

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

О.В. Соколова, В.О. Тищенко, С.І. Караулова

**СУЧАСНІ МЕТОДИ СТАТИСТИЧНОЇ ОБРОБКИ ТА
АНАЛІЗУ НАУКОВИХ ДАНИХ У СФЕРІ ФІЗИЧНОЇ
КУЛЬТУРИ ТА СПОРТУ**

**Навчально-методичний посібник
для здобувачів третього (освітньо-наукового) рівня
вищої освіти
(ступеня доктора філософії)
освітньо-наукової програми «Фізична культура і спорт»**

**Запоріжжя
2022**

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

О.В. Соколова, В.О. Тищенко, С.І. Караулова

**СУЧАСНІ МЕТОДИ СТАТИСТИЧНОЇ ОБРОБКИ ТА
АНАЛІЗУ НАУКОВИХ ДАНИХ У СФЕРІ ФІЗИЧНОЇ
КУЛЬТУРИ ТА СПОРТУ**

**Навчально-методичний посібник
для здобувачів третього (освітньо-наукового) рівня
вищої освіти
(ступеня доктора філософії
освітньо-наукової програми «Фізична культура і спорт»**

Затверджено
вченою радою ЗНУ
Протокол № 5 від 22.12.2022

Запоріжжя
2022

УДК: 37.037.1:519.2(075)
С 594

Соколова О.В., Тищенко В.О., Караулова С.І. Сучасні методи статистичної обробки та аналізу наукових даних у сфері фізичної культури та спорту: навчально-методичний посібник для здобувачів третього (освітньо-наукового) рівня вищої освіти (ступеня доктора філософії) освітньо-наукової програми «Фізична культура і спорт». Запоріжжя: Запорізький національний університет, 2022. 98 с.

Навчально-методичний посібник «Сучасні методи статистичної обробки та аналізу наукових даних у сфері фізичної культури та спорту», підготовлено відповідно до освітньо-наукової програми підготовки здобувачів третього освітньо-наукового рівня вищої освіти ступеня доктор філософії з метою забезпечення систематизації і поглиблення знань аспірантів з використання методів математичної статистики для оволодіння методичними і практичними знаннями, уміннями та навичками щодо здійснення наукових досліджень в галузі фізичної культури і спорту.

Рекомендовано для студентів і аспірантів, які навчаються за спеціальністю 017 «Фізична культура і спорт»

Рецензент

М.В. Маліков, д-р біол. наук, проф., декан факультету фізичного виховання ЗНУ

Відповідальний за випуск


А.П. Конох, д-р пед. наук, проф., завідувач кафедри теорії та методики фізичної культури і спорту

ЗМІСТ

Перелік умовних позначень.....	6
Передмова.....	7
Змістовий модуль 1. Використання математико-статистичних методів у сучасних наукових дослідженнях у сфері фізичної культури та спорту.....	9
Тема 1. Вступ у методи математичної статистики.	9
Тема 2. Основний зміст методів математичної статистики, які застосовуються в фізичному вихованні та спорті.....	10
Тема 3. Використання редактора електронних таблиць Excel в процесі збору і обробки наукових даних у сфері фізичної культури та спорту.....	11
<i>Практичне завдання.....</i>	<i>23</i>
<i>Питання для самоконтролю.....</i>	<i>23</i>
Змістовий модуль 2. Метод середніх величин.....	24
Тема 4. Загальні положення методу середніх величин	24
Тема 5. Представлення статистичних даних	26
Тема 6. Вирішення типових завдань методом середніх величин	26
<i>Практичне завдання.....</i>	<i>45</i>
<i>Питання для самоконтролю.....</i>	<i>45</i>
Змістовий модуль 3. Вибірковий метод математичної статистики.....	46
Тема 7. Загальні положення вибіркового методу.....	46
Тема 8. Перевірка статистичних гіпотез.....	47
Тема 9. Практична реалізація вибіркового методу при обробці й аналізі даних у сфері фізичної культури та спорту.....	52
<i>Практичне завдання.....</i>	<i>66</i>
<i>Питання для самоконтролю.....</i>	<i>66</i>
Змістовий модуль 4. Статистичні методи вимірювання взаємозв'язків....	67
Тема 10. Види взаємозв'язків і поняття про кореляційний аналіз.....	67
Тема 11. Способи визначення взаємозв'язків.....	68
Тема 12. Визначення найбільш ймовірних лінійних залежностей між двома факторами.....	71
Тема 13. Практична реалізація кореляційного аналізу.....	72
<i>Практичне завдання.....</i>	<i>92</i>
<i>Питання для самоконтролю.....</i>	<i>92</i>
Змістовий модуль 5. Аналіз і прогнозування часових рядів.....	92
Тема 14. Загальні положення методу.....	92
Тема 15. Способи побудування лінії тренду.....	94
Тема 16. Вирішення типових завдань за допомогою рядів динаміки.....	95
<i>Практичне завдання.....</i>	<i>98</i>
<i>Питання для самоконтролю.....</i>	<i>99</i>
Змістовий модуль 6. Факторний (дисперсійний) аналіз.....	99
Тема 17. Загальні положення методу.....	99
Тема 18. Види факторного аналізу.....	100
Тема 19. Застосування дисперсійного аналізу при проведенні досліджень в фізичному вихованні та спорті.....	101

<i>Практичне завдання</i>	105
<i>Питання для самоконтролю</i>	105
Використана література.....	106
Додаток А Таблиця граничних значень критерію Стьюдента.....	107
Додаток В Таблиця граничних значень критерію Фішера F.....	108

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

 – теоретичні відомості

 – зверніть увагу

 – прочитайте

 – виконайте вправу

 – питання для самоконтролю

ПЕРЕДМОВА

Курс «Сучасні методи статистичної обробки та аналізу наукових даних у сфері фізичної культури та спорту» належить до вибіркової дисципліни циклу професійної підготовки здобувачів третього (освітньо-наукового) рівня вищої освіти (ступеня доктора філософії) освітньо-наукової програми «Фізична культура і спорт».

Метою викладання навчальної дисципліни «Сучасні методи статистичної обробки та аналізу наукових даних у сфері фізичної культури та спорту» є забезпечити глибоку теоретичну та практичну підготовку аспірантів з фізичної культури і спорту, систематизувати і поглибити знання з методів математичної статистики, що використовуються в практиці фізичного виховання і спорту в процесі збору, аналізу і обробки результатів наукових досліджень.

Дисципліна охоплює всі питання, необхідні для аналізу результатів тренувань, спортивних змагань, контрольних випробувань, підбору вправ для підвищення загальної та спеціальної фізичної підготовленості спортсменів, підтвердження висновків, зроблених шляхом математичних розрахунків на основі спостережень із застосуванням сучасних інформаційних технологій.

Основними завданнями вивчення дисципліни «Сучасні методи статистичної обробки та аналізу наукових даних у сфері фізичної культури та спорту» є: забезпечити підвищення рівня професійної компетентності аспірантів; навчити методам математичної статистики, що набули широкого використання в галузі фізичного виховання і спорту; сформулювати вміння обирати методи математичної статистики для вирішення конкретного дослідницького завдання з практики фізичного виховання і спорту; навчити розраховувати основні характеристики варіаційного ряду, критерії вірогідності, коефіцієнти кореляції, проводити оцінку вірогідності різниць і взаємозв'язку між середньостатистичними показниками вибіркової сукупності. Навчити майбутнього фахівця з фізичної культури, тренера аналізувати результати досліджень на основі використання методів математичної статистики.

Оволодіння методами математичної статистики дозволяє максимально ефективно використовувати та проводити математичну обробку й аналіз експериментальних даних.

Згідно з вимогами освітньо-наукової програми аспіранти повинні досягти таких програмних **компетентностей і програмних результатів навчання**:

Інтегральна компетентність (ІК):

ІК – Здатність продукувати нові ідеї, розв'язувати комплексні проблеми у галузі освіти і педагогіки, професійної та/або дослідницько-інноваційної діяльності у сфері фізичної культури і спорту, що передбачає глибоке переосмислення наявних та створення нових цілісних знань та/або професійної практики, застосовувати новітні методології наукової та педагогічної діяльності, а також проводити власні наукові дослідження, результати яких мають наукову новизну, теоретичне та практичне значення.

Загальні компетентності (ЗК):

ЗК1 – Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2 – Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК3 – Здатність виявляти, ставити та вирішувати проблеми.

ЗК4 – Здатність генерувати нові ідеї (креативність).

ЗК5 – Здатність до критичного мислення.

ЗК9 – Здатність до особистісного та професійного розвитку.

ЗК10 – Здатність розв'язувати комплексні проблеми зі спеціальності фізична культура і спорт на основі системного наукового світогляду та загального культурного кругозору із дотриманням принципів професійної етики та академічної доброчесності.

Спеціальні компетентності (СК):

СК1 – Здатність до розуміння основних концепцій, історичних витоків, сучасного стану та тенденції розвитку спеціальності фізична культура і спорт; оволодіння термінологією з досліджуваного наукового напрямку.

СК3 – Здатність виявляти, ставити та вирішувати проблеми дослідницького та/або інноваційного характеру; оцінювати та забезпечувати якість виконуваних досліджень; комерціалізувати їх результати; здійснювати захист прав інтелектуальної власності.

СК9 – Здатність аналізувати, систематизувати та узагальнювати результати міждисциплінарних наукових досліджень у сфері фізичної культури і спорту.

Програмні результати навчання:

ПРН1 – Мати передові концептуальні та методологічні знання з предметної області та на межі предметних галузей, а також дослідницькі навички, достатні для проведення наукових і прикладних досліджень на рівні останніх світових досягнень з відповідного напрямку, отримання нових знань та/або здійснення інновацій.

ПРН2 – Глибоко розуміти загальні принципи, методи, методології наукових досліджень, застосовувати їх у власних дослідженнях у сфері фізичної культури і спорту та у викладацькій практиці.

ПРН3 – Застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи.

ПРН4 – Формулювати і перевіряти гіпотези; використовувати для обґрунтування висновків належні докази, зокрема, результати теоретичного аналізу, експериментальних досліджень, спостережень, сучасних інструментальних методик, наявні літературні дані з метою розв'язання значущих наукових та науково-прикладних проблем у сфері фізичної культури і спорту.

ПРН5 – Планувати і виконувати експериментальні та/або теоретичні дослідження за напрямом спеціальності та дотичних міждисциплінарних напрямів з використанням сучасних інструментів; оцінювати та забезпечувати якість виконуваних досліджень; комерціалізувати їх результати; здійснювати захист прав інтелектуальної власності.

Змістовий модуль 1. Використання математико-статистичних методів у сучасних наукових дослідженнях у сфері фізичної культури та спорту

Тема 1. Вступ у методи математичної статистики

Математична статистика – розділ математики, присвячений методам збору, аналізу й обробки статистичних даних для наукових і практичних цілей; оперує великою кількістю об'єктів, аналізує масові явища. (це розділ математики, присвячений методам збору, аналізу та обробки статистичних даних для наукових та практичних цілей. *Статистичні дані* являють собою дані, які отримані в результаті досліджень великої кількості об'єктів чи явищ.)

Фахівець із фізичного виховання, студент факультету фізичного виховання, який виконує курсову або кваліфікаційну роботу, аспірант, докторант, науковий співробітник, який аналізує наукові дані, вчитель фізичної культури, тренер, – усі вони в ході своєї роботи отримують фактичний експериментальний матеріал (первинний цифровий масив). Якщо ці дані не будуть коректно оброблені за допомогою методів математичної статистики, то їх робота втратить як теоретичне, так і практичне значення.

Для кращого розуміння теми важливо з'ясувати сутність деяких понять. *Генеральна сукупність* – початкова сукупність (абсолютна кількість об'єктів, яка є в наявності взагалі, наприклад, усі студенти Запорізького національного університету на вказаний рік).

Вибірка – частина об'єктів дослідження, певним чином вибрана з генеральної сукупності (наприклад, студенти факультету фізичного виховання ЗНУ на вказаний рік – це об'єкти дослідження, вибрані з генеральної сукупності за ознакою приналежності до факультету).

Усі об'єкти дослідження повинні мати хоча б одну загальну ознаку, що дозволяє їх класифікувати, порівнювати один з одним (стать, вік, спортивна кваліфікація тощо). В цьому випадку про ці об'єкти можна говорити як про статистичну сукупність.

Фактичний експериментальний матеріал накопичується в ході наукового експерименту. Його традиційна схема така: зазвичай особи, які беруть участь у наукових дослідженнях, поділяються на контрольну й експериментальну групи, важливе значення в яких мають ознаки, що визначають досліджуваних як статистичну сукупність. Ці ознаки мають бути приблизно однаковими за своїми характеристиками, інакше експеримент втрачає свою наукову значущість.

Контрольна група готується за традиційною методикою, а експериментальна – із застосуванням нововведень. До і після експерименту проводяться контрольні випробування (зрізи) і за їх результатами роблять висновок про ефективність нововведень.



Тема 2. Основний зміст методів математичної статистики, які застосовуються в фізичному вихованні та спорті

Уже на етапі відбору в контрольну і експериментальну групи дослідник стикається з низкою проблемних питань: якою має бути чисельність групи, за яким критерієм мають відбиратися кандидати в ці групи, рівень підготовленості учасників експерименту, наскільки істотно відрізняється одна група від іншої за важливими для експерименту показниками тощо? На всі ці питання можна відповісти, застосувавши методи математичної статистики. Наприклад, є методи, що дозволяють однозначно стверджувати, що вибірка є представницькою (репрезентативною) відносно до генеральної сукупності. До них належать:

1) Методи відбору об'єктів з генеральної сукупності у вибірку: жеребкування; механічний відбір; типовий відбір; серійний відбір.

2) Методи точкових та інтервальних оцінок, що дозволяють виявити максимально близькі значення та межі інтервалів, між якими з більшою вірогідністю знаходяться дійсні значення параметрів, що визначаються.

3) Методи, які допомагають визначити той мінімальний обсяг, який би дозволив зробити висновок про середнє значення генеральної сукупності не більше ніж з помилкою на задану величину після проведення контрольних зрізів.

4) Методи описової статистики:

- групування даних і подання їх у вигляді статистичних таблиць з визначенням у них варіаційних рядів;
- графічне представлення експериментальних даних у вигляді гістограм і полігона частот.

5) Методи, що дають уявлення про кількісні числові характеристики:

- характеристики положення: середнє арифметичне; медіана; мода;
- характеристики розсіяння: дисперсія; стандартне відхилення; коефіцієнт варіації;
- характеристики асиметрії емпіричних розподілів: асиметрія; ексцес.

6) Методи перевірки статистичних гіпотез:

- Критерії, засновані на нормальному розподілі: t - критерій Стьюдента; F - критерій Фішера; U - критерії.
- Критерії згоди: χ^2 - критерій (критерій хі-квадрат); критерій Шапіро-Уїлки.
- Непараметричні критерії: критерій Вілкоксона.

7) Дуже часто метою дослідження є встановлення наявності та ступеня зв'язку між спортивним результатом і певним показником тренуваності або фізичного розвитку, між окремими показниками фізичної підготовленості тощо; подібні завдання вирішуються методами кореляційного й регресивного аналізу.


8) Окрім того, в деяких випадках дослідникові необхідно знати ступінь тісноти зв'язку одного показника двома, трьома, чотирма і більш аргументами, що впливають на цей показник. Наприклад, у сфері біомеханічної науки: як впливають на результат стрибка в довжину з розгону: початкова швидкість розгону; швидкість розгону на 3 м до відштовхування; величина кута постановки поштовхової ноги; величина кута згинання в колінному суглобі та швидкість руху махової ноги під час відштовхування, висота польоту в польотній фазі стрибка тощо? Це приклад з багатьма невідомими. Подібні до цього прикладу рівняння вирішуються за допомогою методів множинної кореляції і регресії. Ще більш складні питання вирішують факторний та інші види аналізів.

Тема 3. Використання редактора електронних таблиць Excel в процесі збору і обробки наукових даних у сфері фізичної культури та спорту

Електронні таблиці в більшості випадків застосовуються для автоматизації рутинних операцій. На відміну від текстових редакторів і програм для Інтернету, вони мало схильні до доробок. Тому користувач завжди може розраховувати на те, що, освоївши будь-яку версію електронної таблиці, йому не потрібно буде перенавчатися знову, коли на комп'ютері з'явиться інша аналогічна програма.

Запуск програми Excel

При інсталяції пакету програм Microsoft Office ярлики програм, які його становлять, розміщуються як в Меню „Пуск” Windows, так і можуть розміщуватися безпосередньо на Робочому столі. Тому запуск програми Excel може здійснюватися двома способами:

- шляхом подвійного натискання на ярлик , у випадку, якщо він знаходиться на Робочому столі;
- через вибір потрібної програми в Меню «Пуск», як це показано на Рис.5.

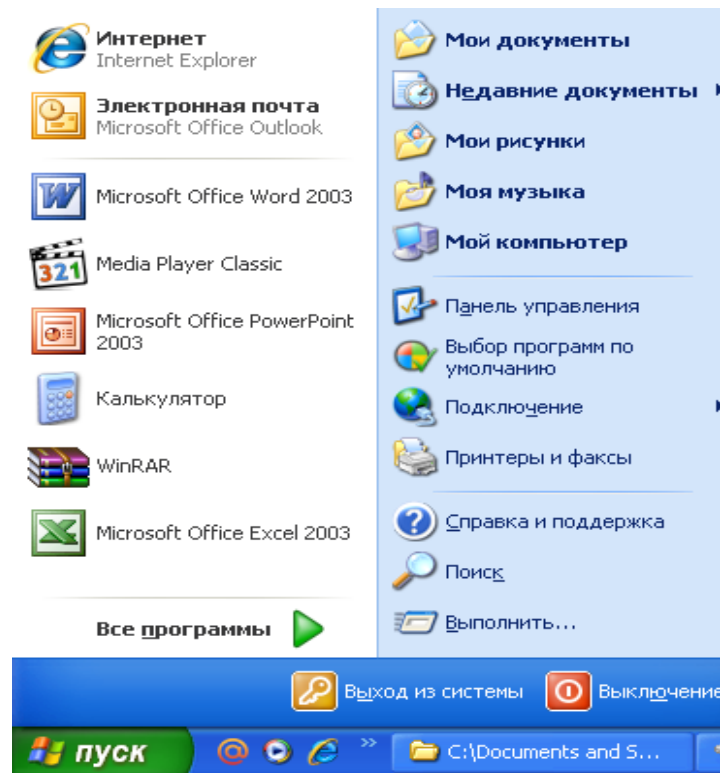



Рис.5. Шлях запуску програми Excel через меню «Пуск»

Загальний вигляд вікна Excel

Після запуску Excel на екрані відкривається його вікно, зовнішній вигляд якого подано на рис. 6.

Верхній рядок вікна інформує про те, що створено нову книгу Microsoft Excel (нову електронну таблицю) – Книга1. Після збереження файлу з таблицею в цьому рядку буде представлена назва, яку користувач присвоїть своєму файлу в процесі збереження. Якщо відкрити файл з якою-небудь раніше створеною таблицею, в рядку буде назва файлу. Наприклад, *Мій файл.xls*. Розширення «xls» указує, що це файл електронної таблиці, створеної за допомогою програми Excel.

Нижче заголовка знаходиться рядок з назвами меню. Кожне меню об'єднує команди, згруповані за будь-якими загальними ознаками: *Файл, Правка, Вид, Вставка, Формат, Сервис, Данные, Окно, Справка*. Натискання мишею на імені меню розкриває вікно зі списком команд, причому напівжирним шрифтом вказані тільки ті команди, які доступні (які можуть бути виконані) в цей момент часу. Біля ряду команд зліва від імені є значки, які також можуть бути знайдені в різних панелях інструментів.

При повсякденній роботі з Excel можна помітити, що перелік доступних команд в меню непостійний. Це результат дії сервісної функції Excel, яка дозволяє бачити тільки ті команди, які частіше за все використовуються. Для відображення всіх доступних команд в меню слід натиснути мишею на значок , який знаходиться внизу вікна меню.

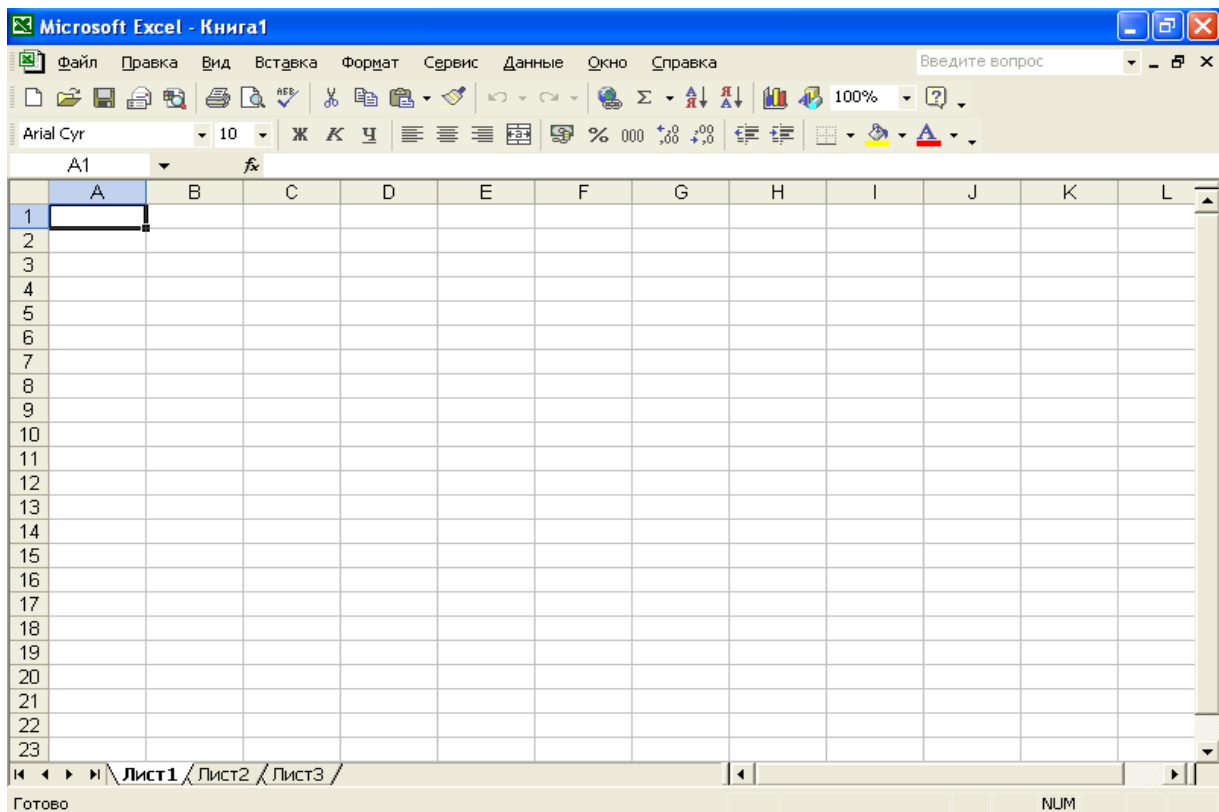



Рис. 6. Зовнішній вигляд вікна Excel

Праворуч від назв багатьох команд є значок , який означає, що при наведенні покажчика миші на цей пункт відкривається підміню, в якому вказуються різні варіанти даної команди.

Нижче рядка головного меню програми Excel розташовані панелі інструментів, на яких знаходяться екранні кнопки-піктограми, що ілюструють виконувани ними команди. Всі кнопки-піктограми на панелях інструментів дублюють аналогічні команди в меню. Опис основних панелей інструментів і команд буде подано нижче.


На рис. 6 видно, що майже весь простір екрану розподілений тонкими лініями на прямокутники – це робочий простір Excel, який називається *Робочим листом*. Тут можна проводити обчислення, створювати графіки та малюнки.

Панелі інструментів

Панелі інструментів призначені для зручності користування командами, що викликаються найбільш часто. Користувач може управляти кількістю відображених панелей інструментів для збільшення ефективності роботи. Для входу в меню управління панелями інструментів необхідно натиснути правою кнопкою миші в межах якої-небудь панелі інструментів. У меню, що з'явилося, доступні у даний момент панелі інструментів помічені прапорцями. У більшості випадків для повсякденної роботи рекомендується залишати налаштування панелей за умовчанням, коли відображаються панелі інструментів *Стандартная* і *Форматирование*.

Програма Excel залежно від виконуваних у цей момент операцій може автоматично відкривати додаткові панелі інструментів.

Кількість кнопок на будь-якій панелі інструментів можна змінювати,

натиснувши мишею на значок . У цьому випадку з'являється продовження відповідної панелі інструментів, де розташовуються кнопки-піктограми, використовувані рідко. Якщо підвести покажчик миші до відповідної кнопки-піктограми і натиснути на ній, то вона з'явиться у відповідній панелі інструментів.

Команди панелі інструментів «Стандартная».


На панелі інструментів *Стандартная* (рис.7) представлені кнопки-піктограми, які управляють командами, що використовуються найбільш часто. Більшість рисунків на кнопках-піктограмах ілюструє дії, які виконуються тією або іншою командою. Далі стисло описуються команди, для яких встановлено кнопки-піктограми за умовчанням. Про сутність решти команд можна дізнатися в довідковій системі Excel, яка викликається натисканням клавіші F1 або з меню *Справка*.





Рис.7. Панель інструментів *Стандартная*


При роботі з Excel команди, що використовуються часто, можна викликати за допомогою клавіш швидкого доступу (клавіш, за допомогою яких у будь-який момент, коли ви працюєте з програмою, можна викликати приписані їм команди). Підказка стосовно клавіші швидкого доступу в меню вказується праворуч від назви команди. Знак «+» в цих командах означає, що клавіші необхідно натискати одночасно.

На цій сторінці клавіші швидкого доступу вказуються в дужках поряд із зображенням кнопки-піктограми відповідної команди.

 Створити (Ctrl + N) – команда забезпечує створення нового документа з «порожньою» електронною таблицею.


 Відкрити (Ctrl + O) – команда забезпечує відкриття документа, який збережений раніше. Для його пошуку у файловій системі відкривається вікно *Проводника*, в якому необхідно знайти потрібний користувачу файл.


 Зберегти (Ctrl + S) – команда слугує для збереження поточної електронної таблиці (з якою наразі працює користувач) на вінчестері або гнучкому диску. Для збереження необхідно натиснути на цій кнопці-піктограмі. Якщо таблиця записується на вінчестер вперше, відкриється вікно, в яке необхідно ввести ім'я файлу або підтвердити пропозицію зберегти файл під іменем «Книга1», вказаним за умовчанням в рядку імені файлу.


 Повідомлення – команда, за допомогою якої запускається алгоритм передачі поточної електронної таблиці по електронній пошті. Користувачеві необхідно просто виконувати вказівки, що з'являються на екрані. Необхідно також мати на увазі, що:




- для передачі файлу з таблицею по електронній пошті має бути активоване з'єднання з Інтернет;


- передача здійснюється тільки на адреси, доступ до яких здійснюється з використанням стандартних поштових програм Microsoft.


 Друк (Ctrl + P) – за цією командою на принтері роздруковується поточний аркуш електронної таблиці в тому вигляді, як це встановлено за умовчанням у налаштуваннях принтера. Оскільки завдання на принтер пересилається негайно, без можливості вплинути на параметри друку, краще користуватися командою *Печать*, доступною в меню *Файл*. У цьому випадку в діалоговому вікні можна встановити параметри.


 Попередній перегляд – дозволяє проглядати зовнішній вигляд таблиці в тому вигляді, який вона матиме після друку.


 Перевірка орфографії (F7) – за допомогою цієї команди можна перевірити орфографію і виправити помилки в тексті.


 | Вирізати (Ctrl + X),  Вставити (Ctrl + V), і  Копіювати (Ctrl + C) – ці три команди працюють з буфером обміну Windows і дозволяють копіювати або вирізати будь-який об'єкт і вставляти його як в інше місце таблиці, так і в інший додаток, наприклад у Word.



 Формат за зразком – команда, яка дозволяє відформувати вміст будь-яких комірок у форматі, заданому в зразковій комірці, що використовується. Щоб скористатися цією командою, необхідно навести курсор на ту комірку, в якій вміст відформатовано так, як потрібно, а потім мишею виділити ті комірки, вміст яких потребує форматування.


 Відмінити (Ctrl + Z) – команда дозволяє відмінити помилкові дії (у тому числі й декілька дій, виконаних підряд).


 Повернути – повертає ті операції, які користувач випадково відмінив.

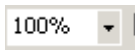
 Додавання гіперпосилання (Ctrl + K) – за допомогою цієї команди створюється зв'язок (гіперпосилання) між коміркою і будь-яким об'єктом, що знаходиться як всередині книги Excel, з якою працює користувач, так і у зовнішньому файлі, який може містити діаграму, таблицю чи інші необхідні елементи. Гіперпосилання також може вказувати на документ, розташований на веб-сайті / мережі Інтернет.

 Автосума – команда дозволяє спростити підсумовування вмісту групи комірок.

 Сортування за зростанням – команда дозволяє провести сортування даних у стовпчику або рядку електронної таблиці. Аналогічно діє команда  Сортування за убаванням.

 Майстер діаграм – виклик діалогового вікна *Мастер диаграмм*, що дозволяє без особливих зусиль створити діаграму.

 Рисування – команда викликає панель інструментів *Рисование*, яка розташовується внизу вікна. За допомогою кнопок, розташованих на ній, можна оформити документ за допомогою рисунків.



Масштаб – це меню використовується для зміни видимих розмірів комірок. Наприклад, щоб не напружувати зір або побачити повністю всю таблицю. Зміна масштабу документа корисна, коли не всі стовпчики або рядки документа видні на екрані.



Довідка Microsoft Excel – команда викликає довідкову систему Microsoft Excel, дублюючи клавішу F1.

Команди панелі інструментів «Форматирование»

На панелі інструментів *Форматирование* (рис.8) представлені найуживаніші команди, які дозволяють надати електронним таблицям Excel необхідної форми та зовнішнього вигляду. Крім того, можна оформити таблиці таким чином, щоб інформація могла легко сприйматися користувачем. Наприклад, виділивши їх за допомогою кольорових ліній або груп комірок.



Рис.8. Панель інструментів *Форматирование*



Шрифт і



Розмір – два меню, які дозволяють

встановити назву й розмір шрифту. Звичайно різні шрифти використовують, коли потрібно сформувати документ для друку. В інших випадках можна використовувати шрифт, встановлений за умовчанням, окрім випадків, коли така зміна необхідна для зорового виокремлення певної інформації, що становить зміст таблиць, або для виокремлення різних заголовків.



Напівжирний (Ctrl + B),



курсив (Ctrl + I) і



підкреслений (Ctrl +

U) – за допомогою цих команд відповідно оформлюють текст. Наприклад, щоб виділити слово, число або всю комірку. Часто напівжирним шрифтом виділяють заголовки таблиць, а курсивом – коментарі.



По лівому краю,



по центру,



по правому краю – ця група команд

потрібна для розміщення тексту всередині комірок. Часто ці команди слугують для форматування вмісту комірок з числами щодо комірок з текстом, оскільки за умовчанням текст розташовується по лівому краю комірки, а числа – по правому.



Об'єднати і розмістити в центрі – команда дозволяє об'єднати декілька суміжних комірок для створення заголовків таблиць або комірок, які займають декілька стовпчиків.



Грошовий формат,



Процентний формат,



Формат з

роздільниками – ці команди використовуються тоді, коли в комірках представлені не просто цифри, а гроші та відсотки. Вони не тільки перетворюють значення в комірці, але й додають символ відсотків або грошової одиниці.






Збільшити розрядність,





зменшити розрядність – ці команди управляють числом відображених після коми десяткових знаків. При цьому

вміст у комірці, що форматується, не змінюється.

 Зменшити відступ,  Збільшити відступ – за допомогою цих команд управляють відступом від початку в першому рядку абзаців у тексті. Ці команди часто використовуються, оскільки клавіша табуляції *Tab* слугує в Excel для переходу до іншої комірки.

 Межі – одна з команд, що використовуються, для рисування рамок таблиць. У підміню можна обрати потрібний варіант рисування.

 Колір заливки – команда дозволяє встановити в комірках кольоровий фон. Наприклад, можна виділяти комірки з формулами, які не вимагають ручного введення даних, щоб уникнути порушення структури таблиці.

 Колір шрифту – ця команда використовується для виокремлення тексту в комірках. У меню, що розкривається, можна обрати бажаний колір.

Користувач може самостійно додавати або прибирати кнопки-піктограми з командами й формувати панелі інструментів за власним бажанням. Для цього необхідно натиснути правою кнопкою миші на ділянці панелей інструментів і у вікні, що з'явилося, вибрати пункт *Настройка*. Після чого відкриється вікно (рис.9), в якому необхідно вибрати закладку *Команды*. Потім необхідно просто перетягувати потрібні кнопки-піктограми мишею на панелі інструментів.

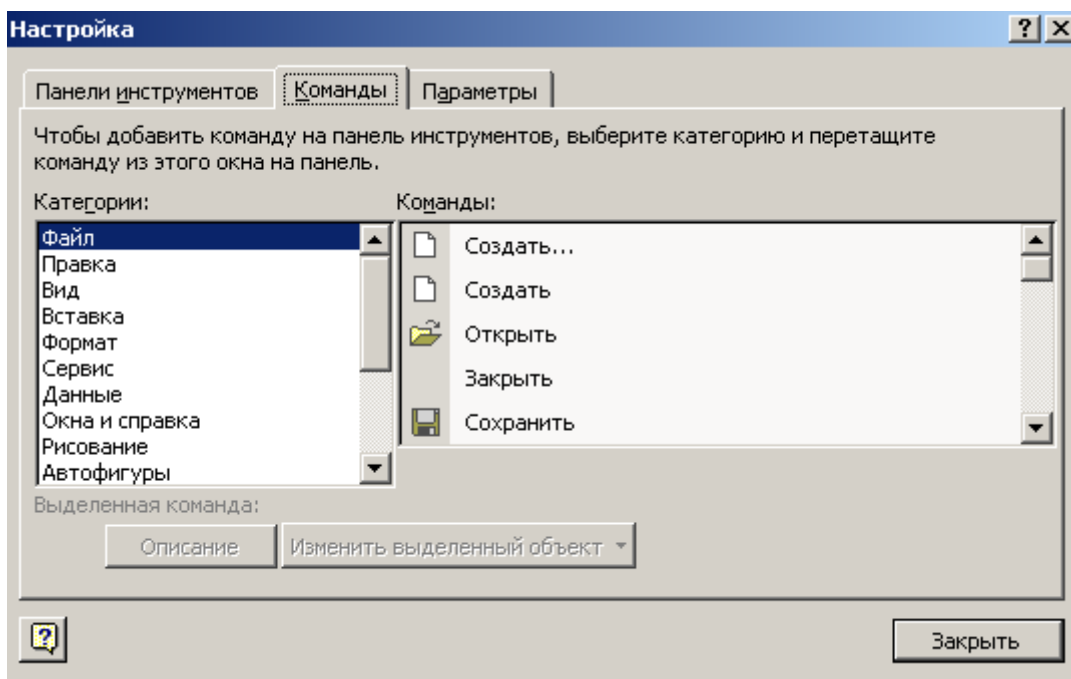


Рис. 9. Вікно налаштування панелей інструментів

Робочий аркуш Excel

Для того щоб можна було визначити адресу комірки, робочий аркуш оточений смугами з цифрами й буквами. Адреса будь-якої комірки складається з буквеного позначення стовпчика й номера рядка, на перетині яких вона знаходиться. Наприклад, A1, C12 і т. д.

