к290

МІНІСТЕРСТВО У СПРАВАХ МОЛОДІ І СПОРТУ

ЛЬВІВСЬКИЙ ДЕРЖАВНИЙ ІНСТИТУТ ФІЗИЧНОЇ КУЛЬТУРИ

КАФЕДРА БІОМЕХАНІКИ ТА СПОРТИВНОЇ МЕТРОЛОГІЇ

СТАТИСТИЧНІ МЕТОДИ ОБРОБКИ РЕЗУЛЬТАТІВ СПОРТИВНИХ ВИМІРІВ

**Методичні вказівки
з виконання контрольної роботи
для студентів факультету заочного навчання**

Львів - 1995

Рекомендовано до друку
кафедрою біомеханіки та
спортивної метрології.
Протокол N 2 від 25.09.1995р.

Уклали: Русіло Петро Олександрович
Заневський Ігор Пилипович

Методичні вказівки містять короткі теоретичні положення, порядок виконання контрольної роботи та зразок її оформлення, а також контрольні запитання і список рекомендованої літератури.

Комп'ютерна верстка: І.Пригода

ВИХІДНІ ДАНІ - результати спортивних вимірів групи спортсменів з двох видів спорту (параметри **X** і **Y**).

МЕТА - набути практичних навиків в проведенні статистичного аналізу спортивних вимірів.

ЗАВДАННЯ 1. Статистична обробка ряду результатів спортивних вимірів.

ТЕОРЕТИЧНІ ВІДОМОСТІ

Предметом математичної статистики є аналіз результатів масових, повторюваних вимірів. Масові виміри однорідних об'єктів, які володіють якісною спільністю, мають певні закономірності. Математична статистика створює методи виявлення цих закономірностей. Виділяють три основні етапи статистичних досліджень.

1. **Статистичне спостереження** - це планомірний, науково обгрунтований збір даних, які характеризують об'єкт, що вивчається. Воно повинно відповідати наступним вимогам:

- об'єкти спостереження повинні бути однаковими (однорідними) з точки зору їх властивостей (кваліфікація, спеціалізація, вік, стаж заняття та ін.);
- число об'єктів спостереження повинно бути достатнім, щоб можна було виявити закономірності і узагальнити їх властивості.

2. **Статистичне зведення і групування** - це важлива підготовча частина до статистичного аналізу даних, яка передбачає:

- систематизацію (групування) даних;
- оформлення певних статистичних таблиць.

3. **Аналіз статистичного матеріалу**, який є завершальним етапом статистичного підходу. Проводять його з використанням відповідних математико-статистичних методів.

В процесі спостереження чи вимірювання якого-небудь показника отримують ряд чисел (вимірів). Ряд результатів вимірів, репрезентованих випадковими числами, називається вибірковою сукупністю або, просто, вибіркою. **Вибірка** - це сукупність однорідних величин, яка доступна для вимірів. Сукупність всіх значень, які можна було б отримати для вибірки, що вивчають, називається генеральною (основною) сукупністю. **Генеральна сукупність** - це безмежна кількість однорідних вимірів.

Однією із основних характеристик вибірки є її **обсяг** (n), який визначається числом об'єктів спостереження. По вибірковим (експериментальним) даним оцінюють параметри, котрі дозволяють описати всю генеральну сукупність, визначають інтервал, в якому з заданим рівнем достовірності (вірогідності) знаходиться істинне значення оцінюваного параметра, а потім перевіряються ті чи інші твердження і робляться висновки про властивості всієї генеральної сукупності.

Вибіркова сукупність має бути **репрезентативною**, тобто обсяг її повинен бути достатньо великим, наприклад, тридцять вимірів або більше.

За результатами вимірів вибірки оцінюють емпіричні середні величини, характеристики розсіювання та однорідність вибірки, порівнюють розподіл результатів з нормальним.

Групування - це процес систематизації первинних даних з метою добування інформації, котра в них міститься.

Ранжуванням називають розміщення результатів вимірів в порядку зростання або спадання.

Вибірку великого обсягу розбивають на інтервали, кожен з яких містить певний діапазон значень ознаки, що вивчається. Наближено число інтервалів k можна знайти за формулою Стерджеса:

$$k = 1 + 3.32 \cdot \lg n ,$$

де n - обсяг вибірки,

lg - символ десяткового логарифма.

Ширину кожного із інтервалів h (**крок інтервалу**) визначають за наступною формулою:

$$h = \frac{X_{max} - X_{min}}{k} ,$$

де X_{max} і X_{min} - максимальне і мінімальне значення результатів вимірів.

Нижня границя групування першого інтервалу $(X_{н.1})$ вибирається рівною X_{min}. Нижня границя наступних інтервалів є одночасно верхньою границею попередніх інтервалів $(X_{н.j} = X_{в.j-1})$. Верхня границя останнього інтервалу дорівнює X_{max}. Значення нижніх і верхніх границь обчислюються за наступними формулами:

$$X_{н.j} = X_{min} + (j - 1) \cdot h ;$$

$$X_{в.j} = X_{min} + j \cdot h ,$$

де j = 1,2,...k - номер інтервалу.

Середина кожного інтервалу $X_{c.j} = X_{min} + \frac{(2 \cdot j - 1)}{2} \cdot h .$

Величина, котра визначається числом результатів вимірів вибірки, які знаходяться в кожному інтервалі групування, називається **частотою інтервалів**. Частота інтервалу позначається символом f_j. Сума частот всіх інтервалів завжди дорівнює обсягу вибірки.

Накопичена частота інтервалу - це число, отримане послідовним додаванням частот в напрямку від першого інтервалу до останнього. Накопичена частота (Σf_j) отримується простим послідовним додаванням частот попередніх інтервалів. Накопичена частота в останньому інтервалі дорівнює обсягу вибірки.

Частістю (відносною частотою) називається відношення частоти до обсягу вибірки: $F_j = f_j / n$. Сума частостей всіх інтервалів завжди дорівнює одиниці.

Накопиченою частістю називається відношення накопиченої частоти до обсягу вибірки: $\Sigma F_j = \Sigma f_j / n$. Накопичена частість в останньому інтервалі дорівнює одиниці.

Таблиця, заповнена відповідними значеннями, котрі були розглянуті вище, являє собою **інтервальний варіаційний ряд** вимірів (табл.1). Варіаційним рядом вимірів називається подвійний числовий ряд, який показує, яким чином чисельні значення ознаки, що вивчається, зв'язані з їх повторністю у вибірці. Варіаційні ряди дають наочне уявлення про характерні особливості змінності ознаки.

Таблиця 1

Таблиця інтервального варіаційного ряду.

N інтервалу j	Границі інтервалів: нижня $X_{н.j}$	Границі інтервалів: верхня $X_{в.j}$	Середина інтервалів $X_{c.j}$	Частота f_j	Накопичена частота Σf_j	Частість F_j	Накопичена частість ΣF_j
1.	X_{min}	$X_{min}+h$	$X_{min}+h/2$	f_1	f_1	f_1/n	F_1
2.	$X_{min}+h$	$X_{min}+2h$	$X_{min}+3h/2$	f_2	f_1+f_2	f_2/n	F_1+F_2
3.	$X_{min}+2h$	$X_{min}+3h$	$X_{min}+5h/2$	f_3	$f_1+f_2+f_3$	f_3/n	$F_1+F_2+F_3$
...	...	...	...	...	...	...	...
k	$X_{min}+(k-1)h$	$X_{min}+kh = X_{max}$	$X_{min}+(2k-1)h/2$	f_k	$f_1+f_2+\ldots+f_{k-1}$	f_k/n	$F_1+F_2+\ldots+F_{k-1}$

Для збільшення наочності та спрощення аналізу варіаційних рядів використовують їх графічне зображення у вигляді гістограми, полігона частот і кумуляти (див.рис.1).

Гістограма - це стовбчата діаграма, яка використовується для графічного зображення розподілу змінних ознак і складається з суміжних прямокутників. Основа кожного прямокутника дорівнює ширині інтервалу групування h, а висота - пропорційна частотам інтервалу.

Полігон частот (полігон розподілу) утворюється ламаною лінією, котра з'єднує точки, що відповідають на осі обсцис - середнім значенням інтервалів групування, а на осі ординат - частоті цих інтервалів.

Кумулята (полігон накопичених частот або частостей) отримується при з'єднанні відрізками прямих точок на координатній площині, абсциса яких відповідає верхнім границям інтервалів групування, а ордината - накопиченим частотам або накопиченим частостям, вираженим у відсотках.

