і вище – група неоднорідна.

Інтервал (амплітуда, варіаційний розмах) – це різниця між максимальним і мінімальним значеннями елементів вибірки. Інтервал є найпростішою і найменш надійною мірою варіації або розсіювання елементів у вибірці.

Стандартна похибка або **похибка середнього** розраховується за формулою:

$$m = \frac{\sigma}{\sqrt{n-1}}.$$

Стандартна похибка характеризує закономірні коливання (варіації) середньої арифметичної величини. Показує варіювання вибірових показників навколо їх генеральних параметрів. Вона має ті ж властивостями, як і стандартне відхилення. Чим більший обсяг вибірки, тим точніше середній результат, тим менша вибіркова середня відрізнятиметься від середньої генеральної сукупності. Отже, зі збільшенням числа випробувань помилка вибіркової середньої зменшуватиметься. Звідси стає ясніше значення вибіркової помилки: вона вказує на точність, з якою визначено середню величину, що супроводжується нею.

Ексцес – це ступінь вираженості «гостровершинності» розподілу, тобто частоти появи віддалених значень від середнього.

Асиметрія – величина, що характеризує несиметричність розподілу елементів вибірки щодо середнього значення. Приймає значення від -1 до 1. У разі симетричного розподілу асиметрія дорівнює 0. Часто значення асиметрії та ексцесу використовують для перевірки гіпотези про те, що дані (вибірка) належать до певного теоретичного розподілу, зокрема,

нормального розподілу. Для нормального розподілу асиметрія та ексцес дорівнює нулю.

Мінімум – значення мінімальної варіанти вибірки.

Максимум – значення максимальної варіанти вибірки.

Сума – сума значень всіх варіантів вибірки.

Рахунок – кількість елементів у вибірці.

Практичне завдання

Практична робота із розрахунку показників описової статистики в програмі MS Excel

Завдання 1. Для оцінки функціональної підготовленості лижників-двоборців спортсмени долають на лижеролерах 30 км на стандартній трасі. Відомий час проходження дистанції – x_i , хв. Проаналізуйте отримані дані з використанням описової статистики. Данні вимірювань представлені у таблиці 13.

Таблиця 13

Вхідні данні	
x_i	p_i
20,2	3
19,9	5
19,5	4
19,0	7
18,7	5
18,5	3
18,0	3
Всього	30

Рішення

1. Введіть експериментальні данні до електронної таблиці.

2. При статистичному аналізі необхідно визначити характеристики вибірки, при цьому найважливішою характеристикою є середнє значення. Для визначення середнього значення в групі необхідно встановити табличний курсор в вільну клітинку (наприклад, A32) і викликати функцію СРЗНАЧ для діапазону значень A2:A31. У відповідній клітинці отримаємо значення 19,1.

3. Наступною за важливістю характеристикою вибірки є міра розкиду елементів вибірки від середнього значення. Такою мірою є середнє квадратичне або стандартне відхилення. Для визначення стандартного відхилення в групі необхідно встановити табличний курсор у вільну клітинку (наприклад, A33) і викликати функцію СТАНДОТКЛОН. У відповідній клітинці отримаємо значення 0,67 (рис. 36).

4. Самостійно визначте значення моди та медіани за наведеним алгоритмом.

28	18,5	
29	18	
30	18	
31	18	
32	19,1	Середня арифметична величина
33	0,67	Середнє квадратичне (стандартне) відхилення
34	3,5	Коефіцієнт варіації
35	19	Мода
36	19	Медіана

Рис. 36. Результат розрахунку показників описової статистики

Висновки. Середній рівень проходження дистанції становить 19,1 хв. Розсіювання вихідних даних щодо середнього рівня є незначним ($v = 3,5\%$). Отже, у можна вважати, що у випробуваних спостерігається однаковий рівень функціональної підготовленості. Величина розсіювання становить $\pm\sigma$, у наведеному прикладі $\pm 0,67$ хв. Отже, загалом група спортсменів характеризується показником часу подолання дистанції $\bar{X} \pm \sigma$, у прикладі $19,1 \pm 0,67$, тобто 18,47...19,81 хв. Виходячи з цих параметрів, можна зробити висновок, що спортсмени, які показали результат 20,2 хв. та 19,9 хв., мають нижчий рівень підготовленості відносно середньогрупових показників; спортсмени, які показали результат 18 хв., мають вищий рівень підготовленості відносно середньогрупових показників.

Проте в цілому група випробовуваних є спортсменами однакової кваліфікації, про що свідчить малий коефіцієнт варіації ($v = 3,5\%$) і збіг значень середньої арифметичної ($\bar{x} = 19,1$ хв.), моди ($Mo = 19$ хв.) і медіани ($Me = 19$ хв.).

Завдання 2. У таблиці наведено антропометричні дані групи дітей 10-11 років. Необхідно визначити основні статистичні характеристики в кожній групі. Данні вимірювань представлені у таблиці 14.

Таблиця 14

Вхідні данні

№	Довжина тіла	Маса тіла	ОГК в спокої	№	Довжина тіла	Маса тіла	ОГК в спокої
1	146	37,6	69	11	146	29,8	61
2	140	50,1	84	12	138	39,5	69
3	144	34	70	13	141	25,9	63
4	137	32,9	62	14	140	36,2	72
5	147	33,3	66	15	148	44,9	74
6	142	43,7	73	16	143	42,9	78
7	140	28,6	59	17	145	35,2	69
8	161	39,1	67	18	153	61,9	85
9	169	52,8	77	19	151	35,5	67
10	142	34	66	20	150	36,0	70

Рішення

1. Введіть експериментальні данні до електронної таблиці.
 2. Для комплексного визначення основних статистичних характеристик груп треба виконати команду **Дані → Аналіз даних**. В списку **Інструменти аналізу** обрати рядок **Описова статистика**.

3. У діалоговому вікні в робочому полі **Вхідний інтервал** вкажіть вхідний діапазон – A1:C21. Активувавши перемикачем робоче поле **Вихідний інтервал**, вкажіть вихідний діапазон – клітинок E1. У розділі **Угруповання** встановіть перемикач в положення **По стовпцях**. Встановіть прапорець в полі **Мітки в першому рядку**. Встановіть прапорець в полі **Підсумкова статистика** та **Рівень надійності**. Натисніть кнопку **ОК** (рис. 37).

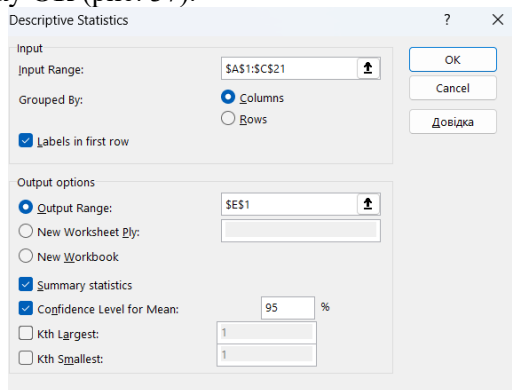


Рис. 37. Вікно Описової статистики

В результаті аналізу в зазначеному вихідному діапазоні для кожного стовпця даних отримаємо відповідні результати (рис. 38).

<i>Довжина тіла</i>		<i>Маса тіла</i>		<i>ОГК в спокої</i>	
Mean	146,15	Mean	38,695	Mean	70,05
Standard Error	1,75660408	Standard Error	1,945419	Standard Error	1,565206
Median	144,5	Median	36,1	Median	69
Mode	140	Mode	34	Mode	69
Standard Deviation	7,855772266	Standard Deviation	8,70018	Standard Deviation	6,999812
Sample Variance	61,71315789	Sample Variance	75,69313	Sample Variance	48,99737
Kurtosis	2,935664936	Kurtosis	1,448886	Kurtosis	0,176545
Skewness	1,613634662	Skewness	1,130136	Skewness	0,653618
Range	32	Range	36	Range	26
Minimum	137	Minimum	25,9	Minimum	59
Maximum	169	Maximum	61,9	Maximum	85
Sum	2923	Sum	773,9	Sum	1401
Count	20	Count	20	Count	20
Confidence Level(95,0%)	3,676614594	Confidence Level(95,0%)	4,07181	Confidence Level(95,0%)	3,276013

Рис. 38. Результат роботи інструменту Описова статистика

Практична робота в програмі SPSS

Завдання 3. В таблиці наведено результати тестування рівня IQ у 21 особи. Проаналізуйте отримані дані з використанням описової статистики. Данні вимірювань представлені у таблиці 15.

Таблиця 15

Вхідні данні

№	IQ	№	IQ
1	107	12	109
2	84	13	90
3	118	14	107
4	102	15	98
5	92	16	91
6	104	17	98
7	97	18	112
8	83	19	94
9	89	20	80
10	93	21	96
11	103		

Рішення

1. Введіть експериментальні данні до електронної таблиці.
2. Перейдіть на вкладку **Variable View** та перейменуйте першу змінну як «IQ». Тип – числовий, шкала – кількісна.
3. Виберіть вкладку **Analyz/Аналіз** → **Descriptive statistics/Описові статистики** → **Frequencies/Частоти**.
4. У діалоговому вікні, що відкрилося, перенесіть у поле **Змінні/Variables** змінну «IQ» та натисніть кнопку **Statistics/Статистики**.
5. У діалоговому вікні, що відкрилося виберіть основні показники описової статистики: **Середнє / Mean**; **Медіана / Median**; **Мода / Mod**; **Сума / Sum**; **Стандартне відхилення / Std. Dev.**; **Мінімум / Minimum**; **Максимум / Maximum**.

Натисніть кнопку **Continue / Продовжити** (рис. 39).

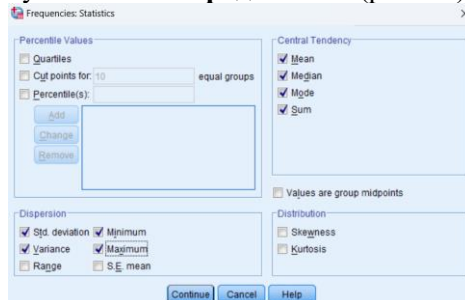


Рис. 39. Вид функції **Frequencies: Statistics**

6. Натисніть кнопку **Charts / Діаграми**. У вікні, що відкрилося виберіть **Histograms / Гістограми** та встановіть прапорець **Show normal curve on histogram / З нормальною кривою** (рис. 40). Натисніть кнопку **Continue / Продовжити**, а потім **OK**.

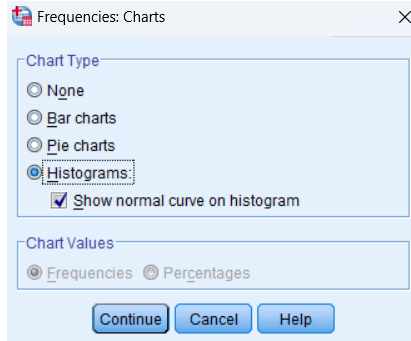


Рис. 40. Вид функції **Frequencies: Charts**

У вікні перегляду результатів з'явиться таблиця з обраними описовими статистиками (рис. 41).

Statistics

IQ

N	Valid	21
	Missing	0
Mean		97,4762
Median		97,0000
Mode		98,00 ^a
Std. Deviation		9,92783
Variance		98,562
Minimum		80,00
Maximum		118,00
Sum		2047,00

a. Multiple modes exist. The smallest value is shown

Рис. 41. Вид таблиці з описовими статистиками

У таблиці буде показано кількість всіх спостережень (обсяг вибірки), кількість пропущених значень та обрані описові характеристики.

У другій таблиці представлена інформація з частотною характеристикою заданої змінної. Це дозволяє проаналізувати частоту ознак, які зустрічаються у вибірці, та дати попередню оцінку їх розподілу (рис. 42).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	80,00	1	4,8	4,8	4,8
	83,00	1	4,8	4,8	9,5
	84,00	1	4,8	4,8	14,3
	89,00	1	4,8	4,8	19,0
	90,00	1	4,8	4,8	23,8
	91,00	1	4,8	4,8	28,6
	92,00	1	4,8	4,8	33,3
	93,00	1	4,8	4,8	38,1
	94,00	1	4,8	4,8	42,9
	96,00	1	4,8	4,8	47,6
	97,00	1	4,8	4,8	52,4
	98,00	2	9,5	9,5	61,9
	102,00	1	4,8	4,8	66,7
	103,00	1	4,8	4,8	71,4
	104,00	1	4,8	4,8	76,2
	107,00	2	9,5	9,5	85,7
	109,00	1	4,8	4,8	90,5
	112,00	1	4,8	4,8	95,2
	118,00	1	4,8	4,8	100,0
Total		21	100,0	100,0	

Рис. 42. Вид частотних таблиць в SPSS

Частоти на гістограмі позначені колонками, які примикають один до одного. Розриви між стовпцями означають відсутність даних. Крім того, на малюнку відображається середнє значення ряду, стандартне відхилення та загальна кількість спостережень (рис. 43).

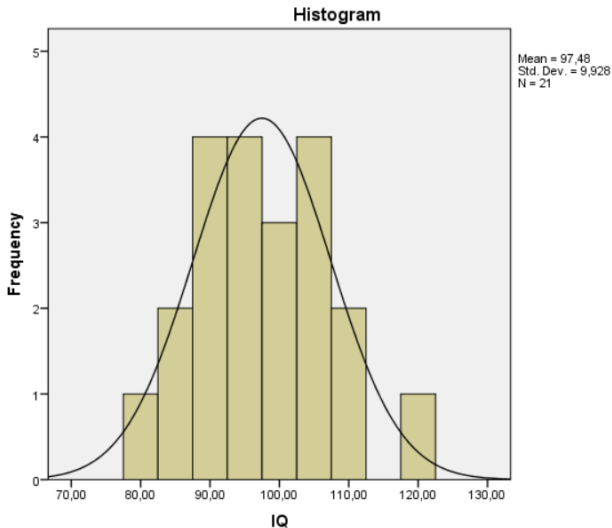


Рис. 43. Вид гістограми з нанесеною кривою нормального розподілу

Самостійна робота

1. Розрахуйте основні показники описової статистики за результатами антропометричних вимірів дітей 7-8 років. Візуалізуйте отримані дані та зробіть висновок.

№	Довжина тіла, (см)	Довжина хребетного стовпа, (см)	Шийний відділ, (см)	Грудний відділ, (см)	Поперековий відділ, (см)	Крижовий та куприковий відділи, (см)
1	129,2	52	5,7	20,0	12,0	14,3
2	123,5	50	5,5	23,0	12,0	9,5
3	138,5	54	6,8	22,0	11,5	13,7
4	117,5	46	5,9	22,0	12,0	6,1
5	124,5	49	6,9	22,0	13,0	7,1
6	117,0	44	5,2	22,0	11,5	5,3
7	112,5	44	5,8	21,0	10,0	7,2
8	122,5	48	6,6	23,0	10,0	8,4
9	125,5	47	6,2	19,5	11,0	10,3
10	116,5	44	6,1	21,0	12,0	4,9
11	119,5	45	5,7	21,0	12,0	6,3

2. У ході тренувального процесу при порівнянні функціонального стану двох боксерів x_i , y_i 25 разів вимірювався час, мс, їх диференційованої (виборчої) реакції. Здійсніть порівняльний аналіз отриманих даних з використанням описової статистики.