Користувач може спеціально виокремити робочу область. Для цього після введення первинного масиву інформації і формування таблиці необхідно скористатися режимом *Предварительный просмотр* (рис.10) і обрати у верхній

частині екрана пункт *Разметка экрана*, після чого робоча область залишиться виокремленою білим кольором, а решта буде забарвлена в сірий колір.

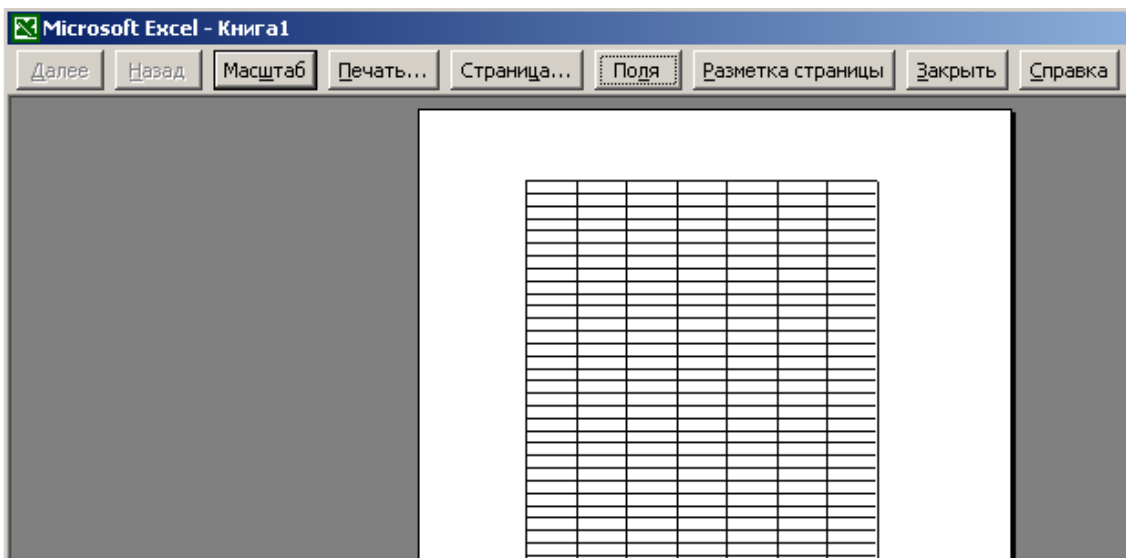


Рис. 10. Вікно попереднього перегляду Excel

Межі робочої області будуть позначені синьою лінією (рис.11).

	C	D	E	F	G	H	I
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							

Рис. 11. Робочий аркуш Excel в режимі розмітки сторінки

У тих випадках, коли необхідно збільшити розміри робочої таблиці (додати нові рядки або колонки), необхідно навести маркер миші на синю смугу межі робочої області й перетягнути її на потрібну кількість рядків або стовпчиків. Так можна встановити бажаний розмір таблиці.

За умовчанням при створенні нової електронної таблиці Excel формує не один робочий аркуш, а три, і всі вони називаються книгою Excel. Назви цих аркушів можна знайти в лівому нижньому куті екрана (рис. 12). За умовчанням вони називаються *Лист 1*, *Лист 2*, *Лист 3*. Ярлик активного робочого аркуша забарвлений в білий колір, ярлики інших – в сірий колір.



Рис. 12. Область управління робочими аркушами

Користувач може змінювати назви робочих аркушів. Для цього необхідно натиснути правою кнопкою миші на ярлик аркуша і в меню (рис. 13), що відкрилося, обрати пункт *Переименовать*.

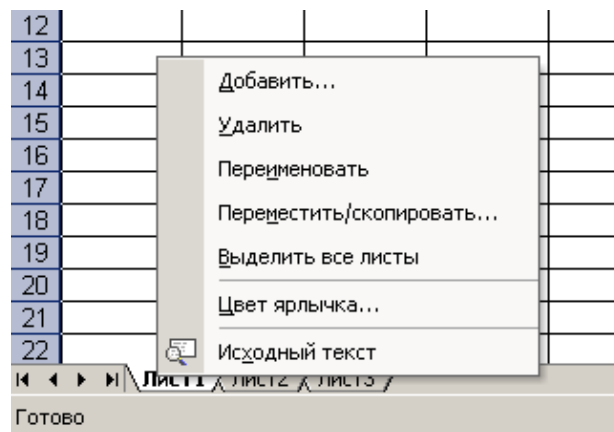


Рис. 13. Меню управління робочими аркушами

Для виокремлення на робочому столі суміжних комірок необхідно натиснути лівою кнопкою миші на першій комірці діапазону і, не відпускаючи, переміщати покажчик миші, виокремлюючи необхідні комірки. Якщо необхідно виокремити декілька окремо розташованих діапазонів комірок, то при їх виокремленні необхідно тримати натисненою клавішу **Ctrl** (рис.14).

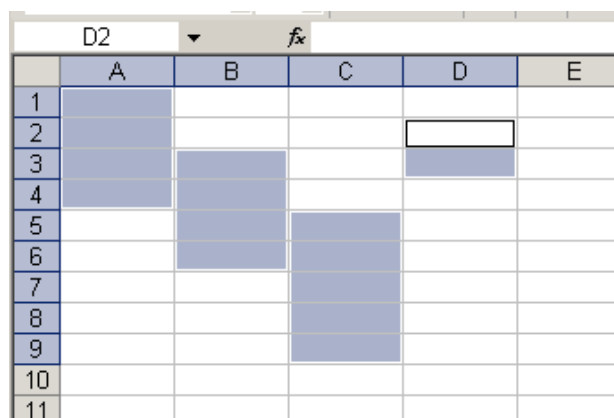


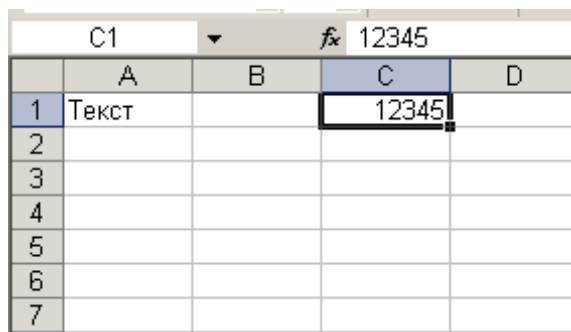
Рис.14. Виокремлення груп комірок

Рядок формул

Коли користувач натискає клавіші на клавіатурі, всі символи, що вводяться, з'являються в комірці, яка наразі є активною (виокремлена рамкою). Одночасно ці ж символи з'являються в полі, розташованому в рядку, що

знаходиться над робочим аркушем, – у рядку формул.

Excel розрізняє, коли вводиться текст, а коли цифри. Текст відцентровано по лівому краю комірки, а число – по правому (рис.15).



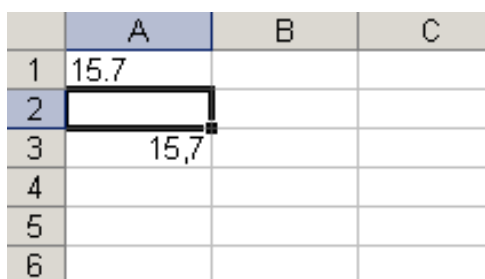
The screenshot shows an Excel spreadsheet with a grid of cells. The formula bar at the top displays '12345'. The active cell is C1, which contains the number '12345'. Cell A1 contains the text 'Текст'. The grid extends to row 7 and column D.

	A	B	C	D
1	Текст		12345	
2				
3				
4				
5				
6				
7				

Рис.15. Введення тексту та числа в комірку

Щодо помилок при введенні чисел, то вони насамперед стосуються введення неправильного знаку, що відділяє цілу частину числа від його дробової частини в десяткових дробах. Річ у тому, що в англійській версії Excel як роздільник використовується крапка, а у версії, розробленій для України – кома. На рис.16 показано, як розрізняється вміст комірки A1, в якому знаходиться текст «15.7» (притиснутий до лівого краю комірки), і комірки A3, куди введено число

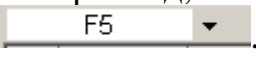
15,7 (відцентроване за правим краєм комірки).



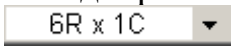
The screenshot shows an Excel spreadsheet with a grid of cells. Cell A1 contains the text '15.7' aligned to the left. Cell A3 contains the number '15,7' centered. The grid extends to row 6 and column C.

	A	B	C
1	15.7		
2			
3	15,7		
4			
5			
6			

Рис.16. Помилки при введенні роздільника в десяткових дробах

Зліва в рядку формул знаходиться поле імені, в якому відображається адреса активної комірки. Адреса складається з імені стовпчика й номера рядка. При цьому для позначення імені стовпчика використовується латинський алфавіт, причому для визначення адреси комірки можна вводити як прописні, так і рядкові букви. Наприклад, f5 або F5. Excel самостійно переводить рядкові символи в прописні. 

Адреса активної комірки в полі імені з'являється автоматично, коли користувач переходить до тієї або іншої комірки. Але можливе і ручне введення адреси в це поле, внаслідок чого після набору конкретної адреси й натискання клавіші Enter активним стає комірка з цією адресою, а вікно огляду таблиці переміщується так, щоб ця комірка була доступною в правому нижньому куті вікна.

Якщо виокремити декілька комірок, то доти, доки кнопка миші не відпущена, у полі імені відображається розмір області захоплених рядків і стовпчиків, наприклад . Символ R позначає рядки (англійською

«рядок» – row), C – стовпчики (англійською «стовпчик» – column), а цифри – кількість рядків і стовпчиків.

Введення формул

Специфічна особливість електронних таблиць полягає в тому, що вони дозволяють автоматизувати обчислення. Для цього у відповідну комірку вводиться формула, по якій обчислюється значення і відображується в комірці.

Щоб програма зрозуміла, що в комірку вводиться формула, необхідно натиснути на клавішу =, а далі набрати вираз, який необхідно обчислити, (наприклад, =15*6-9). Сама формула відображається тільки в рядку формул, а безпосередньо в комірці відображається тільки результат обчислень (рис. 17).

	A	B	C
1	81		
2			
3			

Рис.17. Приклад формули в комірці

Наприклад, при зміні прискорення (комірка A2) Excel автоматично перерахує значення сили за формулою, яка знаходиться в комірці C2. При цьому в цій комірці буде відображений тільки результат обчислень (рис.18). Для полегшення введення адреси комірок у формули необхідно використовувати спеціальну функцію Excel по підстановці у формулу адреси, обраної за допомогою миші комірки, що запобігає грубим помилкам. Для цього потрібно вибрати комірку, в якій буде записана формула, і натиснути клавішу (=). Для введення першої адреси комірки необхідно натиснути мишею на потрібній комірці. Комірка, адреса якої підставляється у формулу, виділяється пунктирною рамкою, а сама адреса з'являється в рядку формул і в активній комірці (рис.18, а). Далі користувач повинен ввести потрібний арифметичний оператор (складання, віднімання, множення, ділення) і натиснути мишею на іншу комірку (рис.18, б). Для завершення введення формули натискають клавішу Enter. У результаті цих дій в комірці з'явиться обчислене за формулою значення. Для того щоб побачити саму формулу, а не її результат, необхідно обрати знову ту саму комірку.

а

б

Рис.18. Введення адреси комірки у формулу

Під час введення даних або формул в комірки, в рядку формул з'являються три кнопки-піктограми .

Вставка функції – викликає віконне меню вибору стандартних формул Excel (рис. 19).

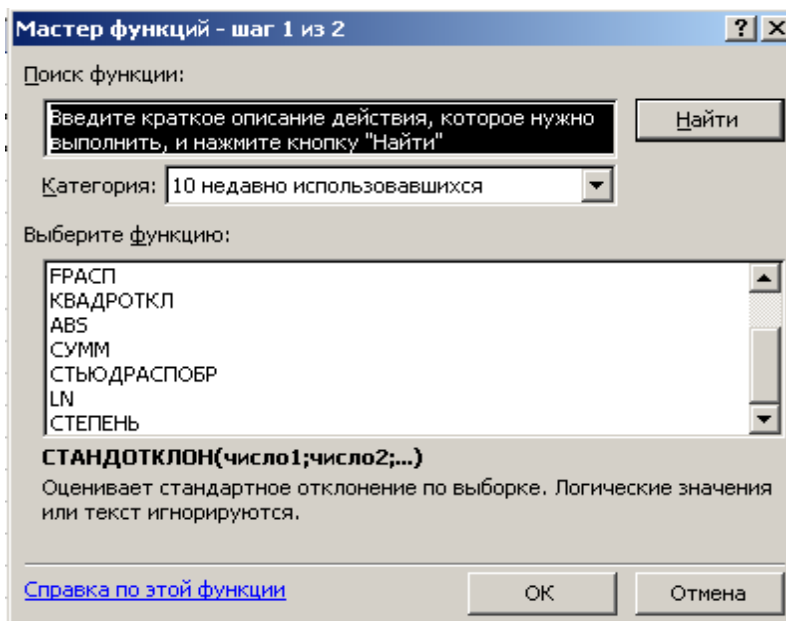


Рис.19. Вікно майстра функцій Excel

✗ Відміна – дія цієї команди дублює клавішу Esc, відмінюючи помилкове введення з клавіатури.

Введення – підтвердження закінчення введення формули (аналогічне натисненню клавіші Enter).

Функція підсумовування

Однією з найчастіше використовуваних стандартних функцій Excel є функція підсумовування. Часто необхідно підсумувати кілька значень, що містяться в суміжних комірках. Можна, звичайно, писати у формулі адреси всіх комірок, які потрібно враховувати, наприклад =B2+C2+D2+E2. Проте це є зайвим. Простіше вказати діапазон комірок і застосувати до них вбудовану функцію підсумовування СУММ.

Щоб не вводити назву цієї функції, можна натиснути мишею на значок Σ , що активує роботу *Мастера автосуммирования*. В результаті в активній комірці з'явиться рядок =СУММ() (рис.20, а).

У круглих дужках необхідно вказати адреси комірок, вміст яких необхідно підсумувати. Для цього на клавіатурі можна через крапку з комою вказати перелік адрес потрібних комірок або вказати діапазон комірок у форматі B2:E2, де B2 – комірka початку, а E2 – комірka кінця діапазону.

Проте є й інший спосіб. Можна виділити мишею необхідні комірki й натиснути клавішу Enter або, тримаючи натиснутою клавішу Ctrl, виділяти необхідні комірki (рис.20, б).

Натиснення клавіші Enter закінчує введення формули. Excel проводить обчислення, показуючи результат у комірці, де знаходиться формула (рис.20, в).

У пропонуваному прикладі в результаті обчислень за формулою, що введена у комірці A2, з'явиться 0, оскільки в комірках B2, C2, D2 і E2 ніяких даних немає. Якщо ж ввести в них числа, в комірці A2 з'явиться результат підсумовування (рис.20, г).

СТАНДОТКЛОН		✖ ✓ fx =СУММ()	
	О	Р	СУММ(число1; [число2]; ...)
1		=СУММ()	
2			
3			

а

СТАНДОТКЛОН		✖ ✓ fx =СУММ(B2:E2)				
	А	В	С	Д	Е	F
1						
2	=СУММ(B2:E2)					
3						
4						

б

A2		fx =СУММ(B2:E2)				
	А	В	С	Д	Е	F
1						
2						
3						
4						
5						

в

A2		fx =СУММ(B2:E2)				
	Имя	А	В	С	Д	Е
1						
2		15	5	7	12	-9
3						
4						
5						

г

Рис. 20. Використовування функції СУММ: а) введення формули в комірку; б) вказівка діапазону підсумовування; в) немає даних для підсумовування; г) у комірці представлений результат підсумовування

Табличний процесор Excel являє собою ідеальне середовище для виконання розрахунків різної складності без особливих затрат на програмування. Крім того, додаток дозволяє виконати такі важливі функції, як створення та вставка в книгу формул, діаграм, організаційних структур; вставка в книгу рисунків, об'єктів з екрана ПК; вставка в книгу об'єктів із середовища текстового процесора Word та пакету підготовки презентацій PowerPoint; імпорт в книгу реляційних таблиць із СУБД Access; сортування даних таблиці; автоматизація обробки даних.

Головна функція додатка – автоматизація обробки інформації, яка представлена у вигляді таблиць. Процес автоматизації включає такі етапи, як побудова формул; використання функцій; створення інформаційних і формульних зв'язків між аркушами та книгами; використання елементів управління; побудова макросів; створення панелей або меню користувача з елементами управління.

Практичне завдання:

Сформулюйте завдання майбутнього власного наукового дослідження. Зазначити які з завдань мають вирішуватися за допомогою методів математичної статистики. Розкрити зміст основних етапів статистичних досліджень.

Питання для самоконтролю:

1. Дайте визначення основних понять математичної статистики.

Змістовий модуль 2. Метод середніх величин

Тема 4. Загальні положення методу середніх величин

Сучасні спортивні дослідження включають великі масиви вимірів. Отримані в ході експериментальної роботи дані представлені у вигляді нерегульованого набору чисел. Для того щоб за ними можна було зробити певні висновки, необхідна їх первинна обробка – угруповання. Спеціальні статистичні операції дають можливість сконцентрувати початковий кількісний матеріал, і при цьому не втрачається корисна інформація. З точки зору математичної процедури, така робота зводиться до формування деяких математичних систем, основні характеристики яких дають уявлення про початковий масив чисел.

Найпопулярнішим методом, що дозволяє проводити такі операції, є метод середніх величин. Початкова кількісна інформація при ранжируванні переходить у варіаційний ряд. Характеристики варіаційного ряду дають уявлення про початковий масив чисел. Концентрація початкового кількісного матеріалу та представлення його декількома параметрами є основою для подальших досліджень, оскільки потім робота проводиться не з усім масивом чисел, а тільки з характеристиками варіаційного ряду.

Робота над методом середніх величин передбачає три основних етапи:

1. Утворення варіаційного ряду.
2. Знаходження основних характеристик варіаційного ряду:
 - обсяг вибірки (N);
 - середнє арифметичне значення (\bar{X});
 - медіана (Me_x);
 - мода (Mo_x);
 - дисперсія (σ^2);
 - середнє квадратичне відхилення (σ);
 - помилка середнього арифметичного значення (m);
 - коефіцієнт варіації (V);
3. Практична реалізація отриманих характеристик.

Метод середніх величин дає можливість визначити середнє значення, припустимі границі досліджуваної вибірки і правильність підбору досліджуваної групи, її однорідність і однотипність. Тільки правильно підібрана група, однорідна й однотипна може брати участь у подальших дослідженнях. Якщо група правильно підібрана, то результати дослідження можуть бути використані при подальшій оцінці й аналізі методики тренування, ефективності навчально-виховного процесу тощо. Прикінцевий результат дослідження можна буде вважати результатом проведення навчально-виховної або навчально-тренувальної роботи, а не наслідком закономірної зміни фізичного розвитку або фізичної підготовленості.

Для того щоб скласти варіаційний ряд, необхідно виконати ранжирування варіант.

Ранжирування – операція розташування варіант у порядку зростання або спадання.

Числа, що входять до складу ранжируваного ряду, називаються *варіантами*, варіанта (X_i) – значення ознаки, що спостерігається.

Кожна варіанта в ранжируваному ряду зустрічається визначену кількість разів – частота варіанти (n_i).

Частота варіанти – кількість варіант у ранжируваному ряду.

Вибірка (вибіркова сукупність) – ряд результатів, поданих випадковими числами.

Генеральна сукупність – сукупність усіх значень, які можна було б отримати для досліджуваної вибірки.

Якщо дослідження охоплена вся генеральна сукупність, то таке дослідження називається *суцільним*. Такі дослідження проводяться рідко.

Наприклад, якщо комусь вдалося обстежити всіх найсильніших спортсменів світу в якомусь виді спорту, тобто провести суцільне дослідження (оскільки інших найсильніших спортсменів світу в досліджуваному виді спорту на момент обстеження не було), вважається, що обстежено всю генеральну сукупність. Усі подальші дослідження будуть вважатися вибірковими (наприклад, дослідження на рівні України).

Наприклад, довжина тіла студентів одного вишу факультету фізичного виховання (якщо нас цікавлять результати студентів щодо показника, який підлягає дослідженню, тільки одного вишу факультету фізичного виховання) – вибіркова сукупність. А довжина тіла студентів усіх факультетів фізичного виховання України – генеральна сукупність.

Обсяг вибірки (N) – одна з основних характеристик варіаційного ряду, що визначається числом об'єктів спостереження або загальної кількості варіант у ранжируваному ряду, що спостерігається.

Варіаційний ряд – ранжируваний ряд з позначкою частоти або відносної частоти.

Накопичена частота (k) – визначається додаванням попередніх частот.

Накопичена частота першої варіанти в ранжируваному ряду дорівнює її ж частоті, а останньої варіанти – загальній кількості варіант досліджуваного ранжируваного ряду, тобто обсягу вибірки (N).

Медіана (Me_x) – основна характеристика варіаційного ряду. Вона визначається як середня варіанта, що розподіляє ранжируваний ряд навпіл.

Залежно від парності або непарності обсягу вибірки медіана розраховується за формулами:

якщо обсяг вибірки непарний: $Me_x = X_{((N+1)/2)}$

якщо обсяг вибірки парний: $Me_x = (X_{(N/2)} + X_{(N/2 + i)})$

Мода (Mo_x) – основна характеристика варіаційного ряду; вона визначається як варіанта з найбільшою частотою.

Якщо дві варіанти ранжируваного ряду мають однакову найбільшу частоту й розташовуються поруч, то мода – середнє арифметичне значення цих двох варіант; якщо розташовані в різних місцях ранжируваного ряду, то

існують дві моди, і вибірка називається бімодальною; якщо всі варіанти зустрічаються однаково кількість разів – моди не існує.

Середнє арифметичне значення (\bar{X}) – основна характеристика варіаційного ряду, що визначається як середній результат досліджуваної вибірки. Середнє арифметичне значення розраховується за формулою:

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

Графік залежності варіанти від частоти дає уявлення про нормальний закон розподілу.



Тема 5. Представлення статистичних даних

Нормальний закон розподілу – закон, при якому переважну більшість варіантів зосереджено в центрі, а в міру віддалення від центру їх кількість поступово зменшується, зберігаючи абсолютну симетрію лівого та правого крила нормальної кривої.

Якщо результати вимірів за своїми характеристиками можуть бути віднесені до нормального розподілу, то вибір статистичного методу для аналізу результатів визначений. Це дуже важливо, оскільки для дослідження зменшується ступінь ризику використання неправильного статистичного методу аналізу.

На основі порівняння медіани, моди, середнього арифметичного значення і побудови графіка залежності варіанти від частоти можливо зробити висновок про правильний добір досліджуваної групи.

Якщо результати вимірів за своїми характеристиками можуть бути віднесені до нормального розподілу, то вибір статистичного методу для аналізу результатів визначений. Це дуже важливо, оскільки для дослідження зменшується ступінь ризику використання неправильного статистичного методу аналізу.

На основі порівняння медіани, моди, середнього арифметичного значення і побудови графіка залежності варіанти від частоти можна зробити висновок про правильний підбір досліджуваної групи.

Наприклад, результати легкоатлетів на дистанції 100 м подано у вигляді варіаційного ряду, а $\bar{X} = Me_x = Mo_x = 12$ (с) (рис. 1):

Якщо значення медіани, моди й середнього арифметичного значення збігаються або відрізняються незначно, і графік залежності варіанти від частоти має форму купола, що дозволяє говорити про нормальний закон розподілу, то досліджувану групу можна вважати правильно підбраною за віком, статтю, фізичною підготовленістю (для осіб, які не займаються спортом), або віком, статтю, спеціалізацією, кваліфікацією (для спортсменів). Така група може брати участь у подальших дослідженнях. Результати дослідження, які були

отримані при правильному підборі групи, можна буде використовувати в подальшій роботі.

X_i, c	n_i
10	1
11	1
12	5
13	2
14	1
15	1

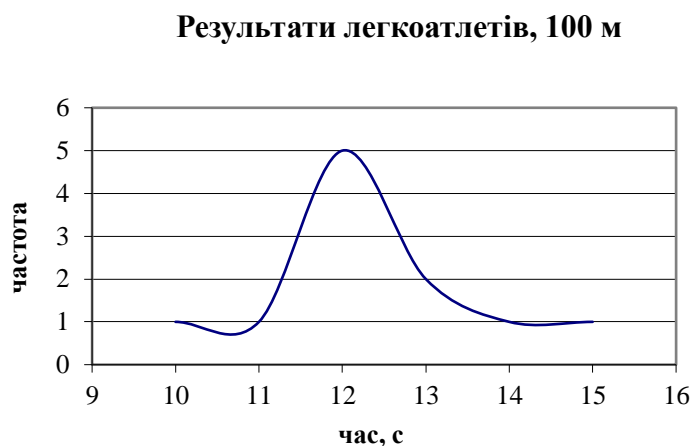


Рис. 1. Графік залежності варіанти від частоти

Якщо значення медіани, моди і середнього арифметичного значно відрізняються, а графік залежності варіанти від частоти не має форми купола, і відзначається лівостороння або правостороння асиметрія (розподіл варіант не відповідає нормальному закону), то групу підібрано неправильно, і вона не може брати участі в подальших дослідженнях. Потрібно переглянути склад групи.

Однорідність і однотипність групи визначається за коефіцієнтом варіації (V), який розраховується за формулою:

$$V = (\sigma / \bar{X}) \times 100\%,$$

де σ – середнє квадратичне відхилення;

\bar{X} – середнє арифметичне значення.

Середнє квадратичне відхилення (σ) – основна характеристика варіаційного ряду, яка показує відхилення від середнього арифметичного значення. Чим менше значення σ , тим більш однорідною й однотипною є група. За σ можна визначити перевагу тієї чи іншої методики тренування, розвитку фізичної якості, підготовленості. За допомогою середнього квадратичного відхилення можна записати припустимі границі ($\bar{X} \pm \sigma$) для досліджуваної вибірки.

Розраховується середнє квадратичне відхилення за формулою:

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

де σ^2 – дисперсія.

Дисперсія показує розсіювання результатів і визначається за формулою:

$$\sigma^2 = \frac{\sum_{i=1}^j (X_i - \bar{X})^2 n_i}{N-1},$$

де X_i – варіанта;

\bar{X} – середнє арифметичне значення;

n_i – частота варіанти;

N – обсяг вибірки.

Помилка середнього арифметичного значення (m) або випадкова помилка вимірів виникає через причини, які можна проектувати, але не можна заздалегідь передбачити (погодні умови; напруга в мережі, від якої працює прилад тощо).

Залежно від обсягу вибірки помилка середнього арифметичного значення може бути розрахована за формулою:

$$m = \sigma / \sqrt{N}, \text{ якщо } N > 20; \text{ або } m = \sigma / \sqrt{N-1}, \text{ якщо } N \leq 20,$$

де N – обсяг вибірки.

Оцінка однорідності й однотипності проводиться за коефіцієнтом варіації:

якщо $V \leq 10\%$, то коливання результатів дослідження невелике, групу можна вважати однорідною й однотипною за своїм складом, вона може брати участь у подальших дослідженнях;

якщо V приймає значення від 10 до 20% – середнє коливання результатів, групу не можна вважати однорідною і однотипною, але склад такої групи можна переглянути;

якщо $V > 20\%$ – коливання результатів істотне; така група є неоднорідною і неоднотипною, а отже, не може брати участі в подальших дослідженнях.



Тема 6. Вирішення типових завдань методом середніх величин



6.1. Визначення правильності добору, однорідності та однотипності групи, яка підлягає дослідженню

Розглянемо послідовність операцій визначення правильності добору, однорідності та однотипності групи, яка підлягає дослідженню на конкретному прикладі.

Приклад: У 14 спортсменів було зафіксовано результати підтягування у висі на перекладині (кількість разів): 10; 10; 15; 14; 14; 15; 12; 12; 12; 9; 13; 13; 16; 16. Визначити правильність добору групи.

Порядок виконання

1. Кожен із результатів, які показали 14 спортсменів, є варіантою – і 10, і 15, і 14

2. Варіанти в ранжируваному ряду записуємо в порядку зростання:
9; 10; 10; 12; 12; 12; 13; 13; 14; 14; 15; 15; 16; 16

3. Обсяг вибірки $N = 14$.

4. Частота (n_i)

$n_i(9) = 1, n_i(10) = 2, n_i(12) = 3, n_i(13) = 2, n_i(14) = 2, n_i(15) = 2, n_i(16) = 2$.

5. Відносна частота (V)

$$V = (n_i/N) \times 100\%$$

$$V_9 = (1/14) \times 100\% = 7,14\%,$$

$$V_{10} = (2/14) \times 100\% = 14,28\%,$$

$$V_{10} = V_{13} = V_{14} = V_{15} = V_{16} = 14,28\%$$

$$V_{12} = (3/14) \times 100\% = 21,42\%$$

6. Варіаційний ряд

6.1 Варіаційний ряд із частотою (n_i)

X_i	9	10	12	13	14	15	16
n_i	1	2	3	2	2	2	2

6.2 Варіаційний ряд з відносною частотою (V)

X_i	9	10	12	13	14	15	16
$V, \%$	7,14	14,28	21,42	14,28	14,28	14,28	14,28

7. Накопичена частота (K)

X_i	9	10	12	13	14	15	16
K	1	3	6	8	10	12	14

8. Медіана (MeX)

У зв'язку з тим, що вибірка парна, медіану розраховуємо за формулою:

$$\overline{MeX} = 1/2 (X_{(N/2)} + X_{(N/2+1)})$$

$$\overline{MeX} = 1/2 (X_{(14/2)} + X_{(14/2+1)}) = 1/2 (X_7 + X_8) = X_6$$

X_7, X_8 – варіанти, які мають у ранжируваному ряду порядкові номери 7 і 8

$$X_7 = 13; X_8 = 13$$

$$\overline{MeX} = 1/2 (13+13) = 13; \overline{MeX} = 13;$$

9. Мода (MoX)

$MoX = 12$, оскільки варіанта 12 зустрічається найбільш часто – 3 рази ($n_i(12) = 3$)

10. Середнє арифметичне значення (\bar{X})

Розраховуємо \bar{X} за формулою зваженого середнього арифметичного значення:

$$\bar{X} = \frac{\sum_{i=1}^j X_i \times n_i}{N}$$

Зважене середнє арифметичне значення, якщо не всі $n=1$,
де \sum - знак підсумовування;
 i – індекс;
 j – кількість варіант у варіаційному ряду.

$$\bar{X} = \frac{9+10+11 \times 2+12 \times 3+13 \times 2+14+15}{11} = 12$$

У середньому кожний з 14 спортсменів виконав 12,9, тобто 13 підтягувань у висі на поперечині.

11. Графік залежності варіанти (X_i) від частоти (n_i) (рис.21).

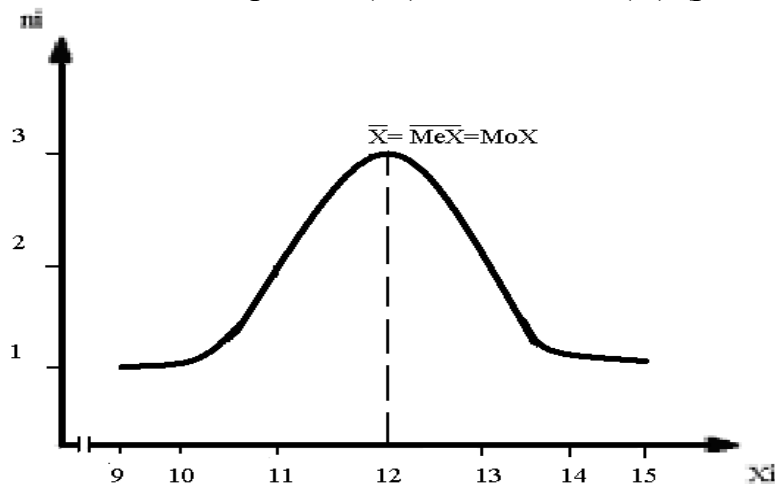


Рис. 21. Графік залежності варіанти (X_i) від частоти (n_i)

Розподіл за цим графіком відповідає нормальному закону розподілу – куполоподібний графік, значення \bar{X} , \overline{MeX} і MoX збігаються; переважне нагромадження варіант зосереджене в центрі, з віддаленням від центру кількість їх поступово зменшується і зберігається абсолютна симетрія лівого та правого крила нормальної кривої.

Висновок: Група з 14 спортсменів за результатами тесту з підтягування у висі на перекладині підібрана неправильно з урахуванням статі, спеціалізації та кваліфікації, оскільки значення \bar{X} , \overline{MeX} і MoX не співпадають. Графік залежності варіанти від частоти не відповідає нормальному закону розподілу. Група може брати участь у подальших дослідженнях тільки після коригування її складу.

Виконаємо практичне завдання за допомогою редактора електронних таблиць Excel.


Приклад: У 14 спортсменів було зафіксовано результати підтягування у висі на перекладині (кількість разів): 10; 10; 15; 14; 14; 15; 12; 12; 12; 9; 13; 13; 16; 16. Визначити правильність добору групи.

Порядок виконання

Крок 1. Відповідності до умов завдання вводимо у стовпчик **В** електронної таблиці результати підтягування у висі на перекладині (кількість разів), тобто ряд спостережень. Таблиця матиме такий вигляд:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Підтягув.													
2		10													
3		10													
4		15													
5		14													
6		14													
7		15													
8		12													
9		12													
10		12													
11		9													
12		13													
13		13													
14		16													
15		16													
16															
17															
18															
19															
20															
21															
22															
23															

Рис. 22. Таблиця результатів підтягування на перекладині (кількість разів)

Крок 2. Проведемо ранжирування результатів. Для цього поставимо курсор на будь-яку комірку стовпчика **В**, в якому вказано результати. Далі натискаємо піктограму сортування  або виконуємо те саме за допомогою команди головного меню *Данные – Сортировка*.

Крок 3. У стовпчику **А** запишемо порядкові номери варіант. У стовпчику **С** запишемо частоту кожної варіанти.

Крок 4. У стовпчику **Д** розрахуємо відносну частоту кожної варіанти. Для цього ставимо курсор на комірку **Д2**, переводимо курсор у *Строку формул*, вводимо з клавіатури позначку = та записуємо формулу: мишею натискаємо на комірку **С2** та ставимо за допомогою клавіатури **/14*100** (Рис. 23). Натискаємо клавішу **Enter**. Після цього проводимо копіювання комірки **С2** уздовж стовпчика даних.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

№	Підтягув.	Частота	Від. Част.
1	9	1	7,142857
2	10	2	
3	10	2	
4	12	3	
5	12	3	
6	12	3	
7	13	2	
8	13	2	
9	14	2	
10	14	2	
11	15	2	
12	15	2	
13	16	2	
14	16	2	


Рис. 23. Таблиця з обчисленим значенням відносної частоти

Крок 5. У стовпчиках **F** та **G** запишемо варіаційний ряд із частотою, а у стовпчиках **I** та **J** – варіаційний ряд з відотною частотою. Отримані результати представлено на рис. 24.

