Письмово виконати індивідуальне науково-дослідницьке завдання за темою:

Тема: Кореляційний аналіз. Оцінка залежності між двома величинами, що вимірюються.

Для виконання завдання необхідно:

1. Провести дослідження зі свого виду спорту та отримати дві пов’язані вибірки числових результатів (не менш 15 результатів у кожній вибірці).

2. Побудувати кореляційне поле і розрахувати коефіцієнт кореляції.

3. За отриманими даними зробити висновок про наявність, напрямок і щільність взаємозв’язку між показниками згідно умов задачі (проведеного дослідження).

Індивідуальне науково-дослідне завдання виконується до екзамену завчасно та оцінюється, виходячи з максимальних 20 балів:

20 балів – завдання виконано бездоганно за змістом, формою та обсягом; 19-15 балів – завдання виконано повно, але містить неточності; 14-10 балів – завдання виконано неповно і містить неточності; 9-5 балів – завдання виконано неповно, містить неточності, порушується послідовність; 4-1 бал – завдання виконано частково; 0 балів – завдання не виконано.

Тема: Кореляційний аналіз. Оцінка залежності між двома величинами, що вимірюються.

Мета: Навчитися розраховувати коефіцієнт кореляції і досліджувати взаємозв'язок між ознаками.

Теоретичні відомості

У спортивних дослідженнях між показниками, що підлягають дослідженню, часто спостерігається взаємозв'язок. Його вид може бути різним. Розрізняють два види взаємозв'язку: функціональний і статистичний.

Функціональний взаємозв'язок - це залежність, при якій кожному значенню одного показника відповідає строго 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 **статистичним**.

Вивченню статистичного взаємозв'язку між різними показниками у спортивних дослідженнях приділяють велику увагу, тому що це дозволяє

розкрити деякі закономірності і 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 в залежності 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 їхньої надійності й інформативності.

За допомогою кореляційного аналізу можна оцінити:

1. Взаємозв'язок між спортивними результатами і функціональними показниками (результат у бігу на 100 м і кількість лактату в крові);
2. Вплив спортивної діяльності на результат (результат човникового бігу 3 × 100 м і результат бігу на 100 м);
3. Взаємний вплив показників тренованості (результат проходження 10 × 200 м і результат проходження середньої дистанції, у плаванні);
4. Надійність тесту (погодженість, стабільність);

4.1 Погодженість - оцінка виступу гімнасток двома суддями;

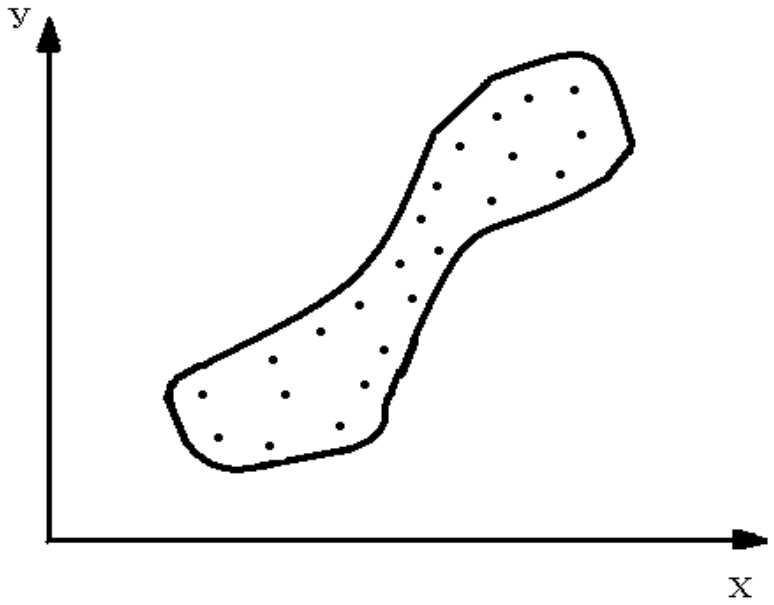
4.2 Стабільність - порівняння результатів тесту (кількість точних кидків із десятьох запропонованих у баскетболі) і ретесту - той же тест, але через рік;

5. Інформативність тесту (залежність між кількістю гребків за 30 секунд і часом проходження дистанції 100 м у плаванні).