№	x_i	n_i
1	600	1
2	580	5
3	570	8
4	550	7
5	520	2
6	500	2
Всього	-	25

№	y_i	n_i
1	650	5
2	620	6
3	600	9
4	580	3
5	570	1
6	560	1
Всього	-	25

Контрольні питання

1. Дайте визначення поняттю «описова статистика».
2. Що називають модою вибірки?
3. Що називають дисперсією вибірки?
4. Яку величину називають середнім квадратичним відхиленням?

Наведіть приклади її застосування.

5. Як розрахувати описову статистику за допомогою Аналізу даних?

6. Які діаграми характеризують вибірку по відношенню до нормального розподілу?

Практична робота № 7

Тема «Параметричні методи перевірки статистичних гіпотез у програмах MS Excel та SPSS»

Мета заняття: ознайомитися із методами перевірки статистичних гіпотез та способами її розрахунку в програмах MS Excel та SPSS.

Зміст і хід заняття

1. Загальні відомості про метод перевірки статистичних гіпотез.
2. Параметричні критерії.

Виконання та захист практичного завдання.

1. Загальні відомості про метод перевірки статистичних гіпотез.

Жодне серйозне медичне або медико-біологічне наукове дослідження не обходиться без статистичного аналізу експериментальних даних. Крім опису початкових вимірювань часто досліднику потрібно довести статистичну значущість результатів проведених дослідів. Тут на допомогу приходять методи, пов'язані з перевіркою статистичних гіпотез. Статистична гіпотеза – це твердження щодо одного або декількох параметрів генеральної сукупності або про форму розподілу генеральної сукупності, зроблене на підставі вибірки. Методи, що дозволяють оцінити статистичну значущість відмінностей, називають критеріями значущості або просто критеріями. Існує велике різноманіття критеріїв, але в основу всіх критеріїв покладені одні й ті ж принципи.

Перед проведенням будь-якого експерименту формують так звані «нульову» та «альтернативну» гіпотези. Нульовою гіпотезою називають таку, за якою між сукупностями, що порівнюються, відсутні значущі відмінності (позначається H_0). Протилежна гіпотеза стверджує, що між сукупностями, які порівнюються, існують значущі відмінності. Вона називається альтернативною (позначається H_1).

Перевірка гіпотез може проводитися за результатами однієї групи досліджуваних. Наприклад, щоб оцінити ефективність нових методик, проводяться вимірювання контрольних показників перед початком та в кінці їх застосування в окремій групі досліджуваних.

Перед початком педагогічного експерименту зазвичай обирають дві або більше однорідні за ознаками групи об'єктів. В одній групі проводять заняття за традиційною методикою (контрольна група), у в другій застосовують нові методики (експериментальна група). В кінці дослідження результати порівнюють, аналізують та роблять висновки щодо доцільності та ефективності нововведень.

Для визначення значущості відмінностей між досліджуваними групами об'єктів спочатку формують нульову та альтернативну гіпотези, задають рівень значущості. Прийнято вважати, що для медико-біологічних досліджень достатній рівень значущості $\alpha = 0,05$. Для підвищення точності рекомендують обирати рівні значущості $\alpha = 0,01$ або $\alpha = 0,001$.

Лише після формулювання гіпотез та вибору рівня значущості, формують вибіркові сукупності. Якщо обчислені параметри не суперечать нульовій гіпотезі, вона підтверджується. У випадку, коли різницю між результатами не вдається пояснити лише дією випадкових факторів, то нульова гіпотеза відхиляється і підтверджується альтернативна.

2. Параметричні критерії.

Параметричні критерії служать для перевірки гіпотез про рівність середніх значень або дисперсій. З параметричних критеріїв найбільшою популярністю при перевірці гіпотез про рівність генеральних середніх (математичних очікувань) користується *t*-критерій Стьюдента. Він використовується для перевірки наступної гіпотези: «Дві вибіркові сукупності належать до одної генеральної сукупності». Критерій дозволяє знайти ймовірність того, що обидва середніх відносяться до однієї генеральної сукупності. Якщо ця ймовірність p нижче рівня значущості ($p < 0,05$), то прийнято вважати, що вибірки належать до двох різних генеральних сукупностей.

t-критерій може бути застосований для перевірки гіпотези про рівність генеральних середніх двох незалежних, незв'язаних вибірок (*двохвибірковий t-критерій*). Наприклад, перевірка значущості різниці між результатами контрольної та експериментальної група; обсяг вибірок може бути різним.

t-критерій може бути застосований також для перевірки гіпотези про рівність генеральних середніх двох залежних, зв'язаних вибірок (*парний t-критерій*). Вибірki при цьому називають залежними, пов'язаними. Наприклад, вимірюється вміст лейкоцитів у крові спортсменів до та після забігу на марафонську дистанцію.

Застосування *t*-критерію передбачає відповідність розподілу у досліджуваних сукупності нормальному закону. Також необхідна попередня перевірка про рівність дисперсій.

Для оцінки достовірності відмінностей за критерієм Стьюдента приймається нульова гіпотеза – середні значення двох вибірок рівні між собою. Далі обчислюється значення ймовірності того, що відмінності між групами зумовлені випадковими чинниками.

В програмі MS Excel для оцінки достовірності відмінностей за критерієм Стюдента використовуються спеціальна функція **СТЮДЕНТ.ТЕСТ** і процедури **Пакету аналізу**. Ці інструменти дозволяють обчислювати ймовірність, відповідну критерію Стюдента. Вони використовуються для визначення ймовірності того, що дві вибірки взяті з однієї генеральної сукупності.

Функція **СТЮДЕНТ.ТЕСТ** має наступний синтаксис:

СТЮДЕНТ.ТЕСТ (массив1; массив2; хвосты; тип) тут:

- массив1 – це перша кількість даних;
- массив2 – це друга кількість даних;
- хвосты – число хвостів розподілу. Зазвичай число хвостів дорівнює 2;
- тип – це вид виконуваного t-тесту. Можливі 3 варіанти вибору:

1 – парний тест, 2 – двохвибірковий тест з рівними дисперсіями, 3 – двохвибірковий тест з нерівними дисперсіями.

Практичне завдання

Використання інструменту «Пакет аналізу» в програмі MS Excel для виявлення відмінностей між вибірками

Завдання 1. Визначити, чи значущі відмінності між результатом вимірювання частоти серцевих скорочень у спортсменів до і після виконання розминки. Таблиця з вхідними даними представлена на рисунку 44.

	А	В
1	до розминки	після розминки
2	70	135
3	66	126
4	61	115
5	72	140
6	80	121
7	59	112
8	64	130

Рис. 44. Таблиця з вхідними даними

Рішення

1. Введіть експериментальні дані до електронної таблиці.
2. Для виявлення достовірності відмінностей встановіть курсор у вільну клітинку (наприклад, A11). Запустіть **Майстер функцій**, оберіть категорію **Статистичні**, функцію **СТЮДЕНТ.ТЕСТ**. У діалоговому вікні функції **СТЮДЕНТ.ТЕСТ** введіть вхідні дані: в поле **массив1** введіть діапазон A2:A8; в поле **массив2** – діапазон даних досліджуваної групи B2:B8. В поле **Хвосты** запишіть з клавіатури цифру 2 (без лапок), а в полі **Тип** – цифру 1. Натисніть кнопку **ОК**. В клітинці A11 з'явиться значення ймовірності - 0,00000363.

3. Оскільки величина ймовірності випадковості різниці між аналізованими вибірками (0,00000363) менше рівня значущості ($\alpha = 0,05$), то нульова гіпотеза відкидається. Отже, відмінності між вибірками не випадкові і середні вибірок вважаються такими, що значущо відрізняються один від одного. Тому на підставі застосування критерію Стьюдента можна зробити висновок статистичну значущість між результатами ЧСС до та після виконання розминки ($p < 0,05$).

Завдання 2. Перевірити ефективність розробленої програми фізичної терапії для чоловіків з посттравматичними контрактурами ліктьового суглоба на довготривалому етапі реабілітації. Ефективність нової програми оцінювалася за показником розгинання ліктьового суглоба (кут відхилення, град.). В основній групі (ОГ) 10 досліджуваних, в контрольній групі (КГ) 11 досліджуваних. Дані вимірювань представлені у таблиці 16.

Таблиця 16

Вхідні дані

ОГ	150	162	155	145	144	170	150	168	165	172	
КГ	124	115	128	130	135	128	122	125	140	135	138

Рішення

Для вирішення завдань такого типу використовуються так звані критерії відмінності, зокрема, **t-критерій Стьюдента**.

1. Введіть дані: ОГ – в діапазон A2:A11; КГ – в діапазон B2:B12.

2. Вибір процедури здійснюється з трьох варіантів **t-тесту**. Необхідно розрахувати дисперсію для кожної групи. Встановіть курсор у вільну клітинку (наприклад, A13). Увійдіть до меню **Майстер функцій**, виберіть категорію **Статистичні** та функцію **ДИСП.В.** У діалоговому вікні функції **ДИСП.В.** введіть діапазон клітинок A2:A11. Натисніть кнопку **ОК**. В клітинці A13 з'явиться значення дисперсії – 158,1. Так само розрахуйте дисперсію для другої групи. Оскільки дисперсії двох груп нерівні. Для реалізації процедури в пункті меню **Дані** виберіть рядок **Аналіз даних** і далі вкажіть курсором миші на рядок **Двохвибірковий t-тест з різними дисперсіями**.

3. У діалоговому вікні задайте **Інтервал змінної 1** – вкажіть діапазон A1:A11.

4. Аналогічно вкажіть **Інтервал змінної 2** – вкажіть діапазон B1:B12.

5. Поставте прапорець у полі **Мітки**.

6. Вкажіть вихідний діапазон. Для цього поставте перемикач в положення **Вихідний діапазон** і введіть посилання на клітинку D1. Клацніть по кнопці **ОК**.

Результати аналізу. У вихідному діапазоні D1:F13 з'являться результати процедури **Двохвибірковий t-тест з різними дисперсіями** (рис. 45).

D	E	F
t-Test: Two-Sample Assuming Unequal Variances		
	ОГ	КГ
Mean	158,1	129,0909091
Variance	111,8777778	56,29090909
Observations	10	11
Hypothesized Mean Difference	0	
df	16	
t Stat	7,184092974	
P(T<=t) one-tail	1,08718E-06	
t Critical one-tail	1,745883676	
P(T<=t) two-tail	2,17437E-06	
t Critical two-tail	2,119905299	

Рис. 45. Результат розрахунку Двохвибіркового t-тесту з різними дисперсіями

Інтерпретація результатів. Оскільки при порівнянні t -статистика $> t$ критичне двостороннє ($7,18 > 2,12$), різницю між вибірками слід вважати статистично значущою. Отже, розроблена програма фізичної терапії для чоловіків з посттравматичними контрактурами ліктьового суглоба на довготривалому етапі реабілітації є ефективною.

Практична робота в програмі SPSS

Завдання 3. Порівняння двох незалежних вибірок. Чи істотно відрізняється дія двох груп медикаментів на осіб різного віку. Дані вимірювань представлені у таблиці 17.

Рішення

1. Введіть експериментальні данні до електронної таблиці.
2. Оберіть **Analyz/Аналіз** → **Порівняння середніх/Compare Means** → **Independent-Samples T Test/T-критерій для незалежних вибірок**. У діалоговому вікні **T-критерій для незалежних вибірок** зі списку вихідних змінних перенесіть змінну «Вік» до списку **Test Variable(s)/Перемінні, що перевіряються** (рис. 46).

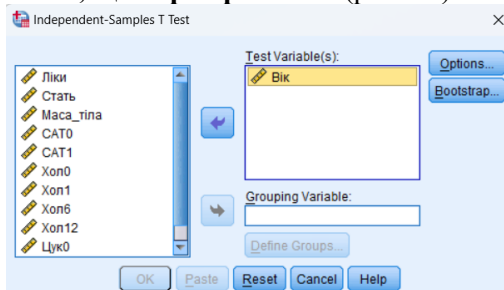


Рис. 46. Діалогове вікно Independent-Samples T Test

Таблиця 17

Антропометричні та біохімічні показники групи пацієнтів поліклініки N

№	Ліки	Стать	Вік	Маса тіла	SAT0	SAT1	Хол0	Хол1	Хол6	Хол12	Цук0	Цук1
1	2	2	66	70	170	135	265	235	245	205	90	85
2	2	2	58	63	160	155	185	185	175	195	85	80
3	2	1	47	70	165	155	225	235	250	260	95	90
4	1	1	80	48	170	165	160	155	135	136	86	80
5	2	2	52	77	170	165	265	260	245	250	104	100
6	1	1	48	82	160	155	195	200	185	210	95	90
7	1	2	54	77	170	160	230	225	235	200	95	100
8	2	1	46	75	160	150	185	180	195	155	90	95
9	1	1	53	83	160	155	220	220	235	200	105	100
10	1	1	71	62	165	155	180	175	175	195	105	100
11	2	2	55	88	170	170	265	260	250	230	145	185
12	2	1	63	73	160	150	185	180	185	170	90	85
13	2	1	52	88	155	150	170	175	175	185	100	95
14	2	2	66	66	165	160	195	190	175	185	120	125
15	1	1	52	78	160	150	185	180	195	170	87	82
16	2	2	51	67	170	170	245	250	245	250	105	110
17	2	1	57	70	160	160	180	170	180	160	75	87
18	2	2	73	55	170	165	205	210	200	200	95	96
19	1	1	47	70	170	160	215	220	200	195	81	88
20	2	2	43	70	170	170	220	225	235	240	96	97

Примітка: Ліки: 1 – Alphasan; 2 – Betasap. Стать: 1 – чоловіки; 2 – жінки. SAT0 – систолічний артеріальний тиск, вихідне значення. SAT1 – систолічний артеріальний тиск, через 1 місяць. Хол0 – холестерин, вихідне значення. Хол1 – холестерин, через 1 місяць. Хол6 – холестерин, через 6 місяців. Хол12 – холестерин, через 12 місяців. Цук0 – вміст цукру в крові, вихідне значення. Цук1 – вміст цукру в крові, через 1 місяць.

3. У такий же спосіб перенесіть змінну «Тип ліків» у поле **Grouping Variable/Групувати по**.

4. Клацніть на кнопці **Define Groups/Задати групи** відкривається вікно, в якому можна ввести значення двох категорій для групуючої змінної. Потрібно порівняти дві групи, які відповідають умовам відповідно Ліки = 1 та Ліки = 2. Тому внесіть у поле **Група 1** значення 1, а в полі **Група 2** — значення 2 → **Continue / Продовжити** → **ОК**. (рис. 47).

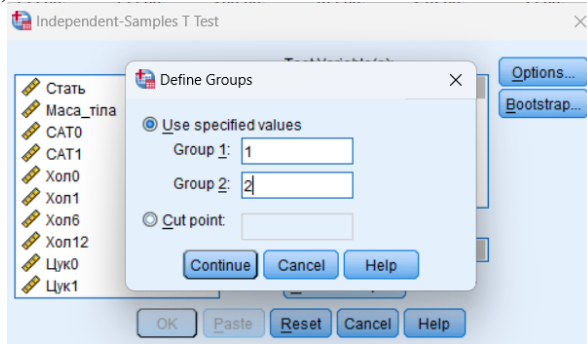


Рис. 47 Діалогове вікно **Define Groups**

У вікні перегляду з'являться такі результати (рис. 48; 49).