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Підтягув.	Частота	Від. Част.		Підтягув.	Частота		Підтягув.	Від. Част.		
2	1	9	1	7,142857		9	1		9	7,14		
3	2	10	2	14,28571		10	2		10	14,28		
4	3	10	2	14,28571		12	3		12	21,43		
5	4	12	3	21,42857		13	2		13	14,28		
6	5	12	3	21,42857		14	2		14	14,28		
7	6	12	3	21,42857		15	2		15	14,28		
8	7	13	2	14,28571		16	2		16	14,28		
9	8	13	2	14,28571								
10	9	14	2	14,28571								
11	10	14	2	14,28571								
12	11	15	2	14,28571								
13	12	15	2	14,28571								
14	13	16	2	14,28571								
15	14	16	2	14,28571								
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												

Рис.24. Таблиця результатів обчислення

Крок 6. У стовпчик **A** запишемо порядкові номери варіант ранжируваного ряду та назви характеристик варіаційного ряду, які необхідно розрахувати: у комірці **A16** – середнє арифметичне; **A17** – медіана; **A18** – мода.

Крок 7. Розраховуємо середнє значення. Для цього виділимо комірку **B16** та скористаємося піктограмою вбудованих функцій  або командою головного меню *Вставка – Функції*. Після цієї дії з’являється вікно *Мастер функций*, у якому обираємо категорію *Статистические* та функцію *СРЗНАЧ* (рис.24).

Натискаємо клавішу *OK*. При цьому з’являється діалогове вікно (рис.25).

Інтервал даних для обчислення середнього арифметичного значення з’являється автоматично, тому що курсор стоїть внизу стовпчика з числовими значеннями. Однак інтервал у вікно діапазону даних можна ввести безпосередньо з клавіатури або натиснувши на піктограму праворуч цього вікна та вибравши мишею діапазон.

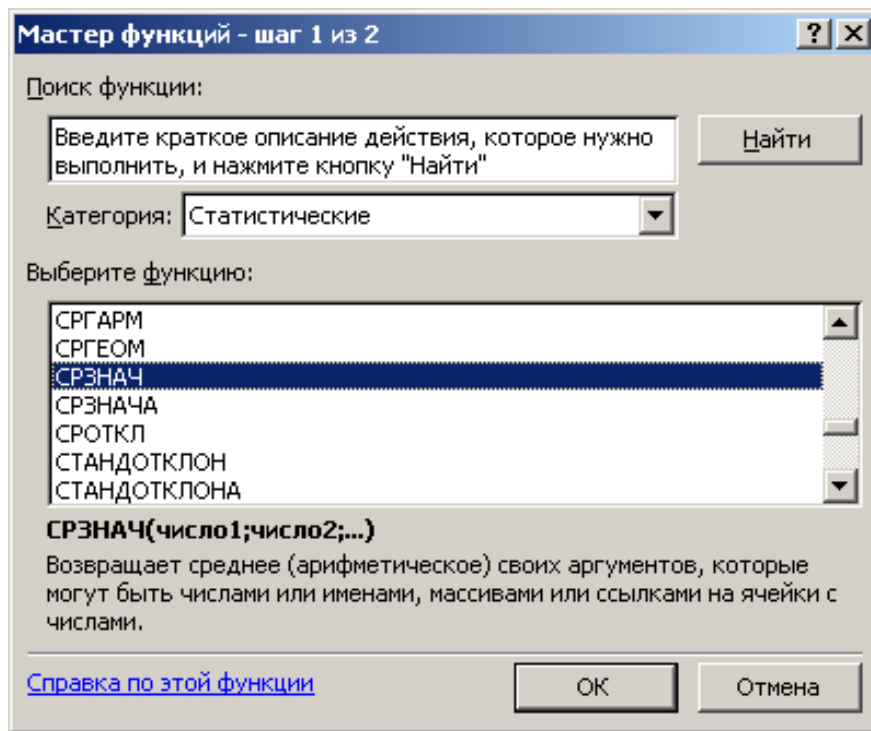


Рис. 24. Вікно виклику функції обчислення середнього значення

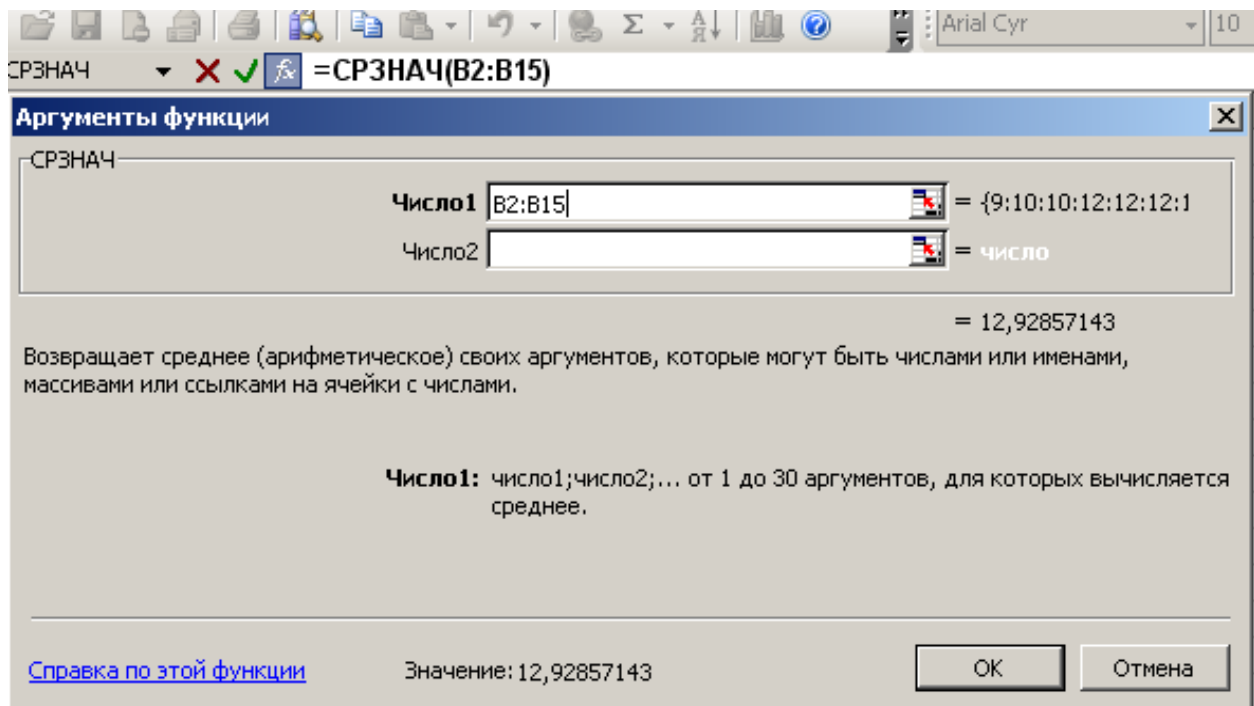


Рис. 25. Діалогове вікно введення даних

Натискаємо клавішу *OK*. У результаті виконаних дій у комірці **B16** відобразиться середнє значення часу долаття відстані 100 м досліджуваними легкоатлетами-спринтерами, як показано на рис. 26.

№	Підтягув.	Частота	Від. Част.	Підтягув.	Частота	Підтягув.	Від. Част.
1	9	1	7,142857	9	1	9	7,14
2	10	2	14,28571	10	2	10	14,28
3	10	2	14,28571	12	3	12	21,43
4	12	3	21,42857	13	2	13	14,28
5	12	3	21,42857	14	2	14	14,28
6	12	3	21,42857	15	2	15	14,28
7	13	2	14,28571	16	2	16	14,28
8	13	2	14,28571				
9	14	2	14,28571				
10	14	2	14,28571				
11	14	2	14,28571				
12	15	2	14,28571				
13	15	2	14,28571				
14	16	2	14,28571				
15	16	2	14,28571				
16	середнє	12,92857					
17	медіана						
18	мода						

Рис. 26. Таблиця з результатами обчислення середнього значення

Крок 8. Обчислимо медіану та моду. Для цього скористаємося піктограмою вбудованих функцій f_x , як і у випадку визначення середнього арифметичного значення. А у вікні виклику функції необхідно обрати *МЕДИАНА* та *МОДА*, відповідно. У вікні вводу даних інтервал масиву даних для обчислення медіани та моди з'являється автоматично, тому що курсор стоїть внизу стовпчика з числовими значеннями. У зв'язку з цим необхідно обов'язково перевірити правильність зазначених комірок масиву та при необхідності провести виправлення номера комірок масиву даних.

Крок 9. Побудуємо графік функції розподілення, тобто відобразимо залежність варіанти від частоти. Для цього виділяємо варіаційний ряд у стовпчиках **F** та **G**. Натискаємо піктограму *Мастер диаграмм* чи виконуємо команду *Вставка – Диаграмма*. У діалоговому вікні, що відкривається, обираємо *Точечная* діаграма (рис. 27). Натискаємо на кнопку *Готово*. Результати обчислення характеристик варіаційного ряду та графік функції розподілення представлені на рис. 28.

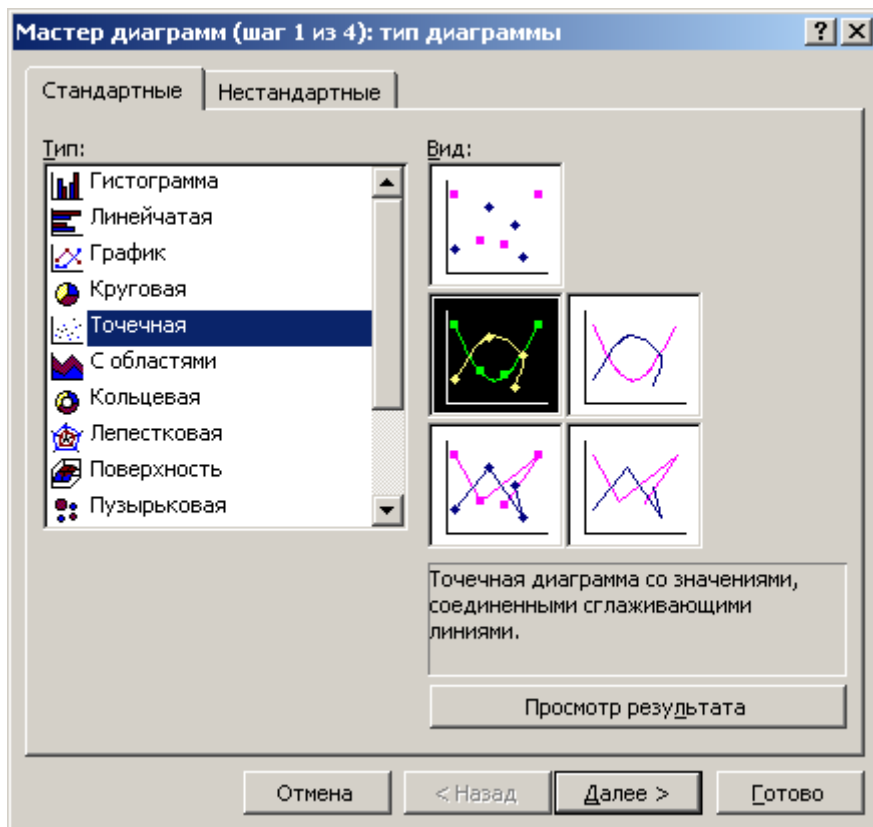


Рис. 27. Діалогове вікно «Мастер диаграмм»

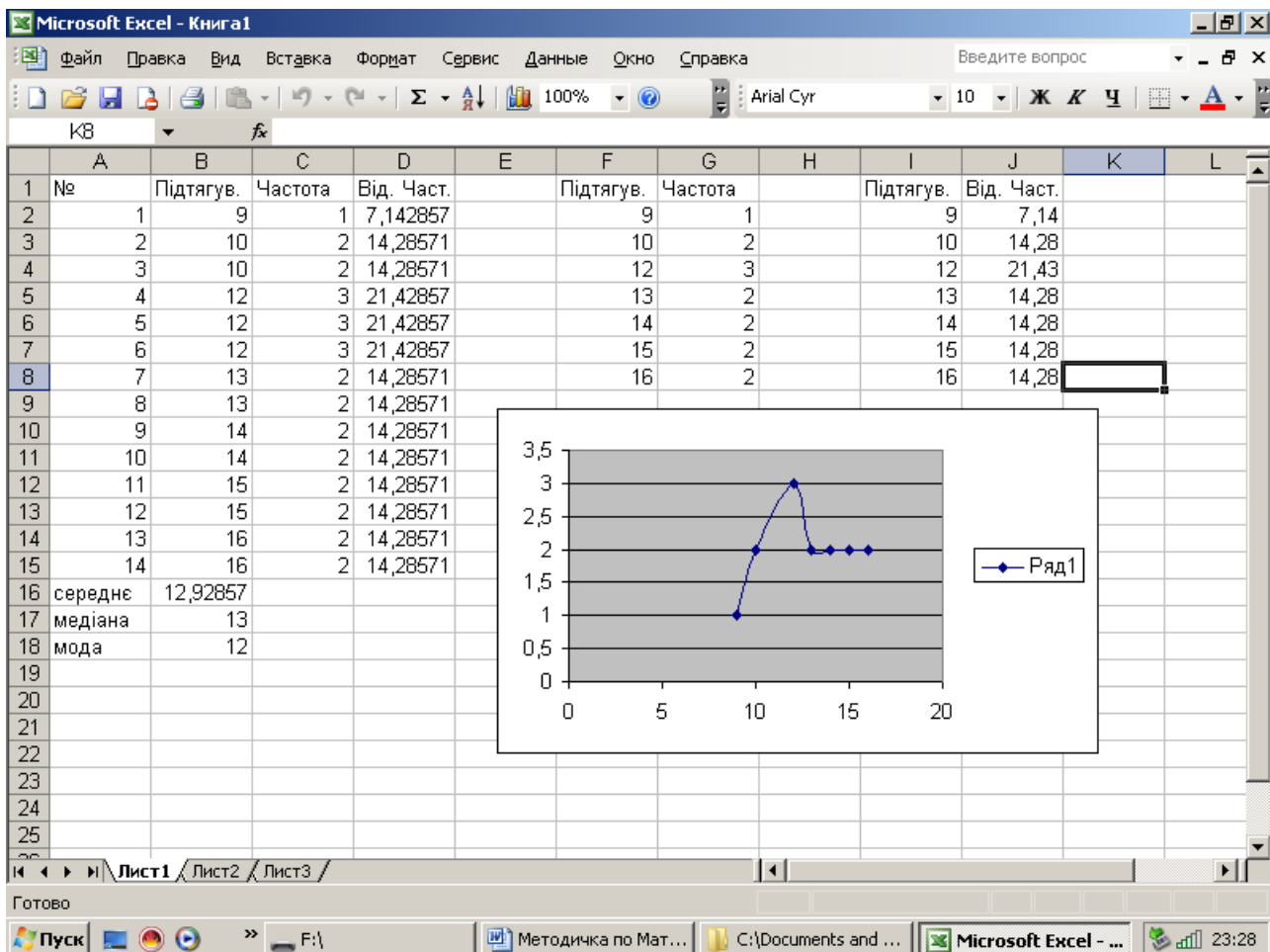


Рис. 28. Результаты обчислення та побудови графіку

Висновок: Група із 14 спортсменів, за результатами виконання тесту з підтягування у висі на перекладині, підібрана неправильно з урахуванням статі, спеціалізації та кваліфікації, оскільки значення \bar{X} , \overline{MeX} і MoX не співпадають. Графік залежності варіанти від частоти не відповідає нормальному закону розподілу. Група може брати участь у подальших дослідженнях тільки після перегляду її складу.

👉 6.2. Визначення однорідності та однотипності групи, яка підлягає дослідженню

Розглянемо послідовність операцій при визначенні однорідності та однотипності групи, яка підлягає дослідженню на конкретному прикладі.

Приклад: У 11 спортсменів було зафіксовано результати підтягування у висі на перекладині (кількість разів): 9 10 12 14 15 11 11 12 12 13 13.

Визначити, наскільки група однорідна й однотипна та чи може вона брати участь у подальших дослідженнях.

Порядок виконання

1. Ранжируваний ряд 9 10 11 11 12 12 12 13 13 14 15
2. Варіаційний ряд із частотою

	X_i	9	10	11	12	13	14	15
n_i	1	2	3	2	1	1		

3. Середнє арифметичне значення (\bar{X}).

Оскільки середнє арифметичне значення зважене, застосовуємо формулу:

$$\bar{X} = \frac{\sum_{i=1}^j X_i \times n_i}{N},$$

де \sum – знак підсумовування;

i – індекс;

j – кількість варіант у варіаційному ряду.

$$\bar{X} = \frac{9 + 10 + 11 \times 2 + 12 \times 3 + 13 \times 2 + 14 + 15}{11} = 12$$

$\bar{X} = 12$ – середній результат у підтягуванні у висі на перекладині для 11 спортсменів (у середньому кожний із 11 спортсменів виконав 12 підтягувань у висі на перекладині).

4. Дисперсія (D):

$$\sigma^2 = \frac{\sum_{i=1}^j (X_i - \bar{X})^2 n_i}{N - 1}$$

4.1 Заповнюємо таблицю для розрахунку σ^2

№ з/п	X_i	n_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 \times n_i$
1	9	1	- 3	9	9
2	10	1	- 2	4	4
3	11	2	- 1	1	2
4	12	3	0	0	0
5	13	2	1	1	2
6	14	1	2	4	4
7	15	1	3	9	9
Σ		11			30

4.2 Розраховуємо σ^2 за формулою:

$$\sigma^2 = \frac{\sum_{i=1}^j (X_i - \bar{X})^2 n_i}{N - 1}$$

$$\sigma^2 = 30 / (11-1) = 30 / 10 = 3$$

5. Середнє квадратичне відхилення (σ)

$$\sigma = \sqrt{\sigma^2} \quad \sigma = \sqrt{3} = 1,73$$

6. Допустимі границі відхилення для нашої вибірки ($\bar{X} \pm \sigma$)

$12 \pm 1,73 \rightarrow [10,27 - 13,73]$ – у дані границі входить 7 результатів з 11

7. Коефіцієнт варіації (V)

$$V = (\sigma / \bar{X}) \times 100\%, \quad V = (1,73 / 12) \times 100\% = 14,4\%$$

8. Помилка середнього арифметичного значення (m).

У даному випадку помилка середнього арифметичного значення розраховується за формулою:

$$m = \sigma / \sqrt{N-1}, \text{ тому що } N < 20 \text{ (} N = 11 \text{)}$$

$$m = 1,73 / \sqrt{11-1} = 1,73 / \sqrt{10} = 1,73 / 3,16 = 0,55$$

Висновок: коефіцієнт варіації дорівнює 14,4% – це середнє коливання результатів. Групу не можна вважати однорідною та однотипною, але група може брати участь у подальших дослідженнях, якщо відкоригувати її склад.

Виконаємо практичне завдання за допомогою редактора електронних таблиць Excel.

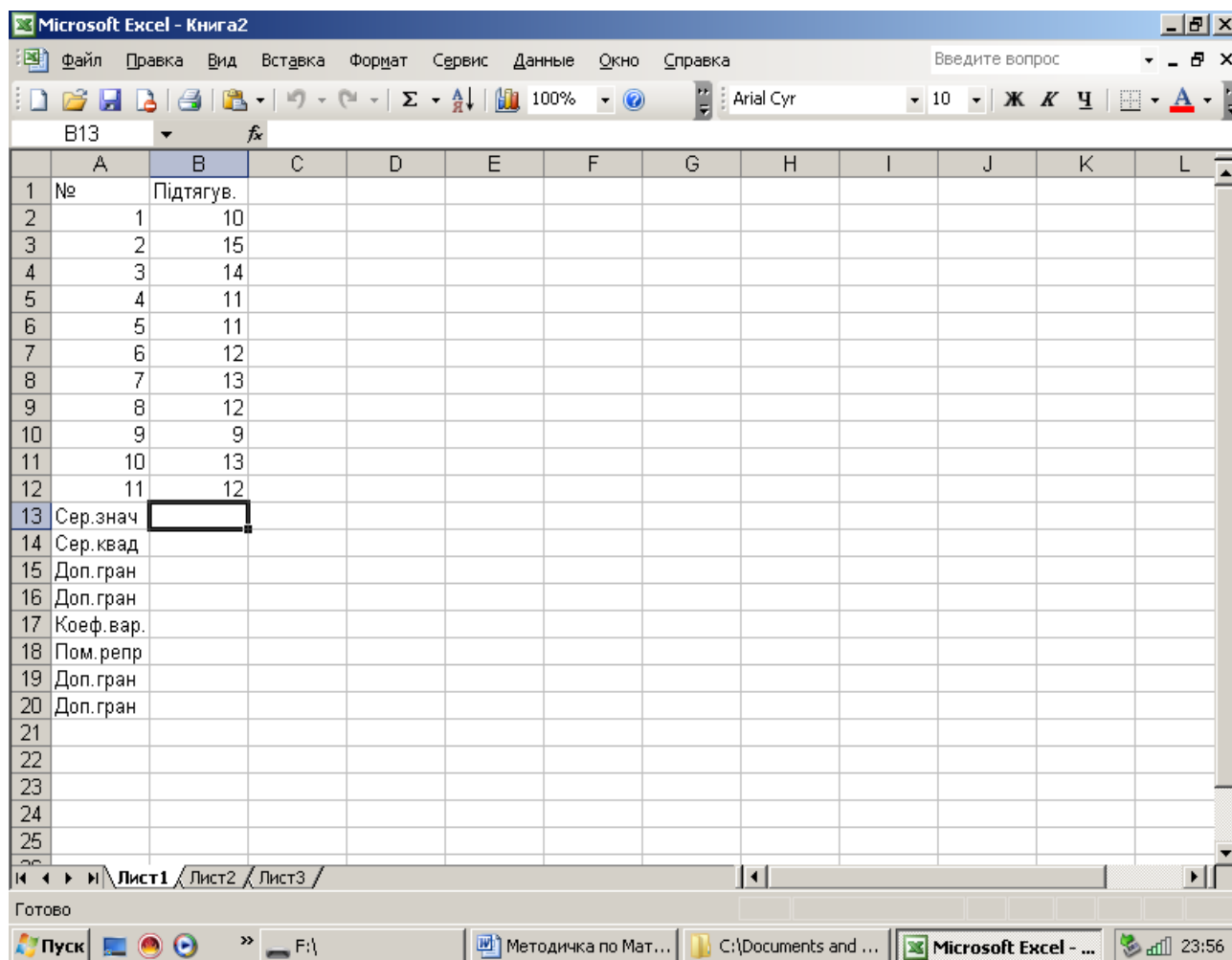
Приклад: У 11 спортсменів було зафіксовано результати підтягування у висі на перекладині (кількість разів): 9 10 12 14 15 11 11 12 12 13 13.

Визначити, наскільки група однорідна й однотипна і чи може вона брати участь у подальших дослідженнях.

Порядок виконання

Крок 1. Вводимо у діапазон **B2:B12** електронної таблиці результати підтягування у висі на перекладині (кількість разів) 11 спортсменів, тобто ряд спостережень.


Крок 2. У стовпчику **A** запишемо назви показників, які необхідно розрахувати: у комірці **A13** – середнє арифметичне; у комірці **A14** – середнє квадратичне відхилення (або стандартне відхилення); у комірках **A15** та **A16** – допустимі границі для вибіркового середнього арифметичного із зазначенням стандартного відхилення ряду спостережень; у комірці **A17** – коефіцієнт варіації; у комірці **A18** – помилку середнього арифметичного (або помилку репрезентативності); у комірках **A19** та **A20** – допустимі границі для вибіркового середнього арифметичного із зазначенням помилки репрезентативності. Таблиця матиме такий вигляд:



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Підтягув.										
2	1	10										
3	2	15										
4	3	14										
5	4	11										
6	5	11										
7	6	12										
8	7	13										
9	8	12										
10	9	9										
11	10	13										
12	11	12										
13	Сер.знач											
14	Сер.квад											
15	Доп.гран											
16	Доп.гран											
17	Коеф.вар.											
18	Пом.репр											
19	Доп.гран											
20	Доп.гран											
21												
22												
23												
24												
25												

Рис. 29. Таблица показателей 11 спортсменов

Крок 3. Обчислимо середнє арифметичне значення. Для цього виділимо комірку **B13** та скористаємося піктограмою вбудованих функцій . У результаті виконаних дій з'явиться вікно *Мастер функций*, у якому обираємо категорію *Статистические* та функцію *СРЗНАЧ* (рис.30).

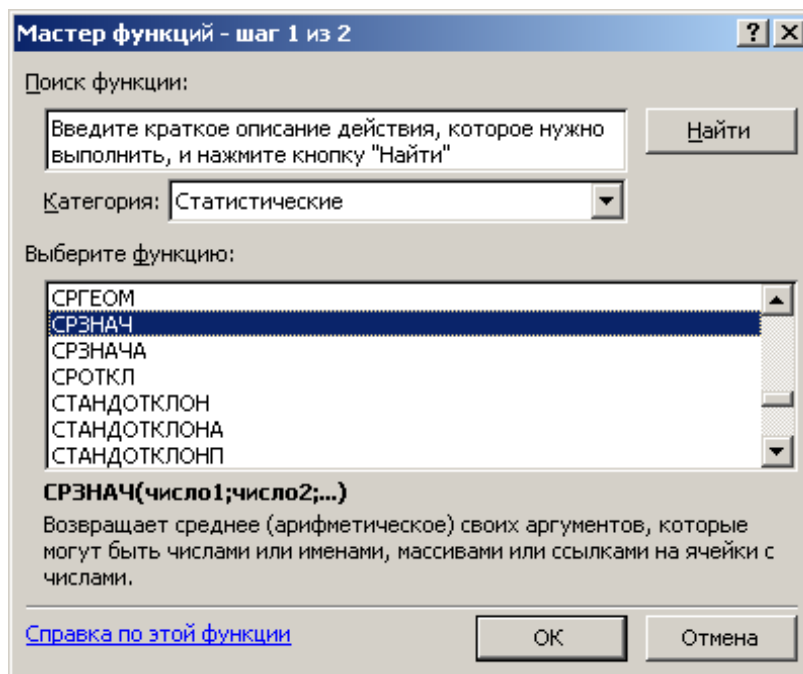


Рис. 30. Вікно виклику функції обчислення середнього значення

Натискаємо клавішу *ОК*. При цьому з'являється діалогове вікно для вводу аргументів функції (рис.31).

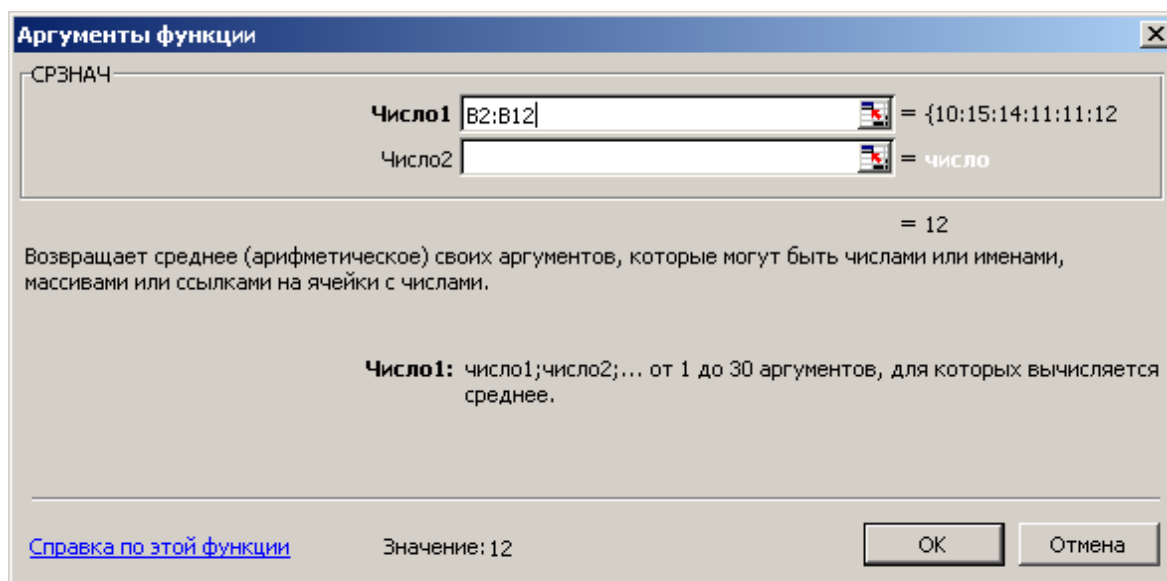


Рис. 31. Діалогове вікно вводу даних

Інтервал даних для обчислення середнього арифметичного з'являється автоматично, тому що курсор стоїть внизу стовпчика з числовими значеннями.

Однак інтервал у вікно діапазону даних можна ввести безпосередньо з клавіатури або натиснувши на піктограму праворуч цього вікна та вибравши мишею діапазон.

Натискаємо клавішу *OK*. У результаті виконаних дій у комірці **B13** відобразиться середнє значення часу додання відстані 100 м досліджуваними легкоатлетами-спринтерами.

Крок 4. Обчислимо стандартне відхилення (або середнє квадратичне відхилення) вибірки. Для цього у комірку **B14** або за допомогою *Мастер функцій*, або із клавіатури вводимо формулу = **СТАНДОТКЛОН(B2:B12)**. Вікно виклику функції обчислення стандартного відхилення зображено на рис.32.

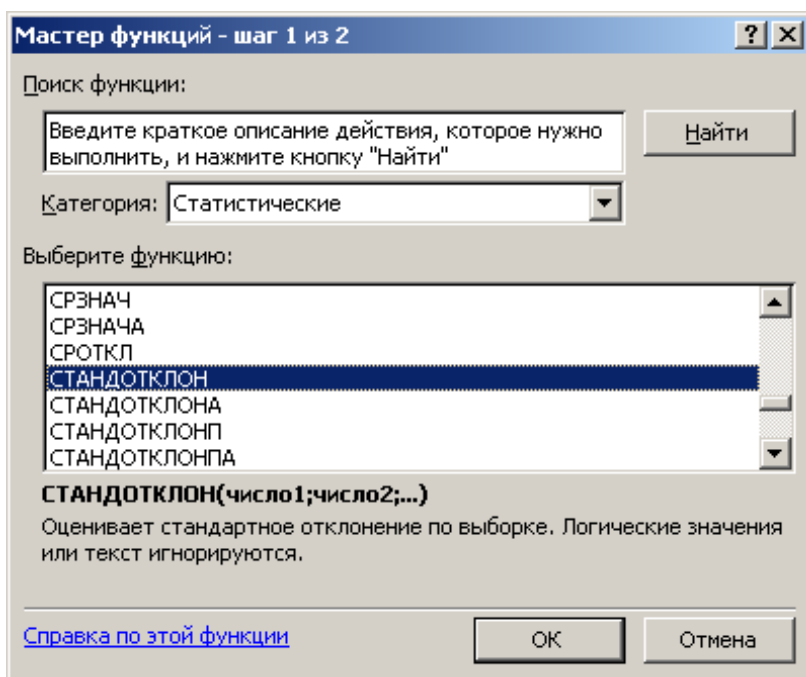


Рис. 32. Вікно виклику функції обчислення стандартного відхилення

Увага! Автоматично в діапазон даних будуть включені всі розташовані вище комірки разом з коміркою **B13** – середнім значенням. Тому завжди треба перевіряти та при необхідності коректувати діапазон даних. У кінцевому підсумку проведених у комірці відобразиться значення стандартного відхилення (рис.33).

Крок 5. Розрахуємо допустимі границі для вибіркового середнього арифметичного із зазначенням стандартного відхилення ряду спостережень. Для цього необхідно від середнього арифметичного відняти стандартне відхилення та до середнього арифметичного додати стандартне відхилення. Виділяємо комірку **B15**, маркером заходимо у *Строка формул*, вводимо формулу =**B13-B14** та натискаємо **Enter**, а у комірку **B16** вводимо формулу =**B13+B14** та натискаємо **Enter**.

Крок 6. Для того щоб розрахувати коефіцієнт варіації, необхідно стандартне відхилення (значення комірки **B14**) поділити на середнє арифметичне значення (комірка **B13**) та помножити на 100% . Виділяємо

комірку **B17**, маркером заходимо у *Строка формул* та вводимо формулу $=B14/B13*100$.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Підтягув.										
2		1	10									
3		2	15									
4		3	14									
5		4	11									
6		5	11									
7		6	12									
8		7	13									
9		8	12									
10		9	9									
11		10	13									
12		11	12									
13	Сер.знач		12									
14	Сер.кв.		1,732051									
15	Доп.гран											
16	Доп.гран											
17	Коеф.вар.											
18	Пом.репр											
19	Доп.гран											
20	Доп.гран											
21												
22												
23												
24												
25												

Рис. 33. Таблиця з обчисленим середнім значенням і стандартним відхиленням

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Підтягув.										
2		1	10									
3		2	15									
4		3	14									
5		4	11									
6		5	11									
7		6	12									
8		7	13									
9		8	12									
10		9	9									
11		10	13									
12		11	12									
13	Сер.знач		12									
14	Сер.квад		1,732051									
15	Доп.гран		10,26795									
16	Доп.гран		13,73205									
17	Коеф.вар.		14,43376									
18	Пом.репр											
19	Доп.гран											
20	Доп.гран											
21												
22												
23												
24												
25												

Рис. 34. Таблиця результатів розрахунків кроків 1-6

Крок 7. Розрахуємо помилку середнього арифметичного. Для цього значення комірки **B14** необхідно поділити на корінь квадратний з кількості значень вибірки. У цій формулі будуть використовуватися дві вбудовані функції *СЧЕТ* та *КОРЕНЬ*. Маркер знаходиться у комірці **B18**, в яку вводимо формулу $= B14/КОРЕНЬ(СЧЕТ(B2:B12)-1)$.

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Підтягув.										
2	1	10										
3	2	15										
4	3	14										
5	4	11										
6	5	11										
7	6	12										
8	7	13										
9	8	12										
10	9	9										
11	10	13										
12	11	12										
13	Сер.знач	12										
14	Сер.квад	1,732051										
15	Доп.гран	10,26795										
16	Доп.гран	13,73205										
17	Коеф.вар	14,43376										
18	Пом.репр	0,547723										
19	Доп.гран											
20	Доп.гран											
21												
22												
23												
24												
25												

Рис. 35. Таблица результатів розрахунків кроків 1-7

Крок 8. Розрахуємо допустимі границі для вибіркового середнього арифметичного із зазначенням помилки репрезентативності. Для цього необхідно від середнього арифметичного відняти помилку репрезентативності, та до середнього арифметичного додати помилку репрезентативності. Виділяємо комірку **B19**, маркером заходимо *Строка формул*, вводимо формулу **=B13-B18** та натискаємо **Enter**, а у комірку **B20** вводимо формулу **=B13+B18** та натискаємо **Enter**.

На рис. 36 представлена підсумкова таблиця практичного заняття №2.

№	Підтягув.
1	10
2	15
3	14
4	11
5	11
6	12
7	13
8	12
9	9
10	13
11	12
12	12
13	Сер.знач
14	Сер.квад 1,732051
15	Доп.гран 10,26795
16	Доп.гран 13,73205
17	Коеф.вар. 14,43376
18	Пом.репр 0,547723
19	Доп.гран 11,45228
20	Доп.гран 12,54772
21	
22	
23	
24	
25	

Рис. 36. Підсумкова таблиця даних

Висновок: коефіцієнт варіації дорівнює 14,4% – це середнє коливання результатів. Групу не можна вважати однорідною та однотипною, але група може брати участь у подальших дослідженнях, якщо переглянути її склад.

