

В.П. Вишневецька

**Основи математичної
статистики з елементами
хмаро орієнтованих технологій**

Київ – 2019

Міністерство освіти і науки України
Національний педагогічний університет
ім. М.П. Драгоманова

Основи математичної статистики з елементами хмаро орієнтованих технологій

Київ – 2019

УДК 519.22:004.77] (076.5)

Рецензенти:

Струтинська О.В. – кандидат педагогічних наук, доцент кафедри інформаційних технологій і програмування, факультету інформатики, заступник декана з наукової роботи та міжнародної діяльності НПУ імені М.П. Драгоманова

Глушак О. М. – кандидат педагогічних наук, доцент кафедри інноваційних технологій і математичних дисциплін Київського університету ім. Б. Грінченко.

Укладачі:

Вишневецька В.П. – старший викладач кафедри інноваційних та інформаційних технологій у фізичній культурі і спорті Національного університету фізичного виховання і спорту України.

Основи математичної статистики з елементами хмаро орієнтованих технологій: Лабораторний практикум для майбутніх фахівців сфери фізичної культури та спорту. / Укладач Вишневецька В.П. – К.: Видавництво НПУ імені М.П. Драгоманова, 2019. – 84 с.

Лабораторний практикум призначений для майбутніх фахівців фізичної культури та спорту. Може бути використаний студентами різних спеціальностей для підвищення рівня сформованості інформатичних компетентностей.

© Вишневецька В. П., 2019

© Вид-во НПУ імені М. П. Драгоманова, 2019

ЗМІСТ

Передмова	3
<i>Лабораторна робота № 1. Елементи описової статистики.</i>	<i>6</i>
Завдання самостійної роботи для перевірки знань та вмінь студентів щодо основ описової статистики.. ..	30
<i>Лабораторна робота № 2. Вибірковий метод.</i>	<i>36</i>
<i>Лабораторна робота № 3. Параметричні критерії.</i>	<i>45</i>
<i>Лабораторна робота № 4. Непараметричні критерії. Критерій Вілкоксона.</i>	<i>55</i>
<i>Лабораторна робота № 5. Кореляційний аналіз. Оцінка та аналіз коефіцієнтів кореляції та детермінації.</i>	<i>65</i>
Додаток А.	76
Додаток Б.	77
Додаток В.	77
Список літературних джерел	78

ПЕРЕДМОВА

*Жодну проблему не можна розв'язати
на тому ж рівні, на якому вона виникла.*

Альберт Ейнштейн

Одним із завдань системи освіти є підвищення рівня конкурентоспроможності випускників, формування самостійної, творчої та відповідальної особистості, яка вирізняється мобільністю, динамізмом, конструктивністю, здатна адаптуватися до бурхливих змін у науці й техніці.

Випускник вищого навчального закладу повинен реагувати на швидкі зміни в сучасних інформаційних та виробничих технологіях, критично їх оцінювати та використовувати для вирішення професійних завдань. Сучасний фахівець повинен вміти розробляти шляхи для досягнення поставленої мети та вміти аналізувати отримані результати.

Практикум присвячений основам математичної статистики. З метою прискорення обчислень розглянуто використання хмарних технологій на прикладі програми *Excel* хмарного сервісу *OneDrive* компанії *Microsoft*.

Лабораторний практикум розроблений для студентів навчальних закладів сфери фізичної культури і спорту, складається зі вступу, п'яти лабораторних робіт, набору завдань до самостійної роботи, додатків, списку рекомендованої літератури.

Кожна лабораторна робота складається з: теми, мети, переліку понять та вмінь, що будуть сформовані під час виконання лабораторної роботи, теоретичного матеріалу, практичних завдань, контрольних запитань, завдань у тестовій формі для самоперевірки, прикладних задач, завдань для систематизації та узагальнення знань.

Систематичне виконання завдань лабораторного практикуму допоможе майбутньому фахівцю підвищити рівень його самостійної роботи, познайомить з основами математичної статистики, сформує вміння аналізувати дані експерименту для визначення ефективності тренувального процесу та його коригування за допомогою програми *Excel* хмарного сервісу *OneDrive* компанії *Microsoft*.

Лабораторна робота № 1.

Тема: елементи описової статистики.

Мета: ознайомитися з основними поняттями описової статистики, сформуванню вміння обчислювати значення статистичних параметрів вибірки, аналізувати вибірку на однорідність, будувати графіки.

Основні поняття: вибірка, основні параметри описової статистики, середня арифметична величина, мода, медіана, дисперсія, стандартне (середнє квадратичне) відхилення, гістограма, полігон частот.

Теоретичний матеріал.

В процесі еволюції людина намагалася пізнати оточуючий світ та явища, що в ньому відбуваються; зрозуміти вплив різноманітних факторів на те чи інше явище, дослідити, які з них є суттєвими, а які ні.

В наш час з метою дослідження певного явища здійснюється велика серія спостережень чи випробувань, за результатами яких, використовуючи спеціальні алгоритми, аналізують досліджуване явище та вплив на нього численних факторів.

В сукупності ці значення називаються *статистичними даними*.

Аналізом статистичних даних, визначенням закономірностей, яким вони відповідають, до планування, проведення та аналізу експерименту займаються в межах *математичної статистики*.

Зазначимо, що слово “статистика” має той самий корінь, що й слово “держава” (state). Раніше це слово позначало *науку про управління*, потім – *збирання даних про державу*, а після – *збирання й аналіз даних*.

Сучасна математична статистика складається з двох частин – *описової та аналітичної статистики*.

Аналітична статистика також називається *теорією статистичних висновків*. Вона займається опрацюванням даних, отриманих в результаті експерименту та формулюванням висновків, що мають значне прикладне значення для різних сфер людської діяльності.

Математичною основою теорії статистичних є інша математична наука – *теорія ймовірностей*.

Використовуючи апарат описової статистики можна узагальнювати дані, отримані в результаті спостереження або експерименту.

Випробування, експерименти, результати яких неможливо передбачити, називають *стохастичними* або *випадковими*.

Дані, отримані в результаті експерименту, називають *статистичними, емпіричними даними, змінними, варіантами, величинами* тощо.

Величина, що в результаті випробування може набувати того чи іншого (але не одного) числового значення, заздалегідь невідомого й зумовленого випадковими причинами, називається *випадковою величиною*.

Випадкові величини позначають великими літерами X, Y, Z, \dots , а їх можливі значення – відповідними малими літерами з індексами.

Розрізняють *дискретні* та *неперервні випадкові величини*.

Дискретною випадковою величиною називається така величина, кількість можливих значень якої або скінченна, або нескінченна зчислена множина (елементи якої можна пронумерувати).

Прикладом дискретних випадкових величин є кількість запитань (X), що задає викладач студенту під час складання заліку: x_1, x_2, x_3, \dots

Величина, можливі значення якої неперервно заповнюють деякий інтервал (скінченний або нескінченний), називається *неперервною випадковою величиною*.

Якщо наслідку стохастичного експерименту ставлять у відповідність деяке число, то отримують *випадкове значення*. Прикладом випадкового значення є число, отримане в результаті суми очок, що випали на обох кубиках, або кількість пострілів із серії випробувань, що влучили в мішень.

Нехай проводиться експеримент, що має певну мету. В результаті його проведення відбулося n випробувань. Отримали набір даних – вибірку $\{x_1, x_2, \dots, x_n\}$. Кожен елемент цієї вибірки називатимемо *варіантою*. Число, що позначає кількість повторів кожної з варіант вибірки будемо називати *частотою* та позначатимемо n_i . Сума частот кожної з варіант вибірки ($n_1 + n_2 + \dots + n_k$, де $i \in [1; k]$ і k – число різних варіант у вибірці) утворюють *обсяг вибірки* n .

Як правило, первинні дані, отримані в результаті експерименту, представлені у вигляді невпорядкованого ряду чисел, що дослідник записав у послідовності їх надходження. Оскільки важко аналізувати та робити якісь висновки, маючи невпорядкований набір даних, тому первинні дані групують, а далі будують варіаційний ряд.

Варіаційним рядом називають впорядковану неспадну послідовність варіант x_i , де $i \in [1; n]$. В залежності від даних розрізняють *інтервальні* та *дискретні* ряди.

Дискретний ряд – це ряд, в якому варіанти виражені цілими числами.

Прикладом дискретного ряду є таблиця “Сила удару правою ногою по м’ячу”, де x_i – це результати футболістів, а n_i – значення частот (кількість повторів значень варіантів вибірки).

Сила удару правою ногою по м’ячу
футболістами (x_i , Н)

x_i	n_i
6,2	1
6,3	2
6,4	7
6,5	5
7,0	2
Разом:	17

Інтервальний ряд розподілу – це ряд, в якому значення ознаки задані у вигляді інтервалів.

В цьому випадку дані впорядковують від мінімального до максимального значення. Весь цей період розбивають на рівні інтервали, так звані *класи*. Як результат, замість варіант матимемо інтервали (класи), а замість частот – накопичену частоту в класі.

Прикладом інтервального ряду розподілу є таблиця, в якій представлена залежність між шкалою оцінювання ECTS (x_i) і кількістю однакових балів, отриманих студентами за результатами складання заліку з певної дисципліни (n_i):

x_i	n_i
90 – 100	1
82 – 89	6
75 – 81	10
67 – 74	7
60 – 66	8
35 – 59	2
1 – 34	1

Разом:	35
--------	----

Впорядкувавши в порядку зростання дані експерименту, обчислюють значення *статистичних параметрів* вибірки – *середню арифметичну величину, моду, медіану, дисперсію, стандартне (середнє квадратичне) відхилення, коефіцієнт варіації вибірки.*

Середня арифметична величина.

Середня арифметична величина – це узагальнююча величина, що характеризує рівень варіюючої ознаки в якісно однорідній сукупності.