	Ліки	N	Mean	Std. Deviation	Std. Error Mean
Вік	1,00	7	57,8571	12,58873	4,75809
	2,00	13	56,0769	8,90189	2,46894

Рис. 48. Статистика групи

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Вік	Equal variances assumed	1,335	,263	,369	18	,716	1,78022	4,81879	-8,34369	11,90413
	Equal variances not assumed			,332	9,328	,747	1,78022	5,36051	-10,28145	13,84189

Рис. 49. Критерій незалежних вибірок

Отримані результати містять:

- кількість спостережень, середні значення, стандартні відхилення та стандартні похибки середніх в обох групах,

- результати тесту Лівіня на рівність дисперсій.

Як правило, гіпотеза про рівність (гомогенність) дисперсій не приймається, якщо тест Лівіня дає значення $p < 0,05$ (гетерогенність дисперсій). При гомогенності (рівності) та гетерогенності (нерівності) дисперсій виводяться такі характеристики:

- результати t-тесту: значення розподілу T, кількість ступенів свободи, ймовірність похибки **p** (позначається **Sig. (2-tailed)/Значимість (2-стороння)**),
- різниця середніх значень, її середньоквадратична похибка та довірчий інтервал.

За результатами обчислень робимо висновки про те, що вплив лікарського препарату на осіб різного віку не має статистично значущої відмінності ($p = 0,716$).

Завдання 4. Порівняння двох залежних вибірок. Необхідно з'ясувати, чи значуще змінюється вміст холестерину через місяць після початку прийому ліків. Для цього порівняйте змінні Хол0 та Хол1 за допомогою t-тесту для залежних вибірок. Вхідними будуть дані всіх пацієнтів, незалежно від групи прийнятих ліків.

Рішення

Analyz/Аналіз → Порівняння середніх/Compare Means → Paired-Samples T Test/T-критерій для парних вибірок. Перенесіть змінні «Хол0» і «Хол1» зі списку вихідних до змінних **Paired Variables/Парні змінні → ОК** (рис. 50).

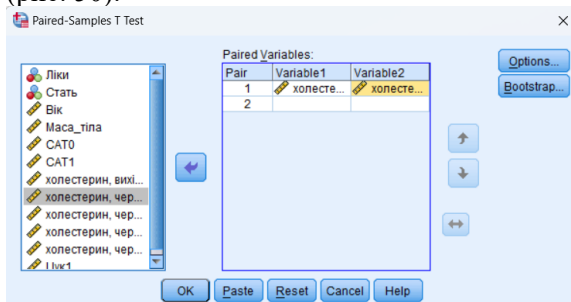


Рис. 50 Діалогове вікно **Paired-Samples T Test**

У вікні перегляду з'являться наступні результати (рис. 51–53):

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	холестерин, вихідне значення	208,7500	20	32,43921	7,25363
	холестерин, через 1 місяць	206,5000	20	31,75150	7,09985

Рис. 51. Статистика парних вибірок

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	холестерин, вихідне значення & холестерин, через 1 місяць	20	,965	,000

Рис. 52. Кореляція парних вибірок

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	холестерин, вихідне значення - холестерин, через 1 місяць	2,25000	8,50310	1,90135	-1,72957	6,22957	1,183	19	,251

Рис. 53. Результат виконання тесту

У вікні перегляду виведені такі результати:

- середні значення, кількість спостережень, стандартні відхилення та стандартні помилки середніх для обох змінних;
- кореляції парних вибірок;
- результати t-тесту: значення розподілу T, кількість ступенів свободи, ймовірність помилки **p** (позначається **Sig. (2-tailed)/Значимість (2-стороння)**).

За результатами обчислень робимо висновки про те, що вміст холестерину протягом місяця вживання ліків не змінився на значущому рівні ($p = 0,251$).

Самостійна робота

1. Група пацієнтів пройшла серію спеціальних відновлювальних занять. Оцінити прогрес досліджуваних пацієнтів за показниками стискання кистьового динамометра Н/кг (ньютон на 10 кг маси) до та після серії спеціальних відновлювальних занять.

До	10,2	10,2	10	10	10	9,8	9,8	9,7	9,7	9,5
Після	12,4	12,4	12,4	12,3	12,3	10	9,3	10,4	11	10,2

2. У двох групах плавців виміряна різниця частоти пульсу (уд./хв) між станом спокою і після максимально швидкого подолання дистанції. Встановити, чи значущі відмінності між двома групами. Зробити висновок.

№	x_i	n_i
1	92	3
2	94	4
3	95	7
4	97	2
5	99	1
6	100	2
Всього	-	19

№	y_i	n_i
1	98	3
2	102	5
3	103	6
4	104	1
5	105	2
Всього	-	17

3. Порівняти показники Гарвардського степ-тесту (ум. од.) у юних футболістів 11–12 років (x_i) і 13–14 років (y_i) і зробити висновок про динаміку їх працездатності.

№	x_i	n_i
1	85	2
2	88	3
3	90	7
4	93	4
5	95	3
6	98	1
Всього	-	20

№	y_i	n_i
1	95	2
2	99	4
3	100	5
4	103	6
5	105	2
6	108	1
Всього	-	20

Контрольні питання

1. Що таке «статистична гіпотеза»?
2. Назвіть види статистичних гіпотез.
3. Яку гіпотезу називають «нульовою»?
4. Яка гіпотеза називається «альтернативною»?
5. Як позначають нульову та альтернативну гіпотези?
6. Охарактеризуйте поняття «статистичний критерій». У чому полягає відмінність між параметричними і непараметричними критеріями?
7. Які види статистичних критеріїв реалізовані в Excel? Як вони застосовуються?
8. Як здійснюється перевірка гіпотез про рівність середніх для двох незалежних груп за допомогою SPSS?
9. Які групи вважаються залежними або парними?
10. Як здійснюється перевірка гіпотез про рівність середніх для двох залежних груп?

Практична робота № 8

Тема «Непараметричні методи перевірки статистичних гіпотез»

Мета заняття: ознайомитися із методами перевірки статистичних гіпотез та способами її розрахунку в програмі SPSS.

Зміст і хід заняття

1. Непараметричні критерії.
2. Критерій Вілкоксона.
3. Критерій знаків.
4. Критерій Манна-Уїтні (U-критерій Вілкоксона-Манна-Уїтні).
5. Критерій серій Вальда-Вольфовіца.

Виконання та захист практичного завдання.

1. Непараметричні критерії.

Оскільки непараметричні критерії не пов'язують аналіз даних із законом їх розподілу, то вони мають достатньо широке застосування у статистичному аналізі. Непараметричні критерії дозволяють досліднику здійснювати аналіз без додаткової перевірки характеру розподілу змінних. Непараметричні критерії переважно застосовують для аналізу даних, що представлені в номінальній чи ранговій шкалі.

2. Критерій Вілкоксона.

Критерій Вілкоксона – це непараметричний тест для порівняння двох залежних вибірок, коли дані не мають нормального розподілу. Його використовують у випадках, коли одні й ті самі об'єкти (наприклад, пацієнти) оцінюються двічі – до і після впливу (тренування, лікування тощо). Метод заснований на ранжуванні модулів різниць між парами і дозволяє виявити статистично значущі зміни. Він часто застосовується у реабілітації для оцінки ефективності вправ, методик чи лікувальних програм у рамках однієї групи. Критерій Вілкоксона можна використовувати при обсязі вибірки до 25 елементів.

3. Критерій знаків.

Критерій знаків – це простий непараметричний метод для порівняння двох пов'язаних вибірок. Замість обчислення числових різниць він враховує лише знак змін: чи збільшилось, чи зменшилось значення. Наприклад, якщо після курсу фізичної терапії у більшості пацієнтів зменшився рівень болю, критерій знаків дозволяє оцінити статистичну значущість цієї тенденції.

4. Критерій Манна-Уїтні (U-критерій Вілкоксона-Манна-Уїтні)

Критерій Манна-Уїтні – це непараметричний тест для порівняння двох незалежних вибірок. Його застосовують, коли дані не підкоряються нормальному розподілу або мають порядковий характер. Метод полягає в ранжуванні всіх спостережень і порівнянні сум рангів між групами. Наприклад, у фізичній терапії можна порівнювати ефективність двох методик на різних групах пацієнтів. Критерій дозволяє визначити, чи значущо відрізняються результати між групами, навіть якщо вибірки невеликі або дані неоднорідні.

5. Критерій серій Вальда-Вольфовіца

Критерій серій Вальда-Вольфовіца являє собою непараметричну альтернативу t-критерію для незалежних вибірок. Дані мають той же вигляд, що і в t-критерії для незалежних вибірок.

Наприклад, для того щоб порівняти чоловіків та жінок за деякою ознакою. Можна впорядкувати дані за зростанням, і знайти ті випадки, коли суб'єкти однієї й тієї самої статі примикають один до одного в побудованому варіаційному ряді (іншими словами, утворюють серію). Якщо немає різниці між чоловіками і жінками, то число і довжина «серій», що відносяться до однієї статі, будуть більш-менш випадковими. В іншому випадку дві групи (чоловіки та жінки) відрізняються одна від одної, тобто не є однорідними.

Критерій передбачає, що змінні, що розглядаються, є безперервними і виміряні, принаймні, в порядковій шкалі. Критерій серій Вальда-Вольфовіца перевіряє гіпотезу про те, що дві незалежні вибірки вилучені з двох популяцій, які в чомусь суттєво різняться між собою, інакше кажучи, відрізняються не лише середніми, а й формою розподілу. Обсяг аналізованих вибірок має бути досить великим.

Практична робота в програмі SPSS

Завдання 1. Виявити, чи існують статистично значущі відмінності між показниками вмісту цукру в крові у пацієнтів різної статі. Таблиця з вхідними даними для практичної роботи представлена в «Практичній роботі 7» таблиця 17.

Рішення.

Analyz/Аналіз → Nonparametric Tests/Непараметричні тести → Independent Samples/Для незалежних вибірок → Settings/Параметри → Customize tests/Налаштувати критерії → Mann-Whitney U (2 samples) / U Манна-Уїтні (для двох вибірок) перенесіть змінну «Цук0» з поля вихідних змінних до **Test Fields/Списку змінних, що**

перевіряються, а змінну «Стать» в поле **Groups/Групувати** → **Run/Запуск** (рис. 54).

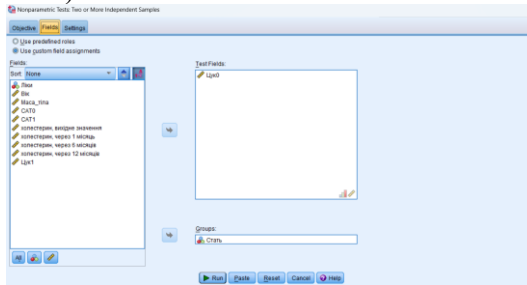


Рис. 54 Діалогове вікно **Nonparametric Tests: to or More Independent Samples**

У вікні перегляду з'явиться наступний результат (рис. 55).

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Цукр0 is the same across categories of Стать.	Independent-Samples Mann-Whitney U Test	,131 ¹	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

¹Exact significance is displayed for this test.

Рис. 55. Результат виконання тесту

Визначено, що між показниками цукру в крові у осіб різної статі відсутні статистично значущі відмінності ($p = 0,131$).

Тест Вілкоксона є традиційним непараметричним тестом для порівняння двох залежних вибірок. Він заснований на побудові рангової послідовності абсолютних різниць пар значень.

Завдання 2. Змінні CAT0 і CAT1 (початкові показники систолічного тиску та показники через місяць після початку прийому медикаменту) є типовим прикладом пов'язаних (залежних) вибірок. Оцініть, чи відрізняються показники систолічного артеріального тиску перед початком лікування та через 1 місяць після проведеного лікування. Таблиця з вхідними даними для практичної роботи представлена в «Практичній роботі 7» таблиця 16.

Рішення

Analyze/Аналіз → **Nonparametric Tests/Непараметричні тести** → **Related Samples/Для залежних вибірок** → **Settings/Параметри** →

Customize tests/Налаштувати критерії → Wilcoxon matched-pair signed (2 samples)/Парний знакових рангів Вілкоксона (для 2-х вибірок) перенесіть змінні «CAT0» і «CAT1» з поля вихідних змінних до **Test Fields/Списку змінних, що перевіряються → Run/Запуск** (рис. 56).

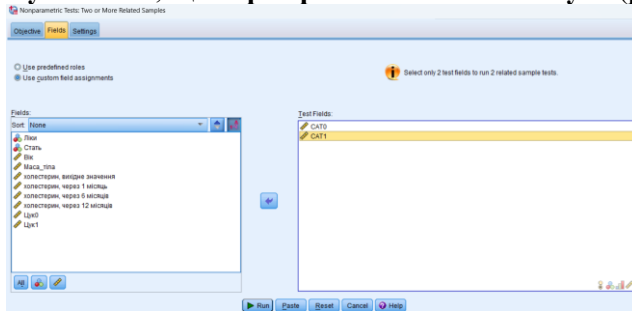


Рис. 56. Діалогове вікно **Nonparametric Tests: to or More Related Samples**

У вікні перегляду з'явиться наступний результат (рис. 57).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between CAT0 and CAT1 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Рис. 57. Результат виконання тесту

Отримана в наведеному прикладі величина $p = 0,000$ свідчить про наявність відмінностей на високому рівні статистичної значущості ($p < 0,001$).

Самостійна робота

1. Порівняти показники витривалості групи велосипедистів до тренувань x_i і після y_i , км. Оцінити ефективність тренувального процесу. Зробити висновок.

№	x_i	y_i	№	x_i	y_i
1	327,4	330,5	7	395	395
2	330,5	352,4	8	398,5	402
3	350,2	360,7	9	399,5	385
4	338,1	344,4	10	399,9	402
5	360,9	358,2	11	405	408,2
6	381,4	390,5	12	410	411,5

2. Перевірити ефективність фізіотерапевтичної програми для чоловіків з посттравматичними нейропатіями малогомілкового нерву на довготривалому періоді реабілітації. Ефективність нової програми оцінювалася за показником аналізу м'язової сили за мануальним м'язовим тестуванням м'язів нижньої кінцівки (бали). В основній групі (ОГ) та контрольній групі (КГ) по 10 досліджуваних. Зробити висновок.

Великогомілковий м'яз (тильне згинання стопи), бали		Малогомілкові м'язи (єверсія стопи), бали		М'язи розгиначі пальців (розгинання пальців), бали		Задній великогомілковий м'яз (інверсія стопи), бали	
ОГ	КГ	ОГ	КГ	ОГ	КГ	ОГ	КГ
2	1	2	2	2	1	3	2
2	2	2	2	2	2	3	3
2	2	3	2	2	2	3	3
2	2	3	2	2	2	3	2
2	2	2	2	2	2	3	3
2	2	2	2	2	2	3	3
2	2	2	2	2	2	3	3
2	2	2	2	2	2	2	3
3	2	2	2	3	2	4	3
2	2	2	2	2	2	3	3

3. Досліджено ефективність 10-денного курсу фізичної терапії для зниження болю у пацієнтів із поперековим синдромом. Показники болю вимірюються за візуально-аналоговою шкалою (ВАШ) до і після терапії. Оцінити ефективність курсу фізичної терапії.