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**:

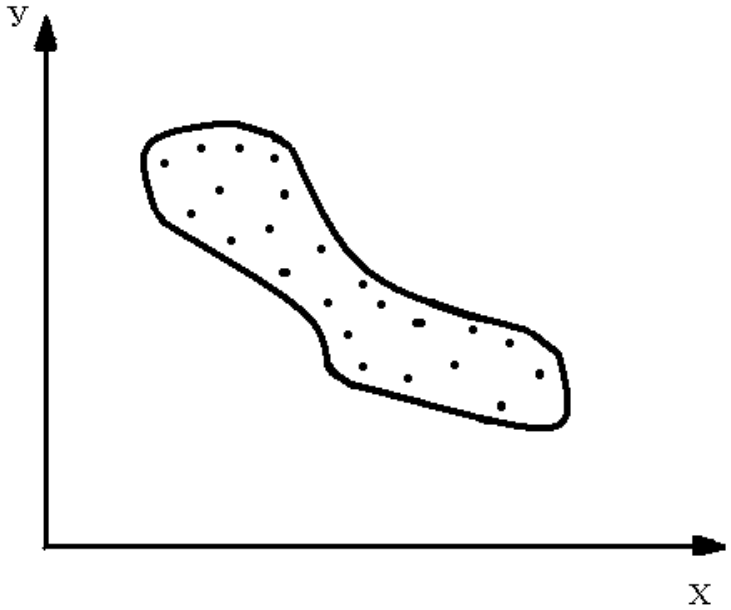
а) пряма позитивна кореляційна статистична залежність (нахил кореляційного поля вправо)



зі зростанням (зменшенням) першої ознаки (Хі) інший (Уі) також зростає (зменшується)

Рисунок 7.1 - Пряма позитивна кореляційна статистична залежність

б) обернено-негативна кореляційна статистична залежність (нахил кореляційного поля вліво)



зі зростанням першої ознаки (Хі) інший (Уі) зменшується і навпаки, зі зменшенням першої ознаки (Хі) інший (Уі) зростає

Рисунок 7.2 - Обернено-негативна кореляційна статистична залежність

Якщо кореляційне поле подане колом, то залежність відсутня

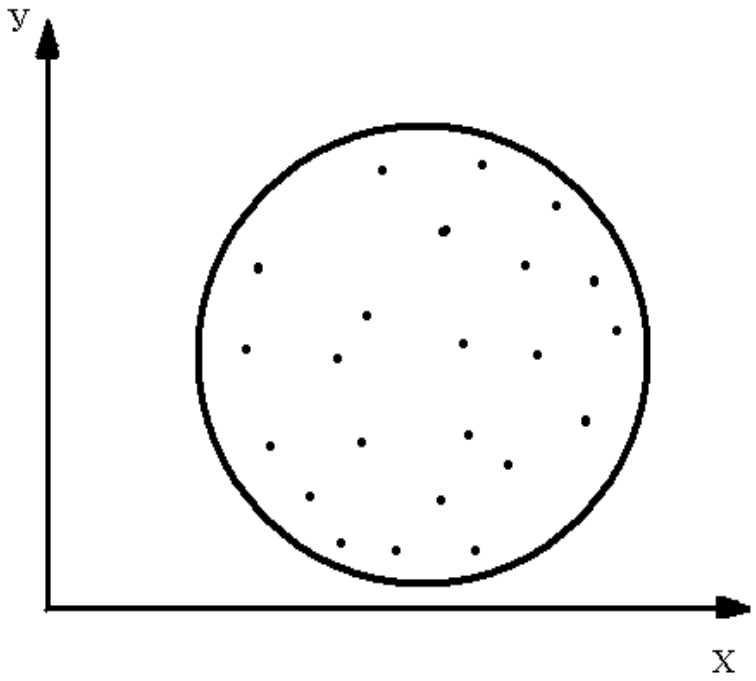


Рисунок 7.3 – Коло - залежність відсутня

Якщо кореляційне поле подане прямою лінією, то має місце функціональна залежність