Для обчислення значення середньої арифметичної величини вибірки, дані якої не впорядковані, використовують формулу:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}, \quad (1)$$

де x_i – різні варіанти вибірки від x_1 до x_n ;

n_i – частоти кожної з варіант вибірки (числа, що вказують на кількість повторень кожного із представників x_i), n – обсяг вибірки.

Якщо дані вибірки є впорядкованими, то для обчислення значення середньої арифметичної величини вибірки доцільніше використовувати формулу:

$$\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{n} \quad (1.1)$$

де x_i – різні варіанти вибірки, впорядковані за зростанням від x_1 до x_k ;

k – кількість різних значень варіант вибірки;

n_i – частоти кожного із варіант вибірки (числа, що вказують на кількість повторів кожного із представників x_i), n – обсяг вибірки.

Розглянемо, як обчислюють значення середньої арифметичної величини вибірки за допомогою табличного процесора *MS Excel* пакету *MS Office* (пункт а) та програми *Excel* хмарного сервісу *OneDrive* (пункт б).

а. Для обчислення значення середньої арифметичної величини у табличному процесорі *MS Excel* пакету *MS Office* версії 2007 і вище доступна функція *AVERAGE* або *СРЗНАЧ* (в залежності від версії пакету *MS Office*), яку можна викликати різними способами. Розглянемо їх:

1. Вказати за допомогою лівої клавіші миші клітинку, де буде результат. На вкладці *Основне* у групі *Редагування* натиснути в правій частині кнопку *СУМА*. У списку обрати *Середнє*. В результаті в

обраній комірці з'явиться запис “=AVERAGE()” або “=CPЗНАЧ()”. У дужках необхідно поставити курсор миші та обрати необхідний діапазон значень вибірки. Натиснувши клавішу клавіатури *Enter*, користувач побачить значення середньої арифметичної величини, обчислене програмою.

2. Обрати за допомогою лівої клавіші миші клітинку, в якій буде результат. В рядку формул натиснути кнопку з назвою *Вставлення функції*. У новому вікні обрати категорію *Статистичні (Усі)*, далі вказати за допомогою лівої клавіші миші функцію *AVERAGE* або *CPЗНАЧ* та натиснути кнопку *OK*. В вікні *Аргументи функції* напроти пункту *Значення1* встановити курсор миші та обрати необхідний діапазон значень вибірки. Ці значення вибірки можуть бути як впорядкованими, так і не впорядкованими.

б. Для обчислення значення середньої арифметичної величини в програмі *Excel Online* хмарного сервісу *OneDrive* необхідно натиснути в правій частині кнопки *Автосума*, що розташована в групі *Редагування* на вкладці *Основне*. У списку, що з'явиться, за допомогою лівої клавіші миші слід обрати кнопку *Середнє*. В комірці, що була обрана за допомогою лівої клавіші миші, перед викликом функції з'явиться запис “=AVERAGE()”. У дужках слід вказати діапазон значень вибірки.

Якщо користувач знає назву необхідної функції у хмарному сервісі *OneDrive*, він може в обраній комірці, де буде знаходитися результат обчислень ввести спочатку знак рівності, після нього ввести з допомогою клавіатури назву певної функції та знак “(”. Далі слід за допомогою курсора миші вказати необхідний діапазон значень вибірки, вписати з клавіатури знак “)” та натиснути клавішу *Enter*.

Медіана.

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

Якщо число, що позначає кількість випробуваних є парним, то числове значення *медіани* (M_e) обчислюється як половина суми двох значень x_i вибірки з порядковими номерами $n/2$ та $(n + 2)/2$, тобто медіана знаходиться за формулою:

$$M_e = \frac{x_{n/2} + x_{(n+2)/2}}{2} \quad (2).$$

Якщо число, що позначає кількість випробуваних, є не парним, то числове значення *медіани* (M_e) обчислюється за формулою:

$$N_{M_e} = \frac{n+1}{2} \quad (2.1).$$

Для обчислення значення медіани за допомогою табличного процесора *MS Excel* необхідно спочатку обрати комірку, де буде результат, а далі слід на вкладці *Основне* у групі *Редагування* натиснути правою клавішею миші в правій частині кнопки Σ . У списку, що з'явиться, обрати пункт *Інші функції*. З'явиться вікно *Вставка функції*, в якому слід обрати категорію *Статистичні*. В нижній частині вікна знайти та обрати функцію *MEDIAN* (або *МЕДИАНА*), далі вказати діапазон значень вибірки та натиснути клавішу *Enter*.

Щоб обчислити значення медіани вибірки в програмі *Excel* хмарного сервісу *OneDrive*, слід спочатку обрати комірку, де буде результат. Далі за допомогою лівої клавіші миші натиснути в правій частині кнопки *Автосума*, що розташована в групі *Редагування* на вкладці *Основне*. У списку, що з'явиться, слід обрати пункт *Додаткові функції*. З'явиться додаткове вікно *Вставка функції* (рис. 15).

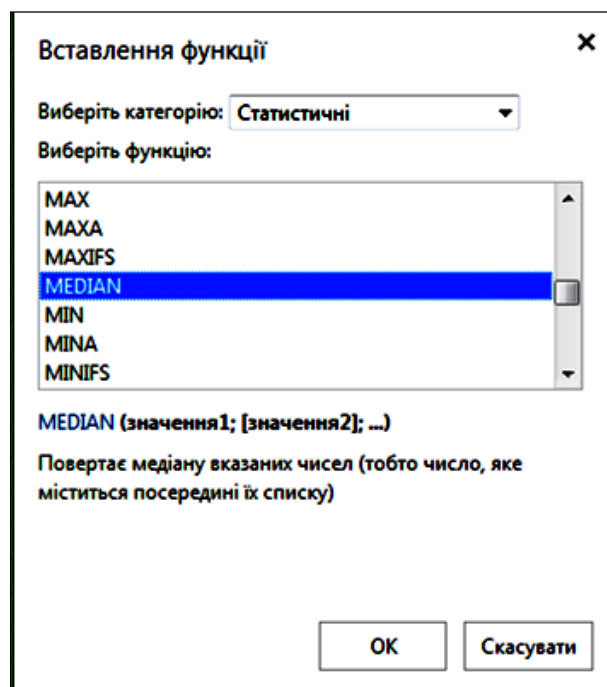


Рисунок 15. Вікно *Вставка функції* програми *Excel* хмарного сервісу *OneDrive*

Те саме вікно можна викликати якщо натиснути кнопку з назвою *Вставка функції*, що розташована зліва від рядка формул. У вікні, що після цього з'явиться, слід обрати категорію *Статистичні* (або *Усі*). У цьому ж вікні в пункті *Виберіть функцію* необхідно знайти та обрати функцію *MEDIAN*.

Мода.

Модою вибірки (M_o) називають значення варіанти, що найчастіше зустрічається у вибірці, *тобто значення* x_i з найбільшою частотою n_i .

Щоб обчислити значення *моди* в табличному процесорі *MS Excel*, необхідно обрати комірку, де буде результат, викликати вікно *Вставка функції* і в категорії *Статистичні* обрати функцію *MODE.SNGL* або *МОДА*; вказати діапазон значень вибірки, натиснути клавішу *Enter*.

Для обчислення значення *моди* в програмі *Excel Online* слід обрати комірку, де буде результат. Далі за допомогою лівої клавіші миші натиснути в правій частині кнопки *Автосума*, що розташована в групі *Редагування* на вкладці *ОСНОВНЕ*. У списку, що з'явиться, слід обрати пункт *Додаткові функції*. З'явиться додаткове вікно *Вставка функції*. У вікні знайти та обрати категорію *Статистичні* (*Усі*), а далі – функцію *MODE.SNGL*. Вказати діапазон значень вибірки та натиснути клавішу *Enter*.

Дисперсія.

Величина, за допомогою якої характеризують розсіювання або скупченість навколо центра розсіювання статистичних даних.

Це середня арифметична квадратів відхилень кожного значення ознаки від середньої величини. Цю величину ще називають *середнім квадратом відхилень*.

Дисперсія вибірки обчислюється за допомогою формули:

$$\sigma^2 = D = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n-1} = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n} = \frac{(x_1 - \bar{x})^2 \cdot n_1 + (x_2 - \bar{x})^2 \cdot n_2 + \dots + (x_k - \bar{x})^2 \cdot n_k}{n} \quad (3),$$

за умови, коли обсяг вибірки $n > 20$. Якщо обсяг вибірки $n \leq 20$, то дисперсія обчислюється за допомогою формули:

$$\sigma^2 = D = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n-1} = \frac{(x_1 - \bar{x})^2 \cdot n_1 + (x_2 - \bar{x})^2 \cdot n_2 + \dots + (x_k - \bar{x})^2 \cdot n_k}{n-1} \quad (3.1).$$

Щоб обчислити значення дисперсії вибірки у програмі *MS Excel*, слід обрати за допомогою лівої клавіші миші комірку, в якій буде

результат. Зліва від рядка формул натиснути кнопку з назвою *Вставлення функції*. У вікні, що з'явиться, обрати категорію *Статистичні (Усі)*, далі вказати функцію *VAR.S* або *ДИСП.В* (в залежності від версії пакету *MS Office*), обрати необхідний діапазон значень вибірки та натиснути кнопку *OK*.

Для обчислення дисперсії вибірки в програмі *Excel Online* спочатку слід обрати комірку, де буде результат. Далі викликати вікно *Вставлення функції*, в якому слід знайти та обрати спочатку категорію *Статистичні (Усі)*, а далі функцію *VAR.S*. Вказати діапазон значень вибірки та натиснути клавішу *Enter*.

Середнє квадратичне (стандартне) відхилення.

Середнє квадратичне (стандартне) відхилення – це величина, за допомогою якої характеризують розсіювання або скупченість навколо центра розсіювання статистичних даних.

Ця величина показує на скільки в середньому кожна варіанта вибірки відрізняється від середньої арифметичної величини цієї вибірки.

Обчислюється за допомогою формули: $\sigma = \sqrt{\sigma^2} = \sqrt{D}$ (4).