Пацієнт	Біль до (ВАШ)	Біль після (ВАШ)	Пацієнт	Біль до (ВАШ)	Біль після (ВАШ)
1	7	5	6	7	4
2	6	4	7	4	3
3	8	8	8	5	2
4	5	3	9	6	5
5	6	6	10	8	6

Контрольні питання

1. Які статистичні критерії називаються непараметричними?
2. Які непараметричні критерії Вам відомі?
3. Чи можна використовувати критерій Вілкоксона для вибірок з різною кількістю елементів?
4. За якої умови різниця між вибірками є статистично значущою при розрахунку критерію Манна-Уїтні?

Практична робота № 9

Тема «Кореляційний аналіз із використанням параметричних методів у програмах MS Excel та SPSS»

Мета заняття: ознайомитися із методами визначення взаємозв'язку між ознаками та способами розрахунку в програмах MS Excel та SPSS.

Зміст і хід заняття

1. Виявлення взаємозв'язків.

2. Кореляція.

Виконання та захист практичного завдання.

1. Виявлення взаємозв'язків.

У різних сферах людської діяльності з метою внесення позитивних змін, для удосконалення існуючих технологій та процесів, проводять аналіз зв'язків між досліджуваними об'єктами, вивчають закономірності та фактори, які на них впливають.

Існують два види прояву зв'язків між ознаками (явищами) – функціональні та кореляційні.

Для функціональних залежностей характерною є відповідність кожного значення змінної величини окремому значенню іншої змінної. Такі залежності спостерігаються в математиці, фізиці.

У кореляційному зв'язку окремому значенню однієї змінної відповідає певне середнє значення іншої змінної.

Функціональний зв'язок визначається у конкретному спостереженні. Кореляція проявляється в середньому для усієї сукупності спостережень. Кореляційний аналіз допомагає виявити взаємодію факторів (ознак), визначати силу та напрямок впливу одних факторів на інші.

Кореляційна залежність визначається за формою, напрямком та силою зв'язку. Форма зв'язку може бути прямолінійною та криволінійною. Прямолінійний зв'язок визначається тоді, коли за незначних відхилень рівномірні зміни однієї ознаки відповідають рівномірним змінам другої ознаки. У випадку криволінійного зв'язку рівномірним змінам однієї ознаки відповідають нерівномірні зміни іншої ознаки.

Напрямок зв'язку може бути прямий (позитивний) або зворотний (негативний).

Залежність між ознаками є *прямою (позитивною)*, якщо при збільшенні однієї ознаки друга теж збільшується (або при зменшенні однієї ознаки друга теж зменшується). Наприклад, при збільшенні температури зростає тиск газу, а при зменшенні – знижується (при постійному об'ємі).

Зворотною (негативною) називається така залежність, за якої при збільшенні однієї ознаки друга зменшується або при зменшенні однієї ознаки друга збільшується. Прикладом зворотної кореляційної залежності може служити зв'язок між температурою повітря на вулиці та кількістю палива, що витрачається на обігрів приміщення.

2. Кореляція.

Кореляція – це вид взаємозв'язку між ознаками.

Коефіцієнт кореляції для генеральної сукупності не відомий, тому він оцінюється, як правило, за допомогою експериментальних даних вибірки з обсягом n пар значень $(x_i; y_i)$, що отримані в результаті вимірювань двох ознак X та Y . Коефіцієнт кореляції називається вибірковим коефіцієнтом кореляції (або просто коефіцієнтом кореляції). Його позначають символом r .

Вибірковий коефіцієнт лінійної кореляції між двома випадковими величинами X і Y розраховується за формулою:

$$r_{xy} = \frac{\sum_{i=1}^{i=k} (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{i=k} (x_i - \bar{X})^2 \sum_{i=1}^{i=k} (y_i - \bar{Y})^2}}$$

Значення коефіцієнта кореляції знаходяться в інтервалі від -1 до $+1$, тобто $-1 \leq r \leq +1$.

Якщо значення коефіцієнта кореляції $r = 0$, то будь-який зв'язок між досліджуваними явищами відсутній.

Якщо значення коефіцієнта кореляції знаходиться в межах від 0 до $+1$, вважають, що зв'язок між ознаками позитивний (прямий); якщо r знаходиться в межах від -1 до 0 , то зв'язок між ознаками вважають негативним (зворотнім).

Чим сильніший зв'язок між ознаками, тим ближче значення коефіцієнта кореляції наближається до ± 1 .

За умови невеликої кількості пар ознак X та Y ($n < 30$) у практичній діяльності використовують таку інтерпретацію значення коефіцієнта кореляції (інтерпретація американського вченого Ч.Г. Чеддока):

- якщо $0,90 \leq r \leq 0,99$, між ознаками існує **дуже високий** ступінь зв'язку;
- якщо $0,70 \leq r < 0,90$, то між ознаками існує **високий** ступінь зв'язку;
- якщо $0,50 \leq r < 0,70$, то між ознаками існує **значний** ступінь зв'язку;
- якщо $0,30 \leq r < 0,50$, то між ознаками існує **помірний** ступінь зв'язку;
- якщо $0,10 \leq r < 0,30$, між ознаками існує **слабкий** ступінь взаємозв'язку.

При великій кількості спостережень коефіцієнти кореляції необхідно послідовно обчислювати з декількох рядів числових даних. Для зручності одержувані коефіцієнти зводять в таблиці, звані кореляційними матрицями.

Кореляційна матриця – це квадратна (або прямокутна) таблиця, в якій на перетині окремих рядка та стовпця знаходиться коефіцієнт кореляції між відповідними параметрами.

Практичне завдання

Практична робота із розрахунку коефіцієнту кореляції в програмі MS Excel

Завдання 1. У групі гімнасток віком 12–14 років досліджено показники серцево-судинної системи залежно від об'єму тренувального навантаження. Протягом одного тренування виміряно кількість підходів до снарядів (x_i) та середній пульс під час тренування (y_i , уд/хв). Встановити ступінь впливу тренувального навантаження на діяльність серцево-судинної системи цих спортсменок. Зробити висновок. Дані вимірювань представлені у таблиці 18.

Таблиця 18

Вхідні дані

x_i	79	80	82	84	85	87	89	90	92	94
y_i	120	120	123	124	125	127	129	131	135	137

Рішення

1. Введіть експериментальні дані до електронної таблиці.
2. Встановіть курсор у вільну клітинку (наприклад A13).

Викличте функцію **КОРРЕЛ**. Введіть в поле **масив1** діапазон даних A2:A11. В поле **масив2** введіть діапазон даних B2:B11. Натисніть кнопку **ОК**. В клітинці A13 з'явиться значення коефіцієнта кореляції – 0,986821.

Значення коефіцієнта кореляції більше ніж 0,95. Це свідчить про тісний кореляційний зв'язок між кількістю підходів до снаряда та середнім пульсом під час тренування. Отже, спостерігається значний вплив тренувального навантаження на діяльність серцево-судинної системи спортсменок.

За даними вибірок побудуйте графік кореляційного поля.

1. Виділіть діапазон клітинок A2:B11.
2. Виберіть команду **Вставка** → **Діаграми** → **Точкова** .
3. Помістіть діаграму в потрібне місце (бажано, щоб вона не закривала таблицю).

4. Використовуючи **Елементи діаграми** додайте **Лінію тренду**.
5. Для нашого прикладу ми отримаємо наступний графік (рис. 58):

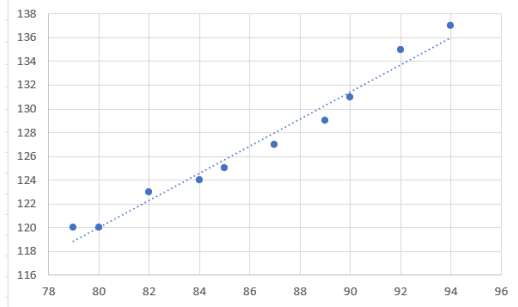


Рис. 58. Графік кореляційного поля

6. Аналіз графіка дозволяє зробити висновок, що в спостережуваній групі гімнасток спостерігається висока ступінь прямого лінійного взаємозв'язку між кількістю підходів до снаряда та середнім пульсом під час тренування. Силу зв'язку можна оцінити за допомогою коефіцієнта кореляції.

Порівняно з величиною коефіцієнта кореляції, графік не несе в собі жодної нової інформації, його особливість – наочність. За допомогою кореляційного поля можна визначити наявність зв'язку між ознаками та його характеристику.

Спосіб оцінки кореляційного поля:

- якщо експериментальні точки розсіяні по полю графіка хаотично, то кореляція відсутня;
- якщо точки групуються вздовж будь-якої уявної лінії, то кореляція є, і вона тим тісніша, чим щільніше розташовані ці точки.
- по напрямку лінії, уздовж якої групуються точки, можна визначити вид кореляції (позитивна чи негативна).

Завдання 2. У таблиці наведено антропометричні дані та результати виконання функціональних проб в групі дітей, що знаходяться на початковому етапі спортивної підготовки. Визначити існуючі взаємозв'язки між досліджуваними ознаками. Обчислити кореляційну матрицю. Зробити висновок. Дані вимірювань представлені у таблиці 19.

Вхідні дані

Довжина тіла	Маса тіла	ОГК	ЖЕЛ	Час затримки дихання на вдику	Час затримки дихання на видиху	Індекс Руф'є
111,0	17,5	56,5	980	7	10	11
115,0	18,8	56,0	1060	10	13	13
112,0	21	57,7	780	10	8	7
108,0	16,6	56,4	860	8	6	15
118,0	19,5	58,0	890	13	9	15
104,0	16,3	54,8	760	6	3	7
110,0	17,7	56,0	1110	14	8	14
112,5	18,5	56,0	760	6	3	7
116,0	19,6	55,0	1090	8	5	14
102,0	15	53,5	870	10	5	14
120,0	23,5	62,0	960	6	8	14
120,0	23,5	55,0	1070	8	6	14
113,0	20	56,0	860	6	8	15
115,0	18,5	63,0	1040	5	3	15
112,5	23	57,0	860	8	6	15
115,0	20	55,0	650	7	7	16
111,0	23,5	56,0	990	6	9	14
116,0	19	53,0	980	18	7	14
118,0	19	57,0	1090	9	9	10

Рішення

3. Введіть експериментальні дані до електронної таблиці.

4. Виберіть команду **Дані** → **Аналіз даних** та оберіть рядок **Кореляція**.

5. У діалоговому вікні вкажіть Вхідний інтервал A1:G20. Вкажіть, що дані згруповані по стовбцях. Вкажіть вихідний діапазон. Для цього поставте прапорець в ліве поле **Вихідний інтервал** і в праве поле введення **Вихідний інтервал** введіть адресу клітинки K1. Натисніть кнопку **ОК**.

У вихідному діапазоні отримуємо кореляційну матрицю (рис. 59).

	Довжина тіла	Маса тіла	ОГК	ЖЕЛ	Час затримки дихання на вдику	Час затримки дихання на видиху	Індекс Руф'є
Довжина тіла	1						
Маса тіла	0,648779	1					
ОГК	0,3823175	0,293102	1				
ЖЕЛ	0,377116	0,124809	0,150948	1			
Час затримки дихання на вдику	0,0820218	-0,18746	-0,39341	0,243123	1		
Час затримки дихання на видиху	0,3083724	0,224131	-0,01225	0,3226	0,293344	1	
Індекс Руф'є	0,2537498	0,216625	0,083231	0,253994	0,137285	0,16782	1

Рис. 59. Результат обчислення кореляційної матриці

Інтерпретація результатів. З таблиці видно, що між довжиною тіла та масою тіла ($r = 0,648779$) спостерігається значний ступінь взаємозв'язку, позитивного характеру. Водночас, між антропометричними даними та результатами виконання функціональних проб взаємозв'язку не виявлено.

Практична робота в програмі SPSS

Завдання 3. На підставі даних дослідження гіпертонії розрахуйте коефіцієнт кореляції за Пірсоном попарно для змінних Хол0, Хол1, Хол6, Хол12 (побудуйте для цих змінних кореляційну матрицю). Таблиця з вхідними даними для практичної роботи представлена в «Практичній роботі 7» таблиця 17.

Рішення

Analyz/Аналіз → Correlate/Кореляції → Bivariate/Парні. Змінні «Хол0», «Хол1», «Хол6», «Хол12» перенесіть по черзі в поле змінних для аналізу. Розрахунок коефіцієнта кореляції за Пірсоном, двостороння перевірка значущості та маркування значимих кореляцій обрано в програмі за замовчуванням. Натисніть **OK** (рис. 60).

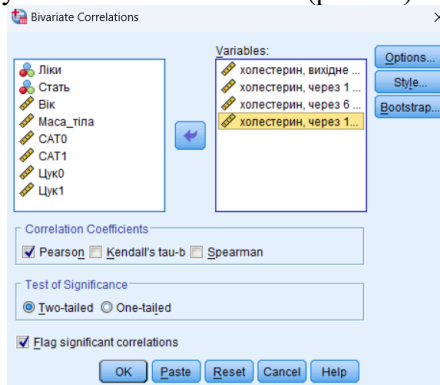


Рис. 60. Діалогове вікно **Bivariate Correlations**

У вікні перегляду з'являться результати, представлені на рис. 61.

Correlations

		холестерин, вихідне значення	холестерин, через 1 місяць	холестерин, через 6 місяців	холестерин, через 12 місяців
холестерин, вихідне значення	Pearson Correlation	1	,965**	,913**	,760**
	Sig. (2-tailed)		,000	,000	,000
	N	20	20	20	20
холестерин, через 1 місяць	Pearson Correlation	,965**	1	,926**	,864**
	Sig. (2-tailed)	,000		,000	,000
	N	20	20	20	20
холестерин, через 6 місяців	Pearson Correlation	,913**	,926**	1	,806**
	Sig. (2-tailed)	,000	,000		,000
	N	20	20	20	20
холестерин, через 12 місяців	Pearson Correlation	,760**	,864**	,806**	1
	Sig. (2-tailed)	,000	,000	,000	
	N	20	20	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

Рис. 61. Кореляційна матриця

Отримані результати містять розрахунок: коефіцієнту кореляції Пірсона r , кількості використаних пар значень змінних та ймовірність похибки p , що відповідає припущенню про ненульову кореляцію. Зробіть висновок про силу кореляції.

Самостійна робота

1. Встановити ступінь залежності між тривалістю паління (роки) та частотою виявлення хронічного бронхіту у молодому віці (до 29 років). Зробити висновок.

Тривалість паління (роки)	3	4	5	6	7	8	9	10
Частота хронічних бронхітів (%)	6,0	9,0	12,0	13,0	14,0	21,0	26,0	35,0

2. У таблиці (ст. 91) наведено антропометричні дані та результати виконання рухових тестів хлопчиків 9-ти років. Встановити, чи існують взаємозв'язки між досліджуваними ознаками. Обчислити кореляційну матрицю. Зробити висновок.