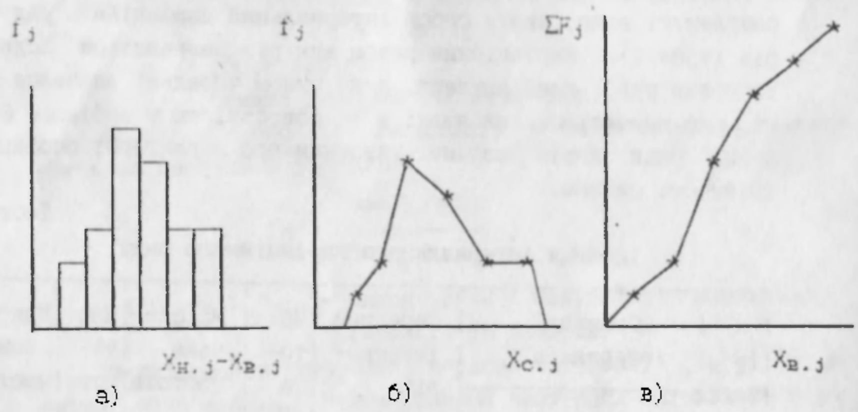


Рис.1. Графічне зображення варіаційних рядів:
а - гістограма; б - полігон; в - кумулята.

Результатом статистичного аналізу є статистичні характеристики і числові показники вибірки.

Варіаційні ряди і графіки емпіричних розподілів не дають повної характеристики вибірки. Числові показники дають кількісне уявлення про емпіричні дані і дозволяють порівняти їх між собою.

Числові показники вибірки поділяються на дві групи:

- характеристики центральної тенденції (положення);
- характеристики розсіяння (варіації).

Центральну тенденцію вибірки описують найбільш імовірні результати вимірів: середнє арифметичне, медіана і мода.

Середнє арифметичне - це одна із основних характеристик вибірки, що визначається діленням суми всіх результатів вимірів на обсяг вибірки:

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n},$$

де $i = 1, 2, \ldots n$.

Медіаною називається значення ознаки X, коли одна половина значень експериментальних даних менша її, а друга половина - більша.

Якщо даних небагато (обсяг вибірки малий), то медіана знаходиться по рангу медіани:

$$R_{Me} = \frac{n + 1}{2}.$$

При непарному числі членів вибірки ранг медіани вказує на номер результату виміру в ранжованому ряді, котрий і буде медіаною. Якщо вибірка має парне число членів, то медіана буде дорівнювати середньому арифметичному двох членів ранжованого ряду з рангами $R_{Me} = n/2$ і $R_{Me} = n/2 + 1$.

Інтервал групування, в якому накопичена частота вперше буде більше половини обсягу вибірки, називається **медіанним**.

Для згрупованих даних медіана визначається за наступною формулою:

$$Me = X_{Me_н} + h \cdot \frac{0.5 \cdot n - f_{Me-1}}{f_{Me}},$$

де $X_{Me_н}$ - значення нижньої границі медіанного інтервалу;

h - ширина (крок) інтервалів групування;

f_{Me} - частота медіанного інтервалу;;

f_{Me-1} - накопичена частота інтервалу, який знаходиться перед медіанним;

n - обсяг вибірки.

Мода являє собою значення ознаки, яка зустрічається у вибірці найбільш часто, тобто значення виміру з найбільшою імовірністю. Інтервал групування варіаційного ряду з найбільшою частотою називається **модальним**. Чисельне значення моди розраховується за наступною формулою:

$$Mo = X_{Mo_н} + h \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} ,$$

де $X_{Mo_н}$ - значення нижньої границі модального інтервалу,

f_{Mo} - частота модального інтервалу,

f_{Mo-1} і f_{Mo+1} - частоти інтервалів, які знаходяться перед і після модального.

Поруч з показниками положення (центральної тенденції) вираховують і характеристики варіації (розсіяння) вибірки.

Розмах вибірки (варіації) вираховується як різниця між максимальним і мінімальним значеннями вибірки, тобто різниця між крайніми значеннями ознаки ранжованого ряду:

$$R = X_{max} - X_{min} .$$

Інформативність розмаху вибірки невелика, він враховує тільки крайні відхилення всіх результатів.

Дисперсія - це середнє значення квадратів відхилень кожного елементу ряду вимірів від середнього арифметичного.

Дисперсію, яку обчислюють за вибірковими даними, називають вибірковою дисперсією:

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

Дисперсію можна розрахувати за формулою, в яку не входить середнє арифметичне. Вона зручніша для складання програми обчислення дисперсії за допомогою ЕОМ:

$$D = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} X_i \right)^2$$

Вибіркова дисперсія характеризує систематичні похибки і її називають зміщеною оцінкою генеральної дисперсії. Незміщена оцінка є точковою оцінкою генеральної дисперсії, яку обчислюють за наступною формулою:

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

Для великих обсягів вибірки різниця між зміщеною і незміщеною оцінкою генеральної дисперсії - невелика, в майбутньому будемо користуватися одним позначенням D.

Зміщену оцінку генеральної дисперсії використовують для характеристики вибірки, а незміщену оцінку - для визначення параметрів генеральної сукупності (інтервалу довіри середнього арифметичного, закону нормального розподілу).

Середнє квадратичне відхилення (стандартне відхилення) - це арифметичний квадратний корінь з дисперсії:

$$\sigma = \sqrt{D}\ .$$

Розмірність стандартного відхилення співпадає з одиницями вимірювання змінної ознаки і характеризує ступінь відхилення результатів від середнього арифметичного значення в абсолютних одиницях.

Стандартне відхилення є абсолютною мірою розсіяння, яку записують в наступному вигляді:

$$\bar{X} \pm \sigma.$$

Наприклад: $\bar{X}$ = 743.43 см, а σ = 42.33 см.

Запис оцінки точності $\bar{X}$ буде мати наступний вигляд:

(743 ± 42) см.

Для порівняння двох і більше сукупностей, які мають різні одиниці вимірів, використовують коефіцієнт варіації.

Коефіцієнт варіації знаходиться як відношення середнього квадратичного відхилення до середнього арифметичного і виражається у відсотках:

$$V = (\sigma / \bar{X}) \cdot 100\%.$$

Коефіцієнт варіації є відносною мірою розсіяння ознаки і показує варіативність вибірки:

до 10% - варіація мала, вибірка однорідна;

від 10% до 20% - варіація середня, вибірка середньої однорідності;

більше 20% - варіація велика, вибірка неоднорідна.

На основі відомих величин вибіркової характеристики можна визначити інтервал, в якому з тою чи іншою вірогідністю визначається параметр генеральної сукупності. Інтервал, в якому з заданою вірогідністю знаходиться оцінюваний генеральний параметр, називається **інтервалом довіри**. Для визначення вірогідності (q) використовується рівень значущості (α):

$$q = (1 - \alpha)\cdot 100\%.$$

В практиці ФВіС проведення суцільних обстежень генеральної сукупності (наприклад, параметрів фізичного стану студентів всієї країни) практично здійснити неможливо. Тому проводять розрахунковим шляхом інтервальну оцінку генеральних характеристик (середнього арифметичного, дисперсії і т.д.) на основі даних вибіркової сукупності.

Оцінка границь інтервалу довіри генерального середнього здійснюється з певною вірогідністю $q = (1 - \alpha)\cdot 100\%$ з використанням t-критерію Стьюдента:

$$t(\alpha,\nu) = \frac{|\bar{X} - \mu|}{S_{\bar{x}}},$$

де μ - генеральне середнє;
$t(\alpha,\nu)$ - t-критерій Стьюдента;
$S_{\bar{x}}$ - стандартна похибка середнього арифметичного.

Величина критерію відповідає t-розподілу Стьюдента, який знаходиться за таблицею 2 з врахуванням числа ступенів свободи ($\nu = n - 1$) і рівня значущості (α).

Стандартна форма запису генерального середнього має наступний вигляд:

$$\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon,$$

де $\varepsilon = S_{\bar{x}}\cdot t(\alpha,\nu)$ - амплітуда інтервалу.

Такий запис дозволяє з вірогідністю $q = (1 - \alpha)\cdot 100\%$ стверджувати, що генеральне середнє μ знаходиться в інтервалі

$[\bar{X} - \varepsilon; \ \bar{X} + \varepsilon]$.

Вибір статистичних характеристик визначається двома основними факторами: шкалою вимірів, якою користується дослідник, і законом розподілу результатів вимірів.

Якщо виміряти зріст (довжину тіла) достатньо великої кількості людей і результати вимірів подати в певній системі координат (рис.2), то отримана крива буде характеризувати розподіл деякої випадкової величини (в даному випадку довжини тіла людені в нашій вибірці. Розподіл, який описується кривою такого типу, називається **нормальним** або **гауссовим** на честь німецького математика Карла Гаусса.