Практичне завдання:

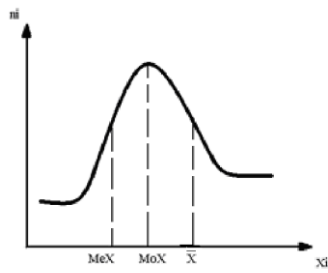
Скласти умови завдання з власного наукового дослідження і вирішити його за допомогою методу середніх величин. Проаналізувати результати, зробити висновок.

Питання для самоконтролю:

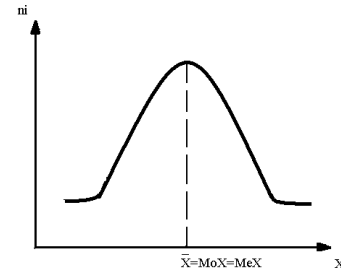
1. Розкрийте сутність методу середніх величин.
2. Назвіть основні статистичні характеристики.
3. Яким символом позначається обсяг вибірки?
4. Як позначається варіанта?
5. Що називається ранжируванням?
6. Що називається частотою варіанти?
7. Що називається варіаційним рядом?
8. Як визначається середнє арифметичне?
9. Що характеризує медіана?
10. Що характеризує мода?

11. Який із запропонованих нижче рисунків відповідає нормальному закону розподілу?

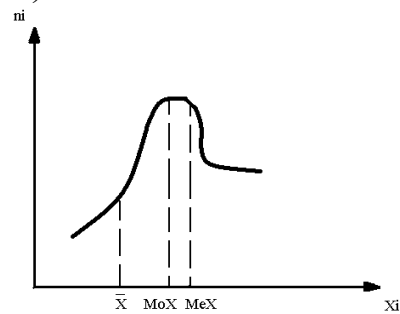
A)



Б)



В)



12. Назвіть характеристики, які відображають варіацію показників.
13. Що показує середнє квадратичне відхилення?
14. Як проводиться обчислення середнього квадратичного відхилення?
15. Назвіть характеристики, які відображають варіацію показників.
16. Що показує дисперсія? За якою формулою вона визначається?
17. Який показник відображає однорідність і однотипність групи?
18. За якою формулою визначається коефіцієнт варіації?
19. Скільком відсоткам повинен дорівнювати коефіцієнт варіації, щоб група була однорідною та однотипною?
20. Яким чином проводиться загальна оцінка масиву чисел за допомогою метода середніх величин?



Змістовий модуль 3. Вибірковий метод математичної статистики

Тема 7. Загальні положення вибіркового методу

Традиційно завдання фізичної культури і спорту розробляються таким чином, що в їх основі лежать ідеї вибіркового методу. Основний зміст методу зводиться до того, що дослідженню підлягає генеральна сукупність не в повному обсязі, а своєю репрезентативною частиною – вибірковою сукупністю. Передбачається, що вибірка з належною вірогідністю відображає генеральну сукупність тільки тоді, коли її елементи обрані з генеральної нетенденційно.

Щодо обсягу вибірки, то відповідно до основних положень математичної статистики вибірка тим більш репрезентативна, чим вона повніша. У кожному

конкретному випадку кількість об'єктів, що відбираються у вибірку, визначається індивідуально.

Основним завданням вибіркового методу є пошук двох показників – середнього арифметичного значення (\bar{X}) генеральної сукупності та середнього квадратичного відхилення (σ) генеральної сукупності.

У практиці спорту прийнято обирати надійність $P = 0,95$ і відповідний до неї рівень значущості $\alpha = 0,05$, які відображають основну групу досліджуваних завдань. У виняткових випадках, коли необхідно значно підвищити надійність розрахунків, приймається $P = 0,99$ і $\alpha = 0,01$.

Велику групу завдань вибіркового методу класифікують як групу порівняльних завдань. Порівнюються дві або більше вибірок сукупностей. При порівнянні встановлюється, належать ці вибірки до однієї і тієї ж генеральної сукупності чи до різних. Це має значення при визначенні вірогідності різниць між двома вибірковими середніми арифметичними.

Отже, при роботі вибірковим методом у практиці фізичної культури і спорту можна знайти основні середньостатистичні показники великої групи досліджуваних за допомогою вивчення незначної кількості їх представників. А також можна виявити, наскільки принциповою є різниця між кількома однотипними групами об'єктів.

За допомогою вибіркового методу математичної статистики можна оцінити:

1. Ефективність навчально-тренувального або навчально-виховного процесу.
2. Рівень підготовленості спортсмена або групи спортсменів.
3. Перевагу або ідентичність методики навчання рухових умінь і навичок, розвитку фізичних якостей.
4. Необхідність введення нового педагогічного чинника в навчально-виховний або навчально-тренувальний процес.

Вибірковий метод дозволяє порівнювати:

- показники спортсмена або групи спортсменів однієї спеціалізації та кваліфікації до і після серії тренувальних занять для виявлення зрушень у цих показниках;
- показники фізичного розвитку, фізичної підготовленості, фізичної роботоздатності основних систем організму у спортсменів і осіб, що не займаються спортом;
- показники спортсменів, які тренуються в різних умовах або за різними методиками.



Тема 8. Перевірка статистичних гіпотез

Для визначення вірогідності різниці між вибірковими середніми арифметичними необхідно обробити числа обох порівнюваних груп способом варіаційних рядів, тобто визначити в обох групах середнє арифметичне

значення (\bar{X}), середнє квадратичне відхилення (σ) і помилку середнього арифметичного значення (m).

Критерієм визначення вірогідності різниць є величина, обчислена за формулою залежно від того, які вибірки порівнюються в процесі дослідження – пов'язані або непов'язані.

Критерій вірогідності різниць – критерій Стьюдента (t_p) – розраховується за формулою:

$$t_p = \frac{|\bar{X} - \bar{Y}|}{\sqrt{m_x^2 + m_y^2}}$$

Критерій Стьюдента (t_p) порівнюється із граничним (табличним) значенням t_{zp} , яке визначається за спеціальною таблицею Стьюдента для конкретної надійності й обсягу вибірки. Для визначення t_{zp} за таблицею необхідно знати k (число ступенів свободи варіації), яке розраховується за формулою:

• у випадку, якщо дорівнюються обсяги вибірок $N_x = N_y$, а середні квадратичні відхилення не дорівнюються $\sigma_x^2 \neq \sigma_y^2$

$$k = 2 * N - 2$$

• у випадку, якщо не дорівнюються обсяги вибірок $N_x \neq N_y$ або дорівнюються середні квадратичні відхилення $\sigma_x^2 = \sigma_y^2$

$$k = N_x + N_y - 2$$

При порівнянні двох вибірових середніх арифметичних, звичайно, перевіряється припущення, що і перша, і друга вибірки належать до однієї генеральної сукупності, а отже, значною мірою не відрізняються одна від одної (порівнюються за одним показником два спортсмени, дві групи). У такому випадку бувають відомі такі статистичні характеристики: \bar{X} , \bar{Y} , σ_x^2 , σ_y^2 і обсяги вибірок N_x і N_y .

Для відповіді на питання щодо вірогідності й невірогідності різниць вибірок, які підлягають дослідженню, необхідно порівняти t розрахункове (t_p), і постійне граничне (t_{zp}).

Якщо $t_p \geq t_{zp}$ – різниця між порівнюваними вибіровими середніми арифметичними вірогідна (невипадкова), істотна і пояснюється впливом визначених чинників, а якщо $t_p < t_{zp}$ – невірогідна (випадкова).

Причини вірогідних різниць при порівнянні двох вибірових середніх арифметичних:

1. Краща підготовленість одного зі спортсменів або однієї із груп.
2. Одна з досліджуваних методик навчання рухових умінь і навичок або розвитку рухових якостей краща, більш ефективна.
3. Ефективно побудований навчально-тренувальний або навчально-виховний процес.
4. При застосуванні тієї самої методики тренування в різних умовах – причина в тому, що одні умови кращі за інші.

5. При введенні в навчально-виховний або навчально-тренувальний процес нового педагогічного чинника з метою виявлення ефективності його використання – причина в ефективності застосування експериментального педагогічного чинника.

Причини невірогідних різниць при порівнянні двох вибірових середніх арифметичних:

1. Неправильний підбір вибірки.
2. Недостатня чисельність вибірки.
3. Однаковий рівень підготовленості обох груп, які підлягають дослідженню, або обох спортсменів.

4. Неefективно побудований навчально-тренувальний або навчально-виховний процес.

5. У випадку порівняння двох методик навчання рухових умінь і навичок або розвитку рухових якостей – ефект однаковий, методики ідентичні, не має значення, за якою методикою працювати.

6. При введенні в навчально-тренувальний або навчально-виховний процес нового педагогічного чинника з метою виявлення ефективності його застосування – причина в негативному впливі нового педагогічного чинника на ефективність навчально-тренувального або навчально-виховного процесу.

7. У випадку застосування однієї методики в різних умовах, причина в однаковому результаті при тренування в різних умовах, немає істотних різниць, у яких умовах застосовувати досліджувану методику.

Для порівняння двох вибірових характеристик варіації необхідно знати розрахункове значення критерію Фішера F і критичне значення теоретичного розподілу Фішера $F_{гр}$.

Основні властивості критерію Фішера:

1. За даним критерієм порівнюються різні за обсягом малі й середні вибірки.

2. При порівнянні за цим критерієм дотримання нормального закону розподілу (правильності підбору групи) необов'язкове, але спеціалізація і стать повинні бути однаковими.

3. За цим критерієм вибірки порівнюються на стабільність.

4. Розрахункове значення критерію Фішера розраховується за формулою:

$$F = \frac{\sigma^2_1}{\sigma^2_2}$$

5. Критичне значення теоретичного розподілу Фішера – $F_{критич}$ визначається за таблицею. Для цього необхідно знати число ступенів свободи більшої та меншої дисперсії k_1 і k_2 .

k_1 і k_2 розраховують за формулами:

$k_1 = N_1 - 1$, де N_1 – обсяг вибірки більшої дисперсії.

$k_2 = N_2 - 1$, де N_2 – обсяг вибірки меншої дисперсії.

За числом ступенів свободи більшої й меншої дисперсії визначаємо за таблицею критичне значення теоретичного розподілу Фішера $F_{гр}$ на перетині k_1 та k_2 .

6. Якщо $F > F_{гр}$, то розходження між двома вибірками вірогідні (невипадкові). Якщо $F \leq F_{гр}$, то розходження між двома вибірками невірогідні (випадкові).

Модельні характеристики – це ідеальні характеристики стану спортсмена, в якому він може показати результати, що відповідають вищим світовим досягненням.

Знати модельні характеристики необхідно для визначення напрямків тренувальної роботи та проведення відбору спортсменів.

Модельні характеристики поділяються на:

а) консервативні (не піддаються тренуванню; наприклад, розміри тіла в довжину) і неконсервативні (змінювані у процесі тренування; наприклад, силові якості);

б) компенсовані й некомпенсовані;

в) етапні.

Компенсовані – показники, низький рівень яких може бути компенсований високим рівнем інших показників. Наприклад, низька ефективність баскетболіста у грі під щитом, яка пов'язана з його невисоким зростом, може компенсуватися точністю кидків з дальньої відстані так, що загальна кількість закинутих м'ячів буде досить високою.

Некомпенсовані – показники, низький рівень яких не можна компенсувати високим рівнем іншого показника. Наприклад, низький рівень будь-якого з показників серцево-судинної системи у лижника-гонщика не може бути компенсований ніякою технікою.

У переважній більшості випадків ми стикаємося із частково компенсованими показниками: невеликі відставання в розвитку однієї якості компенсуються, великі – ні.

Існує три основних шляхи визначення модельних характеристик:

1. Дослідження спортсменів високого класу.
2. Розрахунок так званих належних показників.
3. Прогнозування модельних характеристик.

Якщо немає можливості визначити величини модельних характеристик, використовують метод експертних оцінок.

Середня арифметична генеральної сукупності $\bar{x}_{ген}$ визначається відповідно до нижніх і верхніх меж довірчих інтервалів:

$$\bar{x}_{виб} - mt \leq \bar{x}_{ген} \leq \bar{x}_{виб} + mt, \quad (1)$$

де $\bar{x}_{виб}$ – середня арифметична вибіркової сукупності;

m – помилка репрезентативності;

t – критерій надійності, тобто показник обраної довірчої ймовірності.

Помилка репрезентативності свідчить про відхилення параметрів вибірки, зокрема середньої арифметичної, від відповідних параметрів генеральної сукупності. Про величину цієї помилки можна говорити з певною ймовірністю, на величину якої вказує критерій t .

Величина помилки репрезентативності може бути знайдена в такий спосіб:

$$m = \frac{\sigma_{\text{виб}}}{\sqrt{n-1}} \quad \text{для } n < 20 \quad (2)$$

$$m = \frac{\sigma_{\text{виб}}}{\sqrt{n-1}} \sqrt{\frac{N-n}{N-1}}, \quad (3)$$

де $\sigma_{\text{виб}}$ – середнє квадратичне відхилення вибіркової сукупності;
 n – обсяг вибірки;
 N – обсяг генеральної сукупності.

$$m = \frac{\sigma_{\text{виб}}}{\sqrt{n}} \quad \text{для } n > 20 \quad (4)$$

Помилка репрезентативності відповідно до формули для досить великих N спрощується:

$$m = \frac{\sigma_{\text{виб}}}{\sqrt{n}} \sqrt{\left(1 - \frac{n}{N}\right)} \quad (3a)$$

Значення критерію вірогідності відповідає довірчій ймовірності та визначається за таблицею Стюдента.

У спортивній практиці використовуються такі надійності:

$$P_1 = 0,95 \quad \text{і} \quad P_2 = 0,99$$

Середнє квадратичне відхилення генеральної сукупності ($\sigma_{\text{ген}}$) визначається з нерівності:

$$\sigma_{\text{виб}}(1 - q) \leq \sigma_{\text{ген}} \leq \sigma_{\text{виб}}(1 + q), \quad (5)$$

де $\sigma_{\text{виб}}$ – середнє квадратичне відхилення вибіркової сукупності;

q – величина, яка визначається за таблицею.

Таблиця значень q для визначення довірчих інтервалів σ генерального.

Надійність $P = 0,95$, n – обсяг вибірки.

n	q	n	q	n	q
5	1,37	15	0,46	45	0,22
6	1,09	16	0,44	50	0,21
7	0,92	17	0,42	60	0,188
8	0,80	18	0,40	70	0,174
9	0,71	19	0,39	80	0,161
10	0,65	20	0,37	90	0,151
11	0,59	25	0,32	100	0,143
12	0,55	30	0,28	150	0,115
13	0,52	35	0,26	200	0,099
14	0,48	40	0,24	250	0,089

Тема 9. Практична реалізація вибіркового методу при обробці й аналізі даних у сфері фізичної культури та спорту

9.1 Порівняння двох середніх арифметичних за допомогою критерію Стьюдента

Розглянемо послідовність операцій при порівнянні двох середніх арифметичних за допомогою критерію Стьюдента на конкретному прикладі.

Приклад: У двох групах плавців X_i та Y_i визначили різницю ЧСС, уд/хв після максимально швидкого проходження дистанції та у стані спокою. Встановити, чи вірогідна різниця за показником ЧСС у плавців двох груп.

X_i	92 94 95 97 99 100	$N_x = 19$
n_i	3 4 7 2 1 2	
Y_i	98 102 103 104 105	$N_y = 17$
n_i	3 5 6 1 2	

Порядок виконання

1. Визначаємо середнє арифметичне значення 2-х вибірок

$$\bar{X} = \frac{\sum X_i \cdot n_i}{N_x} \quad \bar{Y} = \frac{\sum Y_i \cdot n_i}{N_y}$$

$$\bar{X} = \frac{92 \times 3 + 94 \times 4 + 95 \times 7 + 97 \times 2 + 99 + 100 \times 2}{19} = 96 \text{ (уд/хв)}$$

$$\bar{Y} = \frac{98 \times 3 + 102 \times 5 + 103 \times 6 + 104 + 105 \times 2}{17} = 102 \text{ уд/хв}$$

2. Визначаємо дисперсії 2-х вибірок

$$\sigma^2_x = \frac{\sum (X_i - \bar{X})^2 n_i}{N_x - 1} \quad \sigma^2_y = \frac{\sum (Y_i - \bar{Y})^2 n_i}{N_y - 1}$$

Для подальшої роботи креслимо таблиці

а) Для показників першої вибірки

X_i	n_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 \times n_i$
92	3	- 4	16	48
94	4	- 2	4	16

95	7	- 1	1	7
97	2	1	1	2
99	1	3	9	9
100	2	4	16	32
Σ				114

$$\sigma^2_x = \frac{114}{19-1} = 6,33$$

б) Для показників другої вибірки

Y_i	n_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^2 \times n_i$
98	3	84	16	48
102	5	0	0	0
103	6	1	1	6
104	1	2	4	4
105	2	3	9	18
Σ				76

$$\sigma^2_y = \frac{76}{17-1} = 4,75$$

$$\sigma^2_x = 6,33$$

$$\sigma^2_y = 4,75$$

3. Визначаємо середнє квадратичне відхилення 2-х вибірок

$$\sigma_x = \sqrt{\sigma^2_x} \quad \sigma_x = \sqrt{6,33} = 2,52$$

$$\sigma_y = \sqrt{\sigma^2_y} \quad \sigma_y = \sqrt{4,75} = 2,18$$

4. Визначасмо t за формулою:

$$t = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma^2_x}{N_x} + \frac{\sigma^2_y}{N_y}}}, \text{ тому що } N_x \neq N_y \text{ і } \sigma^2_x \neq \sigma^2_y$$

$$t = \frac{|96 - 102|}{\sqrt{\frac{6,33}{19} + \frac{4,75}{17}}} = \frac{|-6|}{\sqrt{0,3331 + 0,2794}} = \frac{6}{\sqrt{0,6125}} = \frac{6}{0,78} = 7,69$$

5. Визначаємо число ступенів свободи (K)

$$K = N_x + N_y - 2, \text{ тому що } N_x \neq N_y \text{ і } \sigma^2_x \neq \sigma^2_y$$

$$K = 19 + 17 - 2 = 34$$

При $K = 34$, $t_{\text{гр}} = 2,03$ (див. додаток А).

6. Порівнюємо t розрахункове і t граничне.

$t > t_{гр} (7,69 > 2,03)$ – різниця між двома досліджуваними групами плавців вірогідна (невипадкова).

Висновок: Оскільки $t > t_{гр} (7,69 > 2,03)$ – розбіжність між двома групами плавців за показником різниці ЧСС, уд/хв після максимально швидкого проходження дистанції та у стані спокою статистично вірогідна (невипадкова) і пояснюється кращим рівнем підготовленості спортсменів першої групи, так як $\bar{X} < \bar{Y} (96 < 102)$.

Виконаємо практичне завдання за допомогою редактора електронних таблиць Excel.

Приклад: У двох групах плавців X_i і Y_i визначили різницю ЧСС, уд/хв після максимально швидкого проходження дистанції та у стані спокою. Встановити, чи вірогідна різниця за показником ЧСС у плавців двох груп.

X_i	92	94	95	97	99	100	$N_x = 19$
n_i	3	4	7	2	1	2	
Y_i	98	102	103	104	105		$N_y = 17$
n_i	3	5	6	1	2		

Порядок виконання

Крок 1. Вводимо у стовпчик **В** електронної таблиці результати різниці ЧСС після максимально швидкого подолання дистанції та у стані спокою першої групи плавців (значення X); у стовпчик **С** - результати різниці ЧСС другої групи плавців (значення Y). При введенні даних обов'язково враховуємо кількість повторень кожної варіанти.

Крок 2. У стовпчику **А** запишемо порядкові номери варіант та назви показників, які необхідно розрахувати. У комірці **A21** запишемо середнє арифметичне; у комірці **A22** – середнє квадратичне відхилення (або стандартне відхилення); у комірці **A23** – помилку середнього арифметичного (або помилку репрезентативності); у комірці **A24** – розрахункове значення критерію Стьюдента (t розр.); у комірці **A25** – критичне (табличне) значення критерію Стьюдента (t крит.). Таблиця матиме такий вигляд:

№	ЧСС, уд/х	ЧСС, уд/х
1	92	98
2	92	98
3	92	98
4	94	102
5	94	102
6	94	102
7	94	102
8	94	102
9	95	102
10	95	103
11	95	103
12	95	103
13	95	103
14	95	103
15	95	103
16	97	104
17	97	105
18	99	105
19	100	
20	100	
21	середнє	
22	сер. квадр	
23	помилка	
24	t розрах.	
25	t критич.	

Рис. 37. Таблиця даних різниці ЧСС двох груп плавців

Крок 3. Обчислимо середнє арифметичне значення. Для цього виділимо комірку **B21** та скористаємося піктограмою вбудованих функцій f_x . При цьому з'явиться вікно *Мастер функций*, у якому обираємо категорію *Статистические* та функцію **СРЗНАЧ**.

Натискаємо клавішу **OK**. При цьому з'являється діалогове вікно для вводу аргументів функції. Інтервал даних для обчислення середнього арифметичного з'являється автоматично, оскільки курсор стоїть внизу стовпчика з числовими значеннями. Однак інтервал у вікно діапазону даних можна ввести безпосередньо з клавіатури або натиснувши на піктограму праворуч цього вікна та вибравши мишею діапазон.

Натискаємо клавішу **OK**. У результаті виконаних дій у комірці **B21** відобразиться середнє значення різниці ЧСС, уд/хв після максимально швидкого подолання дистанції та в стані спокою першої групи плавців (значення X).

Крок 4. Для розрахунку середнього арифметичне значення різниці ЧСС, другої групи плавців (значення Y) копіюємо комірку **B21** у комірку **C21**. Обов'язково в *Строка формул* робимо виправлення діапазону даних стовпчика **C**. Масив даних другої вибірки відповідає **C2:C18**.

У результаті виконаних дій у комірці **C21** відобразиться середнє значення різниці ЧСС після максимально швидкого подолання дистанції та в стані спокою другої групи плавців (значення Y).

Крок 5. Обчислюємо стандартне відхилення (або середнє квадратичне відхилення) першої вибірки у комірці **B22**. Для цього в комірку **B22** за

допомогою *Мастер функций* вводимо функцію **СТАНДОТКЛОН** та зазначаємо масив **B2:B20**.

Крок 6. Обчислюємо стандартне відхилення (або середнє квадратичне відхилення) другої вибірки. Для цього копіюємо комірку **B22** у комірку **C22** та обов'язково коректуємо масив. Для даного прикладу він повинен бути **C2:C18**.

Крок 7. Розраховуємо помилку середнього арифметичного (помилку репрезентативності) першої вибірки. Для цього значення комірки **B22** необхідно поділити на корінь квадратний з кількості значень вибірки. У цій формулі будуть використовуватися дві вбудовані функції **СЧЕТ** та **КОРЕНЬ**. Маркер знаходиться у комірці **B23**, в яку вводимо формулу = **B22/КОРЕНЬ(СЧЕТ(B2:B20)-1)**.


Крок 8. Розраховуємо помилку середнього арифметичного (помилку репрезентативності) другої вибірки. Для цього комірку **B23** необхідно скопіювати в комірку **C23** та виправити масив даних функції **СЧЕТ** у рядку формул. Формула повинна мати такий вигляд = **C22/КОРЕНЬ(СЧЕТ(C2:C18)-1)**. Після виконаних дій у комірці **C23** відобразяться значення помилки середнього арифметичного другої вибірки

У результаті виконаних кроків **3-8** отримуємо таблицю з обчисленими середнім арифметичним значенням, стандартним відхиленням (або середнім квадратичним відхиленням) та помилкою репрезентативності (помилкою середнього арифметичного) для двох груп плавців, у яких фіксувалася різниця ЧСС після максимально швидкого подолання дистанції та у стані спокою (рис 38).

№	ЧСС, уд/х	ЧСС, уд/х
1	92	98
2	92	98
3	92	98
4	94	102
5	94	102
6	94	102
7	94	102
8	95	102
9	95	103
10	95	103
11	95	103
12	95	103
13	95	103
14	95	103
15	95	103
16	97	104
17	97	105
18	99	105
19	100	
20	100	
середнє	95,26316	102,1176
сер. квадрат.	2,400049	2,176073
помилка	0,565697	0,544018
t розрах.		
t критич.		

Рис. 38 Таблица результатів виконання кроків 3-8

Крок 9. Обчислюємо *t*-критерій Стьюдента (*t* розр.) за формулою наданою у практичній роботі. Для цього виділяємо комірку **B24**, як показано на рис.38, та в рядку формул записуємо формулу: $= (B21 - C21) / \text{КОРЕНЬ}(B23^2 + C23^2)$. Натискаємо **Enter** і в комірці **B24** з'явиться значення розрахункового критерію Стьюдента для двох вибірок. Слід пам'ятати, що критерій Стьюдента має лише позитивні значення, тому в робочих зошитах знак «-» ніколи не записується. Так, у цьому випадку *t* розрахункове дорівнює 8,98709, або **t=8,99**.

Крок 10. Обчислимо критичне (табличне) значення критерію Стьюдента (*t* крит.) за допомогою засобів **Excel**. Для цього необхідно застосувати функцію **СТЮДРАСПОБР**. Ставимо маркер у комірку **B25**, викликаємо *Мастер функций* за допомогою піктограми  та обираємо функцію **СТЮДРАСПОБР** (рис.39).

У вікні вводу даних аргументів функції вводимо у першому рядку ймовірність **0,05**, а у другому рядку – число ступенів свободи **34**, після чого натискаємо **OK**. У результаті виконаних дій у комірці **B25** з'являється значення критичного (табличного) критерію Стьюдента.

Результати проведених розрахунків подано на рис. 41.

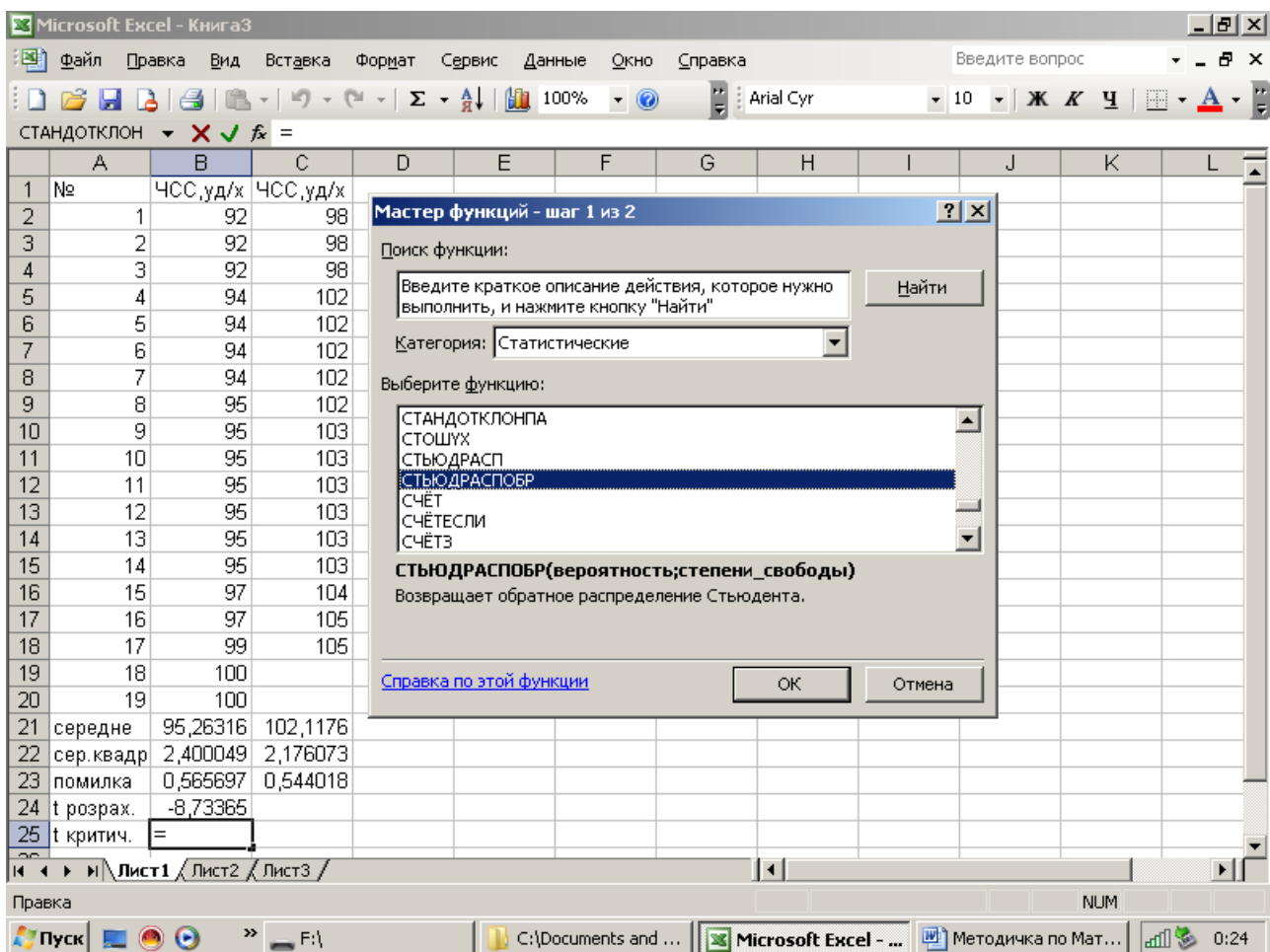


Рис.39. Виклик функції для розрахунку табличного значення критерію Стьюдента

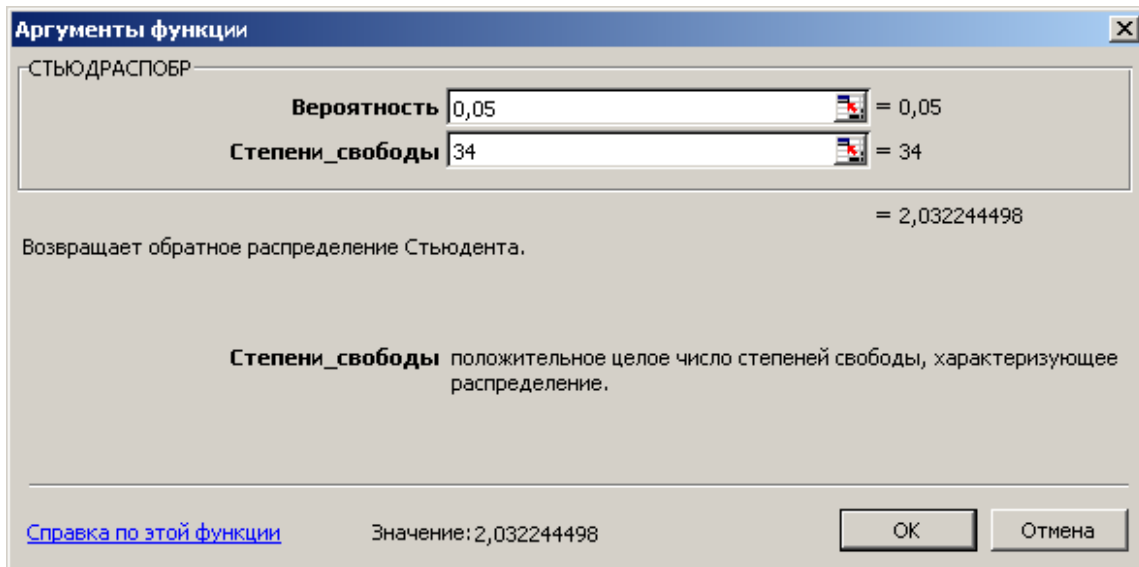


Рис. 40. Диалогове вікно вводу даних для функції *СТЬЮДРАСПОБР*

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	ЧСС, уд/х	ЧСС, уд/х									
2	1	92	98									
3	2	92	98									
4	3	92	98									
5	4	94	102									
6	5	94	102									
7	6	94	102									
8	7	94	102									
9	8	95	102									
10	9	95	103									
11	10	95	103									
12	11	95	103									
13	12	95	103									
14	13	95	103									
15	14	95	103									
16	15	97	104									
17	16	97	105									
18	17	99	105									
19	18	100										
20	19	100										
21	середне	95,26316	102,1176									
22	сер. квадр	2,400049	2,176073									
23	помилка	0,565697	0,544018									
24	t розрах.	-8,73365										
25	t критич.	2,032244										

Рис. 41. Підсумкова таблиця розрахунків для проведення порівняння двох вибірових середніх арифметичних

Крок 11. Порівнюємо t розрахункове і t граничне (критичне).

$t > t_{гр}(8,73 > 2,03)$ – різниця між двома досліджуваними групами плавців вірогідна (невипадкова).

Висновок: Оскільки $t > t_{гр}(8,73 > 2,03)$ – розбіжність між двома групами плавців за показником різниці ЧСС, уд/хв після максимально швидкого проходження дистанції та у стані спокою вірогідна (невипадкова) і пояснюється кращим рівнем підготовленості спортсменів першої групи, так як $\bar{X} < \bar{Y}$ ($96 < 102$).

👉 9.2 Визначення середньостатистичних показників генеральної сукупності (модельних характеристик)

Розглянемо послідовність операцій при визначенні середньостатистичних показників генеральної сукупності (модельних характеристик) на конкретному прикладі.

Приклад: X_i – різниця (с) між звичайним бігом і бігом з бар'єрами на 110 м. Досліджувались $n=36$ спортсменів, які були відібрані з $N=500$ студентів.

Визначити основні середньостатистичні показники студентів (модельну характеристику).

X_i	1,8	1,9	2,0	2,1	2,2
n_i	6	6	10	9	5

Порядок виконання

Заповнюємо таблицю для розрахунку \bar{x} , σ

X_i	n_i	$X_i n_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 n_i$
1,8	6	10,8	- 0,2	0,04	0,24
1,9	6	11,4	- 0,1	0,01	0,06
2,0	10	20,0	0	0	0
2,1	9	18,9	0,1	0,01	0,09
2,2	5	11,0	0,2	0,04	0,2
	36	72,1			0,59

$$\bar{X} = \frac{72,1}{36} = 2,0$$

$$\sigma^2 = \frac{0,59}{36} = 0,016$$

$$\sigma = \sqrt{0,016} = 0,13$$

Беручи до уваги допоміжні обчислення, що були виконані в таблиці, визначаємо основні показники вибірки $\bar{X}_{виб}=2,0$ с; $\sigma_{виб}=0,13$ с. Для розрахунку відповідних генеральних показників скористаємося формулами (1), (3а).

Необхідно знайти середнє арифметичне генеральної сукупності $\bar{X}_{ген}$. Для її визначення знаходимо помилку репрезентативності m :

$$t = \frac{0,13}{\sqrt{36}} \sqrt{\left(1 - \frac{36}{500}\right)} = 0,02c$$

Задаємося надійністю $P = 0,95$, при якій критерій вірогідності $t = 2,02$ (для $n = 40$ – найближчого і $n = 36$ у таблиці Стьюдента (див. додаток А)).

$$2,0 - 0,02 \times 2,02 \leq \bar{x}_{\text{ген}} \leq 2,0 + 0,02 \times 2,02;$$

$$2,0 - 0,04 \leq \bar{x}_{\text{ген}} \leq 2,0 + 0,04;$$

$$1,96 \leq \bar{x}_{\text{ген}} \leq 2,04$$

Якщо для подальших розрахунків потрібно не інтервальне, а дискретне значення $\bar{X}_{\text{ген}}$, то воно визначається як середнє зазначеного інтервалу або як одне з його крайніх значень.