Контрольні питання

1. Дайте визначення поняття «кореляція»?
2. В якому інтервалі можуть знаходитися значення коефіцієнта кореляції?
3. За якої умови значення коефіцієнта кореляції вважають зворотним (негативним)?
4. В якому діапазоні повинно знаходитись значення коефіцієнта кореляції щоб ступінь зв'язку між ознаками був слабким?
5. Чи може значення коефіцієнта кореляції бути більшим за «1»?
6. За яких умов зв'язок між ознаками є прямий (позитивний)?

Обхват, ширина						Довжина тіла				Маса тіла	Швидкість		Координаційні здібності			
голови, см	грудей, см	грудей (види), см	грудей (види), см	плечей, см	таза, см	тіла, см	тіла (сидячі), см	рук, см	ніг, см	маса тіла, кг	біг 20 м, с	піднімання прямих рук в гору за 30 с, разів	проба Ромберга, с	стрибок з поворотом на 360 град., град.	три перекиди вперед, с	човниковий біг 3*10 м, с
55	67	68	65	30	23	125	110	53	62	28	3,7	49	11	360	3,2	8,1
53	67	71	65	28	22	134	113	57	69	28	3,9	38	19	360	3,8	8,6
54	66	66	59	31	20	123	108	53	62	26	4,3	40	60	360	3	9,5
54	69	74	67	31	22	142	116	59	71	33	4,3	40	27	360	3,6	9,6
54	67	62	60	31	22	133	114	55	72	31	3,9	40	10	330	3,1	9,4
55	72	75	70	33	22	133	111	57	69	29	4,3	37	15	340	3,2	11,6
50	63	65	63	24	19	124	110	54	64	24	4,2	42	10	360	3,8	9,9
53	62	67	62	29	21	130	110	57	66	25	3,9	43	27	380	3,2	9,4
53	62	67	61	30	22	132	112	57	68	32	3,8	40	24	420	3,1	9
55	66	68	65	29	23	124	108	53	62	27	4	37	15	380	3,6	9
53	67	71	65	28	22	124	110	54	64	24	4,2	38	18	360	3,5	9,2
54	64	66	59	31	20	130	110	56	66	25	4,3	37	11	360	3,4	8,9
53	69	74	67	31	22	132	112	57	68	32	4,2	40	16	390	3,2	8,6
54	67	62	60	31	22	134	113	57	69	28	4	41	10	420	3,8	9
50	63	65	63	24	19	133	114	55	72	31	4,1	40	27	680	3,4	8,2

Практична робота № 10

Тема «Регресійний аналіз у програмах MS Excel та SPSS»

Мета заняття: ознайомитися із методами визначення взаємозв'язку між ознаками та способами розрахунку в програмах MS Excel та SPSS.

Зміст і хід заняття

1. Основні поняття регресійного аналізу.
 2. Регресійні моделі.
 3. Використання інструменту «Пакет аналізу» для визначення взаємозв'язків між досліджуваними ознаками.
- Виконання та захист практичного завдання.

1. Основні поняття регресійного аналізу.

При дослідженні взаємозв'язків між вибірками окрім кореляції розрізняють також і регресію. *Регресія* використовується для аналізу впливу на окрему залежну змінну значень однієї або більше незалежних змінних. Відповідно, поряд з кореляційним аналізом, ще одним інструментом вивчення стохастичних залежностей є регресійний аналіз. Регресійний аналіз встановлює форми залежності між випадковою величиною Y (залежною) і значеннями однієї або декількох змінних величин (незалежних), причому значення останніх вважаються точно заданими. Така залежність зазвичай визначається деякою математичною моделлю (рівнянням регресії), що містить кілька невідомих параметрів. В ході регресійного аналізу на підставі вибірових даних знаходяться оцінки цих параметрів, визначаються статистичні похибки оцінок або кордону довірчих інтервалів і перевіряється відповідність (адекватність) прийнятої математичної моделі експериментальним даним.

2. Регресійні моделі.

У лінійному регресійному аналізі зв'язок між випадковими величинами передбачається лінійним. У найпростішому випадку в лінійній регресійній моделі є дві змінні X і Y . По n парам спостережень $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ потрібно побудувати (підібрати) пряму лінію (звану лінією регресії), яка найкращим чином наближає спостережувані значення. Рівняння цієї лінії є регресійний рівнянням $Y = bX + b_0$. За допомогою регресійного рівняння можна передбачити очікуване значення залежної величини Y_0 , що відповідає заданому значенню незалежної змінної X_0 .

Таким чином, лінійний регресійний аналіз полягає у підборі графіку і його рівняння для набору спостережень. У регресійному аналізі всі ознаки

(змінні), що входять до рівняння, повинні мати безперервну, а не дискретну природу.

У разі, коли розглядається залежність між однією залежною змінною Y і декількома незалежними змінними $X_1, X_2, \dots, X_n$, говорять про множинну лінійну регресію. В цьому випадку регресійне рівняння має вигляд:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

де $b_1, b_2, \dots, b_n$ – коефіцієнти, що потребують визначення при незалежних змінних $X_1, X_2, \dots, X_n$; b_0 – константа.

Мірою ефективності регресійної моделі є *коефіцієнт детермінації* R^2 (R-квадрат). Коефіцієнт детермінації (R-квадрат) визначає, з яким ступенем точності отримане регресійне рівняння описує (апроксимує) вихідні дані.

Крім того, значущість регресійної моделі можна визначити за допомогою F-критерію (Фішера). Якщо величина F-критерію статистично значуща ($p < 0,05$), то і регресійна модель є значущою. Достовірність відмінності коефіцієнтів $b_1, b_2, \dots, b_n$ від нуля перевіряється за допомогою критерію Стюдента. У випадках, коли $p > 0,05$, коефіцієнт може вважатися нульовим. Це означає, що вплив відповідної незалежної змінної на залежну змінну не є статистично значущим, і ця незалежна змінна може бути виключена з рівняння.

В MS Excel експериментальні дані апроксимуються лінійним рівнянням до 16 порядку:

де Y – залежна змінна, $X_1, \dots, X_{16}$ – незалежні змінні, $b_1, b_2, \dots, b_{16}$ – шукані коефіцієнти регресії.

3. Використання інструменту «Пакет аналізу» для визначення взаємозв'язків між досліджуваними ознаками.

Для розрахунку коефіцієнтів регресії використовується процедура **Регресія** з надбудови **Пакет аналізу**. Крім того, можуть бути використані функції **ЛИНЕЙН** для отримання параметрів регресійного рівняння і **ТЕНДЕНЦИЯ** для отримання передбачених значень Y в необхідних точках.

Для реалізації процедури **Регресія** необхідно:

- виконати команду **Дані → Аналіз даних**;
- в діалоговому вікні **Аналіз даних** в списку **Інструменти аналізу** вибрати рядок **Регресія**;
- в діалоговому вікні задати **Вхідний інтервал Y**, тобто ввести посилання на діапазон даних залежної змінної, що містить один стовпець даних;

- вказати **Вхідний інтервал X**, тобто ввести посилання на діапазон даних незалежної (або незалежних) змінних, що містить до 16 стовпців аналізованих даних;
- вказати вихідний діапазон, тобто ввести посилання на клітинку, в яку будуть виведені результати аналізу (рис. 62);

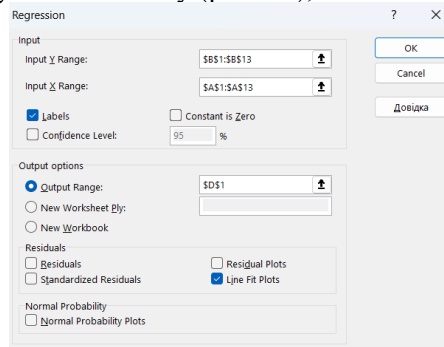


Рис. 62. Приклад заповнення діалогового вікна **Регресія**

- якщо необхідно візуально перевірити відмінність експериментальних точок від передбачених за регресійною моделлю, слід встановити прапорець в полі **Графік підбору**;
- натиснути кнопку **ОК**.

Результати аналізу. Вихідний діапазон буде включати в себе результати дисперсійного аналізу, коефіцієнти регресії, стандартну похибку обчислення $\bar{Y}$, середньоквадратичні відхилення, число спостережень, стандартні похибки для коефіцієнтів.

Інтерпретація результатів. Значення коефіцієнтів регресії знаходяться в стовпці *Коефіцієнти* і відповідають:

- $\bar{Y}$ -перетин – b_0 ;
- змінна X_1 – b_1
- змінна X_2 – b_2 і т. д.

У стовпці **Р-Значення** наводиться достовірність відмінності відповідних коефіцієнтів від нуля. У випадках, коли $p > 0,05$, коефіцієнт може вважатися нульовим; це означає, що відповідна незалежна змінна суттєво не впливає на залежну змінну.

Значення **R-квадрат** (коефіцієнт детермінації) визначає, з яким ступенем точності отримане регресійне рівняння апроксимує вихідні дані. Якщо R-квадрат $> 0,95$, вважають, що точність апроксимації є високою (модель добре описує явище). Якщо R-квадрат знаходиться в діапазоні від 0,8 до 0,95, вважають, що точність апроксимації є задовільною (модель в цілому адекватна відповідно до описуваного явища). Якщо

R -квадрат $< 0,6$, прийнято вважати, що точність апроксимації є недостатньою і модель вимагає поліпшення (введення нових незалежних змінних, обліку нелінійностей тощо).

Практичне завдання

Практична робота з виконання регресійного аналізу в програмі MS Excel

Завдання 1. Визначте взаємозв'язок довжини кроку в бігу (см) від величини споживання кисню ($\text{л}\cdot\text{хв}^{-1}$). На підставі вхідних даних побудуйте регресійне рівняння, що дозволяє по довжині кроку визначати передбачувану величину споживання кисню. Дані вимірювань представлені у таблиці 20.

Таблиця 20

Вхідні дані

Довжини кроку, (см)	132	138	143	141	140	154	159	150	170	135	160	171
Споживання кисню ($\text{л}\cdot\text{хв}^{-1}$)	3,6	3,8	3,9	3,9	4,0	4,1	4,6	4,2	5,0	4,2	4,5	5,1

Рішення

1. Введіть експериментальні дані до електронної таблиці.
2. Виберіть команду **Дані** → **Аналіз даних** → **Регресія**.
3. У діалоговому вікні задайте **Вхідний інтервал Y** – це діапазон клітинок A1:A13 (зверніть увагу, що залежні дані – це ті дані, які передбачається обчислювати).
4. Також вкажіть **Вхідний інтервал X** – задайте діапазон даних незалежної змінної B1:B13 (незалежні дані – це ті дані, які будуть вимірюватися або спостерігатися).
5. Встановіть прапорець в полі **Графік підбору**.
6. Далі вкажіть **Вихідний діапазон**, наприклад, клітинку C1.
7. Натисніть кнопку ОК.

Результати аналізу. У вихідному діапазоні з'являться результати обчислень і графік підбору (рис. 63).

Інтерпретація результатів. За даними таблиці **Дисперсійний аналіз** та **Регресійна статистика** оцінюється загальна якість отриманої моделі. Для ствердження якості моделі рівень значущості критерію Фішера (p) має бути меншим за рівень 0,05. Оцінюємо значущість моделі у прикладі. Рядок *Регресія*, стовпець **Значимість F** – у прикладі $p=0,00001622$; коефіцієнт детермінації в таблиці **Регресійна статистика** – у прикладі **R -квадрат** = 0,856048. Робимо висновок, що модель є статистично значущою і має високу точність апроксимації, тобто модель добре описує спостережуване явище (рис. 63).

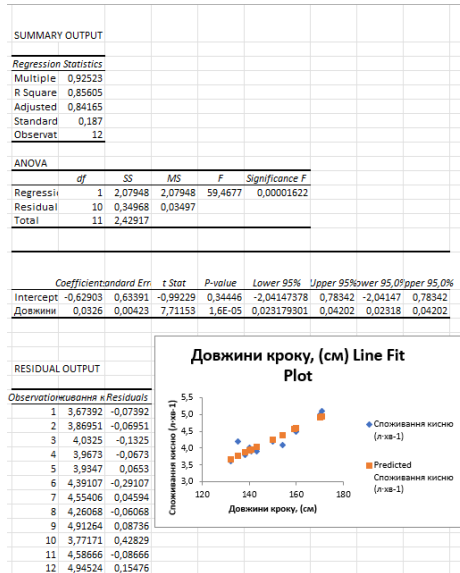


Рис. 63. Результати аналізу і графік відповідності експериментальних даних і тих, що передбачені за регресійною моделлю

Далі необхідно проаналізувати значення коефіцієнтів моделі. Вони представлені у нижній таблиці в стовпці **Коефіцієнти**. У рядку **Y-перетин** наводиться вільний член рівняння; в рядках відповідних змінних наводяться значення коефіцієнтів регресії для цих змінних. У стовпці **p-значення** наводиться достовірність відмінності відповідних коефіцієнтів від нуля. У випадках, коли $p > 0,05$, коефіцієнт може вважатися нульовим. Це означає, що відповідна незалежна змінна практично не впливає на залежну змінну і коефіцієнт може бути прибраний з рівняння.

Отже рівняння для визначення часу бігу на 60 м буде мати наступний вигляд: $\hat{y} = -0,62 + 0,03x$, де x – довжина кроку (см).

Отримана модель в цілому з високою точністю визначає передбачуване споживання кисню ($R^2 = 0,86$).

Скориставшись отриманим рівнянням, можна розрахувати очікувану величину споживання кисню при змінах величини довжини кроку. Наприклад, при довжині кроку 165 см очікуване споживання кисню дорівнює 4,7 (л·хв⁻¹).

Завдання 2. Побудувати регресійну модель для передбачення величини окружності грудної клітини (Y) в залежності від довжини тіла (X_1) і маси тіла (X_2). Дані вимірювань представлені у таблиці 21.