Наявність у формулі знаків + та – вказує на те, що значення цього параметра характеризує мінливість ознаки як в сторону збільшення варіант від значення середньої арифметичної величини, так і в сторону зменшення.

Значення стандартного відхилення вимірюється в тих самих величинах, що й ознака, тобто величина, щодо якої досліджують вибірку.

Властивості середнього квадратичного відхилення:

1. чим більше його значення, тим більше мінливість ознаки;
2. величина стандартного відхилення округлюється на один десятковий знак більше точності обчислення середньої арифметичної величини ознаки;
3. величина стандартного відхилення вимірюється у тих самих величинах, що й ознака;
4. величина стандартного відхилення відхиляється від значення середньої арифметичної величини як в більшу, так і в меншу сторону.

Для обчислення значення стандартного відхилення вибірки у програмі *MS Excel* слід обрати за допомогою лівої клавіші миші

комірку, в якій буде результат. В рядку формул натиснути кнопку з назвою *Вставлення функції*. У вікні обрати категорію *Статистичні (Усі)*, далі – функцію *STDEV.S* або *СТАНДОТКЛОН.В*, в залежності від версії пакету *MS Office*. Вказати необхідний діапазон значень вибірки та натиснути кнопку *OK*.

Щоб обчислити значення стандартного відхилення вибірки у програмі *Excel Online* хмарного сервісу *OneDrive*, слід обрати комірку, де буде результат, викликати вікно *Вставлення функції*, знайти та обрати категорію *Статистичні (Усі)*, а після – функцію *STDEV*. Вказати діапазон значень вибірки, натиснути з клавіатури клавішу *Enter*.

Коефіцієнт варіації.

Значення *коефіцієнта варіації* показує, яку частину середнє квадратичне відхилення складає від значення середньої арифметичної величини і характеризує степінь мінливості у відсотках.

Чим більше величина коефіцієнта варіації, тим більш мінлива, неоднорідна ознака.

В залежності від значення величини розрізняють:

- ознаки з низькою неоднорідністю (будемо вважати вибірку однорідною) за умови, що $V = C_V = 1\% - 15\%$;
- ознаки з середньою неоднорідністю (будемо вважати вибірку з середнім показником розсіювання) $V = C_V = 15,1\% - 25\%$;
- ознаки з високою неоднорідністю, розсіюванням, за умови, що $V = C_V \geq 25\%$.

Для обчислення значення коефіцієнта варіації за допомогою табличного процесора *MS Excel* необхідно створювати формулу. З цією метою спочатку обирають комірку, де буде результат, далі з допомогою клавіатури вводять знак рівності, далі вказують комірку з обчисленим значенням стандартного відхилення, вказують адресу комірки, де міститься обчислене значення середньої арифметичної величини, вводять з клавіатури вираз **100/* і натискають з клавіатури клавішу *Enter*.

Щоб обчислити значення коефіцієнта варіації в програмі *Excel Online* хмарного сервісу *OneDrive*, слід обрати комірку, де буде результат, ввести з клавіатури символ рівності, обрати комірку з обчисленим значенням стандартного відхилення, ввести адресу комірки, де міститься обчислене значення середньої арифметичної

величини, ввести з клавіатури вираз $*100\%$, натиснути з клавіатури клавішу *Enter*.

Вибірку на однорідність аналізують також за обчисленими значеннями *медіани* (M_e), *моди* (M_o) та *середньої арифметичної величини* (\bar{x}). Якщо $\bar{x} \approx M_o \approx M_e$, то група чисел може вважатися *однорідною*. При значних відмінностях між цими значеннями група чисел буде вважатися *неоднорідною*.

Для полегшення аналізу даних вибірки, отриманих в результаті експерименту, їх також представляють у вигляді графіків. Найчастіше будують такі графіки як *гістограма* та *полігон частот*.

Графік гістограма, як і графік полігон частот, будують в прямокутній декартовій системі координат (рис. 16). Він показує залежність значень варіант (або середніх значень класів) та їх частот (частот накопичення у класі).

Для побудови кожного з цих графіків на горизонтальній осі відкладають значення варіант вибірки, на вертикальній осі – відповідні значення частот.

Практичні завдання:

Нехай серед 30 студентів першого року навчання проводилось дослідження, за якого перевіряли кожного студента на максимальну кількість присідань за 30 секунд. Отримали невпорядкований ряд чисел, що записані в порядку їх надходження: 25, 24, 23, 26, 30, 25, 24, 24, 25, 21, 31, 19, 27, 26, 22, 24, 25, 28, 25, 24, 18, 29, 24, 32, 24, 25, 25, 20, 24, 28

Необхідно проаналізувати отримані значення на однорідність.

Розв'яжемо задачу за допомогою можливостей програми *Excel Online* хмарного сервісу *OneDrive*.

1. Завантажити програму браузер, у рядок адреси ввести з клавіатури <https://onedrive.live.com/about/uk-uk/> . В оновленій сторінці натиснути кнопку *Вхід*, що розташована в правій верхній частині вікна.

2. В новому вікні ввести адресу електронної пошти облікового запису, до якого необхідно увійти, натиснути кнопку *Далі*. В оновленому вікні ввести пароль до облікового запису та натиснути кнопку *Увійти*.

3. На *аркуші 1* ввести з клавіатури значення вибірки вздовж стовпчика, починаючи з комірки *A2*.

4. Впорядкувати дані вибірки в порядку зростання (побудувати варіаційний ряд).

З цією метою в комірку *A1* поставити курсор миші, в групі *Редагування* вкладки *Основне* натиснути на кнопку *Сортувати*. У списку, що з'явиться, обрати пункт *Сортувати за зростанням*. Дані вибірки, що містяться у стовпчику *A*, будуть впорядковані за зростанням (рис. 16, а).

5. В комірку з адресою *C3* ввести з клавіатури *Середня арифметична величина*, в комірку *C5* – *Мода*, в комірку *C7* – *Медіана*, в комірку *C9* – *Дисперсія*, в комірку *C11* – *Стандартне (середнє квадратичне) відхилення*.

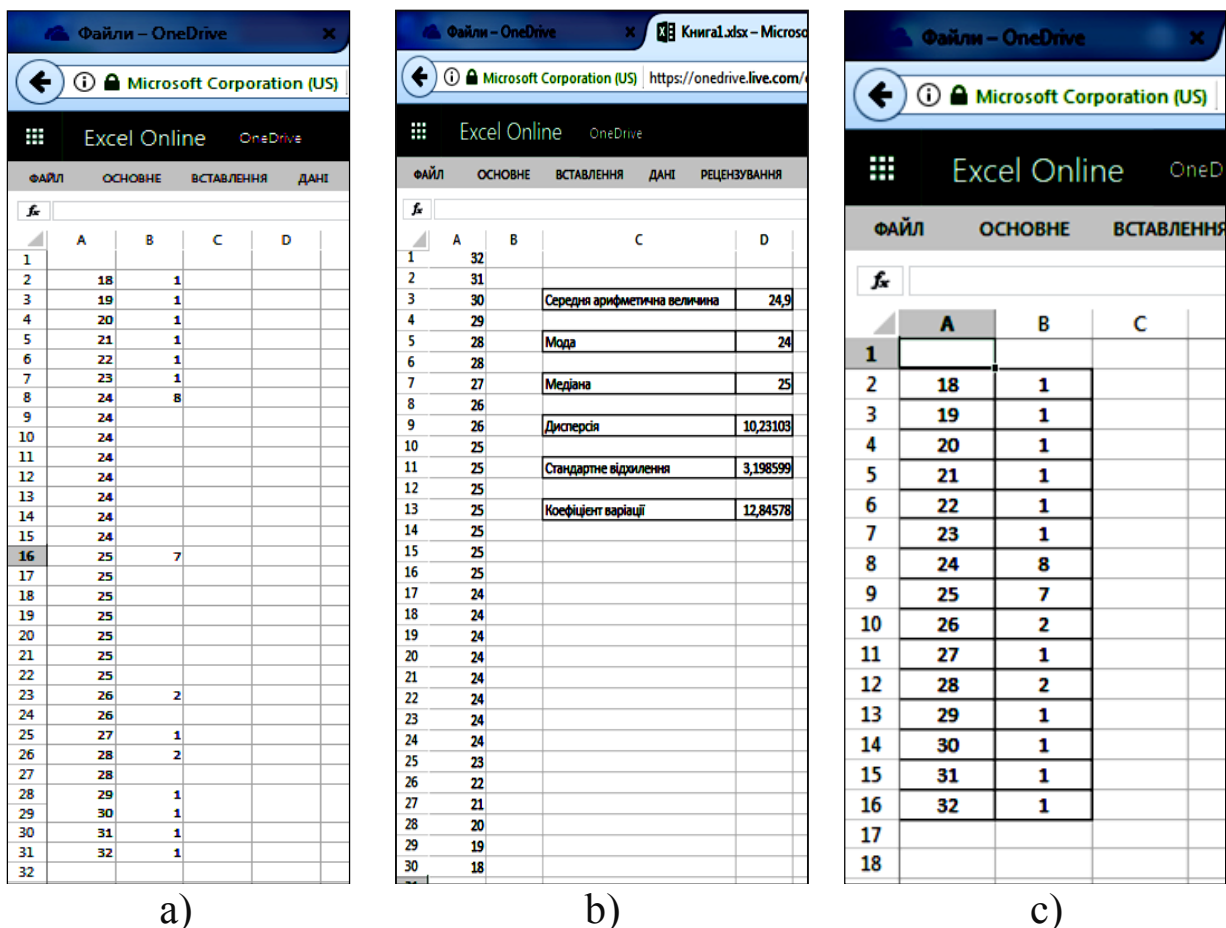


Рисунок 16. Виконання практичного завдання лабораторної роботи “Елементи описової статистики”

6. Обчислити значення статистичних параметрів вибірки у відповідних комірках стовпчика *D*.

1) Обчислити значення середньої арифметичної величини (\bar{x}).