Для визначення $\sigma_{\text{виб}}$ користуємося таблицею значень q .

При надійності $P = 0,95$ і величині вибірки $n = 36$, величина $q = 0,26$. Тому $\sigma_{\text{виб}}$ знаходиться в межах:

$$0,12 (1 - 0,26) \leq \sigma_{\text{ген}} \leq 0,12 (1 + 0,26);$$

$$0,09 \leq \sigma_{\text{ген}} \leq 0,15.$$

Висновок: У межах від 1,96 с до 2,04 с має знаходитися значення певної модельної характеристики – різниці (с) між звичайним бігом і бігом з бар'єрами на 110 м. При остаточному призначенні модельної характеристики обирають те із крайніх значень, яке більш жорстке у виконанні. У цьому випадку, значення модельної характеристики, яке визначається різницею між звичайним бігом і бігом з бар'єрами на 110 м, дорівнює 1,96 с; $\sigma_{\text{ген}}$ знаходиться в межах від 0,09 до 0,15 с.

Виконаємо практичне завдання за допомогою редактора електронних таблиць Excel.

Приклад: X_i – різниця, (с) між звичайним бігом і бігом з бар'єрами на 110 м. Досліджувались $n = 36$ спортсменів, які були відібрані з $N = 500$ студентів.


Визначити основні середньостатистичні показники студентів (модельну характеристику).

X_i	1,8	1,9	2,0	2,1	2,2
n_i	6	6	10	9	5

Порядок виконання


Крок 1. Вводимо у стовпчик **В** електронної таблиці результати показників різниці, (с) між звичайним бігом та бігом з бар'єрами на 110 м у 36 спортсменів, які були відібрані з 500 студентів. При введенні даних обов'язково враховуємо кількість повторень кожної варіанти.

Крок 2. У стовпчику **A** запишемо порядкові номери варіант та назви показників, які необхідно розрахувати. У комірці **A38** запишемо середнє арифметичне; у комірці **A39** – середнє квадратичне відхилення (або стандартне відхилення); у комірці **A40** – помилку середнього арифметичного (або помилку репрезентативності); у комірці **A41** – критичне (табличне) значення критерію Стьюдента (*t* критич.); у комірках **A42** та **A43** – границі для середнього арифметичного генеральної сукупності; у комірках **A44** та **A45** – границі для середнього квадратичного відхилення генеральної сукупності.

Крок 3. Обчислимо вибіркве середнє арифметичне значення. Для цього виділимо комірку **B38** та скористаємося піктограмою вбудованих функцій . При цьому з'явиться вікно *Мастер функций*, у якому обираємо категорію *Статистические* та функцію **СРЗНАЧ** (послідовність дій така сама, що і в попередніх роботах).

Крок 4. Обчислюємо вибіркве стандартне відхилення (або середнє квадратичне відхилення) у комірці **B39**. Для цього в комірку **B39** за допомогою *Мастер функций* вводимо функцію **СТАНДОТКЛОН** та зазначаємо масив **B2:B37**.

Крок 5. Розраховуємо помилку середнього арифметичного (помилку репрезентативності) вибірки. Виділяємо комірку **A40** та, зважаючи на те, що в завданні відомий обсяг генеральної сукупності, у рядку формул записуємо формулу = **B39/КОРЕНЬ(СЧЕТ(B2:B37))*КОРЕНЬ(1-СЧЕТ(B2:B37)/500)**. Натискаємо **Enter** і в комірці **B40** з'явиться значення помилки середнього арифметичного.

Крок 6. Обчислюємо критичне (табличне) значення критерію Стьюдента (*t* табл.) за допомогою засобів **Excel**. Для цього необхідно застосувати функцію **СТЬЮДРАСПОБР**. Ставимо маркер у комірку **B41**, викликаємо *Мастер функций* за допомогою піктограми  та обираємо функцію **СТЬЮДРАСПОБР** (див. практичну роботу № 3, крок 10).

У вікні вводу даних аргументів функції вводимо в першому рядку ймовірність **0,05**, а у другому рядку – число ступенів свободи **36** (відповідно до умов задачі) та натискаємо **OK**. У результаті цих дій у комірці **B41** з'являється значення критичного (табличного) критерію Стьюдента.

Крок 7. Розраховуємо границі для середнього арифметичного генеральної сукупності. Для цього необхідно від вибіркового середнього арифметичного відняти добуток помилки репрезентативності та *t* табл. та до середнього арифметичного додати добуток помилки репрезентативності та *t* табл. Виділяємо комірку **B42**, маркером заходимо у *Строка формул*, вводимо формулу =**B38-B40*B41** та натискаємо **Enter**, а у комірку **B43** вводимо формулу =**B38+B40*B41** і натискаємо **Enter**.

Крок 8. Розраховуємо границі для середнього квадратичного відхилення генеральної сукупності. Відповідно до формули (5), яка надається у практичній роботі в комірку **B44** записуємо формулу = **B39*(1-0,26)**, а в комірку **B45** записуємо формулу = **B39*(1+0,26)**.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Час, с										
2		1,8										
3		1,8										
4		1,8										
5		1,8										
6		1,8										
7		1,8										
8		1,9										
9		1,9										
10		1,9										
11		1,9										
35		2,2										
36		2,2										
37		2,2										
38	Сер.знач	2,002778										
39	Сер.квад	0,129804										
40	Пом.репр	0,020841										
41	t критич.	2,028094										
42	Доп.гран.	1,960511										
43	Доп.гран.	2,045045										
44	Доп.гран.	0,096055										
45	Доп.гран.	0,163554										
46												
47												
48												

Рис. 42. Підсумкова таблиця даних

У результаті всіх виконаних розрахунків отримуємо підсумкову таблицю даних (рис.42).

Крок 9. Записуємо висновок відповідно до результатів розрахунків.

Висновок: У межах від 1,96 с до 2,04 с має знаходитися значення певної модельної характеристики – різниці (с) між звичайним бігом і бігом з бар'єрами на 110 м. При остаточному призначенні модельної характеристики обирають те із крайніх значень, яке більш жорстке у виконанні. У цьому випадку значення модельної характеристики, яке визначається різницею між звичайним бігом і бігом з бар'єрами на 110 м, дорівнює 1,96 с; $\sigma_{\text{ген}}$ знаходиться в межах від 0,09 с до 0,15 с.

👉 9.3 Порівняння двох вибірковоїх характеристик варіації за критерієм Фішера

Розглянемо послідовність операцій при порівнянні двох вибірковоїх характеристик варіації за критерієм Фішера на конкретному прикладі.

Приклад: У двох групах плавців X_i і U_i під час тривалої роботи на витривалість вимірювали кількість поглиненого кисню, л/хв. Визначити, чи принципово є відмінність між цими групами за стабільністю поглинання кисню.

X _i	4,0	4,1	4,2	4,4	4,6	4,7	
n _i	7	9	8	10	4	3	N _x =41
Y _i	4,0	4,3	4,6	4,5	4,7		
n _i	2	4	3	9	1		N _y =19

Порядок виконання

1. Розраховуємо середнє арифметичне значення першої та другої вибірки:

$$\bar{X} = \frac{\sum X_i n_i}{N_x} = \frac{4,0 \times 7 + 4,1 \times 9 + 4,2 \times 8 + 4,4 \times 10 + 4,6 \times 4 + 4,7 \times 3}{41} = \frac{175}{41} = 4,3$$

$$\bar{Y} = \frac{\sum Y_i n_i}{N_y} = \frac{4,0 \times 2 + 4,3 \times 4 + 4,5 \times 9 + 4,6 \times 3 + 4,7}{19} = \frac{84,2}{19} = 4,4$$

2. Розраховуємо дисперсії для першої й другої вибірки

$$\sigma^2_x = \frac{\sum (X_i - \bar{X})^2 n_i}{N_x - 1} = \frac{(4,0 - 4,3)^2 \times 7 + (4,1 - 4,3)^2 \times 9 + (4,2 - 4,3)^2 \times 8 + (4,4 - 4,3)^2 \times 10 + (4,6 - 4,3)^2 \times 4 + (4,7 - 4,3)^2 \times 3}{41 - 1} = \frac{2,01}{40} = 0,05$$

$$\sigma^2_y = \frac{\sum (Y_i - \bar{Y})^2 n_i}{N_y - 1} = \frac{(4,0 - 4,4)^2 \times 2 + (4,3 - 4,4)^2 \times 4 + (4,5 - 4,4)^2 \times 9 + (4,6 - 4,4)^2 \times 3 + (4,7 - 4,4)^2}{19 - 1} = \frac{2,01}{18} = \frac{0,66}{18} = 0,037$$

3. Визначаємо розрахункове значення критерію Фішера за формулою:

$$F = \frac{0,05}{0,037} = 1,4$$

відповідно до того, що $0,05 > 0,037$

4. Визначаємо F_{гр} на підставі числа ступенів свободи більшої й меншої дисперсії:

k₁ – число ступенів свободи більшої дисперсії = 41-1 = 40;

k₂ – число ступенів свободи меншої дисперсії = 19-1 = 18;

при k₁ = 40, k₂ = 18, F_{гр} = 2,1 (див. додаток Б).

5. Порівнюємо розрахункове значення критерію Фішера F та критичне значення теоретичного розподілу Фішера F_{гр}, 1,4 < 2,1, F < F_{гр}.

Висновок: Оскільки F < F_{гр}, розбіжність між двома групами плавців за стабільністю поглинання кисню статистично невірогідна, тобто групи за стабільністю поглинання кисню розрізняються не істотно.

Виконаємо практичне завдання за допомогою редактора електронних таблиць Excel.


Приклад: У двох групах плавців X_i і Y_i під час тривалої роботи на витривалість вимірювали кількість поглиненого кисню, л/хв. Визначити, чи принципово є відмінність між цими групами за стабільністю поглинання кисню.

X_i	4,0	4,1	4,2	4,4	4,6	4,7	
n_i	7	9	8	10	4	3	$N_x = 41$
Y_i	4,0	4,3	4,6	4,5	4,7		
n_i	2	4	3	9	1		$N_y = 19$

Порядок виконання

Крок 1. Згідно з умовами завдання вводимо у стовпчик **B** електронної таблиці показники кількості поглинання кисню, л/хв першою групою плавців (значення X); у стовпчик **C** - показники кількості поглинання кисню, л/хв другою групою плавців (значення Y). При введенні даних обов'язково враховуємо кількість повторень кожної варіанти.

Крок 2. У стовпчику **A** запишемо порядкові номери варіант та назви показників, які необхідно розрахувати. У комірці **A43** запишемо дисперсію; у комірці **A44** – розрахункове значення критерію Фішера ($F_{розр.}$); у комірці **A45** – число ступенів свободи, в комірці **A46** – критичне (табличне) значення критерію Фішера ($F_{крит.}$).

Крок 3. Обчислимо дисперсію першої вибірки. Для цього виділимо комірку **B43** та скористаємося піктограмою вбудованих функцій . При цьому з'явиться вікно *Мастер функций*, у якому обираємо категорію *Статистические* та функцію *ДИСП*.

Натискаємо клавішу *OK*. При цьому з'являється діалогове вікно вводу даних. У цьому випадку інтервал даних для обчислення середнього арифметичного значення з'являється автоматично, оскільки курсор стоїть внизу стовпчика з числовими значеннями. Однак інтервал у вікно діапазону даних можна ввести безпосередньо із клавіатури або натиснувши на піктограму праворуч цього вікна та вибравши мишею діапазон.


Натискаємо клавішу *OK*. У результаті виконаних дій у комірці **B43** відобразиться дисперсія першої вибірки.

Крок 4. Обчислимо дисперсію другої вибірки. Для цього скопіюємо комірку **B43** у комірку **C43** та обов'язково відкоригуємо масив. Для даного прикладу він повинен бути **C2:C20**.

Крок 5. Обчислимо розрахункове значення критерію Фішера ($F_{розр.}$). Для цього необхідно більшу дисперсію поділити на меншу, тобто значення комірки **B43** поділити на значення комірки **C43**. Виділяємо порожню комірку **B44** та у рядку формул записуємо такий вираз: **=B43/C43**.

Крок 6. Розраховуємо число ступенів свободи для більшої (комірка **B45**) та меншої (комірка **C45**) дисперсії. У рядку формул запишемо відповідні формули з використанням вбудованої функції **СЧЕТ**: для комірки **B45** запишемо **СЧЕТ(B2:B42)-1**; для комірки **C45** запишемо **СЧЕТ(C2:C20)-1**.

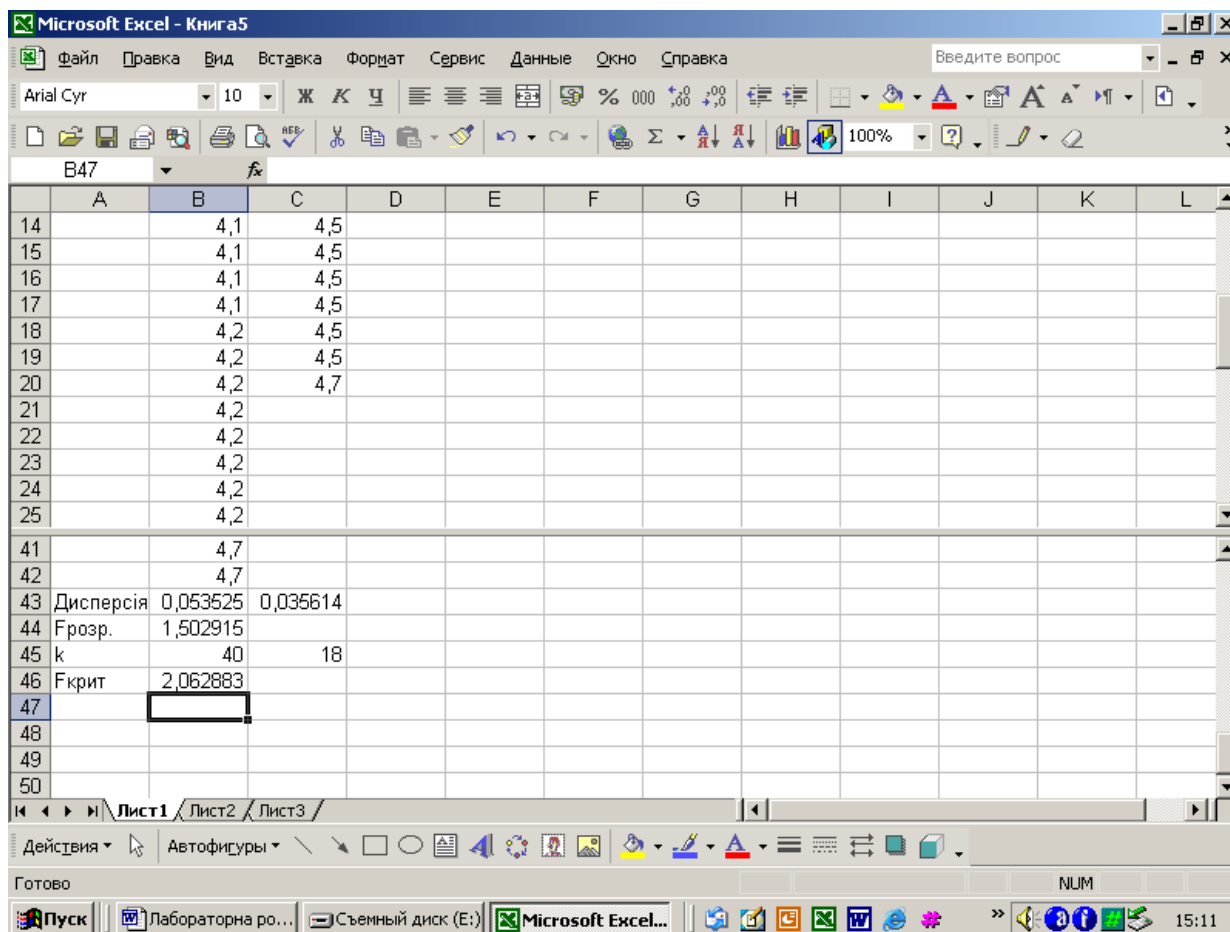
У результаті проведених розрахунків отримаємо підсумкову таблицю даних (рис. 43).

Крок 7. Обчислимо критичне (табличне) значення критерію Фішера за допомогою засобів **Excel**. Для цього необхідно застосувати функцію **FPASPOBR**. Ставимо маркер у комірку **B46**, викликаємо *Мастер функций* за допомогою піктограми  та обираємо функцію **FPASPOBR**.

У вікні вводу даних аргументів функції вводимо у першому рядку ймовірність **0,05**, а у другому та третьому рядках – відповідне число ступенів свободи (**40** та **18**), натискаємо **OK**. У результаті цих дій у комірці **B46** з'являється значення критичного (табличного) критерію Фішера.

Табличне значення критерію Фішера можна визначити за спеціальною таблицею ($P=0.95$) на перехресті значень ступенів свободи для більшої та меншої дисперсії.

Крок 8. Проведемо порівняння розрахункового значення критерію Фішера F та критичного значення теоретичного розподілу Фішера $F_{гр}$, $1,4 < 2,1$, $F < F_{гр}$, запишемо висновок відповідно до отриманих результатів.



	A	B	C	D	E	F	G	H	I	J	K	L
14		4,1	4,5									
15		4,1	4,5									
16		4,1	4,5									
17		4,1	4,5									
18		4,2	4,5									
19		4,2	4,5									
20		4,2	4,7									
21		4,2										
22		4,2										
23		4,2										
24		4,2										
25		4,2										
41		4,7										
42		4,7										
43	Дисперсія	0,053525	0,035614									
44	Грозр.	1,502915										
45	k	40	18									
46	Fкрит	2,062883										
47												
48												
49												
50												

Рис. 43. Підсумкова таблиця даних

Висновок: Оскільки $F < F_{гр}$, розбіжність між двома групами плавців за стабільністю поглинання кисню статистично невірогідна, тобто групи за стабільністю поглинання кисню розрізняються не істотно.

Практичне завдання:

Скласти умови завдання з власного наукового дослідження і вирішити його за допомогою вибіркового методу математичної статистики. Проаналізувати результати, зробити висновок.

? **Питання для самоконтролю:**

1. Розкрийте сутність вибіркового методу.
2. Яка величина є критерієм визначення вірогідності різниць?
3. Поясніть, у чому полягає основне завданням вибіркового методу.
4. За якою формулою визначається критерій Стюдента?
5. За якою формулою визначається помилка середнього арифметичного значення?
6. Що можна оцінити за допомогою вибіркового методу математичної статистики?
7. Що дозволяє порівнювати вибіркового методу математичної статистики?
8. Яке припущення перевіряється при порівнянні двох вибірових середніх арифметичних значень?
9. Назвіть причини вірогідних різниць між двома вибіровими середніми арифметичними.
10. Назвіть причини невірогідних різниць між двома вибіровими середніми арифметичними.
11. При якому співвідношенні розрахованого за формулою t і постійного $t_{гр}$ критеріїв Стюдента різниця між порівнюваними вибіровими середніми арифметичними вірогідна (невипадкова) або невірогідна (випадкова)?
12. Розтлумачте поняття «модельні характеристики».
13. Назвіть види модельних характеристик.
14. Що показує помилка репрезентативності?
15. Назвіть шляхи визначення модельних характеристик.
16. Як визначається середня арифметична генеральної сукупності $\bar{X}_{ген}$?
17. Як визначається середнє квадратичне відхилення генеральної сукупності ($\sigma_{ген}$)?
18. Яким чином може бути знайдена величина помилки репрезентативності при відомому та невідомому обсязі генеральної сукупності?
19. Розкрийте сутність надійності та рівня значущості.
 20. Що показує дисперсія?
 21. За якою формулою визначається дисперсія?
 22. Яка величина є критерієм визначення вірогідності різниці між двома вибіровими характеристиками варіації?
 23. Назвіть основні властивості критерію Фішера.
 24. За якою формулою визначається розрахункове значення критерію Фішера?
 25. Яким чином визначається теоретичне (граничне) значення критерію Фішера?

26. Яким чином визначають число ступенів свободи?

27. Поясніть, за яких умов розбіжність між двома вибірками за показниками варіації буде вірогідною.

28. Поясніть, за яких умов розбіжність між двома вибірками за показниками варіації не буде вірогідною.



Змістовий модуль 4. Статистичні методи вимірювання взаємозв'язків



Тема 10. Види взаємозв'язків і поняття про кореляційний аналіз

У спортивних дослідженнях між досліджуваними показниками часто спостерігається взаємозв'язок. Вид його може бути різним. Розрізняють два види взаємозв'язку – функціональний і стохастичний.

Функціональний взаємозв'язок – це залежність, при якій кожному значенню одного показника відповідає суворо визначене значення іншого і ні якої варіації бути не може.

До другого виду взаємозв'язку належить, наприклад, залежність маси тіла від довжини тіла. Одному значенню довжини тіла може відповідати декілька значень маси тіла і навпаки. У таких випадках, коли одному значенню одного показника відповідає декілька значень іншого показника, взаємозв'язок називається *стохастичним*.

Вивченню статистичного взаємозв'язку між різними показниками в спортивних дослідженнях приділяють велику увагу, оскільки це дозволяє розкрити деякі закономірності й надалі описати їх як за допомогою слів, так і математичних знаків та формул, з метою застосування в практичній роботі тренера і педагога.

Серед статистичних взаємозв'язків найбільш важливі – кореляційні (від латинського *Correlatio* – співвідношення, відповідність).

Кореляція – вид взаємозв'язку між ознаками. Кожна ознака являє собою велику кількість однотипних показників, що варіюють.

Кореляція полягає в тому, що середня величина одного показника змінюється залежно від середньої величини іншого.

Статистичний метод, що застосовується для дослідження взаємозв'язків, називається *кореляційним аналізом*. Головне завдання кореляційного аналізу – визначення форми, щільності, спрямованості досліджуваних показників. Він широко застосовується в теорії тестів для оцінки їх надійності та інформативності.

За допомогою кореляційного аналізу можна оцінити:

1. Взаємозв'язок між спортивними результатами та функціональними показниками (результат у бігу на 100 м і кількість лактату в крові).

2. Вплив спортивної діяльності на результат (результат у човниковому бігу 3×100 м і результат у бігу на 100 м).

3. Взаємний вплив показників тренованості (результат при проходженні 10×200 м і результат при проходженні середньої дистанції, у плаванні).
4. Надійність тесту.
5. Інформативність тесту (залежність між кількістю гребків за 30 секунд і часом проходження дистанції 100 м у плаванні).

Тема 11. Способи визначення взаємозв'язків

Аналіз взаємозв'язку починається із графічного представлення результатів вимірів у прямокутній системі координат. Графічна залежність має назву «діаграма розсіювання» або «кореляційне поле». Візуальний аналіз кореляційного поля дозволяє виявити спрямованість і форму залежності (принаймні, зробити припущення).

За кореляційним полем можна визначити спрямованість:

а) пряма позитивна кореляційна статистична залежність (нахил кореляційного поля праворуч) (рис. 2). Зі зростанням (зменшенням) першої ознаки (X_i) інша (Y_i) також зростає (зменшується).

б) обернено-негативна кореляційна статистична залежність (нахил кореляційного поля ліворуч) (рис. 3). Зі зростанням першої ознаки (X_i) інша (Y_i) зменшується і, навпаки, зі зменшенням першої ознаки (X_i) інша (Y_i) зростає.

Якщо кореляційне поле передається окружністю, то залежність відсутня (рис. 4).

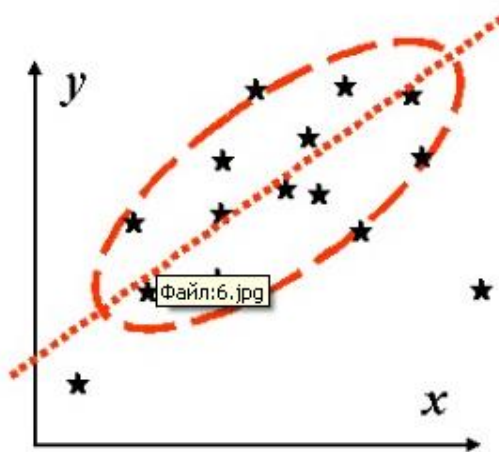


Рис. 2. Пряма позитивна кореляційна статистична залежність

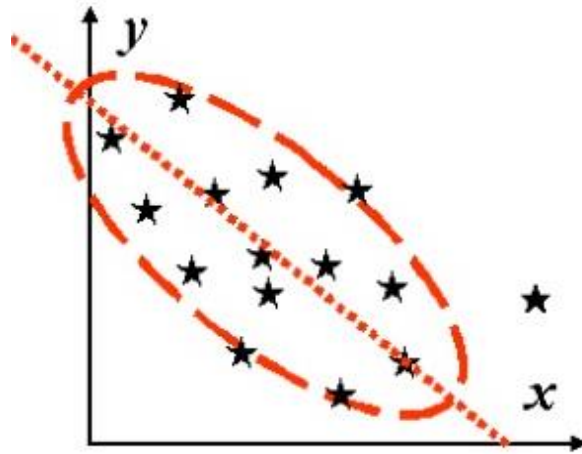


Рис. 3. Обернено-негативна кореляційна статистична залежність

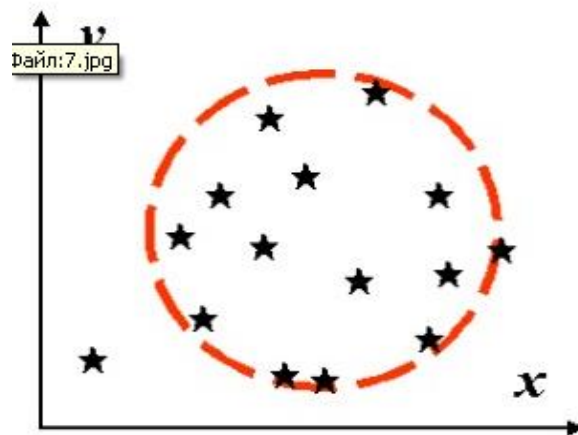


Рис. 4. Статистична залежність відсутня

Якщо кореляційне поле подається прямою лінією, то має місце функціональна залежність.

Форми статистичної залежності:

1. Лінійна форма залежності – форма є близькою до звичайної геометричної фігури (еліпса).
2. Нелінійна форма залежності – будь-яка інша форма, крім еліпса.

Отже, візуальний аналіз кореляційного поля дозволяє виявити форму статистичної залежності – лінійну або нелінійну. Це має істотне значення для подальшого етапу аналізу – вибору й обчислення відповідного коефіцієнта кореляції.

Для більш точної оцінки кореляції, що визначається за формулою, потрібно знати форму залежності:

1. Якщо виміри проводяться за шкалою відношень або інтервалів і форма залежності лінійна, то коефіцієнт кореляції розраховується за формулою Брава-

Пірсона (r_{xy}) :

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{N \times \sigma_x \times \sigma_y}$$

де X_i й Y_i – варіанти 2-х вибірок;

\bar{X} і \bar{Y} – середнє арифметичне значення показників X_i й Y_i ;

σ_x , σ_y – середнє квадратичне відхилення;

N – число вимірів (досліджуваних).

2. Якщо виміри проводяться за шкалою відношень або інтервалів і форма залежності нелінійна, коефіцієнт кореляції (Γ_{xy}) розраховується за формулою:

$$\Gamma_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Для оцінки щільності взаємозв'язку в кореляційному аналізі застосовується значення спеціального показника – коефіцієнта кореляції (Γ_{xy}).

Абсолютне значення коефіцієнта кореляції знаходиться в межах від 0 до 1 – пряма позитивна кореляційна статистична залежність і від (-1) до 0 – обернена негативна кореляційна статистична залежність:

$$-1 \leq \Gamma_{xy} \leq 1$$

Пояснюють значення цього коефіцієнта в такий спосіб:

а) $\Gamma_{xy} = 1$ – зв'язок між ознаками дуже щільний (функціональний взаємозв'язок);

б) $\Gamma_{xy} = 0$ – зв'язок між ознаками X_i й Y_i відсутній;

в) чим ближче значення Γ_{xy} до нуля, тим слабше зв'язок, чим ближче значення Γ_{xy} до одиниці, тим щільніше.

Прийнято вважати, що

$\Gamma_{xy} = 0,2 \dots 0,49$ – слабкий зв'язок;

$\Gamma_{xy} = 0,5 \dots 0,69$ – середній зв'язок;

$\Gamma_{xy} = 0,7 \dots 0,99$ – тісний (сильний) зв'язок.

На кореляційному полі щільність може мати такий вигляд:

а) якщо точки групуються вздовж якої-небудь лінії, то зв'язок є, і він тим щільніший, чим ближче вони групуються.

б) якщо точки розсіяні хаотично, зв'язок між ознаками відсутній або дуже слабкий.

Проте варто пам'ятати, що при роботі з великою точністю (наприклад, при оцінці кореляцій спортсменів високої кваліфікації) коефіцієнт кореляції відображає щільний зв'язок, якщо він не менше 0,9.

У деяких випадках щільність взаємозв'язку визначають на підставі коефіцієнта детермінації (D), що розраховується за формулою:

$$D = r^2 \times 100 \%$$

Цей коефіцієнт визначає частину загальної варіації одного показника, що пояснюється варіацією іншого показника. Залишок відсотків варіації від ста пояснюється впливом інших чинників.



Тема 12. Визначення найбільш ймовірних лінійних залежностей між двома факторами

У випадку лінійного взаємозв'язку між двома ознаками (факторами), тобто коли кореляційне поле має форму еліпса, за допомогою коефіцієнта кореляції можна знайти найбільш ймовірні лінійні залежності, тобто рівняння регресії, які є рівняннями прямої лінії.

Таких рівнянь два:

$$y = a_1 + b_{y/x} \times x - \text{пряме};$$

$$x = a_2 + b_{x/y} \times y - \text{зворотне};$$

де a і b – коефіцієнти або параметри, які необхідно визначити.

Значення коефіцієнтів регресії обчислюється за формулою:

$$b_{y/x} = r \times \frac{\sigma_y}{\sigma_x} \quad \text{і} \quad b_{x/y} = r \times \frac{\sigma_x}{\sigma_y}$$

Коефіцієнт регресії b має розмірність, яка відповідає відношенню розмірностей досліджуваних показників X і Y , і той же знак, що й коефіцієнт кореляції.

Коефіцієнт a визначається за формулою:

$$a_1 = \bar{y} - b_{y/x} \times \bar{x};$$

$$a_2 = \bar{x} - b_{x/y} \times \bar{y}$$

Для того щоб обчислити цей коефіцієнт, треба просто в рівняннях регресії підставити середні значення змінних величин, які корелюють.

Для оцінки точності рівняння регресії розраховується залишкове середнє квадратичне відхилення за формулою:

$$\sigma_{y/x} = \sigma_y \times \sqrt{1 - r^2}$$

$$\sigma_{x/y} = \sigma_x \times \sqrt{1 - r^2}$$

Ці оцінки абсолютні, а отже, не можуть бути порівнянні між собою. Тому застосовують оцінки відносної похибки рівнянь, які мають бути виражені у відсотках і визначаються за формулою:

$$\delta_{y/x} = \frac{\sigma_{y/x}}{\bar{Y}} \times 100\%$$

$$\delta_{x/y} = \frac{\sigma_{x/y}}{\bar{X}} \times 100\%$$

Значення цієї оцінки, якщо $r = \pm 1,00$, дорівнює нулю, а якщо $r = 0,00$, максимальне.

Залишкове середнє квадратичне відхилення характеризує коливання Y щодо лінії регресії з X у прямому рівнянні регресії і, навпаки, в оберненому випадку.

Тема.13. Практична реалізація кореляційного аналізу

13.1. Оцінка залежності між двома величинами, що вимірюються

Розглянемо послідовність операцій при оцінці залежності між двома величинами, що вимірюються на конкретному прикладі.

Приклад: Чи існує залежність і яка саме між результатами бігу на 60 м (X_i), с і швидкістю їзди на велосипеді (Y_i), км/год у 7 велосипедистів?

Чи може біг на 60 м слугувати тестом для цих велосипедистів для визначення швидкості їзди на велосипеді? Наскільки сильно впливає біг 60 м на швидкість їзди на велосипеді?

X_i	9,0	9,1	9,1	9,3	9,5	9,5	9,6
Y_i	18,4	18,0	17,8	17,5	17,1	16,9	16,2

Порядок виконання

1. Обчислимо середнє арифметичне значення 2-х вибірок, X_i та Y_i .

1.1 Оскільки не всі варіанти першої вибірки (X_i) зустрічаються один раз, то середнє арифметичне значення розраховуємо за формулою зваженого середнього арифметичного значення:

$$\bar{X} = \frac{\sum_{i=1}^n X_i \times n_i}{N} \quad \bar{X} = \frac{\sum_{i=1}^7 X_i \times n_i}{7}$$

$$\bar{X} = \frac{9,0 + 9,1 \times 2 + 9,3 + 9,5 \times 2 + 9,6}{7} = 9,3(c)$$

$\bar{X} = 9,3$ (с) – середній час бігу на дистанції 60 м для 7 велосипедистів.

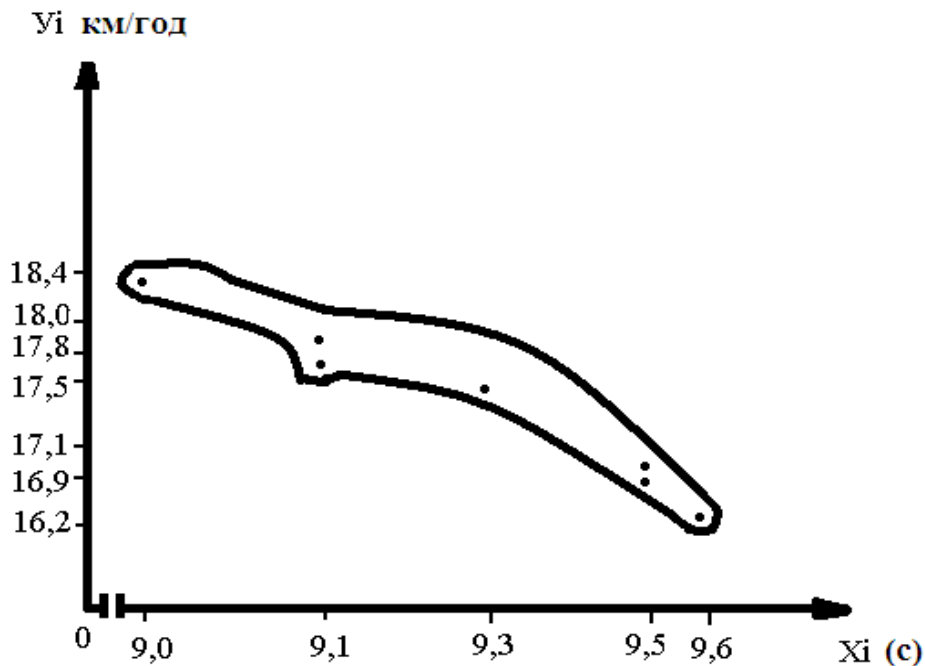
1.2. Оскільки всі варіанти другої вибірки (Y_i) не повторюються, то середнє арифметичне значення розраховуємо за формулою незваженого середнього арифметичного значення:

$$\bar{Y} = \frac{Y_1 + Y_2 + Y_3 + \dots + Y_n}{N} \quad \bar{Y} = \frac{Y_1 + Y_2 + Y_3 + \dots + Y_7}{7}$$

$$\bar{Y} = \frac{18,4 + 18,0 + 17,8 + 17,5 + 17,1 + 16,9 + 16,2}{7} = 17,4 (\text{км/год})$$

$\bar{Y} = 17,4$ (км/год) – середня швидкість їзди на велосипеді для 7 велосипедистів.