Вхідні дані

№	Довжина тіла	Маса тіла	ОГК	№	Довжина тіла	Маса тіла	ОГК
1	147	40	75	11	141	33,9	64
2	147	36	70	12	135	31,2	64
3	148	66,2	86	13	137	36,6	66
4	137	29,4	58	14	145	58	90
5	125	25	64	15	142,5	31,7	62
6	145	35	67	16	151	43,1	71
7	163	58,5	83	17	142	35	68
8	138	27	62	18	135	33	67
9	151	43,1	71	19	135	50,4	83
10	149	40	68	20	129	29,2	66

Рішення

1. Введіть вхідні дані у діапазон A1:C21 робочої таблиці Excel.
2. Виберіть команду **Дані** → **Аналіз даних** → **Регресія**.
3. У діалоговому вікні задайте **Вхідний інтервал Y** – це діапазон клітинок C1:C21 – дані залежної змінної.
4. Також вкажіть **Вхідний інтервал X** – це діапазон даних незалежної змінної A1:B21.
5. Встановіть прапорець в полі **Графік підбору**.
6. Вкажіть **Вихідний діапазон** (наприклад, клітинку E1). Натисніть кнопку **ОК**.
7. У вихідному діапазоні з'являться результати регресійного аналізу та графіки підбору (рис. 64, 65).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Довжина тіла	Маса тіла	ОГК		SUMMARY OUTPUT								
2	147	40	75										
3	147	36	70										
4	148	66,2	86		Regression Statistics								
5	137	29,4	58		Multiple R	0,946113644							
6	125	25	64		R Square	0,895131027							
7	145	35	67		Adjusted R Square	0,882793501							
8	163	58,5	83		Standard Error	2,998752142							
9	138	27	62		Observations	20							
10	151	43,1	71										
11	149	40	68		ANOVA								
12	141	33,9	64			df	SS	MS	F	Significance F			
13	135	31,2	64		Regression	2	1304,877255	652,4386	72,55353	0,000000005			
14	137	36,6	66		Residual	17	152,872745	8,992514					
15	145	58	90		Total	19	1457,75						
16	142,5	31,7	62										
17	151	43,1	71			Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
18	142	35	68		Intercept	66,50632897	12,9007598	5,155226	7,93E-05	39,28810498	93,72455296	39,28810498	93,72455296
19	135	33	67		Довжина тіла	-0,201269152	0,103019946	-1,95369	0,067395	-0,418622239	0,016083935	-0,418622239	0,016083935
20	135	50,4	83		Маса тіла	0,82702414	0,079505306	10,40213	8,66E-09	0,659282608	0,994765673	0,659282608	0,994765673
21	129	29,2	66										

Рис. 64. Результати регресійного аналізу

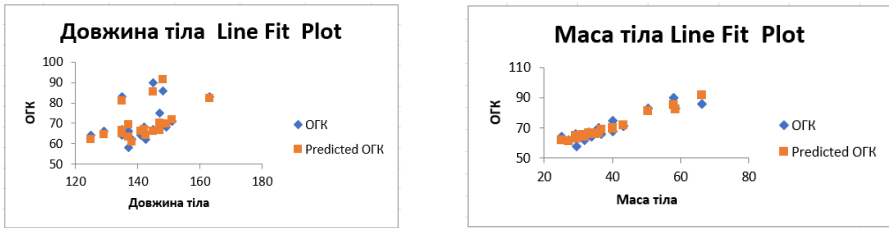


Рис. 65. Графіки розташування фактичних і передбачених точок

Інтерпретація результатів. У таблиці **Дисперсійний аналіз** оцінюється достовірність отриманої моделі за рівнем значущості критерію Фішера (рядок **Регресія**, стовпець **Значимість F**, в прикладі – 0,000000005; оскільки $p < 0,05$, то модель є значущою) і ступінь точності, з якою модель описує процес або явище – **R-квадрат** (другий рядок зверху в таблиці **Регресійна статистика**, в прикладі R-квадрат = 0,89). Отже, робимо висновок про високу точність апроксимації моделі, тобто модель добре описує залежність окружності грудної клітини від довжини тіла і маси тіла.

Далі необхідно визначити значення коефіцієнтів регресійного рівняння. Вони визначаються за даними з нижньої таблиці в стовпці **Коефіцієнти** – в рядку **Y-перетин** наведено вільний член рівняння $b_0 = 66,5$; в рядках відповідних змінних наведено значення коефіцієнтів при цих змінних $b_1 = -0,20$ і $b_2 = 0,82$. У стовпці **p-значення** наведено величина достовірності відмінності відповідних коефіцієнтів від нуля. Оскільки $p < 0,05$, то всі коефіцієнти рівняння є значущими, отже вони можуть вважатися не рівними нулю. Тому рівняння для визначення величини окружності грудної клітини в залежності від довжини тіла і маси тіла матиме вигляд:

$$\hat{y} = 66,5 - 0,20x_1 + 0,82x_2$$

де y – окружність грудної клітини, x_1 – довжина тіла, x_2 – маса тіла.

Практична робота в програмі SPSS

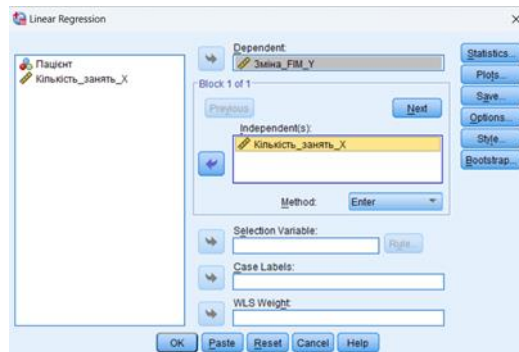
Завдання 2. Дослідити, як кількість занять з фізичної терапії впливає на покращення функціонального стану пацієнтів з ураженням опорно-рухового апарату (оцінка за шкалою FIM – Functional Independence Measure). Дані вимірювань представлені у таблиці 22.

Вхідні дані

Пацієнт	Кількість занять (X)	Зміна FIM (Y)	Пацієнт	Кількість занять (X)	Зміна FIM (Y)
1	5	4	9	13	9
2	6	4	10	14	9
3	7	5	11	15	10
4	8	6	12	16	11
5	9	6	13	17	11
6	10	7	14	18	12
7	11	7	15	20	13
8	12	8			

Рішення

Analyz/Аналіз → Regression/Регресія → Linear/Лінійна → з'явиться діалогове вікно **Linear Regression/Лінійна регресія** → перенесіть змінну «Зміна_FIM_Y» у поле для залежних змінних (**Dependent**) і надайте змінній «Кількість_занять_X» статус незалежної змінної (**Independent(s)**) → **ОК** (рис. 66).

Рис. 66. Діалогове вікно **Linear Regression**

Висновок основних результатів виглядає так (рис. 67–69):

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,995 ^a	,990	,990	,29189

a. Predictors: (Constant), Кількість_занять_X

Рис. 67. Загальні відомості про модель

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	114,626	1	114,626	1345,379	,000 ^b
	Residual	1,108	13	,085		
	Total	115,733	14			

a. Dependent Variable: Зміна_FIM_Y

b. Predictors: (Constant), Кількість_занять_X

Рис. 68. Результат оцінки гомогенності дисперсій

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,611	,218		2,795	,015
	Кількість_занять_X	,623	,017	,995	36,679	,000

a. Dependent Variable: Зміна_FIM_Y

Рис. 69. Зведена таблиця коефіцієнтів регресії

Проаналізуйте таблиці. Тут виводяться коефіцієнт регресії b_1 і зміщення по осі координат b_0 під іменем «константа». Таким чином, рівняння регресії виглядає так: $\text{Зміна_FIM_Y} = 0,611 + 0,623 * \text{Кількість_занять_X}$.

Отже, після 22 занять очікуване покращення становитиме близько 14 балів за шкалою FIM.

Таблиця також відображає два джерела дисперсії: дисперсію, яка описується рівнянням регресії (сума квадратів, обумовлена регресією) та дисперсію, яка не враховується при записі рівняння (залишкова сума квадратів). Частка від суми квадратів, обумовлених регресією та залишкової суми квадратів називається «коефіцієнтом детермінації». У таблиці результатів це частка виводиться під ім'ям «R-квадрат». У цьому прикладі міра визначеності дорівнює: $114,626 / 115,733 = 0,990$.

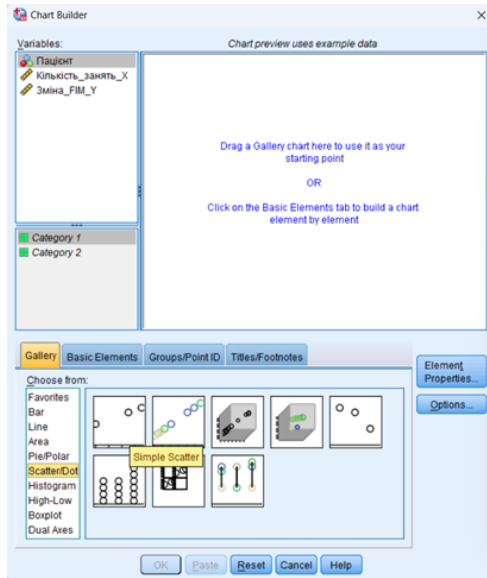
Ця величина характеризує якість регресійної прямої, тобто ступінь відповідності між регресійною моделлю та вихідними даними. Міра визначеності завжди лежить у діапазоні від 0 до 1.

Зобразить на діаграмі розсіювання регресійну пряму.

Рішення

1. **Graphs/Графіка** → **Chart Builder/Майстер діаграм** відкриється вікно **Майстер діаграм**.

2. Виберіть вид діаграми **Scatter/Dot/Розсіювання/Точки** (рис. 70).

Рис. 70. Діалогове вікно **Chart Builder**

3. Перетягніть **Simple Scatter/Просту** діаграму розсіювання у велике вікно.

4. Перенесіть змінну «Зміна_FIM_Y» у поле **осі Y**, а змінну «Кількість_занять_X» у полі **осі X**. Підтвердьте клацанням на **ОК**. У вікні перегляду результатів з'явиться діаграма розсіювання (рис. 71).

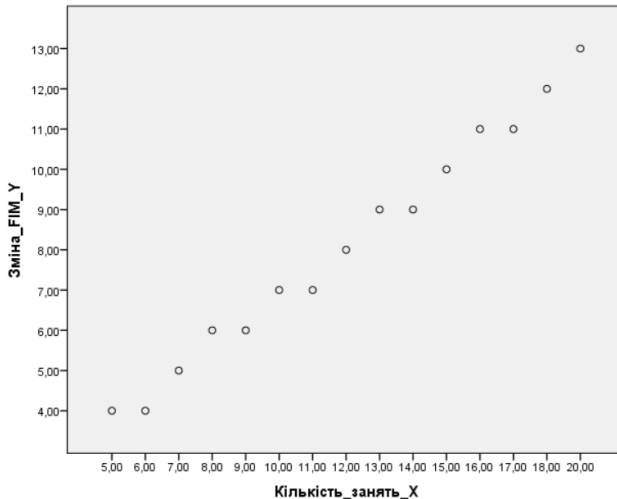


Рис. 71. Діаграма розсіювання у вікні перегляду

5. Клацніть двічі на цьому графіку, щоб перенести його до редактора діаграм.

6. Виберіть у редакторі діаграм меню **Elements/Елементи** → **Fit Line at Total/Лінія апроксимації для підсумку** відкриється діалогове вікно **Properties/Властивості** відмітьте, як показано на рисунку 72.

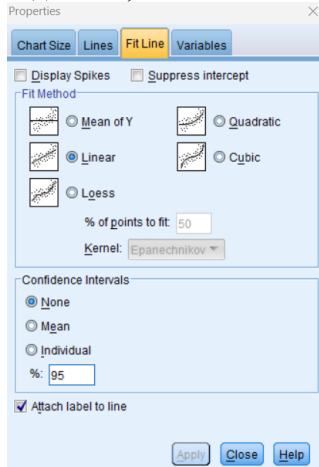


Рис. 72. Діалогове вікно **Properties**

7. Натисніть кнопку **Close/Закрити**.

8. Закрийте редактор діаграм і клацніть один раз де-небудь поза графіком. Тепер у діаграмі розсіювання відображається регресійна пряма (рис. 73).

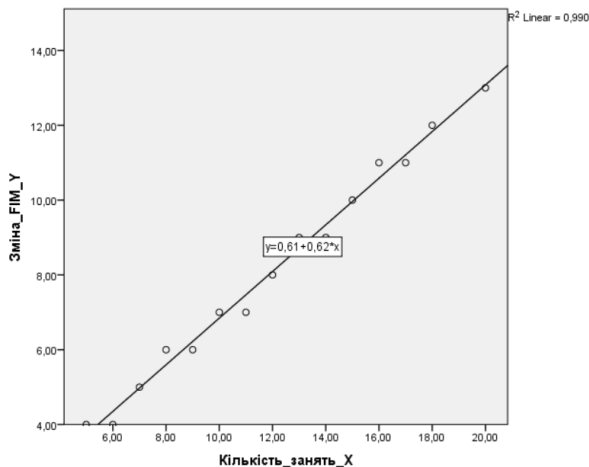


Рис. 73. Діаграма розсіювання з регресійною прямою

Самостійна робота

1. Визначте взаємозв'язок між температурою повітря та кількістю випадків захворювання на бронхіт. На підставі вхідних даних побудуйте регресійне рівняння, що дозволяє по результату температури повітря визначати передбачувану кількість захворювань на бронхіт.

Температурою повітря, °С	К-ть випадків захворювання на бронхіт, абс.
2	253
3	240
4	242
5	188
6	164
7	130
8	87
9	90
10	64

2. Побудувати регресійну модель для передбачення величини життєвої ємності легень спортсмена (Y) в залежності від довжини тіла (X_1) і віку (X_2). У таблиці наведено дані 10 спортсменів.

X_1	X_2	Y
1,85	18	5,4
1,80	25	6,6
1,75	20	4,8
1,70	24	5,1
1,68	21	4,5
1,73	19	4,8
1,77	22	5,1
1,81	23	5,6
1,76	18	4,7
1,85	19	5,2

Контрольні питання

1. В чому полягає сутність регресійного аналізу?
2. Чим визначається кількість незалежних змінних в регресійній моделі?
3. Що визначає коефіцієнт детермінації у регресійній моделі?
4. Як визначити адекватність регресійної моделі?

ТЕСТОВІ ЗАВДАННЯ**Тестові завдання до Практичної роботи №1****1. Що потрібно для створення Google-таблиць?**

- а. Лише комп'ютер
- б. Програма для малювання
- в. Інтернет-з'єднання та обліковий запис Google
- г. Спеціальний принтер

4. Що таке Google-таблиці?

- а. Програма для написання текстів
- б. Програма для створення презентацій
- в. Онлайн-інструмент для роботи з електронними таблицями
- г. Інструмент для редагування зображень

3. Який тип файлу не підтримується для імпорту в Google-таблицях?

- а. .xls
- б. .csv
- в. .doc
- г. .ods

4. Яка функція використовується для логічної перевірки умови в Google-таблицях?

- а. CHECK()
- б. IF()
- в. TEST()
- г. VALIDATE()

5. Яка перевага Google-таблиць у порівнянні з Microsoft Excel?

- а. Відсутність інтернет-доступу
- б. Можливість одночасного редагування декількома користувачами
- в. Робота лише в офлайн-режимі
- г. Вбудоване шифрування таблиць

6. Яка формула використовується для автоматичного перекладу тексту в Google Таблицях?

- а. =TRANSLATE()
- б. =GOOGLETRANSLATE()
- в. =TEXTCONVERT()
- г. =LANGUAGE()

7. Для чого використовується фільтр у Google-таблицях?

- а. Для шифрування документів
- б. Для відображення лише вибраних даних за певними умовами
- в. Для створення графіків

г. Для експорту документа

8. Що таке зведена таблиця в Google-таблицях?

- а. Таблиця з формулами
- б. Інструмент для об'єднання кількох документів
- в. Інструмент для агрегування й аналізу даних
- г. Тип діаграми

9. Який тип діаграми найчастіше використовується для візуалізації розподілу категоріальних даних?

- а. Лінійна діаграма
- б. Кругова діаграма
- в. Точкова діаграма
- г. Гістограма

10. Яку функцію потрібно використати, щоб обчислити середнє значення в стовпці?

- а. =SUM()
- б. =AVERAGE()
- в. =MEDIAN()
- г. =COUNT()

Тестові завдання до Практичної роботи №2

1. Яке головне призначення онлайн-сервісу RawGraphs?

- а. Аналіз великих обсягів тексту
- б. Розробка вебсайтів
- в. Візуалізація структурованих даних
- г. Редагування зображень

2. Який із перелічених форматів даних можна імпортувати до RawGraphs?