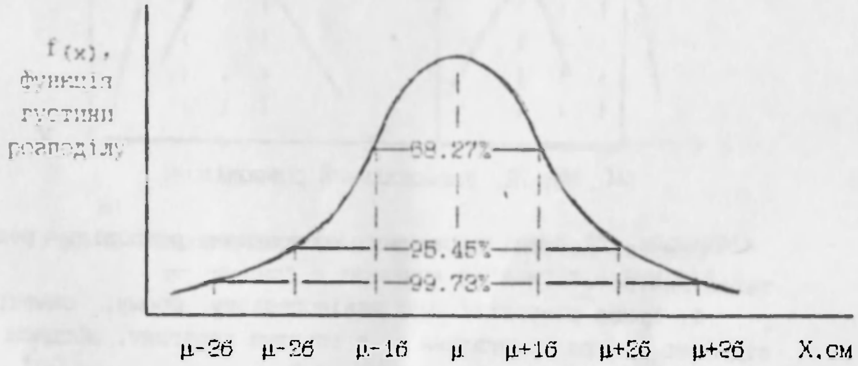


Рис.2. Крива нормального розподілу.

Крива нормального розподілу визначає густину розподілу випадкової величини і описується формулою, яка запропонована англійським математиком Муавром в 1733 р.:

$$f(X_i) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(X_i - \mu)^2}{2 \cdot \sigma^2}}$$

де π і e - математичні константи ($\pi \approx 3.1416$; $e \approx 2.7182$);

μ і σ - генеральне середнє і середнє квадратичне відхилення генеральної сукупності, відповідно;

X_i - результат вимірів.

Розподіл дуже великої вибірки ($N \to \infty$) називають розподілом генеральної сукупності або **теоретичним**, а розподіл експериментального ряду вимірів - **емпіричним**. Криву розподілу можна отримати з графіка полігона розподілу при нескінченно великому числі вимірів і інтервалів. Деколи при емпіричному розподілі отримують криву, яка відрізняється від нормального закону розподілу присутністю двох або більше вершин. Це так звані многомодальні розподіли (рис.3).

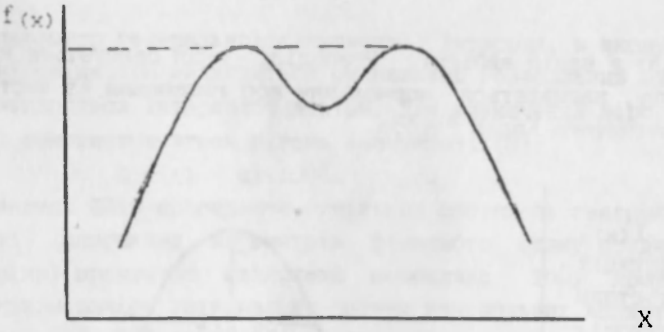


Рис.3. Двомодальний розподіл.

Основні властивості нормального закону розподілу результатів вимірів.

1. Крива розподілу має дзвіноподібну форму, симетричну відносно центра групування μ, з точками перегину, абсциса котрих віддалена від μ на величину $\pm\sigma$.

2. Значення медіани і моди нормального розподілу співпадають і дорівнюють середньому арифметичному (Me = Mo = μ).

3. 99.73% результатів вимірів знаходяться в інтервалі $[\mu - 3\sigma; \mu + 3\sigma]$ (див.рис.2).

Цю властивість для нормального розподілу називають "правилом трьох сигм" і використовують для вилучення сильно віддалених результатів вимірів, які вважають "помилковими" і у подальшому дослідженні не враховують.

Крива емпіричного розподілу не завжди ідеально дзвіноподібна (нормальна) і симетрична. Для багатьох розподілів має місце характерний зсув вершини кривої вліво чи вправо. Тому розрізняють лівосторонню і правосторонню асиметрію (рис.4).

Міра скошеності - це найбільш простий показник асиметрії кривої розподілу. В основу визначення її покладено відхилення середнього арифметичного від моди:

$$Sk = \frac{\bar{X} - Mo}{\sigma}.$$

Якщо знак даного виразу від'ємний ($\bar{X} < Mo$), то асиметрія правостороння; якщо додатній ($\bar{X} > Mo$) - лівостороння.

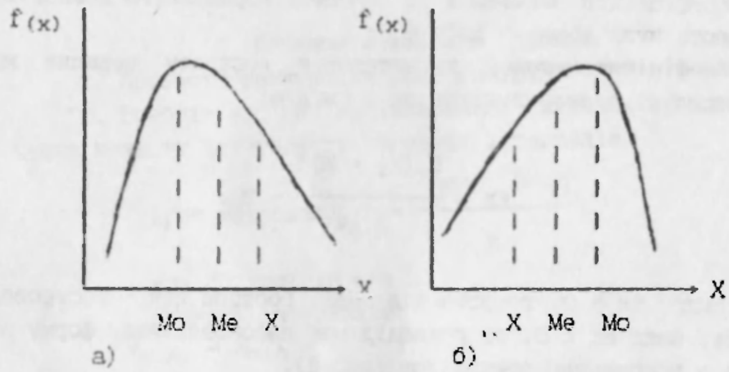


Рис.4. Лівостороння (а) і правостороння (б) асиметрія:
по абсцисі - значення результатів вимірів,
по ординаті - частота.

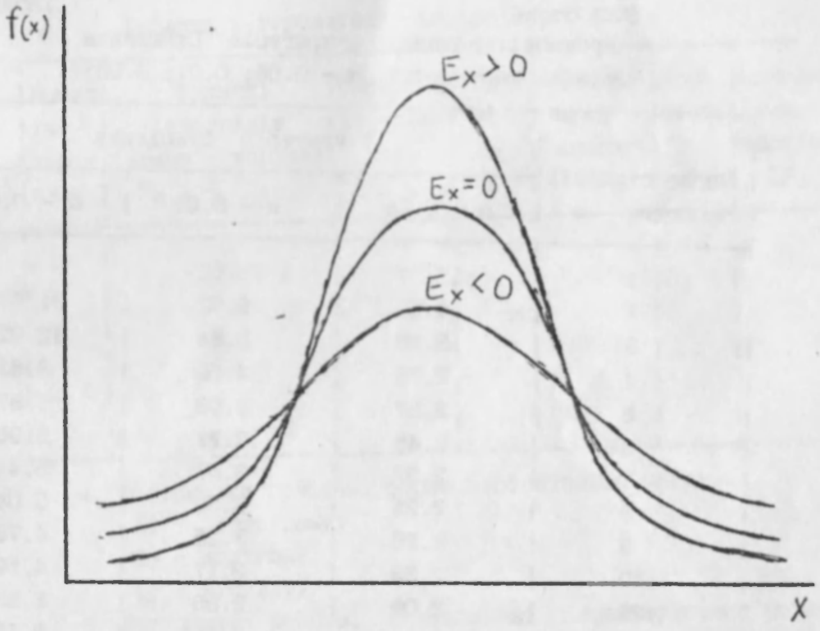


Рис. 5. Криві розподілу:
по абсцисі - значення результатів вимірів,
по ординаті - частота.

Коефіцієнти асиметрії і ексцесу нормального розподілу дорівнюють нулю ($Sk = 0$ і $Ex = 0$).

Коефіцієнт ексцесу характеризує гостроту вершини кривої розподілу і розраховується за формулою:

$$Ex = \frac{\Sigma\,(X_i - \bar{X})^4}{n \cdot \sigma^4} - 3\,.$$

Якщо $Ex > 0$, то розподіл має гострий пік - гостровершинність, якщо $Ex < 0$, то розподіл має плосковершинну форму порівняно з нормальним розподілом (рис.5).

Таблиця 2

Критичні значення t-критерію Стьюдента
(рівень значущості α = 0.05; 0.01; 0.001)

Число ступенів свободи, ν	t-критерій Стьюдента		
	α = 0.05	α = 0.01	α = 0.001
1	12.71	63.66	-
2	4.30	9.92	31.60
3	3.18	5.84	12.92
4	2.78	4.60	8.61
5	2.57	4.03	6.87
6	2.45	3.71	5.96
7	2.37	3.50	5.41
8	2.31	3.36	5.04
9	2.26	3.25	4.78
10	2.23	3.17	4.59
20	2.09	2.85	3.85
30	2.04	2.75	3.65
60	2.00	2.66	3.46

Порядок виконання роботи.

1. Провести ранжування ряду вимірів.
2. Розбити вибірку на інтервали, визначити граничні значення кожного інтервалу і середини інтервалів.