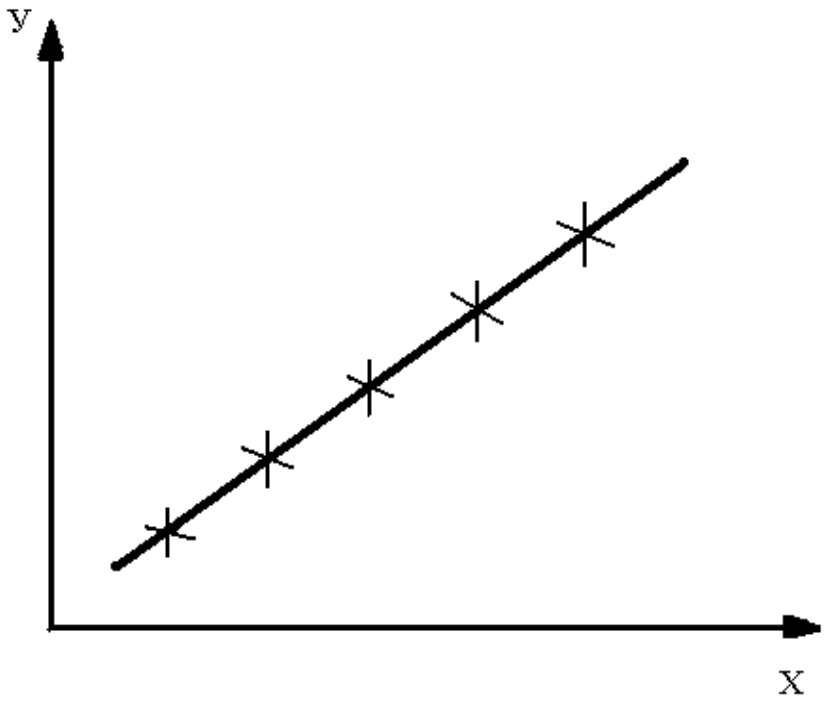


Рисунок 7.4 - Пряма лінія - функціональна залежність

Форми статистичної залежності:

1. Лінійна форма залежності - форма є близькою до звичайної геометричної фігури - еліпсу

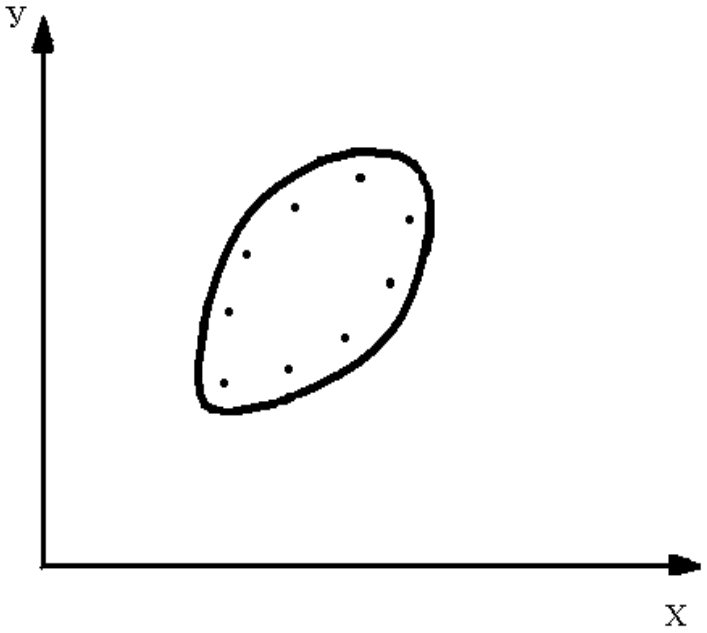


Рисунок 7.5 – Еліпс

2. Нелінійна форма залежності - будь-яка інша форма, крім еліпса.