- а. .docx
- б. .csv
- в. .mp4
- г. .xlsx

5. Яка з наведених візуалізацій присутня серед стандартних типів діаграм у RawGraphs?

- а. Радіальна діаграма
- б. Організаційна діаграма
- в. Картограма
- г. Стовпчаста 3D-діаграма

4. Яку операцію виконує поле Assign dimensions у RawGraphs?

- а. Перетворення файлів у PDF
- б. Збереження діаграми у хмарі

- в. Прив'язка змінних до візуальних елементів діаграми
- г. Підрахунок кількості клітинок

5. Який формат є найбільш придатним для експорту графіка з RawGraphs для подальшого редагування у векторному редакторі?

- а. .jpg
- б. .xlsx
- в. .svg
- г. .txt

6. Яка особливість RawGraphs виділяє його серед інших інструментів візуалізації?

- а. Працює тільки на мобільних пристроях
- б. Не підтримує завантаження даних
- в. Працює повністю в браузері без потреби встановлення
- г. Потребує підключення до бази даних

7. Для чого використовується карта зв'язків (Alluvial Diagram) у RawGraphs?

- а. Показу географічних зв'язків
- б. Візуалізації змін між категоріями або класифікаціями
- в. Створення календаря подій
- г. Аналізу часових рядів

8. Який крок обов'язково потрібно виконати перед створенням графіка в RawGraphs?

- а. Встановити пароль до файлу
- б. Завантажити базу даних з MySQL
- в. Імпортувати або вставити вихідні дані
- г. Додати зображення до проекту

9. Який тип даних найкраще підходить для візуалізації у вигляді хордової діаграми (Chord Diagram)?

- а. Таблиця з двома змінними: X і Y
- б. Дані про зв'язки між групами або категоріями
- в. Хронологічний список подій
- г. Форма запиту користувача

10. Що робить RawGraphs із даними після завершення сесії роботи?

- а. Зберігає дані у хмарному сховищі
- б. Перенаправляє дані на сервер Google
- в. Не зберігає дані та не передає їх на сторонні сервери
- г. Архівує дані автоматично

Тестові завдання до Практичної роботи №3

1. Яка вкладка в MS Excel використовується для доступу до функцій сортування та фільтрації?

- а. Вставлення
- б. Дані
- в. Формули
- г. Головна

2. Яка команда Excel дозволяє відобразити лише ті рядки таблиці, що відповідають певним умовам?

- а. Сортування
- б. Перевірка даних
- в. Фільтр
- г. Умовне форматування

3. Що відбувається з даними, які не відповідають умовам фільтрації?

- а. Вони видаляються з аркуша
- б. Вони копіюються до іншого листа
- в. Вони тимчасово приховуються
- г. Вони змінюють свій формат

4. Що потрібно зробити, щоб відсортувати таблицю за зростанням значень у стовпці?

- а. Використати функцію IF
- б. Натиснути Дані → Сортування → За спаданням
- в. Натиснути Дані → Сортування → За зростанням
- г. Виділити рядки і натиснути Delete

5. Який тип фільтрації в Excel дозволяє обрати записи за вказаним діапазоном чисел?

- а. Автофільтр
- б. Фільтр за кольором
- в. Умовний числовий фільтр
- г. Алфавітне сортування

6. Що дозволяє зробити функція «Сортувати за кольором клітинки»?

- а. Упорядкувати дані за кольором шрифту або фону
- б. Перевірити орфографію в кольорових клітинках
- в. Порахувати кількість клітинок певного кольору
- г. Вставити зображення

7. Як активувати автоматичний фільтр у таблиці?

- а. Виділити клітинки → Вставка → Таблиця
- б. Дані → Фільтр

- в. Вигляд → Макет
- г. Формули → Іменовані діапазони
- 8. **Що відбувається при сортуванні таблиці без закріплення заголовка?**
 - а. Excel автоматично видаляє заголовки
 - б. Заголовок зміщується з даними
 - в. Заголовки копіюються вниз
 - г. Заголовки автоматично ігноруються
- 9. **Для чого використовують команду «Розширений фільтр» в Excel?**
 - а. Для створення діаграм
 - б. Для відбору даних за складними умовами
 - в. Для зміни формату числа
 - г. Для додавання нових рядків
- 10. **Яка дія дозволяє відновити повний перелік записів після фільтрації?**
 - а. Натиснути Ctrl + Z
 - б. Видалити фільтр
 - в. Клікнути на фільтр і вибрати «Очистити фільтр»
 - г. Зберегти файл

Тестові завдання до Практичної роботи №4

- 1. **Що таке зведена таблиця в MS Excel?**
 - а. Таблиця, створена вручну для форматування даних
 - б. Б. Інструмент для автоматичного аналізу і підсумовування великих обсягів даних
 - в. Таблиця з умовним форматуванням
 - г. Таблиця, що містить формули з використанням функції IF
- 2. **З якої вкладки MS Excel починається створення зведеної таблиці?**
 - а. Головна
 - б. Дані
 - в. Вставлення
 - г. Рецензування
- 3. **Який елемент зведеної таблиці використовується для групування записів по категоріях?**
 - а. Фільтри
 - б. Значення
 - в. Рядки та стовпці
 - г. Формули

4. Яку дію слід виконати для створення зведеної таблиці?

- а. Виділити дані → Вставлення → Зведена таблиця
- б. Виділити дані → Головна → Копіювати
- в. Вставлення → Графік
- г. Дані → Умовне форматування

5. Яка функція за замовчуванням застосовується до числових даних у зведеній таблиці?

- а. Підрахунок
- б. Сума
- в. Середнє
- г. Максимум

6. Що дозволяє зробити фільтр у зведеній таблиці?

- а. Додати формули
- б. Відобразити лише вибрані категорії даних
- в. Об'єднати комірки
- г. Вставити діаграму

7. Який об'єкт зведеної таблиці дозволяє змінити вид обчислень (сума, середнє, максимум тощо)?

- а. Рядки
- б. Значення
- в. Стовпці
- г. Заголовки

8. Як називається елемент, що дозволяє швидко фільтрувати дані в зведеній таблиці за окремими полями?

- а. Список
- б. Спарклайн
- в. Роздільник
- г. Випадаючий список

9. Яке розширення має файл Excel, що містить зведені таблиці, при збереженні?

- а. .docx
- б. .xlsx
- в. .pptx
- г. .pdf

10. Що відбудеться, якщо змінити дані у джерелі, на основі якого створена зведена таблиця?

- а. Дані автоматично оновляться
- б. Зведена таблиця видалиться
- в. Потрібно вручну оновити таблицю
- г. Дані не можна змінити після створення

Тестові завдання до Практичної роботи №5**1. У якому вікні SPSS вводяться вихідні дані?**

- а. Variable View
- б. Output Viewer
- в. Data View
- г. Syntax Editor

2. Яка вкладка використовується для визначення змінних у SPSS?

- а. Output
- б. Syntax
- в. Variable View
- г. Analyze

3. Що з наведеного не є обов'язковим при створенні змінної в SPSS?

- а. Ім'я змінної
- б. Тип змінної
- в. Категорія змінної
- г. Довжина змінної

4. Який тип змінної слід обрати для вводу текстових значень у SPSS?

- а. Numeric
- б. String
- в. Date
- г. Currency

5. Що таке Labels у вікні Variable View?

- а. Значення, які вводяться у таблицю
- б. Формули для обчислень
- в. Повні назви змінних, які пояснюють їх зміст
- г. Кодові значення змінних

6. Для чого в SPSS використовується колонка Values у Variable View?

- а. Для введення чисел
- б. Для задання форматування
- в. Для задання числових кодів категоріальним змінним
- г. Для обчислення статистик

7. У якому вікні SPSS формується таблиця частотності або перехресного табулювання?

- а. Data View
- б. Variable View
- в. Output Viewer

г. Chart Builder

8. Яке розширення має файл даних, створений у SPSS?

а. .sav

б. .sps

в. .xlsx

г. .docx

9. Що відбувається при натисканні команди Descriptive Statistics → Frequencies у SPSS?

а. Видаляються пропущені значення

б. Створюється зведена діаграма

в. Формується таблиця частот

г. Створюється нова змінна

10. Яким чином у SPSS можна імпортувати дані з Excel?

а. File → Open → Excel

б. Analyze → Import

в. Edit → Paste

г. View → Table Layout

Тестові завдання до Практичної роботи №6

1. Який інструмент MS Excel використовується для розрахунку описової статистики?

а. Формат клітинок

б. Таблиця

в. Аналіз даних

г. Пошук рішення

2. У якому розділі меню MS Excel знаходиться команда «Аналіз даних» після встановлення надбудови «Пакет аналізу»?

а. Вставлення

б. Дані

в. Формули

г. Головна

3. Яка з наведених статистичних характеристик не входить до стандартного набору описової статистики в MS Excel?

а. Середнє значення

б. Стандартне відхилення

в. Медіана

г. Коефіцієнт варіації

4. Який з варіантів запуску аналізу в SPSS призводить до розрахунку описових характеристик?

- a. File → Open
- б. Analyze → Descriptive Statistics → Descriptives
- в. View → Variables
- г. Graphs → Charts

5. Що з наведеного не є описовим статистичним показником?

- a. Середнє арифметичне
- б. Рангове кореляційне співвідношення
- в. Дисперсія
- г. Мінімум

6. У SPSS для виведення таких характеристик як середнє, мінімум, максимум та стандартне відхилення необхідно обрати:

- a. Analyze → Compare Means
- б. Analyze → Descriptive Statistics → Frequencies
- в. Analyze → Correlate
- г. Analyze → Regression

7. Яке статистичне поняття найкраще характеризує «розсіювання» значень навколо середнього?

- a. Мода
- б. Дисперсія
- в. Частота
- г. Медіана

8. Яке поле необхідно активувати у вікні Descriptives у SPSS, щоб отримати коваріацію?

- a. Save standardized values
- б. Display skewness and kurtosis
- в. Include standard error
- г. Цього не можна зробити через Descriptives

9. Яке з наведених тверджень є вірним щодо медіани?

- a. Це найчастіше значення у вибірці
- б. Це середнє арифметичне значення
- в. Це значення, що ділить вибірку навпіл
- г. Це сума квадратів відхилень

10. У якій формі MS Excel може виводити описову статистику після запуску команди «Аналіз даних» → «Описова статистика»?

- a. У діаграмі
- б. У зведеній таблиці
- в. У новому аркуші або діапазоні клітинок
- г. У надбудові Power Query

Тестові завдання до Практичної роботи №7

1. Для якого закону розподілу випадкової величини справедливе правило трьох сігм (σ)?

- а. Розподіл Стюдента
- б. Бімодальний розподіл
- в. Нормальний розподіл

2. Який параметричний тест використовується для перевірки різниці між середніми двох незалежних вибірок?

- а. χ^2 -критерій
- б. U-критерій Манна–Уїтні
- в. t-критерій Стюдента
- г. Критерій Вілкоксона

3. Оцінка результатів аналізу за критерієм Стюдента: якщо $t \geq t_{\alpha/2}$, то відмінність між порівнюваними вибірками вважається:

- а. Статистично невизначеною
- б. Статистично достовірною
- в. Статистично недостовірною

4. У MS Excel для виконання t-тесту між двома вибірками слід скористатись:

- а. Вставка → Функція → T.TEST
- б. Дані → Аналіз даних → t-тест
- в. Формули → Статистичні → T.TEST
- г. Будь-який з перелічених варіантів

5. Статистична сукупність вважається великою, якщо її об'єм:

- а. Більше 20 елементів
- б. Більше 30 елементів
- в. Більше 40 елементів
- г. Більше 50 елементів

6. Яка функція в MS Excel використовується для розрахунку двостороннього t-тесту?

- а. TINV
- б. TTEST
- в. TDIST
- г. T.TEST

7. Яка з наведених гіпотез є нульовою при t-тесті для двох незалежних вибірок?

- а. Середні значення в групах однакові
- б. Вибірki мають різний розподіл
- в. Вибірki пов'язані між собою
- г. Середнє значення перевищує критичне

8. Якщо р-значення, отримане в результаті t-тесту, становить 0,03, а рівень значущості $\alpha = 0,05$, то:

- а. Нульову гіпотезу слід прийняти
- б. Нульову гіпотезу слід відхилити
- в. Дані некоректні
- г. Тест неможливо виконати

9. У SPSS результат t-тесту відображається у вигляді:

- а. Діаграми
- б. Таблиці середніх значень
- в. Таблиці Levene's Test і Test for Equality of Means
- г. Таблиці розподілу частот

10. Група даних, які належать до однієї ознаки, називається:

- а. Статистична сукупність
- б. Дискриптивна сукупність
- в. Нормальна сукупність

Тестові завдання до Практичної роботи №8

1. Яка з наведених ознак є характерною для непараметричних методів?

- а. Потреба у нормальному розподілі даних
- б. Високі вимоги до масштабів вимірювання
- в. Можливість аналізу рангових даних
- г. Застосування лише для інтервальних змінних

2. Який з наведених методів є непараметричним аналогом t-критерію для незалежних вибірок?

- а. Критерій Вілкоксона
- б. U-критерій Манна-Уїтні
- в. Критерій Фішера
- г. ANOVA

3. Який тест доцільно застосовувати для аналізу зв'язку між двома номінальними змінними?

- а. χ^2 -критерій незалежності
- б. U-критерій
- в. Критерій знаків
- г. t-критерій Стьюдента

4. Для перевірки гіпотези щодо відмінностей у пов'язаних вибірках з порядковими змінними використовується:

- а. U-критерій Манна-Уїтні
- б. t-критерій для залежних вибірок
- в. Критерій Вілкоксона

г. χ^2 -критерій

5. Коли доцільно застосовувати непараметричні методи?

а. Дані мають нормальний розподіл

б. Розмір вибірки дуже великий

в. Рівень вимірювання – порядковий або номінальний

г. Відоме стандартне відхилення

6. Який тест можна використовувати для перевірки гіпотези про розподіл одного порядкового показника у двох незалежних вибірках?

а. t-тест

б. U-критерій Манна–Уїтні

в. χ^2 -критерій

г. Тест Колмогорова–Смірнова

7. Критерій знаків застосовується:

а. Для незалежних вибірок

б. Для порівняння середніх у більше ніж двох групах

в. Для двох пов'язаних вибірок за медіаною

г. Для перевірки зв'язку між двома номінальними змінними

8. Який з наведених тестів перевіряє однорідність розподілів у двох незалежних вибірках?

а. Критерій знаків

б. Критерій Колмогорова–Смірнова

в. t-критерій

г. Критерій Вілкоксона

9. У якому випадку χ^2 -критерій непридатний?

а. При обробці порядкових даних

б. Коли в таблиці спостережень очікувані частоти менші за 5

в. Коли досліджується залежність між змінними

г. При аналізі розподілу ймовірностей

10. Що є нульовою гіпотезою для U-критерію Манна–Уїтні?

а. Групи мають різну дисперсію

б. Групи мають однаковий розподіл

в. Середні значення в групах однакові

г. Групи незалежні

Тестові завдання до Практичної роботи №9

1. Який коефіцієнт кореляції є параметричним і використовується для кількісних змінних з нормальним розподілом?