Для цього потрібно обрати комірку з адресою *D3*, викликати вікно *Вставка функції*, в якому обрати функцію *AVERAGE* з категорії *Статистичні*. Далі за допомогою лівої клавіші миші вказати діапазон *A1:A30* та натиснути з клавіатури клавішу *Enter*. В комірці з адресою *D3* з'явиться значення середньої арифметичної величини, обчислене за допомогою програми *Excel Online* (рис. 16, b).

2) Обчислити значення моди (M_o).

Для цього необхідно обрати комірку *D5*, викликати вікно *Вставка функції*, в якому обрати функцію *MODE.SNGL* з категорії *Статистичні*. Вказати діапазон значень вибірки *A1:A30* та натиснути клавішу *Enter*. Отримаємо результат (рис. 16, b).

3) Підрахувати числове значення медіани (M_e).

З цією метою слід обрати комірку *D7*, викликати вікно *Вставка функції*, в якому обрати функцію *MEDIAN* категорії *Статистичні*. Вказати діапазон значень вибірки *A1:A30* та натиснути клавішу *Enter*. В комірці з адресою *D7* з'явиться значення медіани вибірки (рис. 16, b).

4) Обчислити значення дисперсії (D).

Аналогічно до попередніх пунктів обрати комірку, де буде результат – комірку *D9*. Викликати вікно *Вставка функції*, в якому обрати функцію *VAR.S* з категорії *Статистичні*. Далі вказати за допомогою лівої клавіші миші діапазон *A1:A30* та натиснути з клавіатури клавішу *Enter*.

5) Підрахувати значення стандартного відхилення (середнього квадратичного відхилення, що обчислюється за допомогою формули:

$$\delta = \sqrt{D}^1 \quad (5).$$

Для цього обрати комірку з адресою *D11* та, аналогічно до попередніх двох пунктів, викликати вікно *Вставка функції*, в якому обрати функцію *STDEV.S* з категорії *Статистичні*. Далі вказати за допомогою лівої клавіші миші діапазон *A1:A30* і натиснути з клавіатури клавішу *Enter*. Отримаємо результат (рис. 16, b).

¹ Значення *стандартного відхилення* також можна обчислити, використавши функцію *КОРЕНЬ*, що обрана з категорії *Статистические*.

б) Обчислити значення коефіцієнта варіації

Значення коефіцієнта варіації обчислюють за допомогою формули:

$$V = \frac{\delta}{x} \cdot 100\% \quad (6).$$

Для цього обрати комірку з адресою *D13*, натиснути з клавіатури знак рівності. Лівою клавішею миші обрати комірку з адресою *D11*, в якій міститься значення стандартного відхилення вибірки. Далі ввести за допомогою клавіатури **100/* та за допомогою лівої клавіші миші обрати комірку з адресою *D3*, в якій міститься значення середньої арифметичної величини вибірки. Натиснути з клавіатури клавішу *Enter*. В комірці з адресою *D13* з'явиться значення коефіцієнта варіації вибірки (рис. 16, б).

7. Проаналізувати отримане значення коефіцієнта варіації.

Округлимо значення коефіцієнта варіації, обчислене за допомогою програми: $V = 12.85$. Оскільки значення коефіцієнта варіації знаходиться в межах від 1% до 15%, то досліджувані ознаки вважають ознаками з низькою несталістю. Можна стверджувати, що досліджувана вибірка є однорідною.

8. Проаналізувати обчислені значення медіани (M_e), моди (M_o) та середньої арифметичної величини (\bar{x}).

Вважають, що якщо $\bar{x} \approx M_o \approx M_e$, то група чисел може вважатися однорідною. При значних відмінностях між цими значеннями група чисел буде вважатися неоднорідною.

Для нашого прикладу $\bar{x} = 24.9$, $M_o = 24$, $M_e = 24$. Враховуючи те, що значення ознаки сильно відрізняються, різницю між значенням середньої арифметичної величини, моди й медіани вибірки вважатимемо незначною. Тому зробимо висновок про однорідність ознаки, однорідність даних вибірки.

Про однорідність ознаки вказує ще й значення коефіцієнта варіації $V = 12,85\% < 15\%$.

9. **Документ з усіма змінами завантажити на пристрій.** Для цього на піктограмі документа викликати контекстне меню, в якому обрати пункт *Завантажити*. В результаті з'явиться вікно *Сохранить как*, в якому слід вказати адресу збереження цього документа на пристрої та назву з якою цей документ буде збережений.


10. На основі даних, розміщених у хмарі, побудувати графіки (гістограма та полігон частот) за допомогою табличного процесора *MS Excel*.

Для цього за допомогою лівої клавіші миші обрати діапазон даних вибірки *A2:A31*, що розміщений на вкладці *Аркуш1*. Далі на пристрої (ноутбук, комп'ютер) створити новий документ програми *MS Excel* і вставити скопійовані дані у діапазон *A2:A31*. Обчислити кількість зустрічей кожного із значень вибірки. Ввести за допомогою клавіатури обчислені значення кількості зустрічей напроти кожного із значень вибірки. Скопіювати весь діапазон значень вибірки, створити нову вкладку і вставити скопійований діапазон (рис. 16, а).

Вилучити рядки, в яких зустрічаються значення варіаційного ряду без відповідного значення частоти. Для цього в контекстному меню, що викликане на номерах відповідних рядків, обрати *Видалити*. Отримаємо варіаційний ряд у вигляді таблиці (рис. 16, с).

11. На вкладці *Аркуш 2* побудувати діаграму *Гістограма* (рис. 17).

Для цього:

1) обрати діапазон числових значень стовпця *Частота (n_i)*, на вкладці *Вставлення* у групі *Діаграми* натиснути кнопку . В результаті з'явиться вікно *Вставлення діаграми*;

2) в лівій частині вікна обрати *Стовпчаста*, в правій частині вікна у групі *Стовпчаста* обрати *Гістограма з накопиченням*, натиснути *Ок*;

3) у контекстному меню, що викликається на діаграмі, обрати *Вибрати дані*. У лівій частині вікна *Вибір джерела даних* обрати *<пустий ряд>* та натиснути кнопку *Редагувати*. З'явиться нове вікно *Редагування ряду*. Натиснути у комірці *Значення ряду* та обрати діапазон частот n_i вибірки *B2:B16*. Натиснути *Ок*;

4) у вікні *Вибір джерела даних* в пункті *Підписи горизонтальної осі (категорії)* натиснути кнопку *Редагувати*. У вікні *Підписи осі* вказати діапазон значень вибірки *A2:A16*, двічі натиснути кнопку *Ок*;

5) обрати діаграму. В результаті у головному меню вікна з'явиться додаткова вкладка *Знаряддя для діаграм*. На вкладці *Конструктор* у групі *Макети діаграм* обрати *Макет 9*;

6) **змінити назву діаграми**. Для цього правою кнопкою миші слід натиснути в тій області діаграми, де має бути назва, обрати з

контекстного меню пункт *Змінити текст*. Ввести за допомогою клавіатури слово *Гістограма*. Змінити назви осей: для горизонтальної вісі ввести “*Кількість присідань (x_i)*”, для вертикальної вісі – “*Значення частот (n_i)*”;

7) **вилучити легенду**. З цією метою обрати за допомогою лівої клавіші миші область, що має назву *Ряд1*, натиснути з клавіатури клавішу *Delete*. Той же результат можна отримати, якщо у контекстному меню, викликаному на легенді, обрати пункт *Видалити*.

12. Побудувати діаграму *Полігон частот* (рис. 17).

Для цього:

1) обрати діапазон усіх числових значень частот у таблиці (діапазон *B2:B16*) та на вкладці *Вставлення* у групі *Діаграми* натиснути кнопку *Лінійчаста*. У списку, що з’явиться, обрати *Графік із накопиченнями* та натиснути *Ok*;

2) у контекстному меню, що викликається на діаграмі, обрати *Вибрати дані*. У вікні обрати *Вибір джерела даних* обрати <пустий ряд> та натиснути кнопку *Редагувати* (з’явиться вікно *Редагування ряду*). Поставити курсор миші у пункті *Значення ряду*, далі за допомогою лівої клавіші миші вказати діапазон значень *B2:B16*. Натиснути *Ok*;

3) у вікні *Вибір джерела даних* в пункті *Підписи горизонтальної осі (категорії)* натиснути кнопку *Редагувати*. У вікні *Підписи осі* вказати діапазон *A2:A16*, натиснути двічі кнопку *Ok*;

4) обрати діаграму; в результаті в головному меню вікна з’явиться додаткова вкладка *Знаряддя для діаграм*. На вкладці *Конструктор* у групі *Макети діаграм* обрати *Макет 10*;

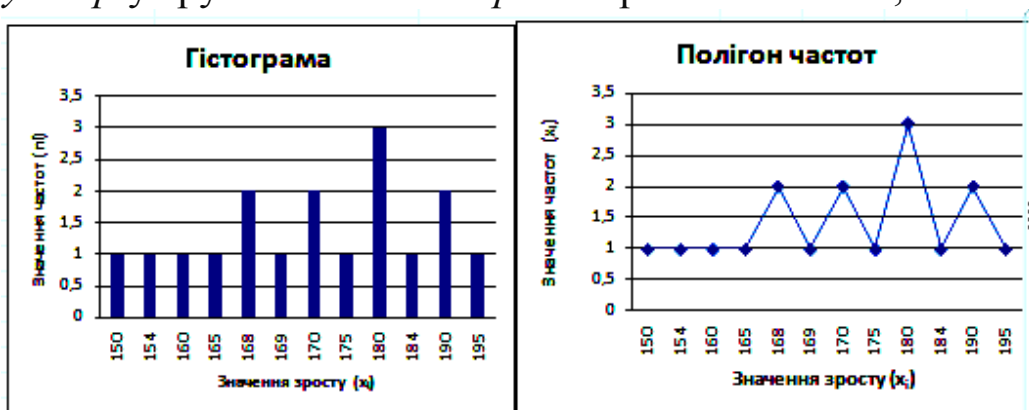


Рисунок 17. Графіки *Гістограма* та *Полігон частот*

5) **змінити назву діаграми.** Для цього лівою кнопкою миші треба двічі натиснути на назві діаграми, вилучити стару назву і ввести за допомогою клавіатури *Полігон частот*. Аналогічно змінити назви осей: горизонтальна вісь отримає назву *Кількість присідань за 30 с (x_i)*, вертикальна вісь – *Значення частот (n_i)*;

6) на підписах до горизонтальної осі викликати контекстне меню і обрати *Формат осі*. У вікні, що з'явиться, на вкладці *Вирівнювання* напроти *Напрямок тексту* обрати за допомогою лівої клавіші миші пункт *Повернути весь текст на 270 градусів*;

7) **прибрати підписи даних на графіку.** Для цього слід за допомогою лівої клавіші миші обрати діаграму. В головному меню з'явиться додаткова вкладка *Знаряддя для діаграм*. На цій вкладці обрати додаткову вкладку *Макет*. У групі *Підписи* натиснути лівою клавішею миші на кнопку *Підписи даних*. У списку, що з'явиться, натиснути *Немає*;

8) **вилучити легенду.** З цією метою слід викликати контекстне меню на елементі діаграми з назвою *Ряд1* і обрати пункт *Видалити*.