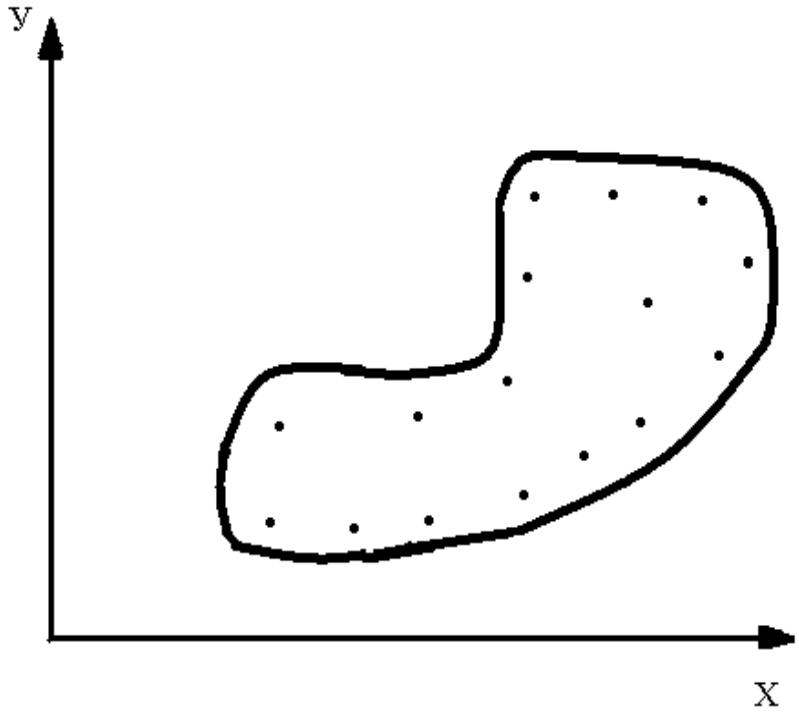


Рисунок 7.6 - Будь-яка інша форма

Отже, візуальний аналіз кореляційного поля дозволяє виявити форму статистичної залежності - лінійну або нелінійну. Це має істотне значення для наступного кроку в аналізі - вибору й обчисленні 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 Браве-Пірсона (r_{xy}) :

$$r^{xy} = \frac{\sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})}{N \times \sigma x \times \sigma y}$$

де

X_i й Y_i - варіанти 2-х вибірок;

$\overline{X}$ і $\overline{Y}$ - середнє арифметичне значення показників X_i й Y_i;

σ_x, σ_y- середнє квадратичне відхилення;

N - число вимірів (випробуваних).

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

$$r_{xy} = \frac{\sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(Xi - \overline{X})^2 \sum_{i=1}^{n}(Yi - \overline{Y})^2}}$$

Для оцінки щільності взаємозв'язку у кореляційному аналізі застосовується значення спеціального показника - коефіцієнта кореляції (r_{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:

$$-1 \le r_{xy} \le 1$$

Пояснюють значення цього коефіцієнта в такий спосіб:

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

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

в) чим ближче значення r_{xy} до нуля, тим зв'язок слабкіше, чим ближче значення r_{xy} до одиниці - тим щільніше.

Прийнято вважати, що

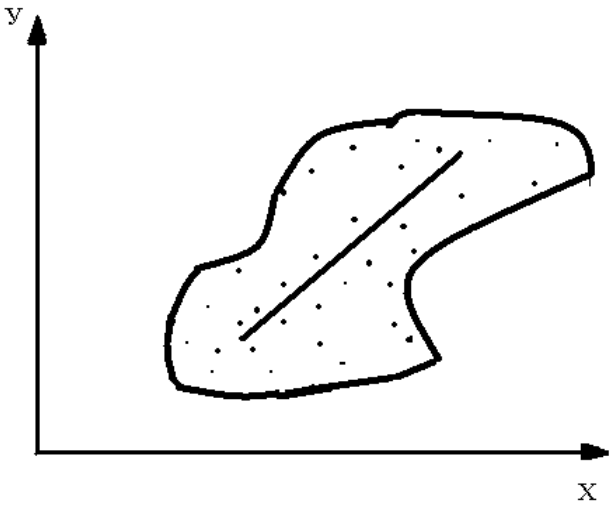
- $r_{xy} = 0,2....0,49$ - слабкий зв'язок;

- $r_{xy} = 0,5....0,69$ - середній зв'язок;

- $r_{xy} = 0,7....0,99$ - тісний (сильний) зв'язок.