Крок інтервалу $h = \dfrac{X_{max} - X_{min}}{k}$.

Для 20 вимірів k = 5.

Границі інтервалів:

$X_{н.k.} = X_{min} + (j - 1)\cdot h$;

$X_{c.k.} = X_{min} + (2j - 1)\cdot h / 2$;

$X_{в.k.} = X_{min} + j\cdot h$.

3. Заповнити таблицю інтервального варіаційного ряду.

Таблиця інтервального варіаційного ряду.

Номер інтервалу	Границі інтервалів (нижня - верхня) $X_{н.j} - X_{в.j}$	Середина інтервалів $X_{c.j}$	Частота f_j	Накопичена частота Σf_j	Частість F_j	Накопичена частість ΣF_j
1.						
2.						
3.						
...	... ...	...	...	...	...	...
k						

4. Побудувати графічні форми варіаційного ряду:
   а) - гістограму;
   б) - полігон;
   в) - кумуляту.
5. Визначити основні статистичні характеристики центральної тенденції розподілу :

а) середнє арифметичне:

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

де X_i - значення i-го результату виміру;

n - обсяг вибірки.

б) границі модального інтервалу і значення моди:

$$Mo = X_{Mo_н} + h \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

де $X_{Mo_н}$ - значення нижньої границі модального інтервалу,

f_{Mo} - частота модального інтервалу,

f_{Mo-1} і f_{Mo+1} - частоти інтервалів, які знаходяться перед і після модального.

в) границі медіанного інтервалу і значення медіани:

$$Me = X_{Me_н} + h \cdot \frac{0.5 \cdot n - f_{Me-1}}{f_{Me}},$$

де $X_{Me_н}$ - значення нижньої границі медіанного інтервалу;

h - ширина (крок) інтервалів групування;

f_{Me} - частота медіанного інтервалу;

f_{Me-1} - накопичена частота інтервалу, який знаходиться перед медіанним;

n - обсяг вибірки.

6. Вирахувати основні статистичні характеристики розсіяння (варіації) вибірки:

а) розмах вибірки:

$$R = X_{max} - X_{min},$$

де X_{max} і X_{min} - крайні значення результатів вимірів ранжованого ряду;

б) дисперсію вибірки:

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

де X_i - значення i-го результату виміру вибірки;

$\bar{X}$ - середнє арифметичне даної вибірки;

в) середнє квадратичне відхилення:

$$б = \sqrt{D} \;;$$

г) коефіцієнт варіації:

$$V = \frac{б}{\bar{X}} \cdot 100\% \;.$$

7. Провести інтервальну оцінку генерального середнього:

а) вирахувати середнє квадратичне відхилення генеральної сукупності за наступною формулою:

$$б = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2} \;,$$

де n - обсяг вибірки;

б) вирахувати стандартну похибку середнього арифметичного генеральної сукупності:

$$S_{\bar{x}} = \frac{б}{\sqrt{n}} \;;$$

в) визначити значення t-критерію Стьюдента при рівні значущості $\alpha = 0.05$ і числі ступенів свободи $\nu = n - 1$;

г) оцінити середню величину спортивних результатів генеральної сукупності:

$$\bar{X} - \varepsilon < \mu < \bar{X} + \varepsilon \;,$$

де $\varepsilon = S_{\bar{x}} \cdot t(\alpha, \nu)$.

8. Визначити границі інтервалу "три сигми":

$$\bar{X} - 3 \cdot б < X < \bar{X} + 3 \cdot б \;.$$

9. Нанести на графік полігона значення середнього арифметичного, моди, медіани і границі інтервалу "три сигми".

10. Заповнити таблицю основних статистичних характеристик.

Основні статистичні характеристики

N п/п	Найменування характеристики	Позначення	Розмірність	Значення
1.	Середнє арифметичне			
2.	Мода			
3.	Медіана			
4.	Розмах			
5.	Дисперсія			
6.	Середнє квадратичне відхилення			
7.	Коефіцієнт варіації			
8.	Стандартна похибка середнього арифметичного			
9.	Границі інтервалу довіри середнього арифметичного генеральної сукупності			
10.	Границі інтервалу "трьох сигм"			

11. Вирахувати характеристику емпіричного розподілу результатів спортивних вимірів - міру скошеності:

$$Sk = \frac{\bar{X} - Mo}{б},$$

де $\bar{X}$, Mo і б - відповідно значення середнього арифметичного, моди і середнього квадратичного відхилення.

12. Зробити висновки:
   а) про розподіл результатів спортивних вимірів і його подібність до нормального закону;
   б) про варіативність вимірів і однорідність групи;
   в) про наявність хибних вимірів за правилом "трьох сигм".

Контрольні питання

1. Що є предметом математичної статистики?
2. Які основні етапи статистичних досліджень?
3. Що таке вибіркова і генеральна сукупність?
4. Що таке групування і ранжування результатів вимірів?
5. Що таке обсяг вибірки?
6. Як вибирають (розраховують) кількість інтервалів?
7. За якою формулою розраховують крок інтервалу?
8. Що таке частота, накопичена частота, частість, накопичена частість?
9. Що являє собою інтервальний варіаційний ряд вимірів?
10. Як побудувати гістограму, полігон і кумуляту?
11. Які показники є характеристиками центральної тенденції вибірки?
12. Які показники є характеристиками розсіяння (варіації) вибірки?
13. Як провести інтервальну оцінку генерального середнього?
14. Які властивості кривої нормального розподілу?
15. Для чого визначають границі інтервалу "три сигми"?

Література.

1. Основы математической статистики.(Под ред. В.С.Иванова). Учебное пособие для институтов физической культуры.-М.: ФиС. 1990, с.12-32 и 72-78.

2. Спортивная метрология./Под ред. В.М.Зациорского/. -М.: ФиС. 1982, с.10-30 и 46-47.

ЗАВДАННЯ 2. Дослідити статистичний взаємозв'язок між результатами спортивних вимірів параметрів **X** і **Y**.

ТЕОРЕТИЧНІ ВІДОМОСТІ

Дослідження в спорті носять комплексний характер. При контролі за ходом тренувального процесу оцінюють цілий ряд фізіологічних, біомеханічних, біохімічних та інших параметрів. При цьому часто виникає питання про взаємозв'язок ознак. Взаємозв'язок, при якому кожному значенню одного показника відповідає чітко визначене значення другого, називається **функціональним**.

Наприклад: середня швидкість на відрізку шляху дистанції бігуна функціонально зв'язана з часом на даному відрізку, закон Фіхнера в психології і закон Хілла в фізіології, рівняння прямої на площині і т.д.

При дослідженнях в спорті частіше зустрічаються показники з іншим видом взаємозв'язку. Взаємозв'язок, при якому одному значенню однієї величини може відповідати декілька значень другої, називається **статистичним**. Наприклад, одному значенню довжини тіла (зріст) людини може відповідати декілька значень її маси і навпаки.

Статистичний метод, який використовується для дослідження взаємозв'язків, називається **кореляційним аналізом**.

Кореляційний аналіз полягає у визначенні ступеня зв'язку між двома випадковими величинами X і Y.

Дослідження характеру взаємозв'язку починається з побудови графічного зображення результатів вимірів в прямокутній системі координат, де кожна пара результатів буде відображатись точкою (рис.7). Така графічна залежність називається діаграмою розсіяння або **кореляційним полем**. Кореляційне поле відображає статистичний взаємозв'язок між результатами вимірів. Візуальний аналіз кореляційного поля дозволяє якісно оцінити форму, спрямованість і тісноту взаємозв'язку.

Спрямованість взаємозв'язку визначається з залежності між результатами вимірів. У випадку покращення одного показника покращується другий - пряма залежність, спрямованість додатня. Збільшення одного результату викликає зменшення другого результату виміру - обернена залежність, спрямованість від'ємна.

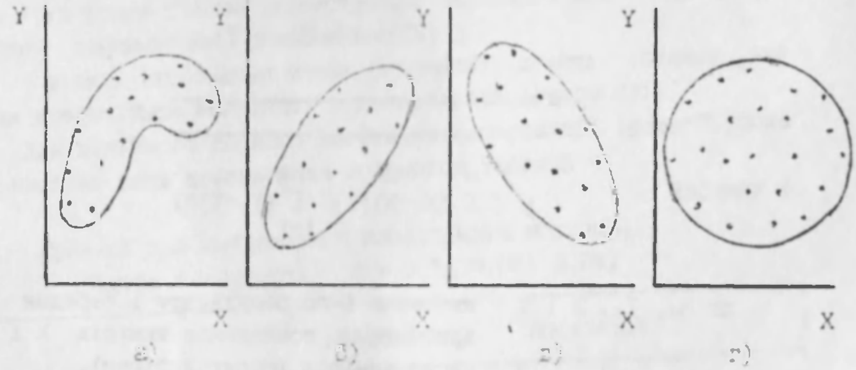


Рис.6. Приклади статистичних взаємозв'язків:
а) - нелінійна форма;
б) - лінійна форма, додатня спрямованість, тіснота середня;
в) - лінійна форма, від'ємна спрямованість, тіснота слабка;
г) - відсутність статистичного взаємозв'язку.