2. Будуємо кореляційне поле, за формою якого визначаємо спрямованість взаємозв'язку.



Зовнішній вигляд кореляційного поля відповідає нелінійній формі залежності й оберненому негативному кореляційному статистичному взаємозв'язку.

Нелінійна форма залежності – кореляційне поле не відповідає формі еліпса. Обернений кореляційний статистичний взаємозв'язок – нахил кореляційного поля ліворуч.

3. Оскільки форма залежності нелінійна, то коефіцієнт кореляції розраховується за формулою:

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

3.1 Для подальшої роботи будемо таблицю:

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
9,0	18,4	-0,3	1	-0,3	0,09	1
9,1	18,0	-0,2	0,6	-0,12	0,04	0,36
9,1	17,8	-0,2	0,4	0,08	0,04	0,16
9,3	17,5	0	0,1	0	0	0,01

9,5	17,1	0,2	-0,3	-0,06	0,04	0,09
9,5	16,9	0,2	-0,5	-0,1	0,04	0,25
9,6	16,2	0,3	-1,2	-0,36	0,09	1,44
Σ	65,1	121,9		-1,02	0,34	3,31

3.2 Коефіцієнт кореляції (r_{xy})

$r_{xy} = -0,96$ – значення коефіцієнта дозволяє говорити про обернений негативний сильний кореляційний статистичний взаємозв'язок.

4. Розраховуємо коефіцієнт детермінації (D)

$$D = r^2 \times 100 \%$$

$$D = (-0,96)^2 \times 100 \% = 0,9216 \times 100 \% = 92,16 \%$$

$D = 92,16 \%$ – значення коефіцієнта дозволяє зробити висновок про те, що 92 % взаємозв'язку результату бігу на 60 м і швидкості їзди на велосипеді пояснюються їх взаємовпливом. Інша частина ($100 \% - 92 \% = 8 \%$) варіації пояснюється впливом інших неврахованих чинників.

Висновок: Розрахований коефіцієнт кореляції ($r_{xy} = -0,96$) вказує на обернений негативний сильний кореляційний статистичний взаємозв'язок між бігом на 60 м і швидкістю їзди на велосипеді. Це означає, що зі зменшенням часу проходження дистанції 60 м швидкість їзди на велосипеді зростатиме. Коефіцієнт детермінації дозволяє стверджувати, що на 92% (з 100%) швидкість їзди на велосипеді залежить від результату бігу на 60 м. Отже, біг на 60 м доцільно застосовувати в тренувальному процесі велосипедистів для підвищення швидкості їзди на велосипеді.

Виконаємо практичне завдання за допомогою редактора електронних таблиць Excel.

Приклад: Чи існує залежність і яка між результатами бігу на 60 м (X_i), с та швидкістю їзди на велосипеді (Y_i), км/год у 7 велосипедистів?

Чи може біг на 60 м слугувати тестом для цих велосипедистів для визначення швидкості їзди на велосипеді? Наскільки сильно впливає біг 60 м на швидкість їзди на велосипеді?

$$X_i \quad 9,0 \quad 9,1 \quad 9,1 \quad 9,3 \quad 9,5 \quad 9,5 \quad 9,6$$

$$Y_i \quad 18,4 \quad 18,0 \quad 17,8 \quad 17,5 \quad 17,1 \quad 16,9 \quad 16,2$$

Порядок виконання

Крок 1. Згідно з умовами завдання вводимо в стовпчик **В** електронної таблиці результати часу долавання дистанції 60м (X_i), с у 7 велосипедистів, а у стовпчик **С** - швидкість їзди на велосипеді (Y_i), км/год у цих же велосипедистів. Таблиця матиме такий вигляд:

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		Час, с	Швидкість																
2		1	9	18,4															
3		2	9,1	18															
4		3	9,1	17,8															
5		4	9,3	17,5															
6		5	9,5	17,1															
7		6	9,5	16,9															
8		7	9,6	16,2															
9																			

Рис. 44. Таблиця вихідних даних

Крок 2. Побудуємо кореляційне поле для залежності між часом долаття дистанції 60 м та швидкістю їзди на велосипеді у 7 велосипедистів. Для цього виконаємо певну послідовність дій.

Крок 3. Натискаємо піктограму *Мастер діаграм* чи виконуємо команду *Вставка – Діаграма*. У діалоговому вікні, що відкривається, обираємо *Точечная* діаграма (рис. 45).

Крок 4. Відзначаємо діапазон даних для створення діаграми (ряди *Час* і *Швидкість*). Для цього натискаємо клавішу *Далее* та мишею вводимо діапазон – виділяємо стовпчики **B** і **C**, як показано на рис. 46. Якщо ряди, за якими будується діаграма, розташовані не суміжно, то відмічаємо кожний з рядів окремо при натиснутій клавіші *Ctrl*.

Крок 5. Натискаємо клавішу *Далее* та вводимо назву діаграми та її осей (рис. 47).

Крок 6. Натискаємо клавішу *Готово* та отримуємо кореляційне поле для залежності між часом долаття дистанції 60 м та швидкістю їзди на велосипеді в 7 велосипедистів (рис. 48).

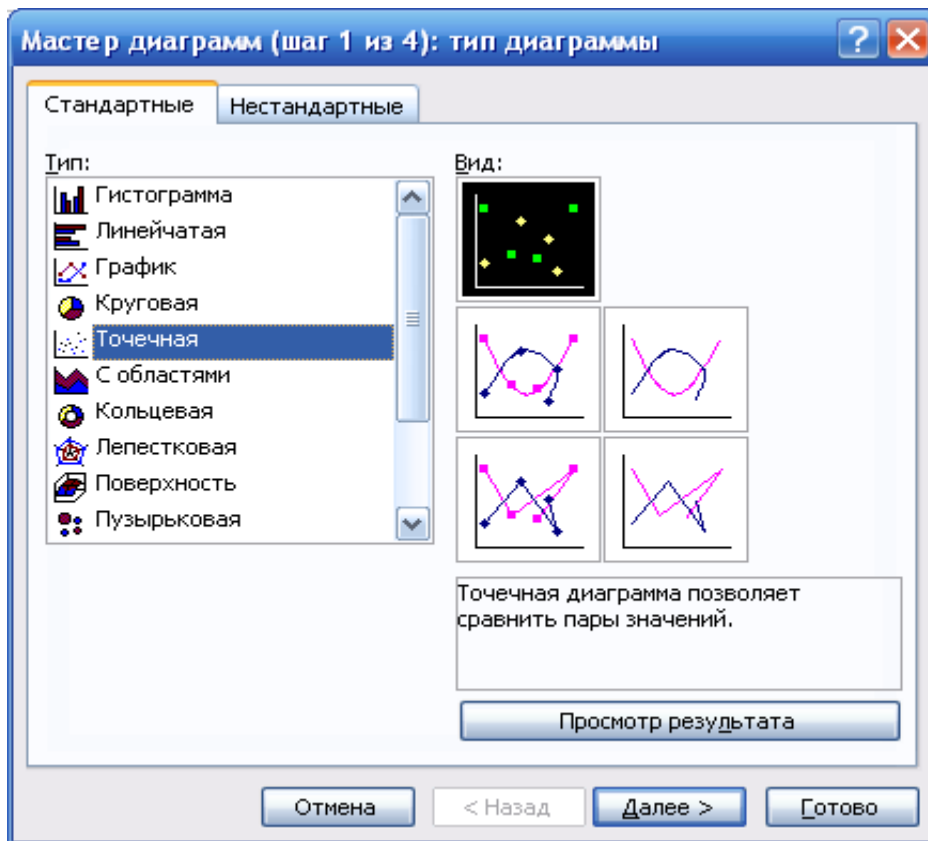


Рис. 45. Диалогове вікно «Мастер диаграмм» (крок 3)

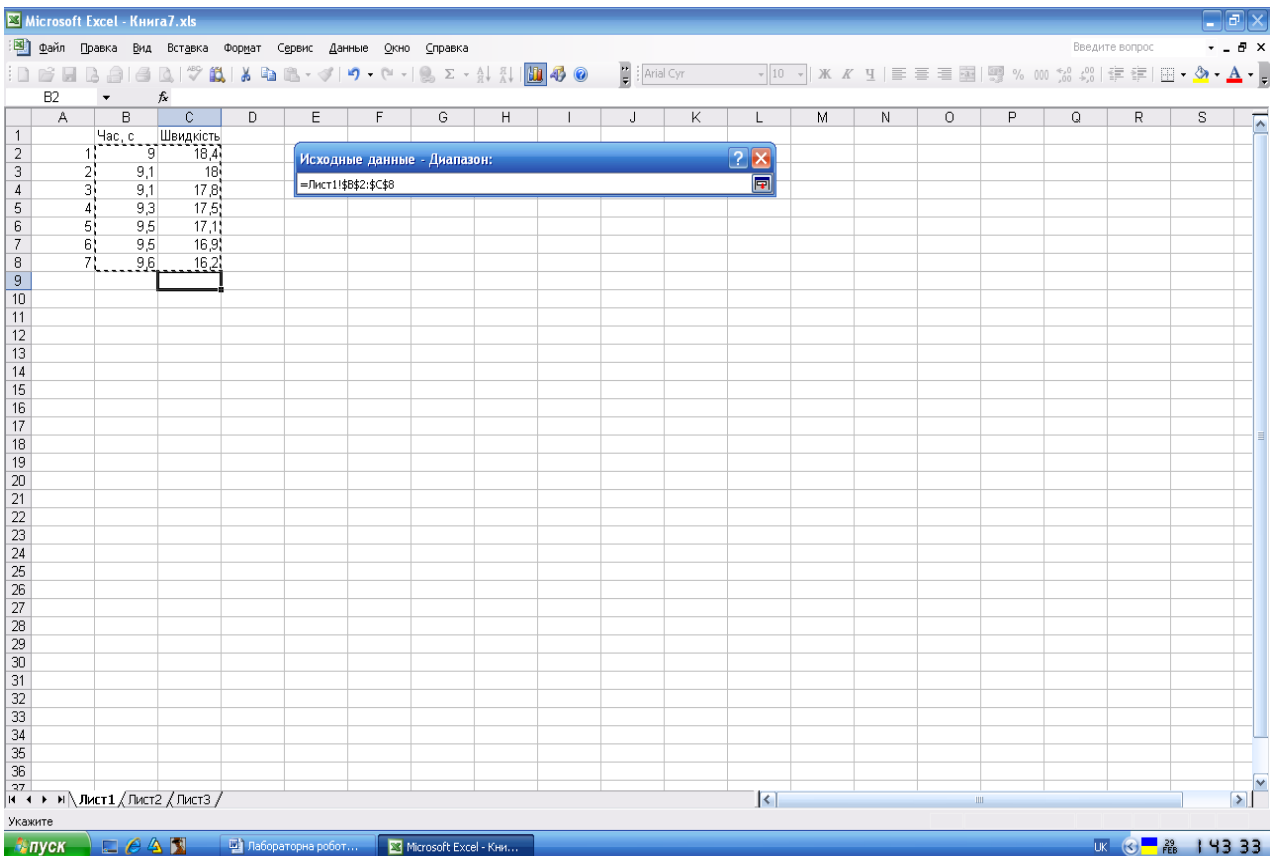


Рис. 46. Операція вводу даних для побудови кореляційного поля

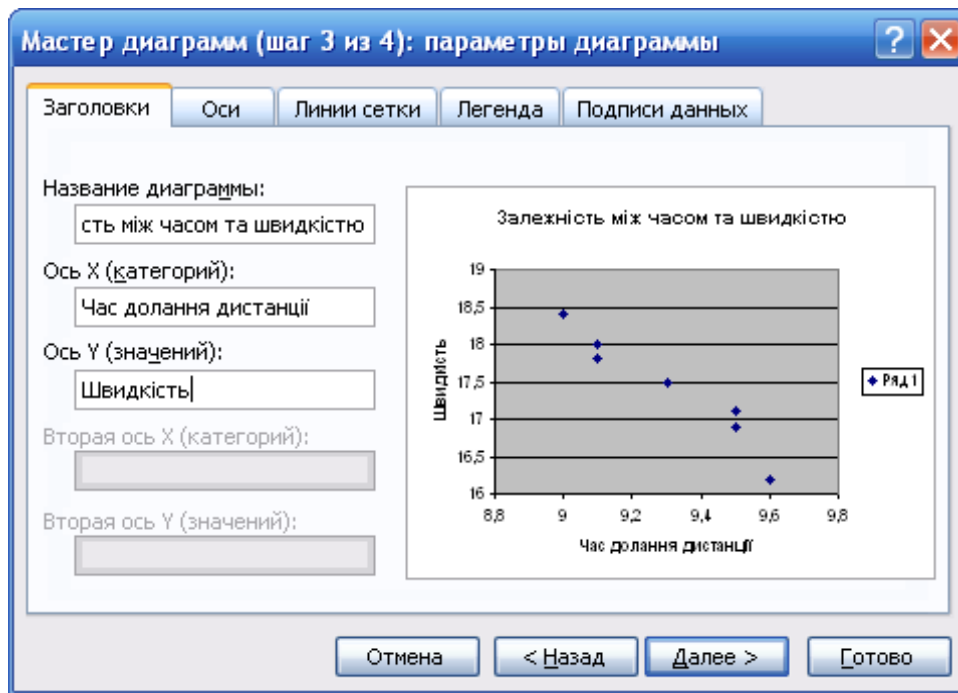


Рис. 47. Діалогове вікно «Мастер диаграмм» (крок 5)

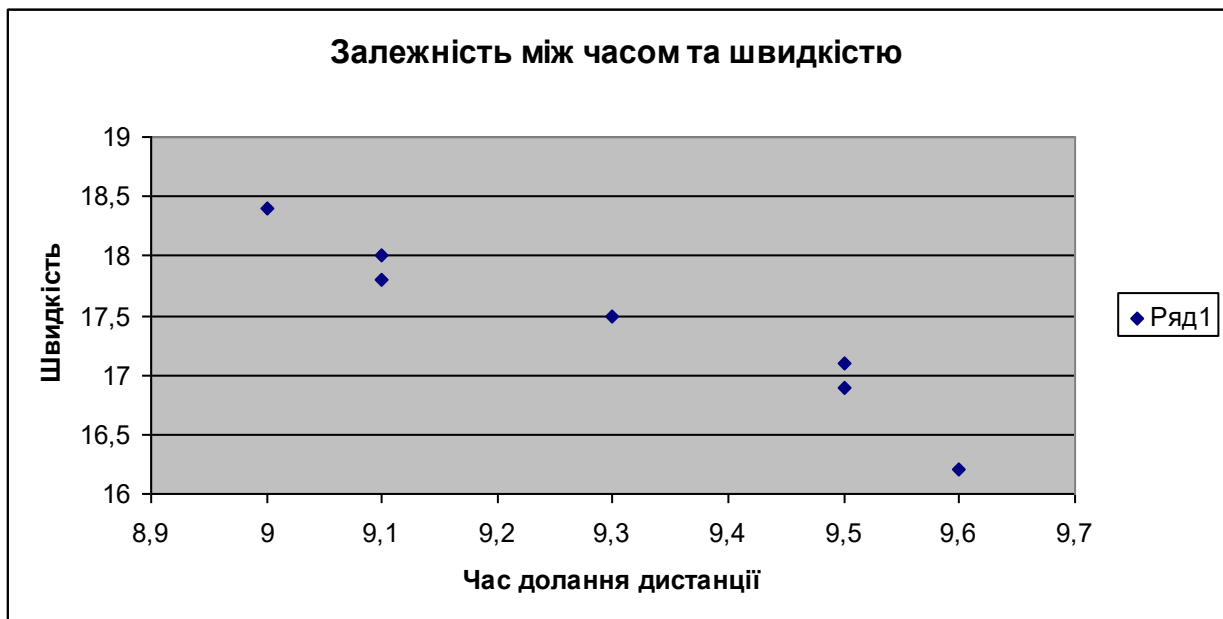


Рис. 48. Кореляційне поле залежності між часом долаття дистанції 60 м та швидкістю їзди на велосипеді у 7 велосипедистів

Крок 7. Обчислюємо коефіцієнт кореляції для визначення залежності між часом долаття дистанції 60 м та швидкістю їзди на велосипеді в 7 велосипедистів. Для цього ставимо маркер миші в порожню комірку C9 (саме в ній у кінцевому підсумку відобразиться коефіцієнт кореляції).

Крок 8. Викликаємо *Мастер функций* через команди *Вставка – Функция* чи натисканням на піктограму *Мастер функций*.

Крок 9. Серед статистичних функцій обираємо функцію **ПИРСОН**. При цьому з'являється діалогове вікно (рис.49), в яке вводимо діапазони даних першої та другої змінних.

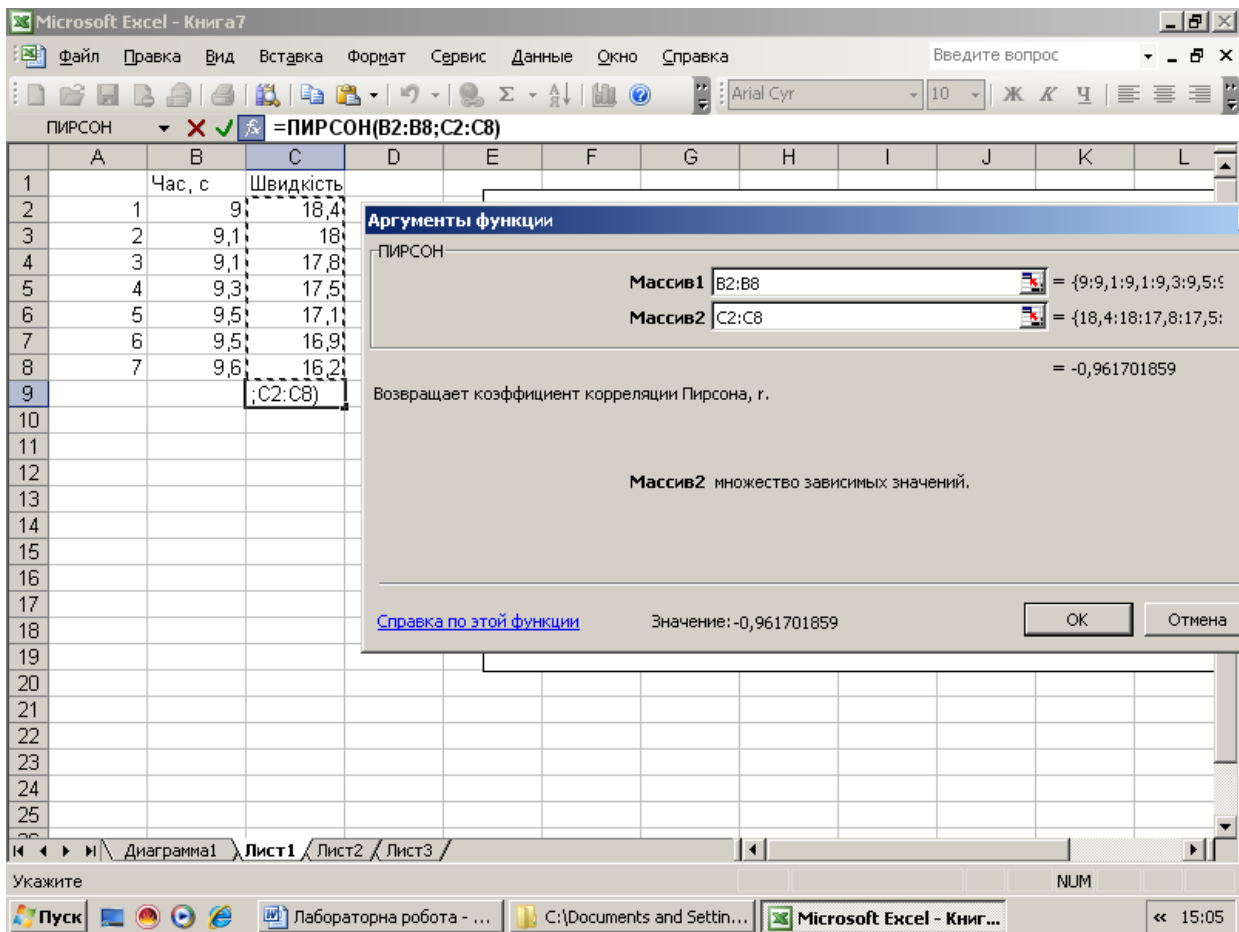


Рис. 49. Ввід даних при використанні функції «ПІРСОН»

Крок 10. Натискаємо кнопку *OK* та в комірці **C9** отримуємо значення коефіцієнта кореляції залежності між часом додання дистанції 60 м та швидкістю їзди на велосипеді в 7 велосипедистів.

Функцію **ПІРСОН**, а також діапазони даних першої та другої змінних можна ввести безпосередньо в рядок формул із клавіатури або мишею.

Крок 11. Розраховуємо коефіцієнт детермінації (D). Для цього ставимо маркер миші в порожню комірку **C10** (саме в ній у кінцевому підсумку відобразиться коефіцієнт детермінації) і в рядку формул записуємо формулу: $=C9^2 \times 100$. Натискаємо **Enter**.

$D = 92,16 \%$ – значення коефіцієнта детермінації дозволяє зробити висновок про те, що 92% взаємозв'язку між результатом бігу на 60 м і швидкістю їзди на велосипеді пояснюються їх взаємовпливом. Інша частина (100 % - 92 % = 8 %) варіації пояснюється впливом інших неврахованих чинників.

Крок 12. Записуємо висновок відповідно до результатів розрахунків.

Висновок: Розрахований коефіцієнт кореляції ($r_{xy} = -0,96$) вказує на обернений негативний сильний кореляційний статистичний взаємозв'язок між бігом на 60 м і швидкістю їзди на велосипеді. Це означає, що зі зменшенням часу проходження дистанції 60 м швидкість їзди на велосипеді зростатиме. Коефіцієнт детермінації дозволяє стверджувати, що на 92% (з 100%) швидкість

їзди на велосипеді залежить від результату бігу на 60 м. Отже, біг на 60 м доцільно застосовувати в тренувальному процесі велосипедистів для підвищення швидкості їзди на велосипеді.

13.2. Знаходження найбільш ймовірних лінійних залежностей між двома факторами (побудова прямих регресій)

Розглянемо послідовність операцій при знаходженні найбільш ймовірних лінійних залежностей між двома факторами (побудуванні прямих регресій) на конкретному прикладі.

Приклад: У групі хлопчиків 10-12 років спецкласу з плавання фіксували вагу X_i , кг і довжину тіла Y_i , см. Обчислити значення параметрів рівняння регресії - прями і обернені. Визначити найбільш імовірні лінійні залежності між двома факторами X і Y .

№	X_i	Y_i	№	X_i	Y_i
1	55	175	1	32	145
2	36	148	3	35	148
3	37	151	4	35	151
4	54	168	5	32	138
5	66	178	6	35	141
6	36	146	7	34	142
7	38	150	8	47	147
8	37	154	9	28	133
9	42	156	0	33	145
10	33	150	1	42	160
11	36	154	2	43	161
12	37	158	3	47	169
			4	41	152
			5		

Порядок виконання

1. Визначаємо середнє арифметичне для двох показників

$$\bar{X} = \frac{\sum X_i}{N}; \quad \bar{X} = \frac{\sum_{i=1}^{25} X_i}{25} = \frac{991}{25} = 39,8 \text{ кг}$$

$$\bar{Y} = \frac{\sum Y_i}{N}; \quad \bar{Y} = \frac{\sum_{i=1}^{25} Y_i}{25} = \frac{3820}{25} = 152,8 \text{ см}$$

2. Для полегшення проведення розрахунків, заповнюємо таблицю:

	i	X _i	X _i - \bar{X}	(X _i - \bar{X}) ²	Y _i	Y _i - \bar{Y}	(Y _i - \bar{Y}) ²	(X _i - \bar{X})(Y _i - \bar{Y})
	5	15,2	-0,4	0,16	22	-30,8	948,64	-12,12
	6	3,8	-3,6	12,96	4,8	-104	10816	-38,88
	7	2,8	-4,6	21,16	1,8	-134	17956	-58,68
	4	14,2	-25,6	655,36	15	-137,8	18988,84	3527,52
	6	26,2	-13,6	184,96	25	-127,8	16332,84	1741,52
	6	3,8	-3,6	12,96	6,8	-104	10816	-38,88
	8	1,8	-4,6	21,16	2,8	-134	17956	-58,68
	7	2,8	-4,6	21,16	1,2	-140,8	19822,4	29,36
	2	2,2	-4,2	17,64	3,2	-129,6	16816,36	22,72
0	3	6,8	-3,2	10,24	2,8	-134	17956	13,12
1	6	3,8	-3,6	12,96	1,2	-140,8	19822,4	-4,56
2	7	2,8	-4,6	21,16	5,2	-106,6	11363,56	-23,52
3	2	7,8	-25,6	655,36	7,8	-125	15625	81,72
4	5	4,8	-3,6	12,96	4,8	-104	10816	-12,12
	3	-	-3,6	12,96	-	-134	17956	8,64

5	5	4,8	4	51	1,8		
6	2	7,8	4	38	14,8	04	219, 115,44
7	5	4,8	4	41	11,8	24	139, 56,64
8	4	5,8	4	42	10,8	64	116, 62,64
9	7	2	4	47	5,8	4	33,6 -41,76
0	8	11,8	24	33	19,8	04	392, 233,64
1	3	6,8	4	45	7,8	4	46,2 - 60,8 53,04
2	2	2	4	60	7,2	4	4,84 51,8 15,84
3	3	2	4	61	8,2	4	10,2 67,2 26,24
4	7	2	4	69	16,	44	51,8 262, 116,64
5	1	2	4	52	0,8		1,44 0,64 0,96
			180				3018 2001,20
			0,40			,04	

3. Обчислюємо середнє квадратичне відхилення для двох показників.

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}} \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}}$$

$$\sigma_x = \sqrt{\frac{1800,40}{25-1}} = 8,66 \text{ кг}; \quad \sigma_y = \sqrt{\frac{3018,04}{25-1}} = 11,21 \text{ см.}$$

4. Знаходимо коефіцієнт кореляції для досліджуваних факторів.

$$r_{x/y} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \times \sum_{i=1}^N (Y_i - \bar{Y})^2}} = \frac{2001,20}{\sqrt{1800,40 \times 3018,04}} = \frac{2001,20}{2331,02} = 0,86$$

5. Знаходимо коефіцієнти регресії для рівняння $y = a_1 + \epsilon_{y/x} \times \bar{X}$

$$\epsilon_{y/x} = r \frac{\sigma_y}{\sigma_x} = 0,86 \times \frac{11,21}{8,66} = 1,11 \text{ см/кг}$$

$$a_1 = \bar{Y} - \epsilon_{y/x} \times \bar{X} = 152,8 - 1,11 \times 39,8 = 108,5 \text{ см}$$

6. Визначаємо коефіцієнти регресії для рівняння $x = a_2 + \epsilon_{x/y} \times y$

$$e_{x/y} = r \frac{\sigma_x}{\sigma_y} = 0,86 \times \frac{8,66}{11,21} = 0,66 \text{ кг/см}$$

$$a_2 = \bar{X} - e_{x/y} \times \bar{Y} = 39,8 - 0,66 \times 152,8 = -61,72 \text{ кг}$$

7. Записуємо отримані рівняння регресії, і підставляємо знайдені параметри.

$$y = a_1 + e_{y/x} \times x \quad x = a_2 + e_{x/y} \times y$$

$$y \text{ (см)} = 108,5 + 1,11 \times x \text{ (кг)}$$

$$x \text{ (кг)} = -61,72 + 0,66 \times y \text{ (см)}$$

8. Визначаємо значення залишкових середніх квадратичних відхилень:

$$\sigma_{y/x} = \sigma_y \times \sqrt{1 - r^2} \quad \sigma_{x/y} = \sigma_x \times \sqrt{1 - r^2}$$

$$\sigma_{y/x} = 11,21 \times \sqrt{1 - 0,86^2} = 5,82$$

$$\sigma_{x/y} = 8,66 \times \sqrt{1 - 0,86^2} = 4,5$$

$$\delta_{y/x} = \frac{\sigma_{y/x}}{\bar{Y}} \times 100\% = \frac{5,82}{152,8} \times 100\% = 3,8\%$$

$$\delta_{x/y} = \frac{\sigma_{x/y}}{\bar{X}} \times 100\% = \frac{4,5}{39,8} \times 100\% = 11\%$$

Отже, похибка рівняння $y \text{ (см)} = 108,5 + 1,11 \times x \text{ (кг)}$ буде найменшою і на практиці воно буде мати перевагу перед рівнянням $x \text{ (кг)} = -61,72 + 0,66 \times y \text{ (см)}$, наприклад, за точністю прогнозування.

Висновок: Отримані рівняння регресії (найбільш імовірні лінійні залежності) мають такий вигляд:

$$y \text{ (см)} = 108,5 + 1,11 \times x \text{ (кг)}$$

$$x \text{ (кг)} = -61,72 + 0,66 \times y \text{ (см)},$$

причому, рівняння $y = 108,5 + 1,11 \times x$ точніше описує залежність показників ваги і довжини тіла для досліджуваної групи хлопчиків 10-12 років спецкласу з плавання.

Виконаємо практичне завдання за допомогою редактора електронних таблиць Excel.

Приклад: У групі хлопчиків 10-12 років спецкласу з плавання фіксували вагу X_i , кг і довжину тіла Y_i , см. Обчислити значення параметрів рівняння регресії - прями і обернені. Визначити найбільш імовірні лінійні залежності між двома факторами X і Y .

№	X_i	Y_i	№	X_i	Y_i
1	55	175	1	32	145
			3		
2	36	148	1	35	148

			4		
3	37	151	1	35	151
			5		
4	54	168	1	32	138
			6		
5	66	178	1	35	141
			7		
6	36	146	1	34	142
			8		
7	38	150	1	47	147
			9		
8	37	154	2	28	133
			0		
9	42	156	2	33	145
			1		
10	33	150	2	42	160
			2		
11	36	154	2	43	161
			3		
12	37	158	2	47	169
			4		
			2	41	152
			5		

Порядок виконання

Крок 1. Для того, щоб визначити найбільш ймовірні лінійні залежності між двома факторами X та Y засобами **Excel** необхідно по-перше побудувати діаграму кореляційного поля для залежності між вагою тіла X_i , кг та показниками довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання, по-друге знайти лінію тренду та рівняння регресії, та по-третє визначити точність прямого та зворотного рівняння регресії .

Крок 2. У відповідності з умовами задачі, для побудування діаграми прямої залежності, вводимо в стовпчик **B** електронної таблиці показники ваги тіла X_i , кг, а в стовпчик **C** – показники довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання. А для побудування діаграми зворотної залежності змінюємо розташування стовпчиків X_i та Y_i та вводимо в стовпчик **E** електронної таблиці показники довжини тіла X_i , см, а в стовпчик **F** – показники ваги тіла Y_i , кг (рис. 8.1).

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Xi, кг	Yi, см		Xi, см	Yi, кг						
2	1	55	175		175	55						
3	2	36	148		148	36						
4	3	37	151		151	37						
5	4	54	168		168	54						
6	5	66	178		178	66						
7	6	36	146		146	36						
8	7	38	150		150	38						
9	8	37	154		154	37						
10	9	42	156		156	42						
11	10	33	150		150	33						
12	11	36	154		154	36						
13	12	37	158		158	37						
14	13	32	145		145	32						
15	14	35	148		148	35						
16	15	35	151		151	35						
17	16	32	138		138	32						
22	21	33	145		145	33						
23	22	42	160		160	42						
24	23	43	161		161	43						
25	24	47	169		169	47						
26	25	41	152		152	41						
27												
28												
29												

Рис. 8.1. Таблица вихідних даних

Крок 3. Побудуємо кореляційне поле для залежності між вагою тіла X_i , кг та показниками довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання. Для цього виконаємо наступні дії:

Крок 4. Натискаємо піктограму «*Мастера диаграмм*» чи виконуємо команду «*Вставка – Диаграмма*». У діалоговому вікні, що відкривається, вибираємо «*Точечная*» діаграма (рис. 8.2).

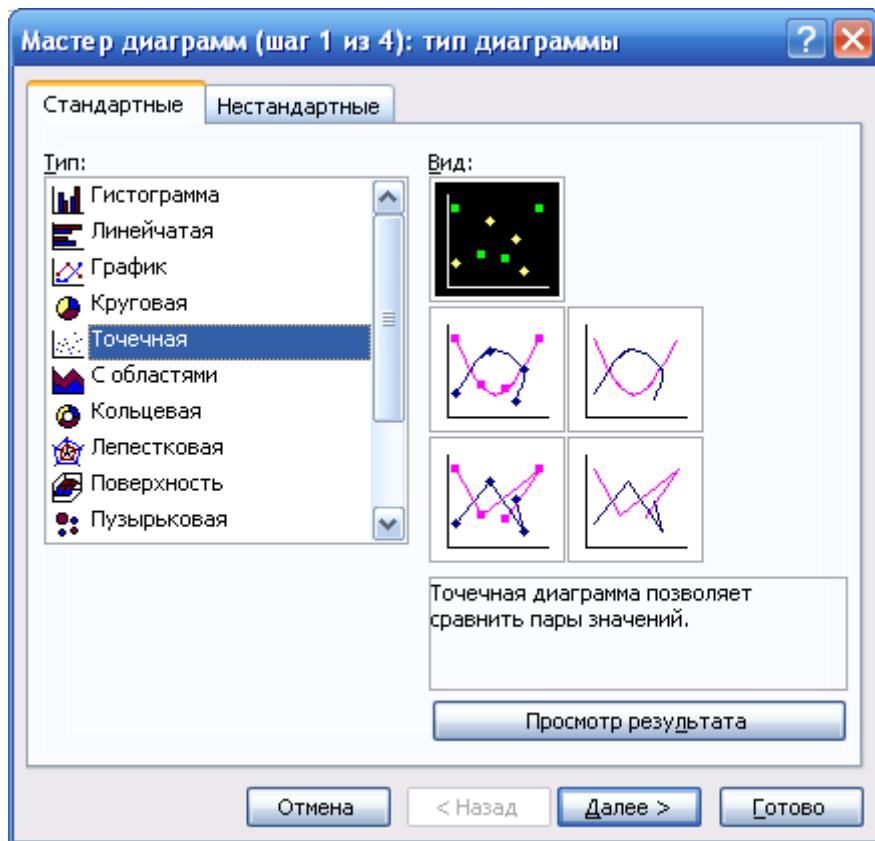


Рис. 8.2. Діалогове вікно «Мастер диаграмм» (крок 4)

Крок 5. Відзначаємо діапазон даних для створення діаграми. Для цього натискаємо клавішу «Далее» та мишею вводимо діапазон – виділяємо стовпчики **В** і **С**. Якщо ряди, за якими будується діаграма, розташовані не суміжно, то відмічаємо кожний з рядів окремо при натиснутій клавіші **Ctrl**.