На кореляційному полі щільність може виглядати так:

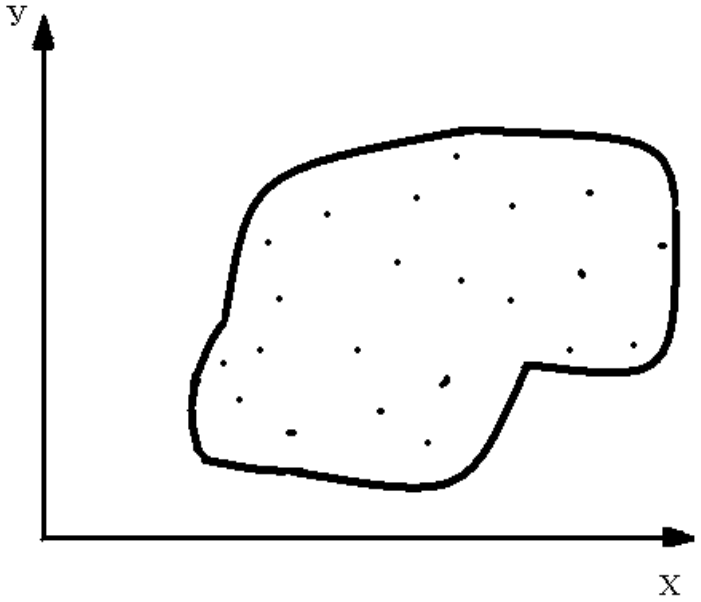
а) якщо точки групуються вздовж якої-небудь лінії, то зв'язок є, і він тим щільніше, чим ближче вони групуються.



r_{xy} = 0,89 - сильна кореляційна статистична залежність

Рисунок 7.7 - Тісний (сильний) зв'язок

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



r_{xy} = 0,09 - дуже слабка кореляційна статистична залежність;

Рисунок 7.8 - Слабкий зв'язок

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 розраховують 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.

Порядок виконання

Розглянемо дану тему на прикладі:

Приклад: Чи існує залежність, і яка між результатами бігу на 60 м (Xi), сек і швидкістю їзди на велосипеді (Уі), км/год. у 7 велосипедистів?

Чи може біг на 60 м служити тестом для цих велосипедистів на виявлення 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
$У_i$	18,4	18,0	17,8	17,5	17,1	16,9	16,2

1. Обраховування середнього арифметичного значення 2-х вибірок, X_i і 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:

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

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

$\overline{X}$ = 9,3 (сек) - середній час бігу на дистанції 60 м для 7 велосипедистів.