13. Зберегти документ з назвою *Diagrams.xlsx*.

14. Завантажити документ до свого акаунту хмарного сервісу *OneDrive*.

15. Оформити завдання цієї роботи у текстовому документі акаунту хмарного сервісу *OneDrive*.

Для цього створити документ у програмі *Word Online* з назвою “Розрахунково-графічна робота”. Першу сторінку оформити за зразком (рис. 18). На другій сторінці розмістити текст завдання свого номера варіанта. Далі вказати значення вашої вибірки, що впорядковані за зростанням. Скопіювати з файла *Primer.xlsx* таблицю *Варіаційний ряд*. Вказати значення середньостатистичних параметрів. Обов'язково зазначити формули, за допомогою яких ці параметри обчислюються (рис. 19). Для цього в головному меню натиснути лівою клавішею миші на пункті *Вставка* і обрати пункт меню *Формула*. З'явиться нова панель інструментів, що містить пункти *Нове рівняння*, *Грецькі літери*, *Різні дії*, *Відношення*, *Математичні дії*, *Стрілки*. Щоб змінити формулу, необхідно на неї натиснути лівою кнопкою миші. Якщо панель формул відсутня на екрані, слід обрати в головному меню

пункт *Вид* та натиснути лівою кнопкою миші *Показати панель формул*.

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ФІЗИЧНОГО ВИХОВАННЯ
І СПОРТУ УКРАЇНИ**

**Кафедра сучасних інноваційних та інформаційних технологій у
фізичній культурі і спорті**

Розрахунково-графічна робота
з дисципліни «_____»
на тему:

(тема)

Виконав студент
_____ групи

(прізвище, ім'я, по-батькові)

Київ – 2017

Рисунок 18. Приклад оформлення титульної сторінки розрахунково-графічної роботи

Середня арифметична величина	$\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{n}$	
Дисперсія	$D = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n} \text{ (або } D = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n - 1} \text{)}$	
Стандартне відхилення	$\delta = \sqrt{D}$	
Коефіцієнт варіації	$V = \frac{\delta}{\bar{x}} \cdot 100 \%$	

Рисунок 19. Заповнення таблиці обчисленими статистичними даними

16. Роздрукувати роботу, продемонструвати викладачу.
17. Вийти з акаунту сервісу *Google*.

Завдання в тестовій формі:

1) Яка функція табличного процесора *MS Excel* пакету *MS Office* дозволяє обчислити значення середньої арифметичної величини:

- a. АРИФМ.
- b. СРЕДН.
- c. СРЗНАЧ.

2) Ця функція допоможе користувачу табличного процесора *MS Excel* пакету *MS Office* обчислити значення дисперсії вибірки:

- a. ДИС.
- b. ДИСП.В.
- c. DISPERS.

3) За допомогою цієї функції у програмі *Excel* хмарного сервісу *OneDrive* не можна обчислити значення стандартного (середнього квадратичного) відхилення:

- a. КОРЕНЬ.
- b. СТАНДОТКЛОН.
- c. STANDVIB.

4) Якщо додати частоти всіх варіант вибірки отримаємо число, що позначає:

- a. Обсяг вибірки.
- b. Суму варіант.
- c. Обсяг сукупності.

5) Для обчислення дисперсії вибірки вручну обирають формулу, в якій у знаменнику дроби вираз $(n - 1)$. Це відбувається за умови, коли обсяг вибірки:

- a. $n \geq 20$.
- b. $n < 20$.
- c. $n < 30$.

6) Назва вікна, за допомогою якого користувач програми *MS Excel* може обрати функцію з певної категорії, наприклад, з категорії *статистические* та обчислити з її допомогою дані:

- a. Мастер функций - шаг 1 из 2.
- b. Функции.
- c. FUNC.

7) Якщо значення коефіцієнта варіації більше 15 %, то вибірку вважають:

- a. Однорідною.

- b. Не однорідною.
- c. Повною.

8) У випадку, коли обчислені значення середньої арифметичної величини, моди та медіани вибірки (сильно) відрізняються одне від одного, то цю вибірку вважають:

- a. Не повною.
- b. Однорідною.
- c. Не однорідною.

9) Впорядковану неспадну послідовність варіант x_i , де $i \in [1; n]$ називають

- a. Варіаційним рядом.
- b. Змінними.
- c. Статистичними параметрами.

Прикладні задачі:

Обчислити значення статистичних параметрів вибірки (середню арифметичну величину, моду, медіану, дисперсію, стандартне (середнє квадратичне) відхилення, коефіцієнт варіації) на основі результатів досліджень:

1. Спортсменів волейболістів тестували на силу удару ведучої руки (Н). Отримали такі результати:

65,0 68,2 69,0 68,3 67,2 70,0 72,2 67,2 69,4 71,2 71,3 67,4 70,2
71,5 70,6.

2. Для визначення функціональної підготовленості лижників-двоборців, спортсмени подолали 30 км на стандартній трасі. Вимірний час проходження дистанції (хв) становить:

95,0 93,5 97,5 90,0 100,5 95,0 93,5 101,0 95,0 97,5 93,5
95,0 97,5 90,0 99,5 95,0 97,5 93,5 95,0 100,5 97,5 93,5
95,0 92,5 99,5.

3. Досліджували силу двоголового м'язу стегна лижників-стрибунів (Н). Отримали наступні числові значення:

161,9 161,2 160,7 161,9 162,3 161,9 161,2 160,7 161,9 159,5
160,7 161,2 161,9 161,2 163,2 160,7 161,9 161,2 160,7 161,2.

4. Під час виконання ривка важкоатлети показали максимальну швидкість руху штанги в фазі тяги (м/с). Отримали дані:

1,44 1,48 1,54 1,56 1,49 1,48 1,52 1,56 1,64 1,52 1,49
1,52 1,54 1,52 1,49 1,54 1,52.

Завдання для систематизації та узагальнення знань.

1. Знайти в мережі *Інтернет* сервіси за допомогою яких можна перевірити вибірку на однорідність.
2. Знайти в мережі *Інтернет* сервіси, що дозволяють будувати графіки *Полігон частот* та *Гістограма*.
3. За допомогою яких програмних засобів можна перевірити вибірку на однорідність?
4. Розписати покроковий алгоритм для перевірки вибірки на однорідність.
5. Проаналізувати формулу для обчислення значення стандартного (середнього квадратичного) відхилення. Назвати кілька формул (функцій), доступних у табличному процесорі *MS Excel*, за допомогою яких можна обчислити значення цієї величини.
6. Описати алгоритм обчислення значення моди та медіани вибірки.
7. Як за допомогою функціональних можливостей табличного процесора *MS Excel* обчислити значення таких статистичних параметрів вибірки:
 - a. середньої арифметичної величини;
 - b. дисперсії;
 - c. стандартного (середнього квадратичного) відхилення;
 - d. коефіцієнта варіації;
 - e. моди;
 - f. медіани;
 - g. обсягу вибірки.
8. Як саме за допомогою можливостей табличного процесора *MS Excel* можна обчислити значення статистичних параметрів:
 - a. обсягу вибірки;
 - b. середньої арифметичної величини;
 - c. стандартного відхилення;
 - d. коефіцієнта варіації;
 - e. моди;
 - f. медіани.
9. Числові значення яких статистичних параметрів використовують для обчислення:
 - a. середньої арифметичної величини;

- b. дисперсії;
- c. стандартного відхилення;
- d. коефіцієнта варіації?

10. Як обчислити значення дисперсії вибірки?

11. В залежності від якої величини обирають формулу для обчислення значення дисперсії вибірки?

12. Як обчислити значення обсягу вибірки?

13. Які числа можуть бути варіантами вибірки?

14. Які числові значення називають частотами вибірки?

15. Яке число називається модою вибірки?

16. Як обчислити медіану вибірки?

17. В якій системі координат будують графік *гістограма*?

18. В яких одиницях вимірюється значення коефіцієнта кореляції?

19. Що впливає на вибір формули, за допомогою якої обчислюють значення дисперсії?

20. Як за допомогою функціональних можливостей табличного процесора *MS Excel* впорядкувати дані, отримані в результаті експерименту?

21. Які дії має виконати користувач щоб переглянути функції категорії *статистические*, доступні в документах табличного процесора *MS Excel*?

22. Значення яких величин беруть участь у побудові графіку:

- a. гістограма;
- b. полігон частот?

23. Розписати алгоритм обчислення значення середньої арифметичної величини вибірки.

24. Описати своїми словами як покроково обчислити значення дисперсії.

25. Назвати кілька способів обчислення значень:

- a. середньої арифметичної величини;
- b. стандартного відхилення;
- c. моди;
- d. медіани;
- e. коефіцієнта варіації.