Крок 6. Натискаємо клавішу «Далее» та вводимо назву діаграми та її осей (рис. 8.3).

Крок 7. Натискаємо клавішу «Готово» та отримуємо кореляційне поле для залежності між вагою тіла X_i , кг та показниками довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання. Після деяких перетворень діаграма буде мати такий вигляд (рис. 8.4).

Крок 8. Виділяємо діаграму, до якої необхідно додати лінію тренда.

Крок 9. З меню «Диаграмма» вибираємо команду «Добавить линию тренда», після чого відкривається діалогове вікно «Линия тренда» (рис. 8.5).

Крок 10. На вкладці «Тип» вибираємо придатний тип лінії тренда (лінійна, логарифмічна, поліноміальна, ступенева, експонентна). У розглянутому вище прикладі виберемо лінійну лінію тренда.

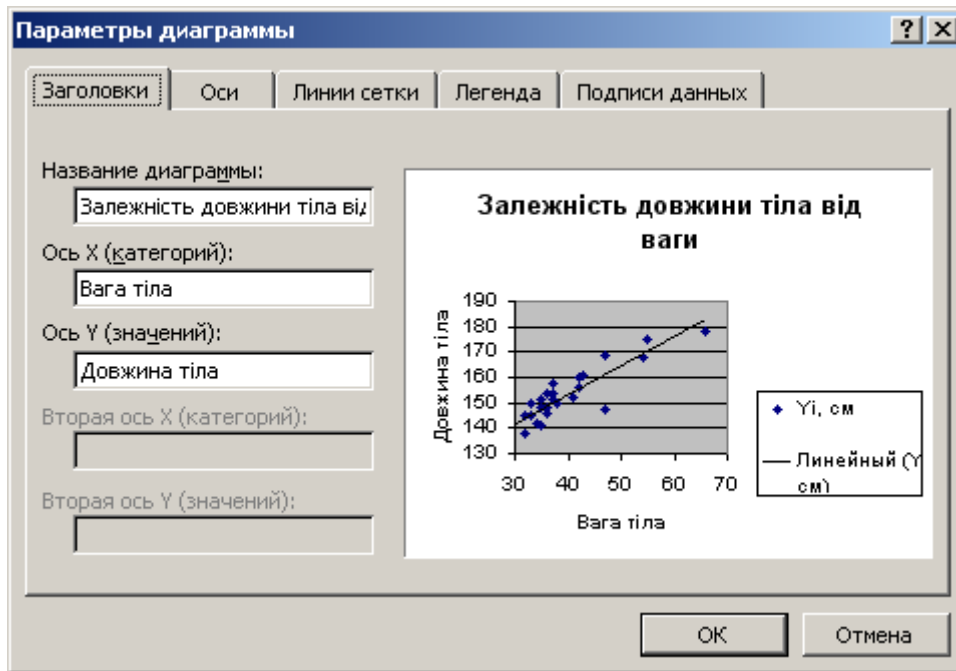


Рис. 8.3. Діалогове вікно «Мастер діаграмм» (крок б)

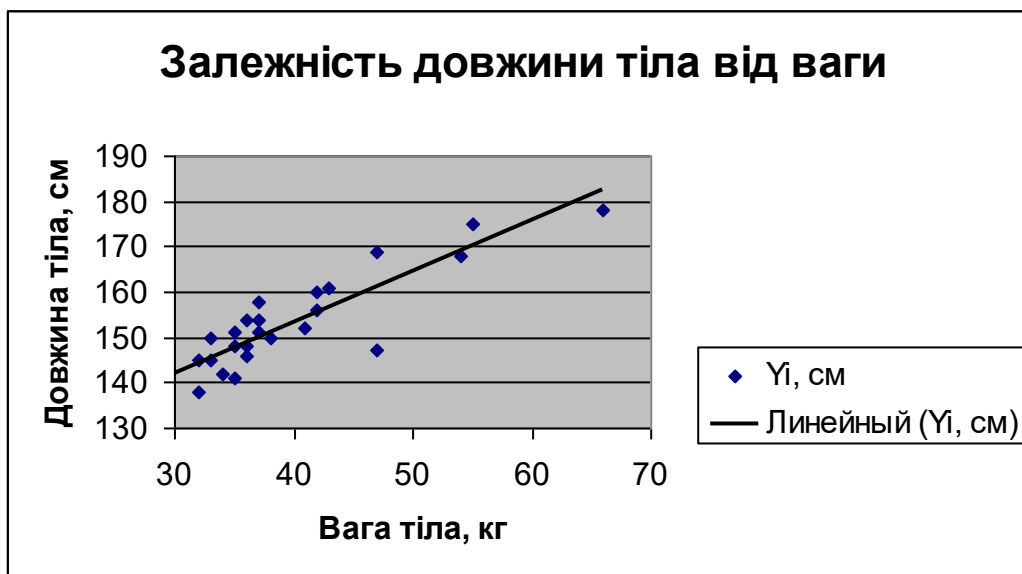


Рис. 8.4. Кореляційне поле залежності між вагою тіла X_i , кг та показниками довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання

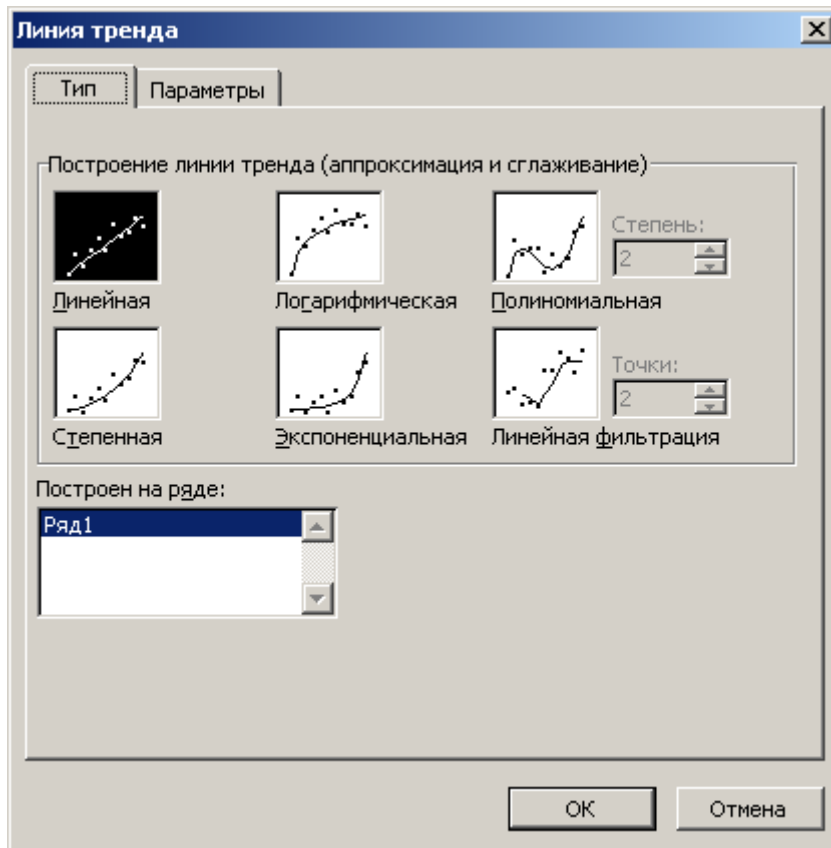


Рис. 8.5. Діалогове вікно команди «*Линия тренда*» (крок 9-10)

Крок 11. Натиснувши на вкладку «*Параметри*», одержуємо наступне діалогове вікно (рис. 8.6). У цьому вікні ставимо відмітку навпроти команд «*Показывать на диаграмме*» та «*Поместить на диаграмму величину достоверности аппроксимации (R^2)*».

Крок 12. Натискаємо кнопку «*ОК*» та одержуємо кореляційне поле залежності між вагою та довжиною тіла з лінією тренда та параметрами цієї лінії. Отримана діаграма має наступний вигляд (рис. 8.7). На ній представлено рівняння лінії регресії, за якою можна прогнозувати значення показників. Статистика R^2 називається коефіцієнтом детермінації та вказує наскільки точно рівняння лінії регресії описує досліджувану залежність (у даному прикладі R^2 дорівнює 0.775, тобто рівняння лінії регресії описує взаємозв'язок показників на 77.5%).

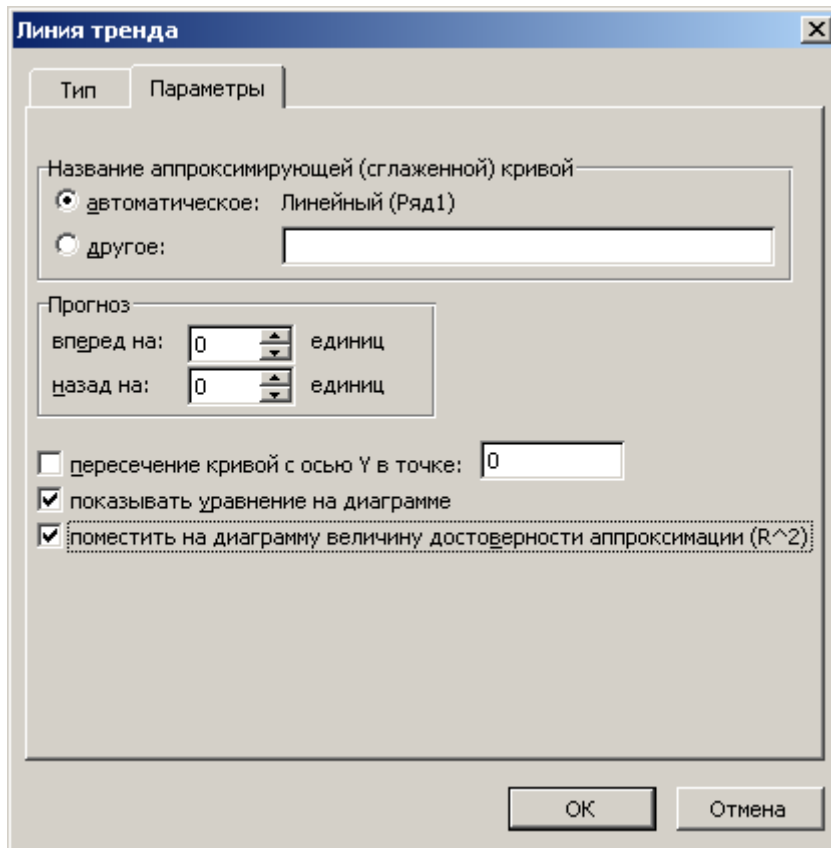


Рис. 8.6. Діалогове вікно команди «Линия тренда -Параметры» (крок 11)

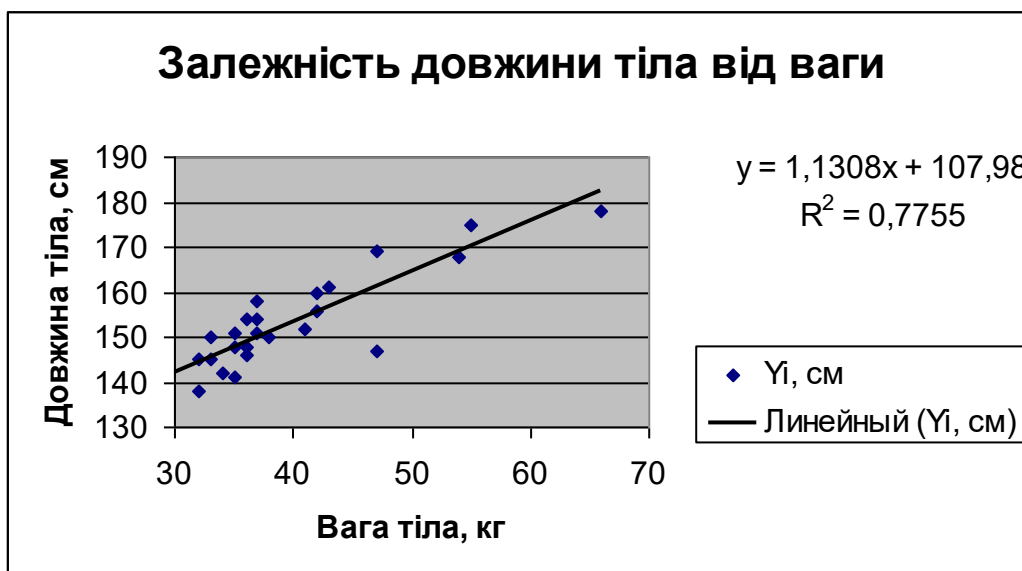


Рис. 8.7. Кореляційне поле з лінією тренду та його параметрами (для прямої залежності)

Крок 13. Тепер побудуємо кореляційне поле для зворотної залежності, тобто залежності між показниками довжини тіла X_i , см та вагою тіла Y_i , кг групи хлопчиків 10–12 років спецкласу з плавання. Для цього необхідно

виконати дії які були описані у кроках 4–12 для даних занесених у стовпчики Е та F. В результаті цього отримуємо наступну діаграму (рис. 8.8)

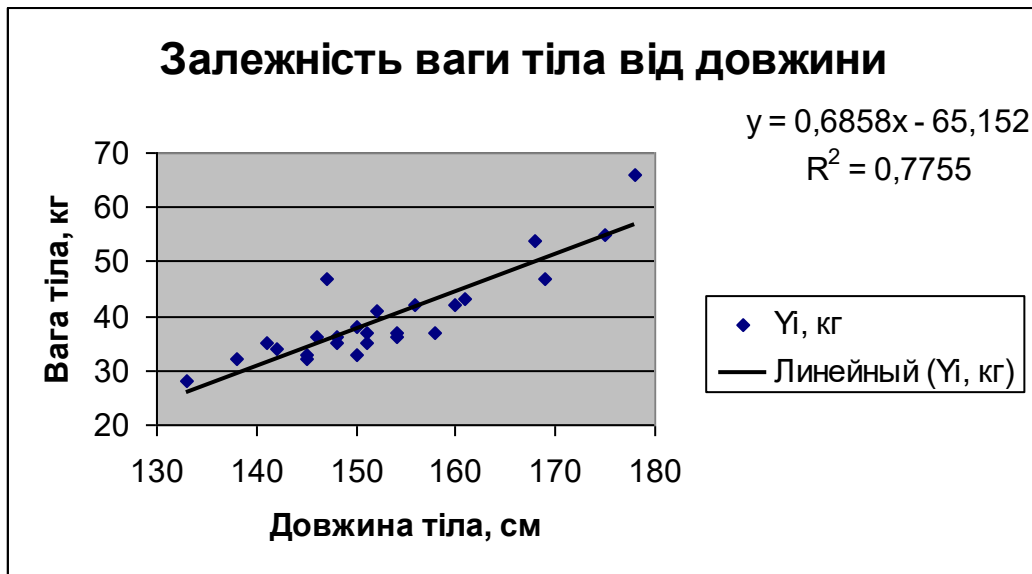


Рис. 8.8. Кореляційне поле з лінією тренду та його параметрами (для зворотної залежності)

Таким чином у результаті дій, які були описані у кроках 4–13 отримуємо такі результати (рис. 8.9)

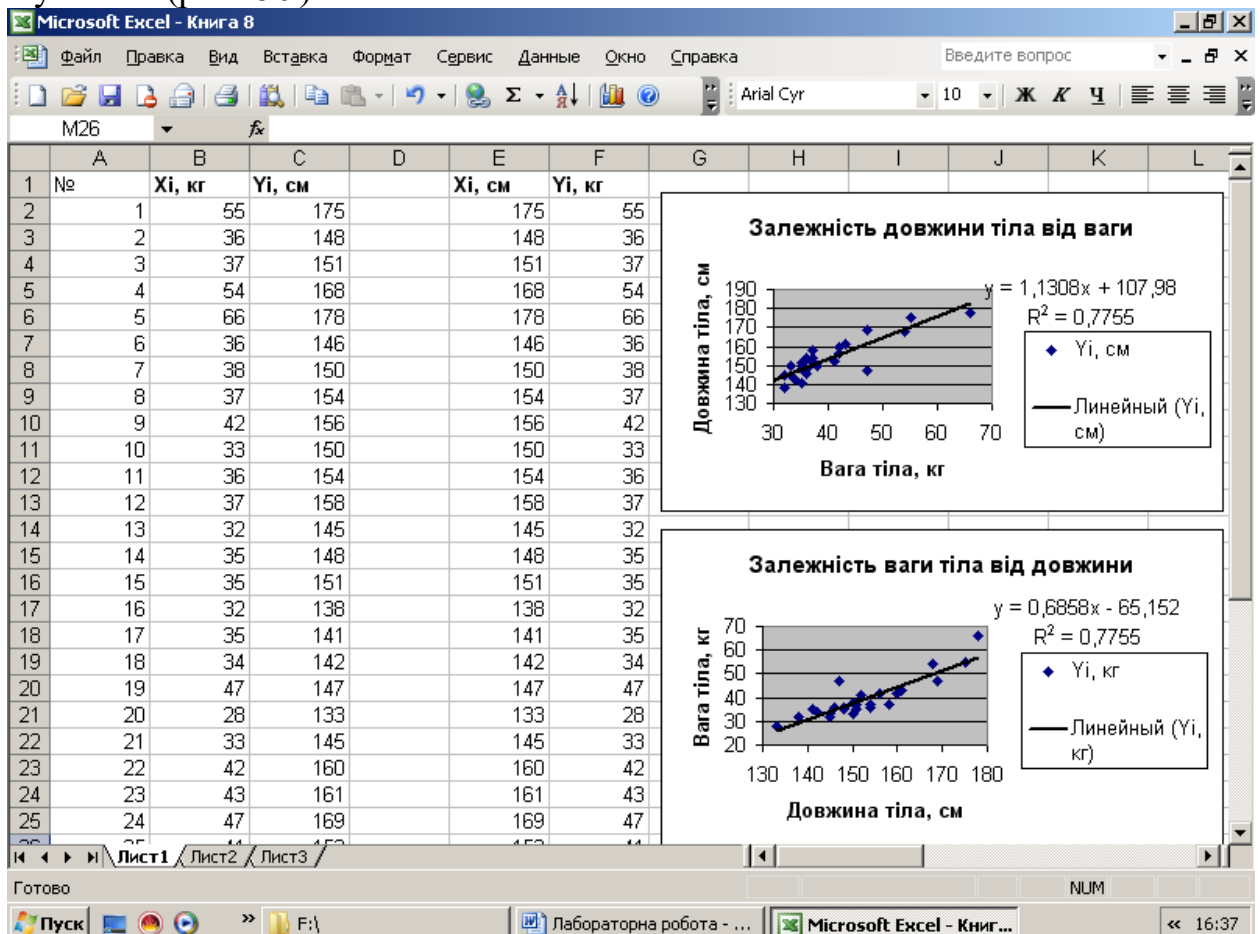



Рис. 8.9. Кореляційні поля та лінії тренду з параметрами для прямого та зворотного рівнянь регресії, які відображають залежність між вагою тіла X_i , кг та показниками довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання

Крок 14. Визначаємо точність прямого та зворотного рівнянь регресії. Для цього необхідно розрахувати: середнє арифметичне значення довжини та ваги тіла, середні квадратичні відхилення, залишкові середні квадратичні відхилення та погрішності прямого та зворотного рівнянь регресії. У стовпчику **A** записуємо назви цих величин: у комірці **A27** запишемо середнє арифметичне; у комірці **A28** – середнє квадратичне відхилення (або стандартне відхилення); у комірці **A29** – коефіцієнт детермінації R^2 ; у комірці **A30** – залишкове середнє квадратичне відхилення; у комірці **A31** – погрішність рівнянь регресії

Крок 15. Обраховуємо середнє значення. Для цього виділяємо комірку **B27** та скористаємося піктограмою вбудованих функцій , або командою головного меню «**Вставка - Функції**». Після цієї дії з'являється вікно «**Мастера функций**», у якому вибираємо категорію «**Статистические**» функції та функцію **СРЗНАЧ**.

Натискаємо клавішу **ОК**. При цьому з'являється діалогове вікно для вводу діапазону даних.

Інтервал даних для обчислення середнього арифметичного значення з'являється автоматично, тому що курсор стоїть внизу стовпчика з числовими значеннями. Однак інтервал у вікно діапазону даних можна ввести як із клавіатури, так і натиснувши на піктограму праворуч цього вікна та вибравши мишею діапазон.

Натискаємо клавішу **ОК**. У підсумку проведених дій у комірці **B27** відобразиться середнє арифметичне значення ваги тіла X_i , кг групи хлопчиків 10–12 років спецкласу з плавання.

Крок 16. Виконуємо такі самі дії у комірці **C27** та отримуємо середнє арифметичне значення довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання.

Крок 17. Обраховуємо стандартне відхилення (або середнє квадратичне відхилення) вибірки. Для цього, за допомогою «**Мастера функций**», або із клавіатури у комірку **B28** вводимо формулу **=СТАНДОТКЛОН(B2:B12)**.

Увага! Автоматично в діапазон даних будуть включені всі вище розташовані комірки разом з коміркою **B27** – середнім значенням. Тому завжди треба перевіряти та при необхідності коректувати діапазон даних. У підсумку проведених дій у комірці відобразиться значення стандартного відхилення ваги тіла X_i , кг групи хлопчиків 10–12 років.

Крок 18. Виконуємо аналогічні дії у комірці **C28** та отримуємо середнє квадратичне відхилення довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання.

Крок 19. Обраховуємо коефіцієнт детермінації R^2 . У зв'язку з тим, що при виконанні кроків 8-12 ми одержали кореляційне поле залежності між вагою

та довжиною тіла з лінією тренда та параметрами цієї лінії. На отриманій діаграмі (рис. 8.7) представлено рівняння лінії регресії та статистика R^2 , яка називається коефіцієнтом детермінації та вказує наскільки точно рівняння лінії регресії описує досліджувану залежність (у даному прикладі R^2 дорівнює 0.775). Тобто ми вже маємо необхідне значення коефіцієнта детермінації R^2 , і тому у комірку **B29** та у комірку **C29** із клавіатури вводимо числове значення R^2 , у даному випадку воно дорівнює 0.775.

Крок 20. Обраховуємо залишкове середнє квадратичне відхилення для прямого рівняння, тобто для залежності довжини тіла Y_i , см від ваги тіла X_i , кг групи хлопчиків 10–12 років спецкласу з плавання. Для цього із клавіатури у комірку **C30** вводимо формулу = **C28*КОРЕНЬ(1-C29)**.

Крок 21 Обраховуємо залишкове середнє квадратичне відхилення для зворотного рівняння, тобто для залежності ваги тіла X_i , кг від довжини тіла Y_i , см групи хлопчиків 10–12 років спецкласу з плавання. Для цього із клавіатури у комірку **B30** вводимо формулу = **B28*КОРЕНЬ(1-B29)**.

Крок 22. Розраховуємо погрішність рівнянь регресії. Для прямого рівняння із клавіатури у комірку **C31** вводимо формулу = **C28*КОРЕНЬ(1-C29)**.

Крок 23.

Крок 1.

Крок 1.

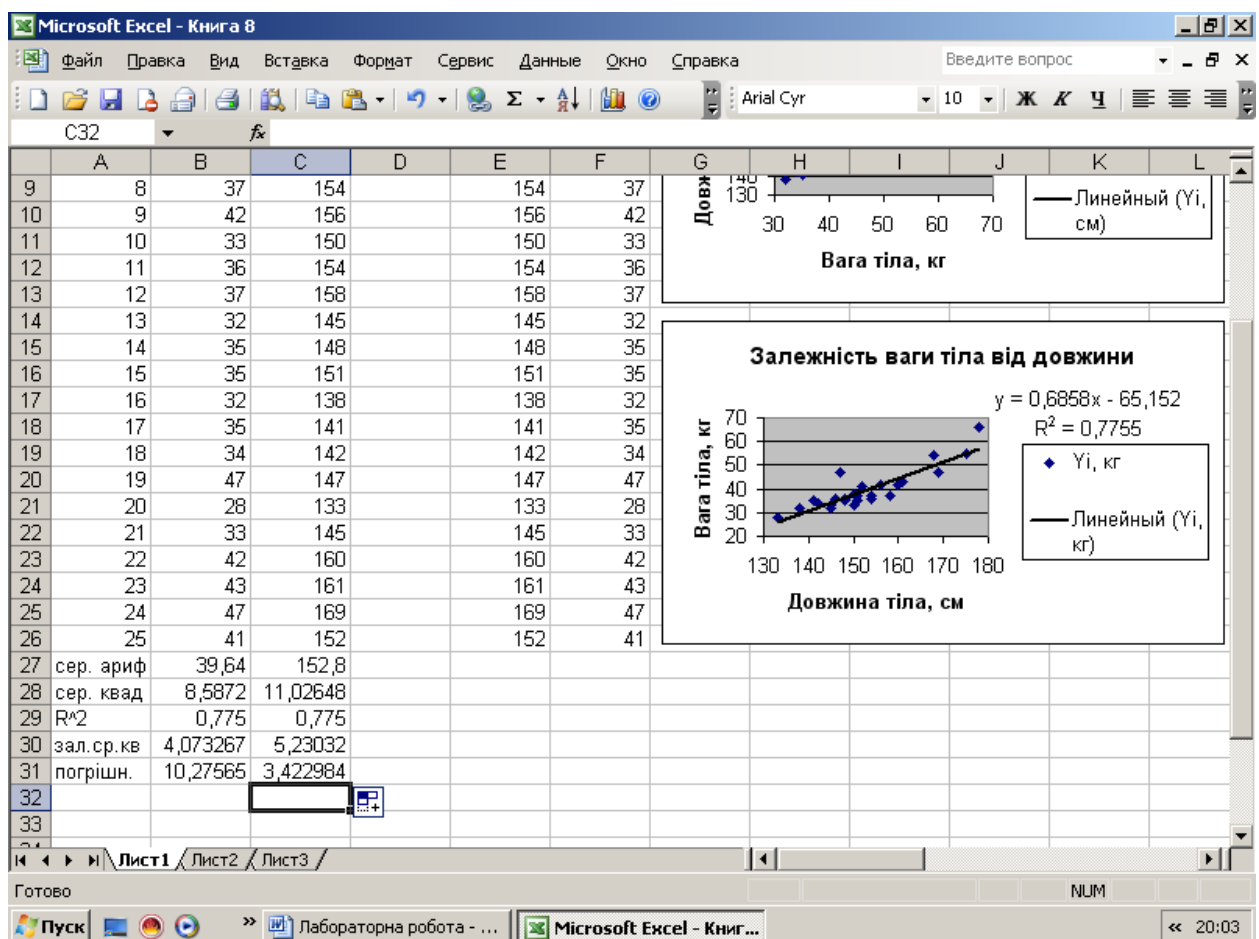


Рис. 8.10. Результати розрахунків кроків 14-21

Висновок: Отримані рівняння регресії (найбільш імовірні лінійні залежності) мають такий вигляд:

$$y(\text{см}) = 108,5 + 1,11 \times x(\text{кг})$$

$$x(\text{кг}) = -61,72 + 0,66 \times y(\text{см}),$$

причому, рівняння $y = 108,5 + 1,11 \times x$ точніше описує залежність показників ваги і довжини тіла для досліджуваної групи хлопчиків 10-12 років спецкласу з плавання.

Практичне завдання:

Скласти умови завдання з власного наукового дослідження і вирішити його за допомогою кореляційного аналізу. Проаналізувати результати, зробити висновок.

Питання для самоконтролю:

1. Що таке функціональний взаємозв'язок між показниками?
2. Що таке статистичний взаємозв'язок між показниками?
3. Поясніть, у чому полягає сутність кореляції.
4. У чому полягає основне завдання кореляційного аналізу?
5. Як називається графічне відображення взаємозв'язку показників?
6. Що дозволяє виявити візуальний аналіз кореляційного поля?
7. Яку розрізняють форму залежності?
8. Яка існує спрямованість залежності між показниками?
9. Який маємо вид залежності, коли кореляційне поле представлене прямою лінією?
10. За якою формулою розраховується коефіцієнт кореляції?
11. Як розраховується і на що вказує коефіцієнт детермінації (D)?
12. Яку розрізняють щільність взаємозв'язку між показниками?
13. У яких межах знаходиться значення коефіцієнта кореляції?

Змістовий модуль 5. Аналіз і прогнозування часових рядів

Тема 14. Загальні положення методу

Ряди динаміки відображають розвиток явища або процесу в часі. Вони являють собою співвідношення рівнів, тобто показників вимірюваного процесу та періодів часу, протягом яких ці рівні були зафіксовані.

На підставі рядів динаміки можливе вирішення двох принципових завдань – аналізу й прогнозу. Аналіз відбувається для будь-якого явища,

представленого вимірами результатів, що змінюються із часом. Особливо важливий подібний аналіз у дослідженнях дитячого та юнацького спорту, де зміни досліджуваних параметрів відбуваються не тільки під впливом тренувальних навантажень, але й відповідно до розвитку організму спортсменів.

Прогноз спортивних процесів відбувається тільки тоді, коли його результати мають безпосереднє відношення до подальшої роботи з досліджуванним: прогноз спортивного результату, стан тренуваності, прогноз розвитку якоїсь якості тощо.

У деяких випадках має сенс на тому самому емпіричному матеріалі поєднувати аналіз та прогноз.

За допомогою рядів динаміки можливо:

1. Аналізувати й прогнозувати властивості досліджуваних.
2. Аналізувати й прогнозувати властивості індивіда.
3. Оцінити й передбачити вікові зміни.
4. Відобразити біомеханічні змінні.

Як правило, ряди динаміки відображають у вигляді графіків, таблиць і формул. Найбільш простий спосіб побудови рядів динаміки – графічний.

Наприклад, школярі були досліджені за величиною середньостатистичної станової сили, N кг (ньютон на 10 кг маси тіла) в 9, 10, 11, 12 і 13 років. Представити ці зміни динамічним рядом.

t_i	9	10	11	12	13
U_i	89,2	91,6	94,4	97,0	98,6

Розглянемо U_i як рівні динамічного ряду, зафіксовані в період часу t_i . За цими даними необхідно побудувати графік $U_i = t(t_i)$, на якому рівні отриманого динамічного ряду будуть відкладені за відповідними періодами часу в системі прямокутних координат. Якщо знайдені точки з'єднати послідовно відрізками прямих, одержимо *статистичну криву* – лінію, що відображає реальну зміну досліджуваного явища у часі. Статистична крива дозволяє оцінювати досліджуваний процес у динаміці.

Основними характеристиками динамічного ряду є:

- абсолютний приріст U ,
- темп росту T ,
- темп приросту $T_{пр}$.

Абсолютний приріст – це різниця між наступним U_i і попереднім U_{i-1} рівнем ряду: $U = U_i - U_{i-1}$

Абсолютний приріст може бути *позитивним* – указувати на зростання досліджуваного явища і *негативним* – указувати на убавання.

Темп росту динамічного ряду являє собою відношення наступного рівня до попереднього. Якщо рівні належать до одного й того ж вихідного (базисного) рівня, вони називаються *базисними*, при послідовному відношенні наступного до попереднього – *ланцюговими*.

$$T_n = \frac{Y_i}{Y_{i-1}} \text{ – ланцюговий темп росту}$$

$$T_B = \frac{Y_i}{Y_0} \text{ – базисний темп росту}$$

Темп росту за весь період – це відношення наступного рівня до базисного:

$$T = \frac{Y_n}{Y_1} \text{ – темп росту за весь період.}$$

Темп приросту – відношення абсолютного приросту ΔY до базисного рівня Y_{i-1} :

$$T_{np} = \frac{\Delta Y}{Y_{i-1}} = \frac{Y_i - Y_{i-1}}{Y_{i-1}}$$

Статистична крива являє собою ламану лінію, вузлові точки якої коливаються щодо якогось напрямку. Визначення цього напрямку – тенденції (тренда) – це визначення тих рівнів, які адекватні вихідним даним. За своєю сутністю тенденція є закономірністю, відповідно до якої змінюється у часі досліджуване явище. Така закономірність дозволяє розв'язати два принципових завдання:

- 1) з її допомогою можна оцінити досліджуване явище;
- 2) можна його прогнозувати.

У такій ситуації рішення зводиться до виявлення тенденції й роботи з нею.



Тема 15. Способи побудування лінії тренду

Виявлення тенденції (тренда) можна здійснити декількома способами, як-от:

- 1) Метод рухомої середньої.
- 2) Метод найменших квадратів.

1. За методом рухомої середньої вихідні показники осереднюються та їх середні значення приймаються за рівні динамічного ряду, на підставі яких будується статистична крива, що відображає зміну досліджуваного явища у часі. Таким чином, визначивши тенденцію, відповідно до якої розвивається у часі досліджуване явище, можна проводити його аналіз і прогноз.

2. Метод найменших квадратів – аналітичний спосіб виявлення тенденції. У ході алгебраїчних рівнянь визначаються коефіцієнти прямої лінії $Y = B_0 + B \times X$, що відображає статистичну закономірність зміни у часі досліджуваного явища. B_0 ; B – коефіцієнти, і вони розраховуються за формулами:

$$B_0 = \frac{\sum_i X_i^2 \times \sum_i Y_i - \sum_i X_i Y_i \times \sum_i X_i Y_i}{n \times \sum_i X_i^2 - (\sum_i X_i)^2},$$

де n – число вимірів

$$B = \frac{n \times \sum_i X_i Y_i - \sum_i X_i \sum_i Y_i}{n \times \sum_i X_i^2 - (\sum_i X_i)^2}$$

Отже, метод найменших квадратів дає відповідь на питання, як провести криву через експериментально отримані точки, щоб ця крива ближче всього підходила до істинної кривої шуканої функції.

Не можна проводити криву за отриманими точками. Її треба провести так, щоб сума квадратів вертикальних відхилень експериментальних точок від проведеної кривої вийшла мінімальною.

Після одержання статистичної кривої можна проводити аналіз і прогнозування досліджуваного показника.



Тема 16. Вирішення типових завдань за допомогою рядів динаміки



Розглянемо послідовність операцій при відображенні у часі явища, що підлягає дослідженню; його оцінці й прогнозуванні за допомогою рядів динаміки на конкретному прикладі.

Приклад: У веслярів виміряли імпульс сили U_i , Н/с (Ньютон за 1с на 10кг маси), протягом багатьох тренувальних занять.

Виміри проводилися щомісяця - 6 місяців поспіль (t_i). Оцінити динаміку імпульсу сили у даних веслярів. Зробити прогноз на сьомий місяць роботи.

Вихідні дані в таблиці:

t	0	1	2	3	4	5	6
U_i	7.8	7.	7.9	8.1	8.	8	8
U_i		9			0	.2	.3

Порядок виконання

1. Вихідні дані дозволяють виявити тенденцію у вигляді прямої АВ. Ця тенденція вказує на поступове збільшення імпульсу сили з часом.