Для оцінки тісноти взаємозв'язку в кореляційному аналізі використовується абсолютна величина (модуль) спеціального показника - **коефіцієнта кореляції**. Абсолютне значення коефіцієнта кореляції знаходиться в межах від 0 до 1. Коефіцієнт кореляції дає кількісну оцінку статистичного взаємозв'язку між результатами вимірів (див.табл. 3).

Таблиця 3

Границі і значення коефіцієнта кореляції.

$r = 0$	- кореляція відсутня;
$0 < \|r\| \leq 0.25$	- дуже слабкий взаємозв'язок;
$0.25 < \|r\| \leq 0.5$	- слабкий взаємозв'язок;
$0.5 < \|r\| \leq 0.75$	- середній взаємозв'язок;
$0.75 < \|r\| < 1$	- сильний статистичний взаємозв'язок;
$\|r\| = 1$	- функціональний взаємозв'язок.

Для оцінки статистичного взаємозв'язку, коли вимірювання проведено в шкалі відношень або інтервалів і форма взаємозв'язку лінійна, використовується коефіцієнт кореляції Браве -

Пірсона, який обчислюється за наступною формулою:

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{[\sum_{i=1}^{n} (X_i - \bar{X})^2] \cdot [\sum_{i=1}^{n} (Y_i - \bar{Y})^2]}},$$

де X_i, Y_i, $\bar{X}$ і $\bar{Y}$ - значення і-го результату і середнє арифметичне показників вимірів X і Y;

n - число вимірів (обсяг вибірки).

Знак коефіцієнта кореляції показує спрямованість зв'язку. Коефіцієнт кореляції зі знаком мінус показує обернену пропорційну залежність і навпаки - додатній знак коефіцієнта кореляції характеризує пряму залежність.

Коли виміри виконані в шкалі порядку, то визначення взаємозв'язку показників проводять з використанням рангового коефіцієнта кореляції Спірмена, який обчислюють за наступною формулою:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n} \Delta d_i^2}{n \cdot (n^2 - 1)},$$

де $\Delta d_i = R_{x_i} - R_{y_i}$ - різниця рангів даної пари показників X і Y;

R_{x_i} і R_{y_i} - місце виміру в ранжованому ряді;

n - число вимірів.

Квадрат коефіцієнта кореляції називається **коефіцієнтом детермінації**:

$$D = r^2 \cdot 100\%.$$

Коефіцієнт детермінації визначає міру лінійної залежності, тобто ту частину загальної варіації одного показника, яка зумовлена варіацією іншого показника. Стверджують, що тільки D (%) взаємозв'язку спортивних результатів пояснюється їх взаємним впливом, а решта (100 - D)% варіативності залежить від інших незумовлених факторів.

При оцінці вірогідності коефіцієнтів взаємозв'язків потрібно встановити наступне:

- чи існує дійсно статистичний взаємозв'язок між двома явищами (параметрами, показниками)?

- в яких вірогідних межах (границях) лежить істинне значення коефіцієнта кореляції в генеральній сукупності?

Для лінійного парного коефіцієнта кореляції Браве-Пірсона це питання розв'язується за допомогою таблиці 4.

Таблиця 4

Границі для вибіркового коефіцієнта кореляції (рівень значущості α - 0.1; 0.05; 0.01)

Число ступенів свободи, ν	Коефіцієнт кореляції		
	α - 0.1	α - 0.05	α - 0.01
2	0.900	0.950	0.990
3	0.805	0.878	0.959
4	0.729	0.811	0.917
5	0.669	0.754	0.874
6	0.622	0.707	0.834
7	0.582	0.666	0.798
8	0.549	0.632	0.765
9	0.521	0.602	0.735
10	0.497	0.576	0.708
20	0.360	0.423	0.537
30	0.296	0.349	0.449
50	0.231	0.273	0.354

Використання цієї таблиці дозволяє визначити значення коефіцієнта кореляції генеральної сукупності $r(\alpha,\nu)$ при заданому рівні значущості α і числі ступенів свободи ν - n - 2.

Порівнюючи абсолютне значення вибіркового (розрахункового) коефіцієнта кореляції $r_{роз}$ з табличним $r(\alpha,\nu)$, можна з вірогідністю σ - $(1 - \alpha)\cdot 100\%$ говорити про існування статистичного взаємозв'язку:

$|r_{роз}| < r(\alpha,\nu)$ - взаємозв'язок не існує;

$|r_{роз}| \geqslant r(\alpha,\nu)$ - взаємозв'язок існує.

Оцінка вірогідності рангового коефіцієнта кореляції Спірмена (ρ) виконується на основі t-критерію Стьюдента, котрий розраховується за наступною формулою:

$$t_{роз} = \frac{|\rho| \cdot \sqrt{n - 2}}{\sqrt{1 - \rho^2}} ,$$

де n - число пар вимірів.

Коли розрахункове значення t-критерію Стьюдента менше критичного $t_{роз} < t(\alpha, \nu)$, то статистичний взаємозв'язок відсутній, і навпаки. Число ступенів свободи, як і у випадку лінійного парного коефіцієнта кореляції, також на два менше числа обсягу вибірки ($\nu = n - 2$). Критичне значення t-критерію Стьюдента визначається за таблицею 2.

В практичних дослідженнях виникає необхідність апроксимувати (описати наближено) діаграму розсіяння математичним рівнянням. Для лінійної залежності необхідно кореляційний еліпс замінити прямою лінією. В прямокутній системі координат рівняння прямої лінії записується в наступному вигляді:

$$Y = a + \mathbf{b} \cdot X.$$

Такий наближений запис статистичної залежності називається рівнянням регресії.

Регресія - це залежність середніх значень випадкової величини Y від величини X.

Кореляція - це залежність між двома випадковими величинами Y і X, котра характеризується за допомогою коефіцієнтів кореляції.

В рівнянні регресії статистичної залежності **b** називають коефіцієнтом регресії, **a** - вільним членом рівняння регресії.

Значення коефіцієнта регресії обчислюється за наступною формулою:

$$\mathbf{b} = r \cdot \frac{\sigma_y}{\sigma_x} ,$$

де r - значення коефіцієнта кореляції;

σ_x і σ_y - середні квадратичні відхилення для показників X і Y.

Коефіцієнт регресії **b** має розмірність, яка дорівнює відношенню розмірностей показників, що вивчаються, і той же знак, що і коефіцієнт кореляції.

Вільний член рівняння регресії знаходять за наступною формулою:

$$a = \bar{Y} - b\cdot\bar{X} ,$$

де $\bar{Y}$ і $\bar{X}$ - вибіркові середні арифметичні показників Y і X.

Рівняння лінії регресії $Y = a + b\cdot X$ наближено замінює статистичний взаємозв'язок між показниками Y і X на функціональний.

Лінію рівняння регресії будують на кореляційному полі по двох точках ([illegible], при X_{min} і X_{max}), отримуючи відповідні значення Y_1 і Y_2 (див.рис.7):

$$Y_1 = a + b\cdot X_{min};$$
$$Y_2 = a + b\cdot X_{max}.$$

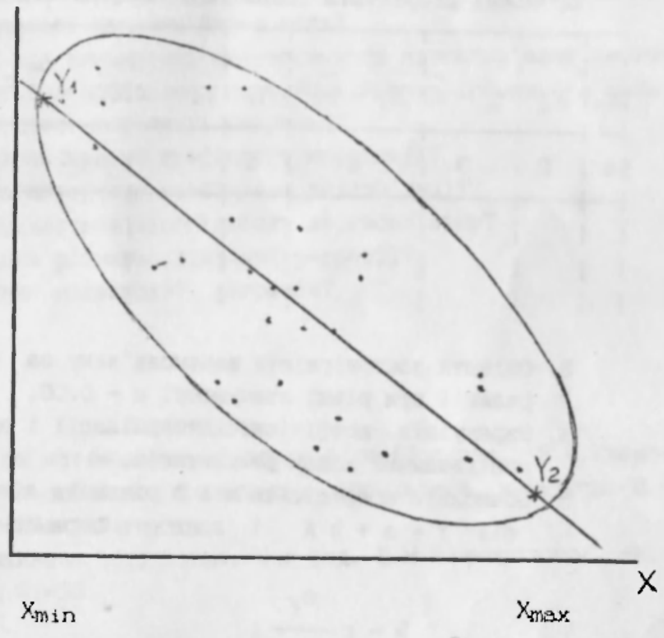


Рис.7. Кореляційне поле з нанесеною лінією рівняння регресії

Порядок виконання роботи.