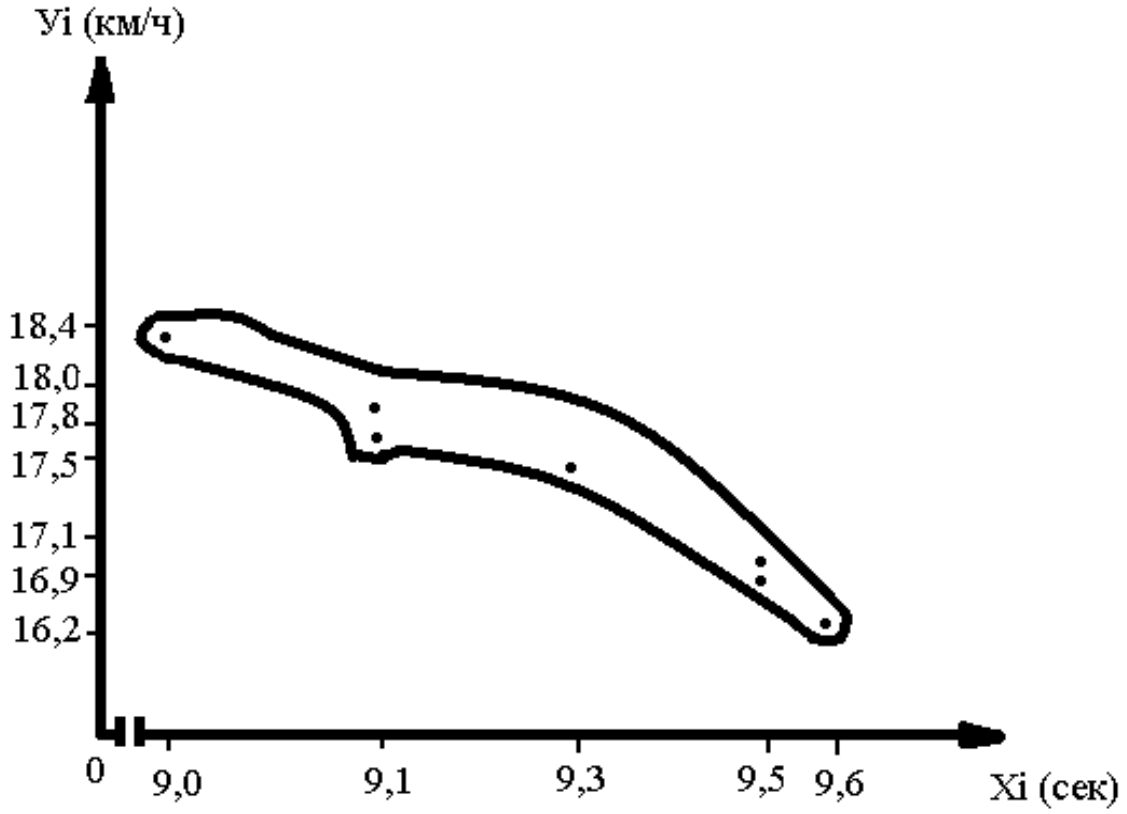
1.2 Оскільки всі варіанти другої вибірки (Уі) не повторюються, то середнє арифметичне значення розраховуємо за формулою незваженого середнього арифметичного значення:

$$\overline{Y} = \frac{Y_1 + Y_2 + Y_3 + ... + Y_n}{N} \qquad \overline{Y} = \frac{Y_1 + Y_2 + Y_3 + ... + Y_7}{7}$$

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

$\overline{Y}$ =17,4 (км/ч) - середня швидкість їзди на велосипеді для 7 велосипедистів.

2. Побудова кореляційного поля, за яким визначаємо форму і спрямованість взаємозв'язку.



Побудова кореляційного поля відповідає нелінійній формі залежності і зворотному кореляційному статистичному взаємозв'язку.

Нелінійна форма залежності - кореляційне поле подане не у формі еліпса. Обернений кореляційний статистичний взаємозв'язок - нахил кореляційного поля вліво.

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

$$r_{xy} = \frac{\sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(Xi - \overline{X})^2 \sum_{i=1}^{n}(Yi - \overline{Y})^2}}$$

3.1 Для подальшої роботи будуємо таблицю

Xi	Yi	Xi - $\overline{X}$	Yi - $\overline{Y}$	$(X_i - \overline{X})(Y_i - \overline{Y})$	(Xi - $\overline{X}$)²	(Yi - $\overline{Y}$)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
∑ 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. Розрахунок коефіцієнта 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 говорити про те, що 92 % взаємозв'язку результату бігу на 60 м і швидкістю їзди на велосипеді пояснюються їх взаємовпливом. Інша частина (100 % - 92 % = 8 %) варіації пояснюється впливом інших неврахованих чинників.

Висновок: Розрахований коефіцієнт кореляції ($r_{xy} = -0,96$) зазначає 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 буде зростати. 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 збільшення швидкості їзди на велосипеді.