26. Значення яких величин необхідно знати щоб обчислити значення:

- a. середньої арифметичної величини;

- b. дисперсії;
- c. стандартного (середнього квадратичного) відхилення;
- d. коефіцієнта варіації;
- e. моди;
- f. медіани.

27. Встановити відповідності між символьним позначенням статистичних параметрів та відповідними назвами функцій в табличному процесорі *MS Excel* пакету *MS Office*:

M_o	СТАНДОТКЛОН.В
δ	ДИСП.В
\bar{x}	МЕДИАНА
$D = \sigma^2$	МОДА.ОДН
M_e	СРЗНАЧ

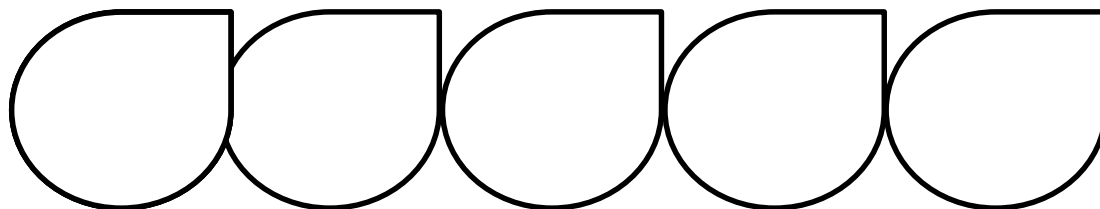
28. Встановити відповідність між назвою і позначенням статистичних параметрів вибірки:

Середня арифметична величина	M_o
Дисперсія	M_e
Середнє квадратичне (стандартне) відхилення	$D = \sigma^2$
Коефіцієнт варіації	\bar{x}
Мода	V
Медіана	σ

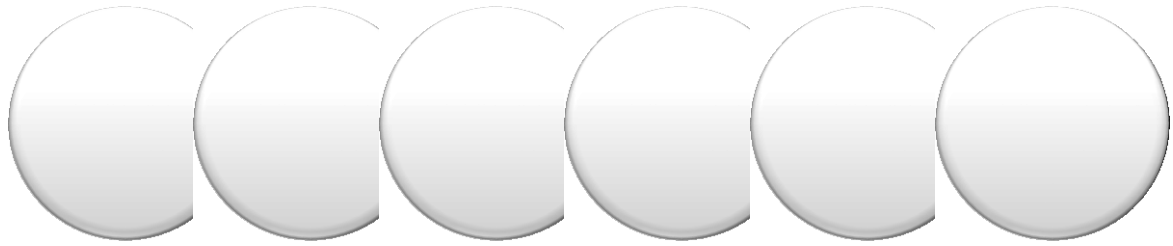
29. Встановити відповідність між символьним та текстовим позначеннями статистичних параметрів вибірки:

M_o	Значення варіанти, що найчастіше зустрічається у варіаційному ряді
M_e	Ця величина вимірюється у відсотках
$D = \delta^2$	Цю величину вважають показником розкиду (розсіювання) випадкової величини
\bar{x}	Це значення варіанти, що ділить впорядкований ряд розподілу на дві рівні частини, знаходиться у середині варіаційного ряду.
V	Ця величина знаходиться в результаті ділення числа, що позначає суму всіх варіант вибірки $\sum x_n$ на обсяг вибірки (n).

30. Знайти онлайн-сервіси, за допомогою яких можна побудувати графіки *Полігон частот* та *Гістограма*. Заповнити схему відповідними назвами веб-сервісів.



31. Знайти онлайн-сервіси, за допомогою яких можна обчислити основні параметри описової статистики.



32. Закінчити речення:

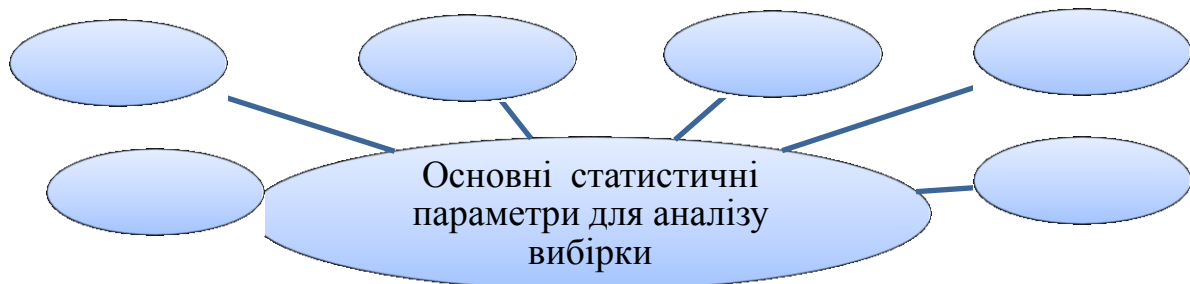
1. Числові значення цих величин використовуються для аналізу вибірки на однорідність (неоднорідність)...
2. Фактором розсіювання вважають...
3. Значення коефіцієнта варіації вимірюється у...
4. Вибірку на однорідність можна дослідити за допомогою такого програмного забезпечення:...
5. Графік гистограма показує...
6. Гистограма будується за допомогою значень...
7. Графік “Полігон частот” характеризує вибірку...
8. Для побудови графіка “Полігон частот” використовують значення...
9. Дисперсія вибірки показує...

33. Перелічити величини, що вказують на однорідність (неоднорідність) вибірки:

34. Дати відповідь на запитання:

1. За яких умов вибірку вважають однорідною?
2. В яких випадках вибірка є не однорідною?
3. Як обчислити в програмі *Excel* хмарного сервісу *OneDrive* значення стандартного (середнього квадратичного) відхилення?

35. Заповнити схему (вказати назву та їх символічне позначення):



36. Встановити відповідність між позначеннями статистичних величин та формулами, за допомогою яких вони обчислюються:

D для вибірки з обсягом до 30 елементів	$= \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$
D для вибірки з обсягом 30 елементів або більше	$= \frac{\delta}{\bar{x}} \cdot 100\%$
\bar{x} для впорядкованої вибірки	$= \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n - 1} = \frac{(x_1 - \bar{x})^2 \cdot n_1 + (x_2 - \bar{x})^2 \cdot n_2 + \dots + (x_k - \bar{x})^2 \cdot n_k}{n - 1}$
\bar{x} для неупорядкованої вибірки	$= \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{n} = \frac{(x_1 - \bar{x})^2 \cdot n_1 + (x_2 - \bar{x})^2 \cdot n_2 + \dots + (x_k - \bar{x})^2 \cdot n_k}{n}$
V	$= \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{n}$

Завдання самостійної роботи для перевірки знань та вмінь студентів щодо основ описової статистики.

В результаті проведення експерименту були отримані статистичні дані. Здійснити аналіз і опрацювання експериментальних даних за допомогою двох хмарних сервісів (на Ваш вибір). Їх пошук здійснюється у мережі *Інтернет*. Варіант завдання визначається викладачем або обирається відповідно до номера в журналі (студентського квитка, залікової книжки).

1. Ранжувати вихідні дані, знайти середньостатистичні параметри вибірки, побудувати графіки, проаналізувати вихідні дані за допомогою знайдених параметрів.

2. Записати адреси сервісів онлайн. Вписати обчислені значення параметрів, що були обчислені за допомогою першого та другого сервісу в мережі *Інтернет*:

Адреса сервісу в мережі <i>Інтернет</i> :		Адреса сервісу в мережі <i>Інтернет</i> :	
Значення середньої арифметичної величини (\bar{x})		Значення середньої арифметичної величини (\bar{x})	
Дисперсія (D)		Дисперсія (D)	
Стандартне відхилення (σ)		Стандартне відхилення (σ)	
Коефіцієнт варіації (V)		Коефіцієнт варіації (V)	
Мода (M_o)		Мода (M_o)	
Медіана (M_e)		Медіана (M_e)	
Оцініть значення коефіцієнта варіації ($V > 15\%$ або $V < 15\%$). На що це значення може вказувати?		Оцініть значення коефіцієнта варіації ($V > 15\%$ або $V < 15\%$). На що це значення може вказувати?	

Порівняйте значення:				Порівняйте значення:			
\bar{x}		M_o		M_e		M_o	

3. Перевірити значення параметрів вибірки, що Ви обчислювали за допомогою кількох сервісів для обчислення статистичних даних онлайн. Чи за однаковими формулами проводилося обчислення? Зробити висновки.

Варіант 1.

Проводилося дослідження, в якому фіксувався час (с), за який велосипедисти долали відрізок спринтерської дистанції 200 м. Отримали такі результати:

12,6 12,8 12,0 12,4 12,0 12,8 12,6 12,5 12,0 12,4 12,2 12,6 12,2
 12,2 12,2 12,4 12,4 11,8 11,9 12,0 12,4 12,6 11,9 12,6 12,4 12,8
 12,2 12,5 11,9 12,0.

Варіант 2.

25 футболістів високої кваліфікації тестували на величину максимального споживання кисню (X_i , л/хв). Отримали такі результати:

3,41 3,32 3,62 3,47 3,42 3,29 3,58 3,42 3,47 3,32 3,42 3,47 3,58
 3,32 3,42 3,58 3,25 3,29 3,62 3,42 3,47 3,32 3,41 3,29 3,47.

Ранжувати вихідні дані, знайти середньостатистичні параметри вибірки, побудувати графіки, проаналізувати вихідні дані за допомогою знайдених параметрів.

Варіант 3.

Фіксувався час (с), за який спортсменки долали дистанцію 100 м. Отримали такі результати:

11,47 12,50 11,48 11,52 11,48 11,48 11,53 11,48 11,52 11,57 11,58
 11,59 11,60 11,62 11,63 11,62 11,59 11,67 11,63 11,68 11,62 11,69
 11,68 11,67 11,64 11,65 11,65 11,67 12,67 11,56 12,60 12,68 12,56
 12,57 11,57 12,65

Варіант 4.