1. Побудувати кореляційне поле.
2. Підрахувати коефіцієнт кореляції Браве-Пірсона:

$$r = \frac{\Sigma(X_i - \bar{X})\cdot(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i-\bar{X})^2\cdot\Sigma(Y_i-\bar{Y})^2}} .$$

де X_i, Y_i, $\bar{X}$ і $\bar{Y}$ - значення і-го результату і середні арифметичні показників вимірів X і Y.

Результати проміжних розрахунків занести в таблицю.

Проміжні результати розрахунку коефіцієнта кореляції.

n_i	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	Y_i	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$	$(X_i-\bar{X})\cdot(Y_i-\bar{Y})$
1	2	3	4	5	6	7	8

3. Оцінити достовірність взаємозв'язку за коефіцієнтом кореляції при рівні значущості $\alpha = 0.05$.
4. Вирахувати коефіцієнт детермінації і зробити висновок про взаємний вплив результатів.
5. Обчислити коефіцієнти **a** і **b** рівняння лінійної регресії $Y = a + b\cdot X$ і записати отримане рівняння:

$$b = r\cdot\frac{\sigma_y}{\sigma_x} ;$$

$$a = \bar{Y} - b\cdot\bar{X} .$$

6. Побудувати лінію регресії на кореляційному полі по двох точках X_{min} і X_{max}.

7. Зробити загальний висновок про виявлений статистичний взаємозв'язок.

Контрольні питання.

1. Що таке функціональний взаємозв'язок?
2. Що таке статистичний взаємозв'язок?
3. Що називають кореляційним аналізом і в чому він полягає?
4. Що таке кореляційне поле?
5. Які характеристики якісно можна оцінити при візуальному аналізі кореляційного поля?
6. Коли для визначення статистичного взаємозв'язку застосовується коефіцієнт кореляції Браве-Пірсона, а коли коефіцієнт кореляції Спірмена?
7. Що характеризує коефіцієнт кореляції?
8. Що характеризує коефіцієнт детермінації?
9. Як оцінюється вірогідність взаємозв'язку?
10. Що таке рівняння лінійної регресії?
11. Що таке кореляція? регресія?

Література.

1. Основы математической статистики.(Под ред. В.С.Иванова). Учебное пособие для институтов физической культуры.-М.: ФиС. 1990, с.124-150.

2. Спортивная метрология./Под ред. В.М.Зациорского/. -М.: ФиС. 1982, с.31-53.

ПРИКЛАДИ ВИКОНАННЯ РОЗРАХУНКОВО-ГРАФІЧНИХ РОБІТ

Тема: Статистична обробка ряду результатів спортивних вимірів.

Мета: Набути практичних навиків в проведенні статистичного аналізу.

Завдання 1. Провести попередню статистичну обробку результатів вимірів в бігу на 100 м.

Вихідні дані: X, с - результати бігу на 100 м

11.1	11.1	13.0	12.4	10.9	11.2	12.5	11.1	10.7	10.2
14.0	14.5	12.3	12.3	11.8	14.7	11.8	12.6	12.5	12.5
10.1	11.9	12.7	10.3	14.9	14.6	10.3	13.1	10.1	14.7

Виконання роботи:

I. Проводимо ранжування ряду вимірів.

Ранжуванням називається розміщення результатів вимірів у порядку зростання чи спадання.

1.	2.	3.	4.	5.
10.1	11.1	12.3	13.0	14.0
10.1	11.1	12.3	13.1	14.5
10.2	11.1	12.4		14.6
10.3	11.2	12.5		14.7
10.3	11.8	12.5		14.7
10.7	11.8	12.5		14.9
10.9	11.9	12.6		
		12.7		

2. Визначаємо границі інтервалів і розбиваємо ранжований ряд на інтервали.

Для 30 вимірів $k = 5$.

Границі інтервалів:
1. 10.10 - 11.06
2. 11.06 - 12.02
3. 12.02 - 12.98
4. 12.98 - 13.94
5. 13.94 - 14.90

$$\text{Крок інтервалу } h = \frac{X_{max} - X_{min}}{k} = \frac{14.9 - 10.1}{5} = 0.96 \text{ (с)}.$$

Вибір числа інтервалів: Таблиця 1

Обсяг вибірки, n	Число інтервалів
25 - 40	5 - 6
40 - 60	6 - 8
60 - 100	7 - 10
100 - 200	8 - 12
більше 200	10 - 15

3. Заповнюємо таблицю інтервального варіаційного ряду.

Частота (f_j) - кількість значень вимірів в кожному інтервалі.

Накопичена частота $\sum_{j=1}^{k} f_j$ - це число, отримане послідовним додаванням частот.

Частість $F_j = \frac{f_j}{n}$ - відношення частоти до обсягу вибірки.

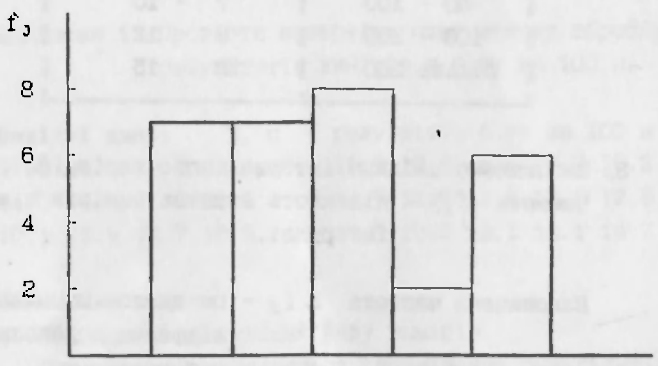
Накопичена частість $\sum_{j=1}^{k} F_j$ - відношення накопиченої частоти до обсягу вибірки.

ВАРІАЦІЙНИЙ РЯД ВИМІРІВ Таблиця 2

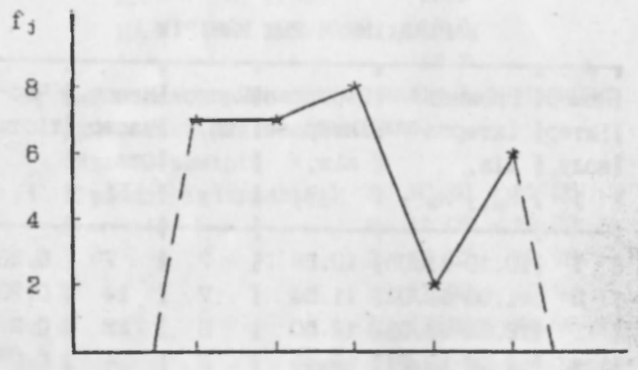
Номер інтервалу, j	Границі інтервалів, $X_{н.j}-X_{в.j}$	Середина інтервалів, $X_{c.j}$	Частота, f_j	Накоп. частота, Σf_j	Частість, F_j	Накоп. частість, ΣF_j
1	10.10-11.06	10.58	7	7	0.23	0.23
2	11.06-12.02	11.54	7	14	0.23	0.46
3	12.02-12.98	12.50	8	22	0.27	0.73
4	12.98-13.94	13.46	2	24	0.07	0.80
5	13.94-14.90	14.42	6	30	0.20	1.00

4. Будуємо графічні форми варіаційного ряду.

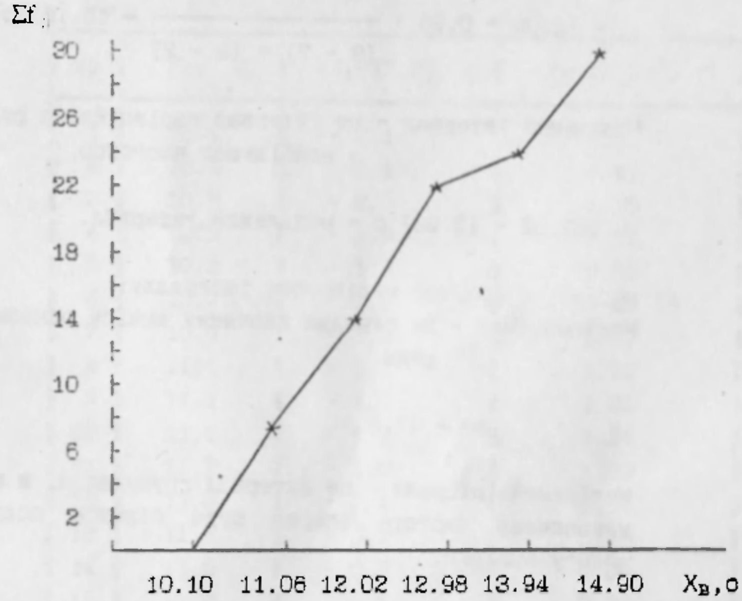
Гістограма - це стовбчата діаграма, що відображає розподіл частот по границях інтервалів.