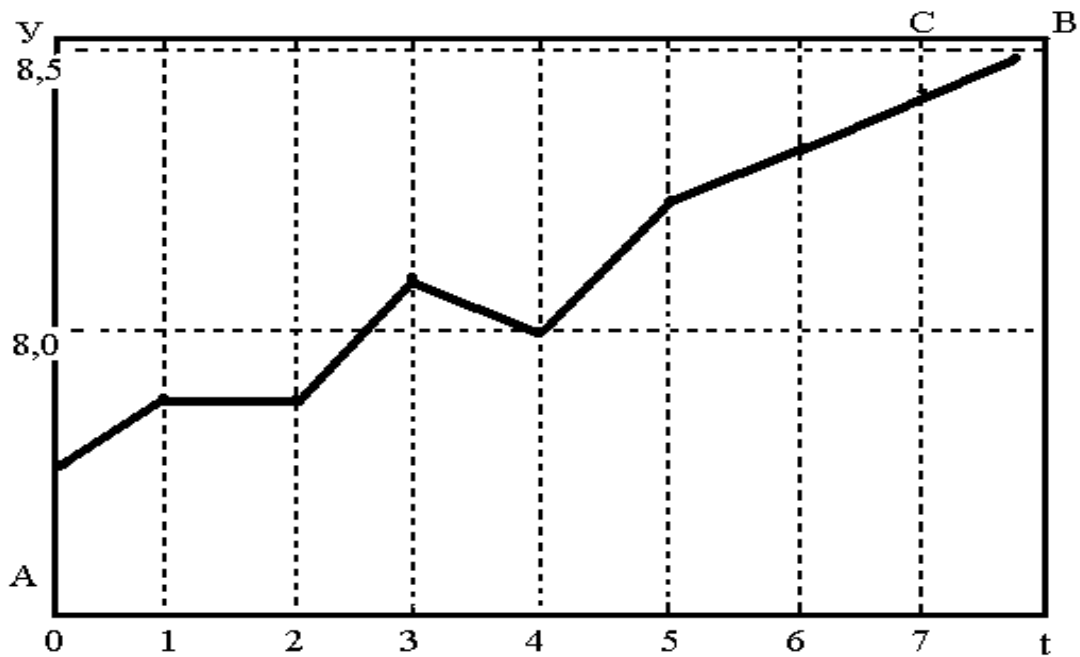


Рисунок 6.2 - статистична крива

Для прогнозу з $t_i = 7$ проводимо вертикаль до точки С (до перетинання з прямою), горизонталь із якої приводить до шуканого прогнозу $Y_7 = 8,4$ ум. од.

Виявлення тенденції можна зробити методом рухомої середньої.

Розглянемо даний приклад таким чином, що кожні 3 місяці показники імпульсу осереднюються:

$$\bar{Y}_1 = \frac{7,8 + 7,9 + 7,9}{3} = \frac{23,6}{3} = 7,87 ;$$

$$\bar{Y}_2 = \frac{7,9 + 7,9 + 8,1}{3} = \frac{23,9}{3} = 7,97 ;$$

$$\bar{Y}_3 = \frac{7,9 + 8,1 + 8,0}{3} = \frac{24}{3} = 8,00 ;$$

$$\bar{Y}_4 = \frac{8,1 + 8,0 + 8,2}{3} = \frac{24,3}{3} = 8,10 ;$$

$$\bar{Y}_5 = \frac{8,0 + 8,2 + 8,3}{3} = \frac{24,5}{3} = 8,17$$

t_i	Y_i	\bar{Y}_i
0	7,8	-
1	7,9	7,87
2	7,9	7,97
3	8,1	8,00
4	8,0	8,10
5	8,2	8,17

6	8,3	-
---	-----	---

Будуємо графік на підставі середніх показників

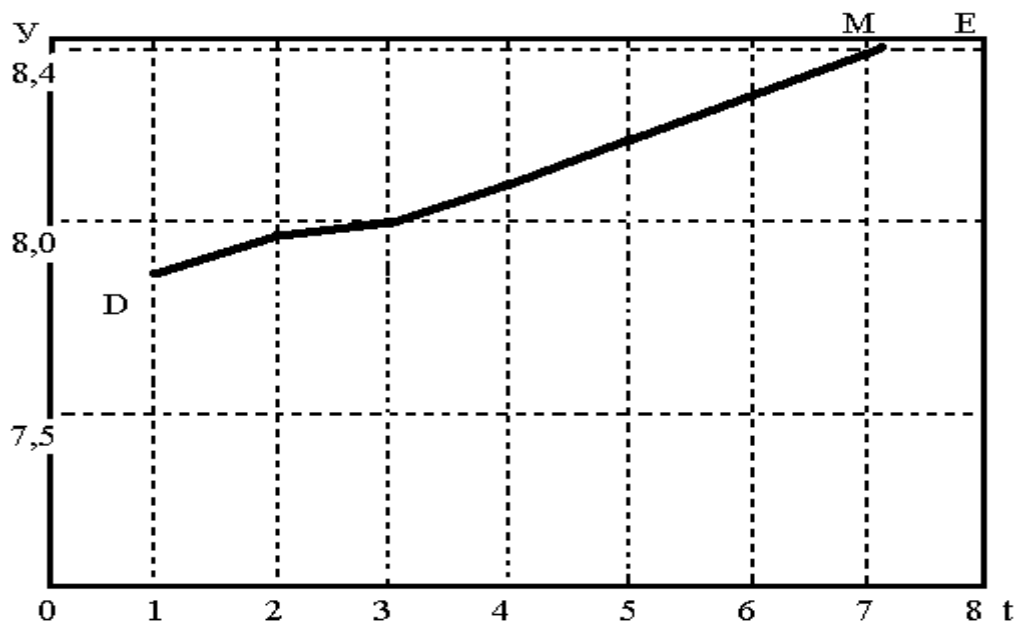


Рисунок 6.3 - статистична крива

Відповідно до методу рухомої середньої, визначення вихідних даних може йти не тільки за трьома крапками, але й за 5, 6 і тощо.

Для прогнозу з $t_i = 7$ продовжуємо пряму DE до вертикалі М (до перетинання із прямою) і горизонталь із точки М приводить до шуканого прогнозу $U_7 = 8,4$ ум. од., що збігається з прогнозом наведеним вище.

2. Тепер визначаємо основні характеристики динамічного ряду:

1) Абсолютний приріст: $\Delta Y = Y_i - Y_{i-1}$

$$\Delta Y_1 = Y_1 - Y_0 = 7,9 - 7,8 = 0,1 \quad \Delta Y_2 = Y_2 - Y_1 = 7,9 - 7,9 = 0,0$$

$$\Delta Y_3 = Y_3 - Y_2 = 8,1 - 7,9 = 0,2 \quad \Delta Y_4 = Y_4 - Y_3 = 8,0 - 8,1 = -0,1$$

$$\Delta Y_5 = Y_5 - Y_4 = 8,2 - 8,0 = 0,2 \quad \Delta Y_6 = Y_6 - Y_5 = 8,3 - 8,2 = 0,1$$

2) Темп росту динамічного ряду:

а) $T_{лi} = \frac{Y_i}{Y_{i-1}}$ - ланцюговий темп росту

$$T_{л1} = \frac{Y_1}{Y_0} = \frac{7,9}{7,8} = 1,013$$

$$T_{л2} = \frac{Y_2}{Y_1} = \frac{7,9}{7,9} = 1,0$$

$$T_{л3} = \frac{Y_3}{Y_2} = \frac{8,1}{7,9} = 1,025$$

$$T_{л4} = \frac{Y_4}{Y_3} = \frac{8,0}{8,1} = 0,988$$

$$T_{л5} = \frac{Y_5}{Y_4} = \frac{8,2}{8,0} = 1,025$$

$$T_{л6} = \frac{Y_6}{Y_5} = \frac{8,3}{8,2} = 1,012$$

б) $T_б = \frac{Y_i}{Y_0}$ - базисний темп росту

$$T_{B1} = \frac{Y_1}{Y_0} = \frac{7,9}{7,8} = 1,013$$

$$T_{B3} = \frac{Y_3}{Y_0} = \frac{8,1}{7,8} = 1,039$$

$$T_{B5} = \frac{Y_5}{Y_0} = \frac{8,2}{7,8} = 1,051$$

$$T_{B2} = \frac{Y_2}{Y_0} = \frac{7,9}{7,8} = 1,013$$

$$T_{B4} = \frac{Y_4}{Y_0} = \frac{8,0}{7,8} = 1,026$$

$$T_{B6} = \frac{Y_6}{Y_0} = \frac{8,3}{7,8} = 1,064$$

T_{B6} - темп росту за весь період.

3) Темп приросту

$$T_{np} = \frac{\Delta Y}{Y_{i-1}} = \frac{Y_i - Y_{i-1}}{Y_{i-1}} = \frac{Y_i}{Y_{i-1}} - 1$$

$$T_{np1} = \frac{Y_1}{Y_0} - 1 = \frac{7,9}{7,8} - 1 = 1,013 - 1 = 0,013$$

$$T_{np2} = \frac{Y_2}{Y_1} - 1 = \frac{7,9}{7,9} - 1 = 1 - 1 = 0$$

$$T_{np3} = \frac{Y_3}{Y_2} - 1 = \frac{8,1}{7,9} - 1 = 1,025 - 1 = 0,025$$

$$T_{np4} = \frac{Y_4}{Y_3} - 1 = \frac{8,0}{8,1} - 1 = 0,988 - 1 = -0,012$$

$$T_{np5} = \frac{Y_5}{Y_4} - 1 = \frac{8,2}{8,0} - 1 = 1,025 - 1 = 0,025$$

$$T_{np6} = \frac{Y_6}{Y_5} - 1 = \frac{8,3}{8,2} - 1 = 1,012 - 1 = 0,012$$

Висновок: Отримана статистична крива АВ (тенденція) зазначає на поступове збільшення імпульсу сили під впливом тренувальних навантажень. Абсолютний приріст позитивний (за винятком рівня динамічного ряду, що відповідає $t=4$) і вказує на поступове збільшення якості, що вимірюється. Однак, темп приросту нерівномірний, і на 4-ому місяці тренування імпульс сили зменшується у порівнянні з 3-м місяцем.

Відповідно до тенденції динамічного ряду на 7-й місяць тренувань імпульс сили у даної групи веслярів буде дорівнювати 8,4 Н/с (Ньютон за 1с на 10 кг маси).

Практичне завдання:

Скласти умови завдання з власного наукового дослідження і здійснити аналіз і прогноз якостей об'єктів, що підлягають дослідженню.

? Питання для самоконтролю:

1. Який метод математичної статистики застосовується для знаходження найбільш імовірних лінійних залежностей між двома факторами?
2. За яких умов ми маємо можливість побудувати пряму регресію?
3. Який вигляд має рівняння прямої лінії?
4. Наведіть пряме та зворотне рівняння регресії.
5. Як розраховуються коефіцієнти рівнянь регресії?
6. Який показник характеризує точність рівняння регресії?
7. Як розраховується похибка рівняння залежності між двома факторами?
8. Поясніть, як визначити точність рівнянь регресії.
9. Що таке ряди динаміки?
10. У якому вигляді можуть бути представлені ряди динаміки?
11. Які завдання можна вирішити за допомогою рядів динаміки?
12. Назвіть основні характеристики рядів динаміки.
13. Розкрийте сутність основних характеристик динамічного ряду.
14. На що вказує позитивний абсолютний приріст?
15. Яким чином змінюється явище у часі, якщо абсолютний приріст має негативне значення?
16. Як змінюється явище у часі, якщо темп росту динамічного ряду дорівнює одиниці?
17. Як називається лінія, що відображає реальну зміну досліджуваного явища у часі?
18. Назвіть способи виявлення тенденції (тренда) динамічного ряду.
19. Яким чином розраховується рухоме середнє арифметичне значення?
20. Розкрийте сутність методу найменших квадратів.
21. Як здійснити прогноз за допомогою методу рухомої середньої?

Змістовий модуль 6. Факторний (дисперсійний) аналіз

Тема 17. Загальні положення методу

Дисперсійний аналіз – це статистичний метод, за допомогою якого можна встановити вплив однієї ознаки на іншу. На відміну від кореляційного аналізу, де також визначається вплив ознак, розглядається їх більш складна структура.

Дисперсійний аналіз має перевагу над кореляційним: тут можна врахувати вплив ознаки, що не має кількісного вираження.

Ознака, що підлягає дослідженню, називається *результативною* (R).

Вона випробовує на собі вплив факторної ознаки F , чия структура підрозділяється на ряд підгруп, так званих *градацій фактора*.

Для дослідження залежності результативної ознаки від факторної обирається спеціальна вибірка, яка називається *дисперсійним комплексом*, із загальним числом елементів N .

Вихідні позиції методу базуються на сумі квадратів відхилень варіантів від їх середньої:

$$S = \sum_{i=1}^{i=k} (X_i - \bar{X})^2 \times n_i, \quad (1)$$

де S – сума квадратів відхилень;

X_i – варіанта вихідного ряду;

n_i – частота вихідного ряду;

\bar{X} – середнє арифметичне значення вихідного ряду.

Встановлено, що сума квадратів відхилень на всій результативній ознаці S_y складається із двох додатків:

$$S_y = S_x + S_z \quad (2)$$

де S_x – сума квадратів відхилень, обумовлена варіацією середніх арифметичних усередині градації щодо середньої арифметичної всього комплексу.

Ця варіація враховує вплив досліджуваного фактора на результативну ознаку.

S_z – сума квадратів відхилень, обумовлена варіацією вихідних даних щодо середніх арифметичних усередині кожної градації.

Ця варіація обумовлена впливом так званих випадкових причин, тобто причин, не пов'язаних з дією фактора.

Рівняння (2) є *основним рівнянням дисперсійного аналізу*.

Отже, після визначення значення S_x , S_y , S_z можна оцінити загальну варіацію результативної ознаки (S_y) та її складові частини: варіацію, яка обумовлена факторною ознакою (S_x), і варіацію від випадкових причин (S_z).

Надалі необхідно перейти до порівняльного аналізу суми квадратів відхилень S_x і S_z . Аналіз проводиться за допомогою критерію Фішера.

Якщо значення варіацій, обумовлених факторною ознакою й випадковими причинами, відрізняються статистично - вірогідно, то робимо висновок, що розглянута факторна ознака впливає на результативну ознаку.



Тема 18. Види факторного аналізу

У практичному використанні градації факторної ознаки помітні між собою кількісно, у якісному ж відношенні вони являють собою однозначну характеристику.

Наприклад, оцінюючи вплив вікового розвитку дітей на який-небудь показник, дисперсійний комплекс складається з результативної ознаки (досліджуваний показник) і факторної (вік дітей). При цьому вік дітей, які

підлягають дослідженню, доречно розподілити за віковими категоріями: 10-12 років; 13-14 років; 14-16 років тощо. Вони й будуть являти собою градації фактора.

Якщо завдання ускладнюється і факторна ознака змінює свою структуру в якісному відношенні, розглядають дисперсійний аналіз двох, трьох і більше багатofакторних комплексів.



Тема 19. Застосування дисперсійного аналізу при проведенні досліджень в фізичному вихованні та спорті

Практична реалізація дисперсійного аналізу:

1. Оцінка тренувальних впливів: якщо розглядати спортивний результат або яку-небудь кількісну характеристику результату як результативну ознаку, то обсяг тренувального навантаження, його інтенсивність або співвідношення обсягу й інтенсивності можна розглядати як факторну ознаку.

Градаціями фактора в цьому випадку є різні кількісні межі обсягу і інтенсивності навантажень або їх співвідношень.

Використовуючи дисперсійний аналіз, можна оцінити вплив ознаки на кінцевий результат спортивної діяльності.

2. Вплив умов спортивної діяльності: під умовами спортивної діяльності будемо розуміти різноманітні види робіт з учасниками досліджень: різні програми, методики, режими, види робіт спортивних спеціалізацій, умови експерименту.

Параметр спортивної діяльності приймається як результативна ознака.

Факторною ознакою є умови роботи спортсмена, що відрізняються різними кількісними межами – градаціями фактора.

Застосування дисперсійного аналізу дозволяє оцінити вплив різноманітних умов роботи на параметр спортивної діяльності, який підлягає дослідженню.

3. Вплив біологічного й соціального факторів: як факторну ознаку зручно представити різноманітні біологічні й соціальні фактори: вік, стать, різні медичні групи досліджуваних, ставлення до спорту, соціальні групи населення, професійні особливості.

Вихідними даними дисперсійного аналізу будемо вважати X_i – вимір практичних спортивних величин та їх частоти n_i .

Величини X_i вимірюються за умов дії на досліджуваних факторної ознаки F , що має градації F_1, F_2, \dots, F_n .

Градації помітні за кількісними характеристиками. При виконанні статистичних операцій дисперсійного аналізу насамперед визначаються суми квадратів відхилень за формулами:

$$S_y = \sum_{i=1}^{i=N} X_i^2 - \frac{(\sum_{i=1}^{i=N} X_i)^2}{N}, \quad (3)$$

де S_y – загальна сума квадратів відхилень варіантів результативної ознаки;

X_i – варіанти результативної ознаки;

N – обсяг усього дисперсійного комплексу.

$$S_x = \sum_{j=1}^{j=f} \frac{(\sum_{i=1}^{i=k} X_i)^2}{n_k} - \frac{(\sum_{i=1}^{i=N} X_i)^2}{N}, \quad (4)$$

де S_x – сума квадратів відхилень варіантів результативної ознаки, обумовлена впливом фактора;

$i = 1, 2, \dots, N$;

N – обсяг дисперсійного комплексу;

$j = 1, 2, \dots, f$ – кількість градацій;

k – обсяг кожної градації дисперсійного комплексу.

$$S_z = \sum_{i=1}^{i=N} X_i^2 - \sum_{j=1}^{j=f} \times \frac{(\sum_{i=1}^{i=k} X_i)^2}{n_k}, \quad (5)$$

де S_z – сума квадратів відхилень варіантів результативної ознаки, що обумовлена впливом випадкових причин;

$j = 1, 2, \dots, f$;

f – кількість градацій.

Для визначення S_x , S_y , S_z у практичних задачах необхідно вихідні дані й проміжні обчислення подати в таблиці (див. табл. у прикладі).

Визначивши S_x , S_y , S_z , необхідно переконатися у виконанні основного рівняння дисперсійного аналізу:

$$S_y = S_x + S_z$$

Після цього можна приступити до загальної оцінки впливу факторної ознаки на результативну за допомогою загальної суми квадратів відхилень і її складових.

Надалі приступаємо до оцінки впливу факторної ознаки на результативну за допомогою оцінки на статистичну вірогідність.

Критерій Фішера, що аналізує статистичну вірогідність (невірогідність), визначається як частка від ділення порівнюваних дисперсій.

Отже, порівнюючи вплив факторної ознаки й випадкових причин, необхідно визначити дисперсії результативної ознаки, що залежать від фактора й випадку.

Для їхнього визначення необхідно знайти відповідні ступені свободи:

$k_y = N - 1$;

$k_x = j - 1$;

$$k_z = (N - 1) - (j - 1) = N - J, \quad (6)$$

де k_y – ступені свободи загальної дисперсії;

k_x – ступені свободи дисперсії, яка обумовлена фактором;

k_z – ступені свободи дисперсії, яка обумовлена випадковими причинами. Дисперсії визначаються як:

$$\sigma_y^2 = \frac{S_y}{k_y}; \quad \sigma_x^2 = \frac{S_x}{k_x}; \quad \sigma_z^2 = \frac{S_z}{k_z} \quad (7)$$

Тепер визначимо критерій Фішера:

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \quad (8)$$

де σ_1^2 і σ_2^2 – дисперсії від факторної ознаки й від випадкових причин. Відповідно до вимог критерію Фішера в чисельнику знаходиться більша із цих дисперсій ($\sigma_1^2 > \sigma_2^2$)

Критерій, що розраховали, порівнюється з $F_{гр}$ – граничним значенням критерію, який визначається за таблицею Фішера за ступенями свободи k_x і k_z та надійністю P .

Проводимо порівняння F і $F_{гр}$ та робимо висновок:


1 якщо $F > F_{гр}$ – розходження між впливом від факторної ознаки і від випадкових причин статистично вірогідне, тобто фактор, який підлягає дослідженню, впливає на результативну ознаку;

2 якщо $F < F_{гр}$ – розходження статистично невірогідне, і фактор, який підлягає дослідженню, принципово не впливає на результативну ознаку.

Оцінити ступінь впливу фактора, який підлягає дослідженню, на результативну ознаку можна за допомогою такого показника:

$$\eta = \frac{S_x}{S_y} \times 100\%$$

Величина, розрахована за цією формулою, показує, який відсоток варіації результативної ознаки визначається впливом фактора, що підлягає дослідженню.

 Розглянемо послідовність операцій при вивченні впливу факторної ознаки, залежно від її градацій, на результативну ознаку в практиці спортивної діяльності на конкретному прикладі.

Приклад: Вивчаємо характеристику бігу, опорний показник бігу G .

$$G = \frac{t_R}{t_A}$$

де t_A та t_R - час, відповідно, реактивної та активної фази опори.

Припускаємо, що опорний показник бігу перебуває під впливом різних видів дистанцій забігу:

F_1 - короткі; F_2 - середні; F_3 - довгі.

Досліджуємо $N = 10$ спортсменів, з яких $n_1=3$ є бігунами на короткі дистанції, $n_2 =4$ - бігунами на середні дистанції, $n_3=3$ - бігунами на довгі дистанції.

Розраховуємо G для всіх 10 спортсменів і вносимо її як вихідні дані у два рядки таблиці - X_i , n_i .

У таблиці робимо всі допоміжні обчислення:

	F1	F2	F3	Σ
X_i	2,0; 2,1; 2,2	1,9; 1,9; 2,0; 2,0	1,7; 1,8; 1,9	19,5
n_i	3	4	3	10
Σ	6,3	7,8	5,4	19,5
$(\Sigma X_i)I$	36,69	60,84	29,16	-
$(\Sigma X_i)I/n_i$	13,23	15,21	9,27	38,16
I	X_i 4,0; 4,41; 4,84	X_i 3,61; 3,61; 4,0; 4,0	X_i 2,89; 3,24; 3,61	38,21

Визначаємо суми квадратів відхилень:

За формулою (3) визначаємо S_y :

$$S_y = 38,21 - 19,5^2/10 = 38,21 - 38,024 = 0,19;$$

За формулою (4) визначаємо S_x :

$$S_x = 38,16 - 19,5^2/10 = 38,16 - 38,02 = 0,14;$$

За формулою (5) визначаємо S_z :

$$S_z = 38,21 - 38,16 = 0,05;$$

Визначаємо ступінь свободи дисперсій за формулою (6):

$$k_y = 10 - 1 = 9;$$

$$k_x = 3 - 1 = 2;$$

$$k_z = 10 - 3 = 7.$$

За формулою (7) знаходимо дисперсії:

$$\sigma^2 = \frac{0,14}{2} = 0,07 \quad ; \quad \sigma^2 = \frac{0,15}{7} = 0,007$$

Обчислюємо критерій Фішера: $F = 0,07 / 0,007 = 10$.

При надійності $P=0,95$ і ступенях свободи $k_x = 2$ і $k_z = 7$ за таблицею Фішера знаходимо значення критерію $F_{гр}(0,95; 2; 7) = 4,7$.

За формулою (9) визначаємо ступінь впливу факторної ознаки на результативну:

$$\eta = \frac{0,14}{0,19} \times 100\% = 73,6\%$$

Висновок: Оскільки $F > F_{гр}$ ($10 > 4,7$) розходження варіацій, яке обумовлене фактором і випадковими причинами, статистично вірогідно. Це свідчить про те, що у 10 спортсменів опорний показник бігу істотно залежить від спортивної спеціалізації легкоатлетів за дистанцією.

Причому, 73,6 % варіації визначається впливом фактору, який підлягає дослідженню.

✍ Практичне завдання:

Скласти умови завдання з власного наукового дослідження і вирішити його за допомогою однофакторного дисперсійного аналізу. Проаналізувати результати, зробити висновок.

? Питання для самоконтролю:

1. Що можна встановити за допомогою дисперсійного (факторного) аналізу?
2. Доведіть перевагу дисперсійного аналізу перед кореляційним.
3. Як називається ознака, що підлягає дослідженню?
4. Що називається дисперсійним комплексом?
5. Наведіть основне рівняння дисперсійного аналізу.
6. За допомогою якого критерію проводиться порівняльний аналіз суми квадратів відхилень?
7. Охарактеризуйте однофакторні та багатфакторні дисперсійні комплекси.
8. У чому полягає практична реалізація дисперсійного аналізу?
9. За якими формулами визначаємо ступені свободи дисперсій?
10. За допомогою якого показника можна оцінити ступінь впливу фактора, що підлягає дослідженню, на результативну ознаку?

ВИКОРИСТАНА ЛІТЕРАТУРА

1. Бондаренко І. Г. Спортивна метрологія : [методичні рекомендації] / І. Г. Бондаренко. Миколаїв : Вид-во ЧДУ імені Петра Могили, 2012. 104 с.
2. Гаркуша С.В. Г 20 Методи математичної статистики в педагогічних дослідженнях. Навчально-методичний посібник для аспірантів. Чернігів, 2019. 72 с.
3. Горкавий В.К., Ярова В.В. Математична статистика: навчальний посібник. Київ: ВД Професіонал, 2004. 384 с.
4. Костюкевич В.М. Спортивна метрологія: навчальний посібник для студентів факультетів фізичного виховання пед. університетів. Вінниця : ДОВ "Вінниця".ВДПУ, 2001. 183 с.
5. Сергієнко Л.П. Спортивна метрологія: теорія і практичні аспекти: підручник. Київ : КНТ, 2010. 776 с.
6. Соколова О.В., Омеляненко Г.А., Тищенко В.О. Методи математичної статистики у фізичному вихованні (з використанням електронних таблиць): навчально-методичний посібник для здобувачів ступеня вищої освіти бакалавра спеціальності «Фізична культура і спорт» освітньо-професійних програм «Фізичне виховання» і «Спорт». Запоріжжя: Запорізький національний університет, 2018. 90 с.
7. Спортивна метрологія: навчальний посібник для здобувачів ступеня вищої освіти бакалавра спеціальності «Фізична культура і спорт» освітньо-професійних програм «Фізичне виховання» і «Спорт» / уклад.: О.В. Соколова, Г.А. Омеляненко, В.О. Тищенко. Запоріжжя : Запорізький національний університет, 2018. 86 с.
8. Соколова О.В. Спортивна метрологія: методичні рекомендації до лабораторних занять для студентів освітньо-кваліфікаційного рівня "бакалавр" напрямів підготовки "Фізичне виховання", "Спорт", "Здоров'я людини". Запоріжжя: Запорізький національний університет, 2013. 57 с.
9. Статистика [Електронний ресурс] : навчальний посібник / О. В. Раєвнева, І. В. Аксьонова, О. І. Бровко ; за заг. ред. д-ра екон. наук, професора О. В. Раєвневої. Харків : ХНЕУ ім. С. Кузнеця, 2019. 389 с.
10. Осипов О.Ю. Теорія ймовірностей та математична статистика: навчальний посібник. Запоріжжя, ЗДУ, 2002. 190 с.
11. Руденко В. М. Математична статистика. Навч. посіб. К.: Центр учбової літератури, 2012. 304 с.
12. Тищенко В. О., Соколова О. Педагогічні дослідження у фізичному вихованні: навч. посібник. Запоріжжя: ЗНУ, 2019. 96 с.
13. Худолій О. М., Іващенко О. В Основи науково-дослідної роботи у фізичному вихованні і спорті : навч. посіб. Харків: ОВС, 2014. 320 с.
14. Цьось А. В., Балахнічова Г. В., Заремба Л. В. Сучасні технології викладання спортивних дисциплін: навч. посіб. Луцьк: ВНУ ім. Лесі Українки, 2010. 132 с.

ДОДАТОК А

Таблиця граничних значень критерію Стьюдента

Надійність $P=0.95$, K – число ступенів свободи

K	$t_{\alpha/2}$	K	$t_{\alpha/2}$
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	2,21	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 Б

Таблиця граничних значень критерію Фішера F

Надійність P=0.95, k – число ступенів свободи

k1	k2 – ступінь свободи для більшої дисперсії											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,1	19,0	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,6	9,3	9,1	9,0	8,9	8,9	8,8	8,8	8,8	8,8	8,7
4	7,7	6,9	6,6	6,4	6,3	6,2	6,1	6,0	6,0	6,0	5,9	5,9
5	6,6	5,8	5,4	5,2	5,1	5,0	4,9	4,8	4,8	4,7	4,7	4,7
6	6,0	5,1	4,8	4,5	4,4	4,3	4,2	4,2	4,1	4,1	4,0	4,0
7	5,6	4,7	4,4	4,1	4,0	3,9	3,8	3,7	3,7	3,6	3,6	3,6
8	5,3	4,5	4,1	3,8	3,7	3,6	3,5	3,4	3,4	3,3	3,3	3,3
9	5,1	4,3	3,9	3,6	3,5	3,4	3,3	3,2	3,2	3,1	3,1	3,1
10	5,0	4,1	3,7	3,5	3,3	3,2	3,1	3,1	3,0	3,0	2,9	2,9
11	4,8	4,0	3,6	3,4	3,2	3,1	3,0	3,0	2,9	2,9	2,8	2,8
12	4,8	3,9	3,5	3,3	3,1	3,0	2,9	2,9	2,8	2,8	2,7	2,7
13	4,7	3,8	3,4	3,2	3,0	2,9	2,8	2,8	2,7	2,7	2,6	2,6
14	4,6	3,7	3,3	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,6	2,5
15	4,5	3,7	3,3	3,1	2,9	2,8	2,7	2,6	2,6	2,6	2,5	2,5
16	4,5	3,6	3,2	3,0	2,9	2,7	2,7	2,6	2,5	2,5	2,5	2,4
17	4,5	3,6	3,2	3,0	2,8	2,7	2,6	2,6	2,5	2,5	2,4	2,4
18	4,4	3,6	3,2	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,4	2,3
19	4,4	3,5	3,1	2,9	2,7	2,6	2,6	2,5	2,4	2,4	2,3	2,3
20	4,4	3,5	3,1	2,9	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,3
21	4,3	3,5	3,1	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,3
22	4,3	3,4	3,1	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,3	2,3
23	4,3	3,4	3,0	2,8	2,6	2,5	2,5	2,4	2,3	2,3	2,2	2,2
24	4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2
25	4,2	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,2
26	4,2	3,4	3,0	2,7	2,6	2,6	2,5	2,4	2,3	2,2	2,2	2,2
27	4,2	3,4	3,0	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,2	2,1
28	4,2	3,3	3,0	2,7	2,6	2,4	2,4	2,3	2,3	2,2	2,2	2,1
29	4,2	3,3	2,9	2,7	2,5	2,4	2,4	2,3	2,2	2,2	2,1	2,1
30	4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,1	2,1
40	4,1	3,2	2,8	2,6	2,5	2,3	2,3	2,2	2,1	2,1	2,0	2,0
50	4,0	3,2	2,8	2,6	2,4	2,3	2,3	2,1	2,1	2,0	2,0	2,0
100	3,9	3,1	2,7	2,5	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,9
150	3,9	3,1	2,7	2,4	2,3	2,2	2,1	2,0	1,9	1,9	1,9	1,8
200	4,0	3,0	2,7	2,4	2,3	2,1	2,1	2,0	1,9	1,9	1,8	1,8
400	3,9	3,0	2,6	2,4	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,8
1000	3,9	3,0	2,6	2,4	2,2	2,1	2,0	2,0	1,9	1,8	1,8	1,8
∞	3,8	3,0	2,6	2,4	2,2	2,1	2,0	1,9	1,9	1,8	1,8	1,8

k1	k2 – ступінь свободи для більшої дисперсії											
	14	16	20	24	30	40	50	75	100	200	500	∞
1	245	246	248	249	250	251	252	253	253	254	254	254
2	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5
3	8,7	5,8	5,8	5,8	5,8	5,7	5,7	5,7	5,7	5,7	5,6	5,6
4	5,9	5,8	5,8	5,8	5,8	5,7	5,7	5,7	5,7	5,7	5,6	5,6
5	4,6	4,6	4,6	4,5	4,5	4,5	4,4	4,4	4,4	4,4	4,4	4,4
6	4,0	3,9	3,9	3,8	3,8	3,8	3,8	3,7	3,7	3,7	3,7	3,7
7	3,5	3,5	3,4	3,4	3,4	3,3	3,3	3,3	3,3	3,3	3,2	3,2
8	3,2	3,2	3,2	3,1	3,1	3,1	3,0	3,0	3,0	3,0	2,9	2,9
9	3,0	3,0	2,9	2,9	2,9	2,8	2,8	2,8	2,8	2,7	2,7	2,7
10	2,9	2,8	2,8	2,7	2,7	2,7	2,6	2,6	2,6	2,6	2,6	2,5
11	2,7	2,7	2,7	2,6	2,6	2,5	2,5	2,5	2,5	2,4	2,4	2,4
12	2,6	2,6	2,5	2,5	2,5	2,4	2,4	2,4	2,4	2,3	2,3	2,3
13	2,6	2,5	2,5	2,4	2,4	2,3	2,3	2,3	2,3	2,2	2,2	2,2
14	2,5	2,4	2,4	2,4	2,3	2,3	2,2	2,2	2,2	2,2	2,1	2,1
15	2,4	2,4	2,3	2,3	2,3	2,2	2,2	2,2	2,1	2,1	2,1	2,1
16	2,4	2,3	2,3	2,2	2,2	2,2	2,1	2,1	2,1	2,0	2,0	2,0
17	2,3	2,3	2,2	2,2	2,2	2,1	2,1	2,0	2,0	2,0	2,0	2,0
18	2,3	2,3	2,2	2,2	2,1	2,1	2,0	2,0	2,0	2,0	1,9	1,9
19	2,3	2,2	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,9	1,9
20	2,2	2,2	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,9	1,9	1,8
21	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,9	1,8	1,8	1,8
22	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,9	1,8	1,8	1,8	1,8
23	2,1	2,1	2,0	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,8	1,8
24	2,1	2,1	2,0	2,0	1,9	1,9	1,9	1,8	1,8	1,8	1,7	1,7
25	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7
26	2,1	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7
27	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7	1,7
28	2,1	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7	1,7
29	2,1	2,0	1,9	1,9	1,9	1,8	1,8	1,7	1,7	1,7	1,7	1,6
30	2,0	2,0	1,9	1,9	1,8	1,8	1,8	1,7	1,7	1,7	1,6	1,6
40	2,0	1,9	1,8	1,7	1,7	1,7	1,6	1,6	1,6	1,5	1,5	1,5
50	1,9	1,9	1,8	1,7	1,7	1,6	1,6	1,6	1,5	1,5	1,5	1,4
100	1,8	1,8	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,3
150	1,8	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,3	1,2
200	1,7	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,2	1,2
400	1,7	1,7	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,2	1,2	1,1
1000	1,7	1,7	1,6	1,5	1,5	1,4	1,4	1,3	1,3	1,2	1,1	1,1
∞	1,7	1,6	1,6	1,5	1,5	1,4	1,4	1,3	1,2	1,2	1,1	1,0

Навчально-методичне видання
(українською мовою)

Соколова Ольга Валентинівна

Тищенко Валерія Олексіївна

Караулова Світлана Іванівна

СУЧАСНІ МЕТОДИ СТАТИСТИЧНОЇ ОБРОБКИ ТА
АНАЛІЗУ НАУКОВИХ ДАНИХ У СФЕРІ ФІЗИЧНОЇ
КУЛЬТУРИ ТА СПОРТУ

Навчально-методичний посібник
для здобувачів третього (освітньо-наукового) рівня
вищої освіти
(ступеня доктора філософії)
освітньо-наукової програми «Фізична культура і спорт»

Рецензент *М.В. Маліков*
Відповідальний за випуск *А.П. Конох*
Коректор *О.В. Соколова*