Фіксувався час (хв), за який спортсмени велосипедисти долали дистанцію 25 км. Отримали такі результати:

38,2 37,8 36,6 37,6 38,9 40,0 37,6 37,8 39,0 40,0 37,6 40,0 36,8
 38,5 39,0 39,0 36,6 37,8 40,5 37,6 37,8 39,0 36,6 39,0 40,0
 40,0 37,6 40,0 37,8 39,0.

Варіант 5.

Проводилося дослідження, за якого тестували спортсменів чоловіків з бар'єрного бігу на 110 м. Отримані наступні дані (час у секундах):

15,9 17,6 17,4 16,0 18,4 15,9 18,7 18,4 17,4 17,6 19,2 19,4 18,2
17,5 18,4 15,9 17,4 18,4 17,9 20,0 18,7 17,4 18,4 18,4.

Варіант 6.

Проводилося дослідження, під час якого спортсменів волейболістів тестували на силу удару ведучої руки (Н). Отримали наступні результати:

65,0 68,2 69,0 68,3 67,2 70,0 72,2 67,2 69,4 71,2 71,3 67,4 70,2
71,5 70,6.

Варіант 7.

Отримані результати запливу спортсменів дистанцію 25 м (в перехідному періоді (с)):

14,5 15,3 15,0 14,8 15,0 15,3 14,8 15,7 14,8 15,0 15,3 15,7 15,3
15,7 14,5 15,0 15,3 14,5 15,7 15,3 14,8 15,7 15,0

Варіант 8.

Проводилось дослідження, за якого 17 веслярів дослідили на темп веслування. Вимірювали кількість гребків за хвилину. Отримали такі результати:

93,5 90,0 91,5 91,5 92,0 90,5 90,0 89,5 91,5 91,0 92,5 91,0 90,5
91,5 90,0 90,0 90,5

Варіант 9.

Проводились змагання з метання ядра. Отримали наступні результати 22 легкоатлетів (м):

17,5 20,2 20,0 21,0 21,5 22,0 21,6 20,8 20,5 21,0 21,5 20,5 22,0
20,8 22,0 20,8 21,6 20,2 20,5 21,0 20,0 21,5.

Варіант 10.

Для визначення функціональної підготовленості лижників-двоборців, спортсмени подолали 30 км на стандартній трасі. Вимірний час проходження дистанції (хв) становить:

95,0 93,5 97,5 90,0 100,5 95,0 93,5 101,0 95,0 97,5
93,5 95,0 97,5 90,0 99,5 95,0 97,5 93,5 95,0 100,5 97,5
93,5 95,0 92,5 99,5.

Варіант 11.

В якості показника тренуваності обрали максимальну силу тяги плавця у воді (Н). Отримані наступні дані:

12,5 12,9 10,0 12,5 12,9 12,5 13,0 12,5 11,0 13,5 12,5 11,0 12,9
12,5 13,0 11,0 12,9 12,5 11,0 13,5.

Варіант 12.

Отримані результати запливу спортсменів дистанцію 25 м (с):

14,5 14,5 12,8 15,0 14,7 14,3 14,5 14,7 14,5 14,7 14,8 14,5 14,3
14,7 14,5.

Варіант 13.

Досліджували силу чотириглавого м'язу стегна лижників-стрибунів (Н). Отримали наступні числові значення:

196,5 199,5 193,7 196,5 195,1 199,5 196,5 193,7 196,5 200,0
195,1 196,5 199,5 200,0 196,5 199,5 196,5 195,1 199,5 196,5.

Варіант 14.

Досліджували силу двоголового м'язу стегна лижників-стрибунів (Н). Отримали наступні числові значення:

161,9 161,2 160,7 161,9 162,3 161,9 161,2 160,7 161,9 159,5
160,7 161,2 161,9 161,2 163,2 160,7 161,9 161,2 160,7 161,2.

Варіант 15.

Отримали результати запливу спортсменів на дистанції 25 м (с):

14,5 14,5 12,8 15,0 14,7 14,3 14,5 14,5 14,7 14,5 14,7 14,8
14,5 14,3 14,7.

Варіант 16.

Борців досліджували на час простої латентної реакції (м/с). Отримані наступні результати.

230 235 230 225 240 230 235 240 250 235 240 235 225
230 235.

Варіант 17.

Гандболісти виконували серію спеціальних тренувальних вправ. Після закінчення роботи частота пульсу становила:

165 188 172 175 175 180 165 172 186 175 172 175 175
165 186 180 180 186 172 180

Варіант 18.

Під час виконання ривка важкоатлети показали максимальну швидкість руху штанги в фазі тяги (м/с). Отриманні наступні дані:

1,44 1,48 1,54 1,56 1,49 1,48 1,52 1,56 1,64 1,52 1,49
1,52 1,54 1,52 1,49 1,54 1,52.

Варіант 19.

Кожного з 20 веслярів досліджували на темп веслування (кількість гребків за хвилину). Отримали числові значення:

94,0 93,5 93,0 94,0 92,5 96,0 94,5 93,5 94,0 93,0 92,5 95,0 94,0
94,0 93,5 94,0 94,5 95,0 94,5 93,5.

Варіант 20.

Час виконання вправи “сальто назад, прогнувшись” у 25 спортсменів становив (с):

1,07 1,05 1,08 1,01 0,98 1,11 1,05 0,95 1,06 1,01 1,05 0,98 1,08
1,05 1,06 1,04 1,02 1,04 0,95 1,01 1,04 1,07 1,08 0,98 1,01.

Варіант 21.

25 спортсменів виконували вправу “сальто назад, прогнувшись”. В результаті тестування на час виконання вправи у секундах, отримали числові значення:

1,06 1,04 1,08 1,01 0,98 1,12 1,04 0,95 1,06 1,01 1,04 0,9 1,08
1,04 1,01 1,04 1,01 1,04 0,95 1,01 1,04 1,06 1,08 0,98 1,01.

Варіант 22.

Тестуючи велосипедистів на час спринту в секундах (дистанція 200 м з ходу), отримали результати:

12,2 11,8 12,0 12,4 12,0 12,2 12,6 12,5 11,9 12,4 12,0 12,6
12,2 12,0 12,2 12,4 11,8 11,9 12,0 12,4 12,2 11,9 12,6 12,4
11,8 12,2 12,5 11,9.

Варіант 23.

Під час тестування велосипедів був фіксований час спринту 200 м з ходу (с):

12,2 11,8 12,0 12,4 12,0 12,2 12,6 12,5 11,9 12,4 12,1 12,6
12,2 12,0 12,2 12,4 11,8 11,9 12,0 12,4 12,2 11,9 12,6
12,4 11,8 12,2 12,5 11,9 12,0 12,4.

Варіант 24.

25 футболістів високої кваліфікації дослідили на величину максимального споживання кисню (л/хв).

3,4 3,32 3,62 3,47 3,40 3,29 3,58 3,40 3,47 3,32 3,4 3,47 3,58
3,25 3,29 3,62 3,40 3,47 3,32 3,40 3,29 3,47 3,32 3,40
3,58.

Варіант 25.

Проводилося тестування спортсменів на час (хв) подолання дистанції 15 км. Отримали наступні результати:

52,0; 51,36; 50,1; 51,5; 51,68; 51,36; 50,42; 51,5; 51,37; 51,37; 50,73;
51,5; 51,68; 50,73; 50,1; 51,36; 50,42; 51,5; 51,36; 50,42; 51,5; 50,73;
51,37; 51,68; 51,5; 51,2.

Варіант 26.

Проводилося дослідження, за якого фіксувався час (хв), за який велосипедисти долали дистанцію 25 км. Отримали такі результати:

37,2 37,8 37,6 37,6 38,9 40,0 37,6 37,8 39,0 40,0 37,6 40,0 37,8
38,5 39,0 39,0 37,6 39,0 40,0 40,0 37,6 40,0 37,8 39,0 37,6 37,8
40,5 37,6 37,8 39,0.

Варіант 27.

Проводилося тестування спортсменів чоловіків щодо бігу з бар'єрами на 110 м. В результаті отримали числові значення (с):

14,9 17,6 19,4 16,0 18,4 14,9 18,7 18,4 19,4 17,6 19,2 19,4 19,2
17,5 18,4 14,9 19,4 18,4 17,9 20,0 18,7 19,4 18,4 18,4.

Лабораторна робота № 2.

Тема: вибірковий метод.

Мета: ознайомитися з основними поняттями вибіркового методу, з алгоритмом дослідження та аналізу даних.

Сформувані вміння: за допомогою основних положень вибіркового методу формувати модельні характеристики для сукупності об'єктів, що об'єднані за спільною ознакою (ознаками).

Основні поняття: вибірковий метод, вибірка, генеральна сукупність, помилка репрезентативності, нижня довірча границя, верхня довірча границя, граничне значення, довірчий інтервал.

Теоретичний матеріал.

Під час проведення експериментів важко здійснити дослідження на основі всієї генеральної сукупності.

Генеральна сукупність – це найбільш загальна сукупність об'єктів, що об'єднана за однією ознакою. Генеральну сукупність розглядають не у всій кількості, а за допомогою її представницької (репрезентативної) частини – вибіркової сукупності (вибірки). Чим вибірка повніша, тим більш репрезентативна. Дослідник в той же час зацікавлений в мінімальному обсязі вибірки.

Основною метою використання вибіркового методу – знаходження середньої арифметичної величини ($\bar{x}_{ген}$) та стандартного (середнього квадратичного) відхилення ($\sigma_{ген}$) генеральної сукупності. Вважають, що генеральна сукупність розподілена нормально і тому ці величини повністю відображають сукупність.