Полігон частот розподілу - це графік залежності між серединами інтервалів та частотами.



Кумулята - це графік розподілу накопиченої частоти по верхніх границях інтервалів.



5. Визначаємо характеристики центральної тенденції (положення) вибіркової сукупності:

   а) **середнє арифметичне** значення (див. табл.3) визначається діленням суми всіх результатів вимірів на обсяг вибірки:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{365.9}{30} = 12.20\ (\text{с}).$$

   б) моду і границі модального інтервалу:

   Мода (Mo) - це значення виміру з найбільшою імовірністю появи.

$$Mo = X_{Mo_н} + h \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} =$$

$$= 12.02 + 0.96 \cdot \frac{8 - 7}{(8 - 7) + (8 - 2)} = 12.16 \text{ (с)}.$$

Модальний інтервал - це інтервал варіаційного ряду з найбільшою частотою.

[12.02 - 12.98] с - модальний інтервал.

в) медіану і границі медіанного інтервалу:

Медіана (Me) - це середнє значення виміру ранжованого ряду.

Me = 12.8 с.

Медіанний інтервал - це інтервал групування, в якому накопичена частота вперше буде більшою половини обсягу вибірки.

[12.02 - 12.98] с.

6. Визначаємо характеристики розсіяння (варіації) вибіркової сукупності:

а) **розмах** вибірки характеризує крайні відхилення граничних результатів вимірів:

$$R = X_{max} - X_{min} = 14.9 - 10.1 = 4.8 \text{ (с)}.$$

б) **дисперсія** вибірки - це середнє значення квадратів відхилень кожного елементу ряду вимірів від середнього арифметичного.

$$D = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n} = \frac{65.29}{30} = 2.176 \text{ (с}^2\text{)}.$$

Таблиця 3

Дані для розрахунку статистичних характеристик

No	X_i, c	$(X_i-12.2)$, c	$(X_i-12.2)^2$, c^2
1	10.1	- 2.1	4.41
2	10.1	- 2.1	4.41
3	10.2	- 2.0	4.00
4	10.3	- 1.9	3.61
5	10.3	- 1.9	3.61
6	10.7	- 1.5	2.25
7	10.9	- 1.3	1.69
8	11.1	- 1.1	1.21
9	11.1	- 1.1	1.21
10	11.1	- 1.1	1.21
11	11.2	- 1.0	1.00
12	11.8	- 0.4	0.16
13	11.8	- 0.4	0.16
14	11.9	- 0.3	0.09
15	12.3	0.1	0.01
16	12.3	0.1	0.01
17	12.4	0.2	0.04
18	12.5	0.3	0.09
19	12.5	0.3	0.09
20	12.5	0.3	0.09
21	12.6	0.4	0.16
22	12.7	0.5	0.25
23	13.0	0.8	0.64
24	13.1	0.9	0.81
25	14.0	1.8	3.24
26	14.5	2.3	5.29
27	14.6	2.4	5.76
28	14.7	2.5	6.25
29	14.7	2.5	6.25
30	14.9	2.7	7.29
Σ	365.9	~ 0	65.29

в) **середнє квадратичне відхилення** (стандартне відхилення) характеризує ступінь відхилення результатів від середнього значення в абсолютних одиницях.

$$\sigma = \sqrt{D} = \sqrt{2.176} = 1.47 \text{ (с)}.$$

г) **коефіцієнт варіації** - це відносна міра розсіяння ознаки, що характеризує варіативність вибірки і показує ступінь відхилення результатів від середнього значення у відсотках.

$$V = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{1.47}{12.2} \cdot 100\% = 12.04\%$$

Отже, варіація результатів в групі - середня.

7. Проводимо інтервальну оцінку генерального середнього:

а) вираховуємо середнє квадратичне відхилення генеральної сукупності:

$$\sigma = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{65.29}{30 - 1}} = 1.50 \text{ (с)}.$$

б) вираховуємо стандартну похибку середнього арифметичного:

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.50}{\sqrt{30}} = \frac{1.50}{5.48} = 0.27 \text{ (с)}.$$

в) визначаємо величину інтервалу довіри при рівні значущості $\alpha = 0.05$:

$$\bar{X} - \varepsilon < \mu < \bar{X} + \varepsilon.$$

Згідно таблиці підручника ("Спортивная метрология" под ред.В.М.Зациорского, с.240) або таблиці 2 теоретичної частини цих методичних вказівок визначаємо критичне значення t-критерію Стьюдента при $\alpha = 0.05$ і числі ступенів свободи $\nu = 30 - 1 = 29$.

$t(\alpha,\nu) = 2.04$

$q = (1 - 0.05) \cdot 100\% = 95\%$ - значення вірогідності.

$$\varepsilon = S_{\bar{x}} \cdot t(\alpha,\nu) = 0.27 \cdot 2.04 = 0.55 \text{ (с)}.$$

г) визначаємо границі інтервалу довіри для генерального середнього:

$$12.20 - 0.55 < \mu < 12.20 + 0.55$$

[11.65 - 12.75] с.

8. Визначаємо границі інтервалу "три сигми":

Правило 3-ох сигм:

якщо результат є відмінним від середнього на величину більшу ніж 3-сигми, то його вважають випадковим (невірним) і в подальшому аналізі не використовують.

$$\bar{X} - 3\cdot\sigma < X < \bar{X} + 3\cdot\sigma$$

$$12.20 - 3\cdot1.47 < X < 12.20 + 3\cdot1.47$$

[7.79 - 16.61] с.

9. Наносимо на вісь абсцис графіка полігону значення середнього арифметичного, моди, медіани і границі інтервалу "три сигми".

10. Заповнюємо таблицю основних статистичних характеристик.

ОСНОВНІ СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ Таблиця 4

Номер	Найменування характеристики	Позначення	Значення	Розм.
1.	Середнє арифметичне значення	$\bar{X}$	12.20	с
2.	Мода	Mo	12.16	с
3.	Медіана	Me	12.30	с
4.	Розмах вибірки	H	4.80	с
5.	Дисперсія	D	2.176	c^2
6.	Середнє квадрат. відхилення	σ	1.47	с
7.	Коефіцієнт варіації	V	12.04	%
8.	Стандартна похибка середнього арифметичного	$S_{\bar{x}}$	0.27	с
9.	Границі інтервалу довіри середнього арифметичного генеральної сукупності	$[\bar{X}-\varepsilon]$ $[\bar{X}+\varepsilon]$	11.65 - 12.75	с
10.	Границі інтервалу "3-сигми"	$[\bar{X}-3\sigma]$ $[\bar{X}+3\sigma]$	7.79 - 16.61	с

11. Вираховуємо міру скошеності - характеристику емпіричного розподілу результатів спортивних вимірів.

$$Sk = \frac{\bar{X} - Mo}{\delta} = \frac{12.20 - 12.16}{1.47} = 0.03.$$

Висновки:

Розподіл результатів спортивних вимірів має подібність до нормального з слабовираженою лівосторонньою асиметрією ($Sk > 0$).

Виходячи з аналізу результатів в бігу на 100м можна вважати:

а) - рівень результатів бігу на 100м - середній і дорівнює 12.20 с;

б) - дана група за даним показником варіації є середньої однорідності ($10\% < V = 12.04\% < 20\%$);

в) - з вірогідністю 95% можна стверджувати, що генеральне середнє арифметичне знаходиться у межах [11.65 - 12.75] с;

г) - за правилом "трьох сигм" всі результати групи належать до одної генеральної сукупності і з вірогідністю 99.7% можна стверджувати, що хибних вимірів немає, оскільки всі виміри знаходяться в межах [7.79 - 16.61] с, що відповідає границі інтервалу $[\bar{X} \pm 3\cdot\delta]$.

Завдання 2. Дослідити статистичний взаємозв'язок між результатами у стрибках в довжину та бігу на 100 м.