а. Коефіцієнт Спірмена

б. Коефіцієнт Пірсона

в. Коефіцієнт Кендалла

г. χ^2 -критерій

2. У якій вкладці в SPSS можна знайти інструмент для розрахунку коефіцієнта кореляції Пірсона?

- а. Analyze → Descriptive Statistics
- б. Analyze → Correlate → Bivariate
- в. Analyze → Compare Means
- г. Transform → Compute Variable

3. Яка величина обчислюється за формулою:

$$r_{xy} = \frac{\sum_{i=1}^{i=k} (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{i=k} (x_i - \bar{X})^2 \sum_{i=1}^{i=k} (y_i - \bar{Y})^2}}$$

- а. Розмах варіації
- б. Коефіцієнт кореляції
- в. Помилка репрезентативності
- г. Коефіцієнт варіації

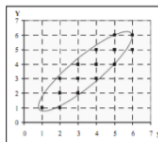
4. Для обчислення коефіцієнту кореляції необхідно обрати категорію вбудованих функцій Excel:

- а. Математичні
- б. Логічні
- в. Статистичні
- г. Текстові

5. Яке з наступного найкраще інтерпретує значення $r = 0,95$?

- а. Слабкий позитивний зв'язок
- б. Сильний позитивний зв'язок
- в. Відсутність зв'язку
- г. Негативний зв'язок

6. Яка форма статистичної залежності представлена на кореляційному полі:



- а. Строго позитивна кореляція
- б. Сильна позитивна кореляція
- в. Нульова кореляція
- г. Помірна негативна кореляція

7. Яке значення кореляції Пірсона свідчить про повну відсутність зв'язку між змінними?

- а. 1

- б. -1
- в. 0
- г. 0,5

8. У якому діапазоні можуть лежати значення коефіцієнта кореляції Пірсона?

- а. Від -100 до 100
- б. Від -10 до 10
- в. Від -1 до 1
- г. Від 0 до 1

9. У SPSS для того щоб отримати матрицю кореляцій з рівнями значущості, слід:

- а. Вибрати Correlate → Partial
- б. У вікні Bivariate Correlations позначити «Pearson» і «Two-tailed»
- в. Використати Analyze → Regression
- г. Створити діаграму розсіювання

10. Який з наведених висновків коректний, якщо значення $p < 0,05$ при обчисленні кореляції Пірсона?

- а. Вибірка замала
- б. Кореляція статистично незначуща
- в. Існує статистично значущий зв'язок між змінними
- г. Змінні незалежні

Тестові завдання до Практичної роботи №10

1. Яку функцію в MS Excel використовують для виконання лінійного регресійного аналізу?

- а. =LINEST()
- б. =REGRESS()
- в. =TRENDLINE()
- г. =CORREL()

2. У SPSS лінійну регресію можна здійснити за допомогою команди:

- а. Analyze → Compare Means → Means
- б. Analyze → Correlate → Bivariate
- в. Analyze → Regression → Linear
- г. Analyze → Descriptive Statistics → Explore

3. У лінійній регресійній моделі змінна, яку ми передбачаємо, називається:

- а. Незалежною
- б. Залежною
- в. Побічною
- г. Системною

4. Який з наведених показників у регресії характеризує точність прогнозу?

- а. Коефіцієнт детермінації (R^2)
- б. Коефіцієнт кореляції
- в. Стандартна похибка середнього
- г. р-значення

5. У регресійному аналізі р-значення менше 0,05 свідчить про:

- а. Невірогідність моделі
- б. Відсутність зв'язку
- в. Статистично значущий зв'язок
- г. Сильну автокореляцію

6. Яка умова повинна бути виконана для застосування лінійної регресії?

- а. Наявність лише номінальних змінних
- б. Незалежні змінні мають бути корельовані
- в. Лінійний зв'язок між змінними
- г. Залежна змінна повинна мати бінарний формат

7. До якого порядку в MS Excel експериментальні дані апроксимуються лінійним рівнянням:

- а. до 10
- б. до 16
- в. до 18
- г. до 20

8. Яке з наведених тверджень є вірним щодо коефіцієнта b у рівнянні лінійної регресії $Y = a + bX$?

- а. b – це точка перетину з віссю Y
- б. b – це похибка моделі
- в. b – це зміна Y при зміні X на одиницю
- г. b – це зворотне значення X

9. У MS Excel для побудови графіка регресії слід:

- а. Вставити гістограму
- б. Побудувати точкову діаграму та додати трендову лінію
- в. Створити таблицю частот
- г. Використати функцію =FORECAST

10. Який з наведених показників регресійного аналізу оцінює залишкову дисперсію?

- а. Стандартна похибка оцінки
- б. Коефіцієнт Пірсона
- в. F-критерій
- г. Коваріація

Список рекомендованої літератури

1. Ахметов Р. Ф. Спортивна метрологія : навчальний посібник. Житомир: Вид-во ФОП Євенок О. О., 2017. 176 с.
2. Ашанін В. С., Жерновнікова Я. В., Пятисоцька С. С. Комп'ютерна обробка даних експериментальних досліджень : навч. посіб. Харків : ХДАФК, 2024. 116 с.
3. Ашанін В. С., Пятисоцька С. С., Жерновнікова Я. В. Статистична обробка та аналіз інформації [навчальний посібник для здобувачів вищої освіти спеціальності 014 Середня освіта (Фізична культура)]. Харків : ХДАФК, 2022. 112 с.
4. Ашанін В. С., Пятисоцька С. С., Жерновнікова Я. В., Петренко Ю. І. Системно-інформаційні основи наукових досліджень у фізичній культурі та спорті : методичні рекомендації для здобувачів другого магістерського рівня вищої освіти спеціальності 017 «Фізична культура і спорт». Харків : ХДАФК, 2024. 144 с.
5. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для студентів. Запоріжжя : КПУ, 2011. 268 с.
6. Василенко О. А., Сенча І. А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. Одеса : ОНАЗ ім. О. С. Попова, 2011. 166 с.
7. Горбачик А. П., Сальнікова С. А. Аналіз даних соціологічних досліджень засобами SPSS: Навч. посіб. Луцьк: РВВ «Вежа» Волин, нац. ун-ту ім. Лесі Українки. 2008. 64 с.
8. Грицюк П. М., Остапчук О. П. Аналіз даних : навчальний посібник. Рівне : НУВГП, 2008. 218 с.
9. Гусак О. М., Гусак В. В. Сучасні інформаційні технології та медична статистика : навч.- метод. посіб. Чернівці : Чернівець. нац. ун-т ім. Ю. Федьковича, 2022. 160 с.
10. Донченко В. С., Сидоров М. В. Теорія ймовірностей та математична статистика для соціальних наук : навч. посіб. К. : ВПЦ «Київський університет», 2015. 400 с.
11. Жерновнікова Я. В. Підготовка майбутніх психологів до вирішення дослідницьких завдань за допомогою регресійного аналізу. *Науковий часопис НПУ імені М. П. Драгоманова. Сер. 15 : Науково-педагогічні проблеми фізичної культури (фізична культура і спорт) : зб. наук. пр. / за ред. О. В. Тимошенка*. 2023. Вип. 8 (168)23. С. 62–65.
12. Жерновнікова Я. В., Долгополова Н. В. Використання можливостей онлайн-сервісу RawGraphs для візуалізації даних при вивченні курсу «Комп'ютерна обробка даних експериментальних досліджень». *Освіта. Інноватика. Практика*. 2024. Том 12, №7. С. 36–42.

13. Жерновнікова Я. В. Формування змісту курсу «Статистична обробка та аналіз інформації» для майбутніх вчителів фізичної культури. *Proceedings of the scientific and pedagogical internship. Directions of improving the professional training of specialists in physical culture and sports.* Włocławek, 2023. P. 86-89.

14. Жерновнікова Я. В., Алексеєва І. А., Алексєнко Я. В. Використання електронних таблиць Microsoft Excel для обробки статистичних даних в галузі фізичного виховання. *Фізична культура, спорт і здоров'я : стан, проблеми та перспективи* : матеріали XXIII Міжнар. наук.-практ. конф. (м. Харків, 6 груд. 2023 р.). Харків, 2023. С. 177–178.

15. Корда М. М., Кашуба М. О. Основи медичної статистики та проведення комп'ютерного статистичного аналізу даних статистичними програмами : навч.-метод. посіб. Ч. 2. Кореляція та регресія. Тернопіль : ТНМУ, 2022. 212 с.

16. Кушлик-Дивульська О. І., Поліщук Н. В., Орел Б. П., Штабалюк П. І. Теорія ймовірностей та математична статистика : навчальний посібник. К. : НТУУ «КПІ», 2014. 212 с.

17. Петренко О. Я. Розумні та зведені таблиці : навчальний посібник. К. : ІПДО, 2023. 43 с.

18. Пятисоцька С. С., Жерновнікова Я. В. Інформаційні технології у фізичній культурі та спорті [методичні рекомендації для студентів I-го курсу спеціальності 017 Фізична культура і спорт]. Харків : ХДАФК, 2022. 52 с.

19. Руденко В. М. Математична статистика : навчальний посібник. К. : Центр учбової літератури, 2012. 304 с.

20. Телейко А. Б., Чорней Р. К. Матиматико-статистичні методи в соціології та психології : навчальний посібник. К. : МАУП, 2007. 424 с.

21. Філатова Г. Є., Толстанов О. К. Застосування математичної статистики в медичних і медико-біологічних наукових дослідженнях. *Клінічна генетика і перинатальна діагностика*. 2013. № 1(2). С. 153–157.

ДОДАТКИ**Додаток 1**

Граничні значення критерію Стьюдента $t_{гр.}$. Надійність $P=0,95$, k – число ступенів свободи

k	$t_{гр.}$	k	$t_{гр.}$
1	12,71	18	2,10
2	4,30	19	2,09
3	3,18	20	2,09
4	2,78	21	2,08
5	2,57	22	2,07
6	2,45	23	2,07
7	2,36	24	2,06
8	2,31	25	2,06
9	2,26	26	2,06
10	2,71	27	2,10
11	2,20	28	2,05
12	2,18	29	2,05
13	2,16	30	2,04
14	2,14	40	2,02
15	2,13	60	2,00
16	2,12	120	1,98
17	2,11	∞	1,96

Додаток 2

Граничні значення критерію Вілкоксона $W_{гр.}$. Надійність $P=0,95$,
 n – кількість парних спостережень

n	6	7	8	9	10	11	12	13	14	15
$W_{гр.}$	1	3	5	7	9	12	15	18	22	26

n	16	17	18	19	20	21	22	23	24	25
$W_{гр.}$	31	36	41	49	53	60	67	74	82	90

Додаток 3

Граничні значення критерію Уайта $W_{гр}$. Надійність $P=0,95$,
 n – кількість парних спостережень

Більша кількість спостережень	Менша кількість спостережень														
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
4			11												
5		6	11	17											
6		7	12	18	26										
7		7	13	20	27	36									
8	3	8	14	21	29	38	49								
9	3	8	15	22	31	40	51	63							
10	3	9	15	23	32	42	53	65	78						
11	4	9	16	24	34	44	55	68	81	96					
12	4	10	17	26	35	46	58	71	85	99	115				
13	4	10	18	27	37	48	60	73	88	103	119	137			
14	4	11	19	28	38	50	63	76	91	106	123	141	160		
15	4	11	20	29	40	52	65	79	94	110	127	145	164	185	
16	4	12	21	31	42	54	67	82	97	114	131	150	169		
17	5	12	21	32	43	56	70	84	100	117	135	154			
18	5	13	22	33	45	58	72	87	103	121	139				
19	5	13	23	34	46	60	74	90	107	124					
20	5	14	24	35	48	62	77	93	110						
21	6	14	25	37	50	64	79	95							
22	6	15	26	38	51	66	82								
23	6	15	27	39	53	68									
24	6	16	28	40	55										
25	6	16	28	42											
26	7	17	29												
27	7	17													

Додаток 4

Границі критичної області критерія знаків $z_{гр}$

n	$z_{гр}$	n	$z_{гр}$	n	$z_{гр}$	n	$z_{гр}$
1	2	3	4	5	6	7	8
5	0-5	34	11-23	63	24-39	92	37-55
6	1-5	35	12-23	64	24-40	93	37-56
7	1-6	36	12-24	65	25-40	94	38-56
8	1-7	37	13-24	66	25-41	95	38-57
9	2-7	38	13-25	67	25-41	96	38-58
10	2-8	39	13-26	68	26-42	97	39-58
11	2-9	40	14-26	69	26-43	98	39-59
12	3-9	41	14-27	70	27-43	99	40-59
13	3-10	42	15-27	71	27-44	100	40-60
14	3-11	43	15-28	72	28-44		

1	2	3	4	5	6	7	8
15	4-11	44	16-28	73	28-45		
16	4-12	45	16-29	74	29-45		
17	5-12	46	16-30	75	29-46		
18	5-13	47	17-30	76	29-47		
19	5-14	48	17-31	77	30-47		
20	6-14	49	18-31	78	30-48		
21	6-15	50	18-32	79	31-48		
22	6-16	51	19-32	80	31-49		
23	7-16	52	19-33	81	32-49		
24	7-17	53	19-34	82	32-50		
25	8-17	54	20-34	83	33-50		
26	8-18	55	20-35	84	33-51		
27	8-19	56	21-35	85	33-52		
28	9-19	57	21-36	86	34-52		
29	9-20	58	22-36	87	34-53		
30	10-20	59	22-37	88	35-53		
31	10-21	60	22-38	89	35-54		
32	10-22	61	23-38	90	36-54		
33	11-22	62	23-39	91	36-55		

Додаток 5

Критичні значення коефіцієнтів лінійної кореляції

Критичні значення коефіцієнта лінійної кореляції ($r_{кр}$) визначаються кількості ступенів свободи, пов'язаного з обсягом вибірки (n) співвідношенням $\nu = n - 2$ різних рівнів значимості.

ν	$p=0,05$	$p=0,01$
2	0,95	0,99
3	0,88	0,96
4	0,81	0,92
5	0,75	0,87
6	0,70	0,83
7	0,67	0,80
8	0,63	0,77
9	0,60	0,74
10	0,58	0,71
13	0,51	0,64
15	0,48	0,61
18	0,44	0,56

ν	$p=0,05$	$p=0,01$
20	0,42	0,54
25	0,38	0,49
30	0,35	0,45
35	0,32	0,42
40	0,30	0,39
50	0,27	0,35
60	0,25	0,33
70	0,23	0,30
80	0,22	0,28
90	0,21	0,27
100	0,19	0,25
200	0,14	0,18

Навчальне видання

Жерновнікова Яна Вікторівна

доцент кафедри інформатики та біомеханіки Харківської державної академії
фізичної культури, кандидат педагогічних наук, доцент

Пятисоцька Світлана Сергіївна

в.о. завідувача кафедри інформатики та біомеханіки Харківської державної
академії фізичної культури, доктор наук з фізичного виховання та спорту,
доцент

Ольховий Олег Михайлович

професор кафедри теорії та методики фізичного виховання Харківської
державної академії фізичної культури, доктор наук з фізичного виховання та
спорту, професор

Статистична обробка та аналіз наукових даних

Навчальний посібник

Харківська державна академія фізичної культури
Вул. Клочківська, 99, Харків, 61058

Підписано до друку 10.06.2025

100 примірників