Вихідні дані: n - 30

X, см - результати стрибків в довжину:

808 728 735 759 710 808 727 704 751 804
652 639 726 733 736 698 748 688 715 677
756 722 760 759 653 655 743 707 824 712

Y, с - результати бігу на 100 м:

11.1 11.1 13.0 12.4 10.9 11.2 12.5 11.1 10.7 10.2
14.0 14.5 12.3 12.3 11.8 14.7 11.8 12.6 12.5 12.5
10.1 11.9 12.7 10.3 14.9 14.6 10.3 13.1 10.1 14.7

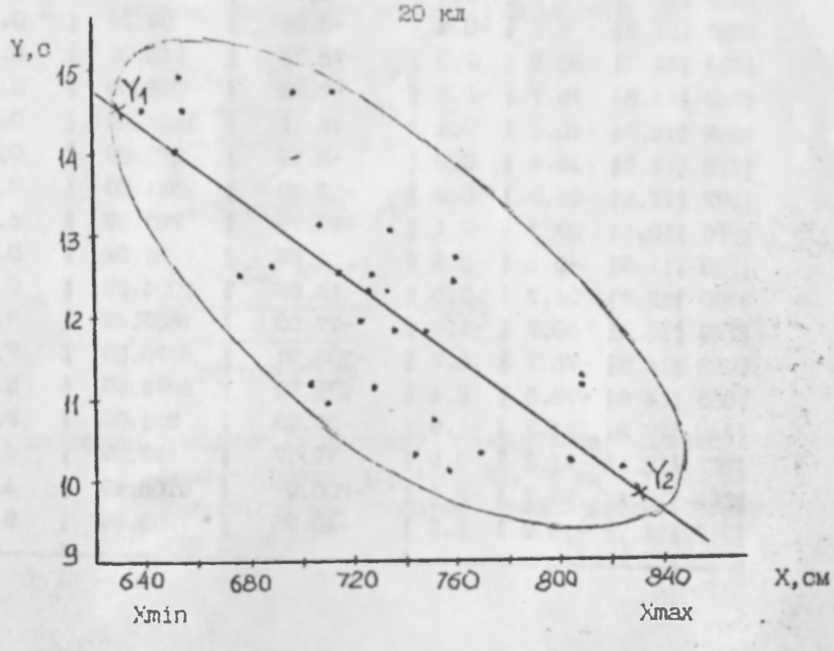
Виконання роботи:

1. Будуємо на площині кореляційне поле.

Обираємо масштаб (1 клітинка в зошиті - 5 мм):

по осі ординат $\dfrac{(14.9 - 10.1)\ с}{20\ кл} \approx 0.25$ с/кл.

по осі абсцис $\dfrac{(824 - 639)\ см}{20\ кл} \approx 10$ см/кл.



Дані для розрахунку коефіцієнта кореляції Браве-Пірсона.

X, см	Y, с	$X_i-\bar{X}$, см	$Y_i-\bar{Y}$, с	$(X_i-\bar{X})\cdot(Y_i-\bar{Y})$	$(X_i-\bar{X})^2$, см^2	$(Y_i-\bar{Y})^2$, с^2
808	11.1	79.7	-1.1	-87.67	6352.09	1.21
728	11.1	- 0.3	-1.1	0.33	0.09	1.21
735	13.0	6.7	0.8	5.36	44.89	0.64
759	12.4	30.7	0.2	6.14	942.49	0.04
710	10.9	-18.3	-1.3	23.79	334.89	1.69
808	11.2	79.7	-1.0	-79.70	6352.09	1.00
727	12.5	-1.3	0.3	-0.39	1.69	0.09
704	11.1	-24.3	-1.1	26.73	590.49	1.21
751	10.7	22.7	-1.5	-34.05	515.29	2.25
804	10.2	75.7	-2.0	-151.40	5730.49	4.00
652	14.0	-76.3	1.8	-137.34	5821.69	3.24
639	14.5	-89.3	2.3	-205.39	7974.49	5.29
726	12.3	- 2.3	0.1	-0.23	5.29	0.01
733	12.3	4.7	0.1	0.47	22.09	0.01
736	11.8	7.7	-0.4	-3.08	59.29	0.16
698	14.7	-30.3	2.5	-75.75	918.09	6.25
748	11.8	19.7	-0.4	-7.88	388.09	0.16
688	12.6	-40.3	0.4	-16.12	1624.09	0.16
715	12.5	-13.3	0.3	-3.99	176.89	0.09
677	12.5	-51.3	0.3	-15.39	2631.69	0.09
756	10.1	27.7	-2.1	-58.79	767.29	4.41
723	11.9	-5.3	-0.3	1.59	28.09	0.09
760	12.7	31.7	0.5	15.85	1004.89	0.25
769	10.3	40.7	-1.9	-77.33	1656.49	3.61
653	14.9	-75.7	2.7	-203.31	5670.09	7.29
655	14.6	-73.3	2.4	-175.92	5372.89	5.76
743	10.3	14.7	-1.9	-27.93	216.09	3.61
707	13.1	-21.3	0.9	-19.17	453.69	0.81
824	10.1	95.7	-2.1	-200.97	9158.49	4.41
712	14.7	-16.3	2.5	-40.75	265.69	6.25

Кореляційне поле відображає статистичний взаємозв'язок між результатами вимірів і являє собою графічне зображення результатів вимірів в прямокутній системі координат.

Візуальний аналіз кореляційного поля дозволяє стверджувати, що форма взаємозв'язку наближається до лінійної, спрямованість - зворотня, тіснота - середня.

Табличні дані:

ΣX_i - 21849 см, $\bar{X} = \Sigma X_i / n$ - 21849 / 30 - 728.3 см.

ΣY_i - 365.9 с, $\bar{Y} = \Sigma Y_i / n$ - 365.9 / 30 - 12.20 с.

$\Sigma (X_i - \bar{X})\cdot(Y_i - \bar{Y})$ - -1541.67 см·с.

$\Sigma (X_i - \bar{X})^2$ - 65070.9 см^2, $б_x$ - 46.57 см.

$\Sigma (Y_i - \bar{Y})^2$ - 65.29 с^2, $б_y$ - 1.47 с.

2. Обчислюємо коефіцієнт кореляції Браве-Пірсона:

$$r_p = \frac{\Sigma(X_i - \bar{X})\cdot(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i-\bar{X})^2\cdot\Sigma(Y_i-\bar{Y})^2}} = \frac{-1541.67}{\sqrt{65070.9\cdot 65.29}} = -0.75 .$$

Висновок: виходячи із значення коефіцієнта кореляції Браве-Пірсона (r_p = -0.75) констатуємо, що залежність має зворотню спрямованість і середню тісноту.

3. Оцінюємо достовірність статистичного взаємозв'язку.

Вибираємо рівень значущості α - 0.05, що відповідає вірогідності q - 95% .
Число ступенів свободи ν - n - 2 - 30 - 2 - 28.
Згідно табл.4, яка подана в теоретичній частині, граничне значення коефіцієнта кореляції $r(\alpha,\nu)$ - 0.396.

Висновок: оскільки $|r_{в}| \geqslant r(\alpha,\nu)$, то з вірогідністю 95% можна вважати, що між результатами існує статистичний взаємозв'язок.

4. Обчислюємо коефіцієнт детермінації.

Коефіцієнт детермінації визначає міру лінійної залежності.

$$D - r^2\cdot 100\% - (-0.75)^2\cdot 100\% - 56\%.$$

Висновок: отже 56% взаємозв'язку пояснюється їх взаємним впливом, а решта результатів (44%) - залежить від інших незумовлених причин.

5. Обчислюємо коефіцієнти рівняння лінійної регресії.

Регресія - це залежність середніх значень випадкової величини Y від величини X.
Рівняння лінії регресії $Y - a + b\cdot X$ наближено замінює статистичний взаємозв'язок між показниками Y і X на функціональний.

$$\mathbf{b} - r\cdot\frac{б_y}{б_x} - -0.75\cdot\frac{1.47}{46.57} \approx -0.0237.$$

$$\mathbf{b} - \frac{\Sigma(X_i-\bar{X})\cdot(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2} - \frac{-1541.67}{65070.9} \approx -0.0237.$$

$a = \bar{Y} - b \cdot \bar{X} = 12.2 - (-0.0237) \cdot 728.3 = 29.46.$

$Y = 29.46 - 0.0237 \cdot X$ - рівняння регресії.

6. Будуємо лінію регресії на кореляційному полі.

Приймаємо: $X_1 = 620$ см, $X_2 = 830$ см,
тоді $Y_1 = 29.46 - 0.0237 \cdot 630 = 14.53$ (с);
$Y_2 = 29.46 - 0.0237 \cdot 830 = 9.79$ (с).

Висновок: з вірогідністю 95% стверджуємо, що між результатами в бігу на 100м та у стрибках в довжину в цій групі спортсменів існує лінійний, зворотній, середній статистичний взаємозв'язок.