Середня арифметична величина генеральної сукупності знаходиться між значеннями верхньої ($\bar{x}_{виб} + m \cdot t_{гр}$) та нижньої ($\bar{x}_{виб} - m \cdot t_{гр}$) довірчої межі:

$$\bar{x}_{виб} - m \cdot t_{гр} \leq \bar{x}_{ген} \leq \bar{x}_{виб} + m \cdot t_{гр} \quad (7)$$

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

m – помилка репрезентативності,

t – критерій надійності

Помилка репрезентативності m – величина, що характеризує різницю між значеннями генерального та вибіркового показників. Це значення вказує на відхилення параметрів вибірки, в особливості середньої арифметичної величини вибірки від середньої арифметичної величини генеральної сукупності. Значення цієї величини обчислюється за формулою:

$$m = \frac{\sigma_{\text{виб}}}{\sqrt{n-1}} \quad (8)$$

якщо обсяг вибірки $n < 20$. Якщо об'єм вибірки $n \geq 20$, то формула має вигляд:

$$m = \frac{\sigma_{\text{виб}}}{\sqrt{n}} \quad (8.1)$$

Формули правильні лише за умови, коли об'єкти у вибірці не використовуються більше одного разу. Значення критерію надійності t відповідає довірчій вірогідності й визначається з таблиці Стьюдента (додаток А до цього лабораторного практикуму).

Значення *надійності* чи *довірчої вірогідності* (P) відображає коректність репрезентативності генеральної сукупності. Значення $P = 0,95$ означає, що в 95 % випадках із загальної кількості досліджень прийнята гіпотеза буде вірною.

Величина, що доповнює значення довірчої вірогідності до одиниці, називається *рівнем значущості* (α). У спортивних дослідженнях прийнято вважати $P = 0,95$, $\alpha = 0,05$. В ексклюзивних випадках значення довірчої вірогідності прийнято вважати $P = 0,99$, $\alpha = 0,01$.

Середнє квадратичне відхилення генеральної сукупності $\sigma_{\text{ген}}$ знаходиться з нерівності:

$$\sigma_{\text{виб}} \cdot (1 - q) \leq \sigma_{\text{ген}} \leq \sigma_{\text{виб}} \cdot (1 + q) \quad (9)$$

Ця формула застосовується для всіх $q > 0$. Величина q знаходиться з таблиці (додаток Б цього лабораторного практикуму).

Якщо $q < 0$, то $\sigma_{\text{ген}}$ знаходиться з наступної нерівності:

$$0 \leq \sigma_{\text{ген}} \leq \sigma_{\text{виб}} \cdot (1 - q) \quad (10)$$

Контрольні запитання:

1. Які існують вимоги до генеральної сукупності?
2. Яка основна задача вибіркового методу?
3. За допомогою якої формули обчислюють значення середньої арифметичної величини вибірки?
4. Як обчислити значення помилки репрезентативності?
5. Яка (і) умова (и) впливає (ють) на вибір формули, за допомогою якої обчислюють значення помилки репрезентативності?
6. Як знаходиться граничне значення $t_{\text{гр}}$?
7. В якому вигляді можна представити довірчий інтервал?
8. Як знаходиться рівень значущості?

9. За допомогою якої формули обчислюють значення стандартного відхилення генеральної сукупності?

10. Що впливає на вибір формули за допомогою якої обчислюють значення середнього квадратичного відхилення генеральної сукупності?

11. Як впорядкувати дані вибірки у програмі *Google Таблиці* хмарного сервісу *Диск*?

12. Як обчислити значення середньої арифметичної величини вибірки у програмі *Google Таблиці* хмарного сервісу *Диск*?

13. Як обчислюють значення дисперсії вибірки у програмі *Google Таблиці* хмарного сервісу *Диск*?

14. Як обчислити значення помилки репрезентативності за допомогою функціональних можливостей програми *Google Таблиці*?

15. Назвати основні кроки алгоритму обчислення значення критерію надійності $t_{гр}$?

Практичні завдання.

Розв'яжемо приклад за допомогою програми *Google Таблиці* хмарного сервісу *Диск*:

x_i – це різниця в секундах між гладким бігом і бігом з бар'єрами на 110 м. Досліджені $n = 36$ спортсменів, які обрані з генеральної сукупності $N = 500$ студентів. Визначити основні середньостатистичні показники студентів-спортсменів.

2,0 2,2 1,8 2,1 1,8 2,1 1,8 1,9 2,1 1,9 1,9 2,2 1,9 2,1
2,0 1,8 2,0 2,2 2,0 1,8 2,0 2,0 2,0 2,1 2,0 2,1 2,0 2,1 1,9
1,8 2,1 2,1 2,0 2,2 1,9 2,2.

Для цього необхідно:

1) Зайти до свого акаунту сервісу компанії *Google*. Для цього у програмі браузер створити нову вкладку, в рядку адреси ввести: *gmail.com*.

2) В оновленому вікні ввести адресу своєї електронної скриньки, далі вказати свій пароль, натиснути *Ok*.

3) Створити файл "*Primer.xlsx*" у програмі *Excel*. На вкладці *Лист1*, починаючи комірки *A2* до комірок першого стовпчика, ввести за допомогою клавіатури дані вибірки.

4) **Впорядкувати дані вибірки за зростанням.** Для цього на вкладці *Лист1* у комірку з адресою *C2* ввести: " $= \text{sort}(**:**; b; c)$ ", де

замість “**:*” вказують діапазон значень вибірки, замість b – вказують номер стовпчика, дані якого слід впорядкувати за зростанням чи спаданням; замість c – вказати *TRUE*, якщо дані необхідно впорядкувати за зростанням. (Якби необхідно було впорядкувати дані у порядку спадання, слід було у записі замість c вказати *FALSE*). Натиснути клавішу *Enter*.

5) **Обчислити значення середньої арифметичної величини вибірки у програмі Google Таблиці.** Для цього необхідно обрати комірку з адресою *E3* та на панелі інструментів натиснути лівою клавішею миші справа від кнопки *Функції*. У списку, що з’явиться, необхідно обрати *AVERAGE*, вказати діапазон значень вибірки та натиснути клавішу *Enter*. Для нашого прикладу значення середньої арифметичної величини $\bar{x} \approx 2,0$.

6) **Обчислити значення дисперсії.** Для цього обрати комірку з адресою *E5* і набрати з клавіатури: “=VAR(**:*)”, де “**:*” – діапазон значень вибірки. Ми отримали значення дисперсії $\sigma^2 = 0,016$.

7) **Обчислити значення стандартного відхилення.** Для цього у програмі *Google Таблиці* обрати комірку з адресою *E7* і набрати з клавіатури: “=STDEV(**:*)”. Для прикладу значення стандартного квадратичного відхилення вийшло рівним $\sigma = 0,13$.

8) **Обчислити значення помилки репрезентативності:**

$$m = \frac{0.13}{\sqrt{36}} \cdot \sqrt{\left(1 - \frac{36}{500}\right)} \approx 0.02 \text{ с.}$$

9) **Знайти значення критерію надійності $t_{гр}$.**

Задаючи надійність $P = 0,95$ для вибірки з обсягом $n = 36$ знаходимо значення критерію надійності $t_{гр} = 2,02 \approx 2,0$. ($t_{гр}$ знаходимо з таблиці, що міститься в додатку А до цього лабораторного практикуму; для вибірки $n = 40$ (це найближче значення до $n = 36$) значення $t_{гр} = 2,02$).

10) **Вказати числовий діапазон, в якому знаходяться значення середньої арифметичної величини генеральної сукупності.**

Знаходимо в яких межах знаходиться значення середньої арифметичної величини генеральної сукупності:

$$2,0 - 0,02 \cdot 2,0 \leq \bar{x}_{ген} \leq 2,0 + 0,02 \cdot 2,0,$$

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

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

Якщо для подальших обчислень потрібне не інтервальне, а дискретне значення, то воно знаходиться як середня арифметична значень цього інтервалу або як одне з його крайніх значень.

11) Обчислити значення стандартного відхилення генеральної сукупності $\sigma_{\text{ген}}$.

За надійності $P = 0,95$ для вибірки з обсягом $n = 36$, $q = 0,26 > 0$ (q знаходимо з таблиці додатку Б до лабораторного практикуму), обчислюємо значення $\sigma_{\text{ген}}$:

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

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

За необхідності створити модельні характеристики знаходять середню арифметичну генеральної сукупності $\bar{x}_{\text{ген}}$.

Інструкції щодо виконання розрахунково-графічної роботи:

1. Виконати завдання свого номера варіанта у програмі *Excel* хмарного сервісу *Onedrive* за зразком, описаним вище. Номер варіанта завдання визначає викладач і обирає з пункту “Завдання самостійної роботи для перевірки знань та вмінь студентів щодо основ описової статистики”.

2. Оформити завдання розрахунково-графічної роботи в такій послідовності:

- a. титульна сторінка;
- b. текст завдання, що відповідає номеру варіанта в списку журналу;
- c. розв’язання завдання свого номера варіанта за алгоритмом, поданим вище.

Завдання в тестовій формі:

1. Основною метою використання вибіркового методу вважають знаходження величин:

- a. $\sigma_{\text{ген}}, \bar{x}_{\text{виб}}$.
- b. $\bar{x}_{\text{виб}}, \bar{x}_{\text{ген}}$.
- c. $\sigma_{\text{ген}}, \bar{x}_{\text{ген}}$.

2. Значення помилки репрезентативності обчислюється за допомогою формули $m = \frac{\sigma_{\text{виб}}}{\sqrt{n}}$ за умови, коли:

- a. $n > 20$.

b. $n \geq 20$.

c. $n < 20$.

3. Формула $t = \frac{\sigma_{\text{виб}}}{\sqrt{n-1}}$ використовується для обчислення помилки репрезентативності за умови, коли:

a. $n > 20$.

b. $n \geq 20$.

c. $n < 20$.

4. Якщо величина надійності (довірчої вірогідності) $P = 0,99$, то значення рівня значущості α дорівнює:

a. 0,1.

b. 0,01.

c. 1.

5. В спортивних дослідженнях прийнято вважати значення надійності P рівним:

a. 0,97.

b. 0,99.

c. 0,95.

6. Між значеннями верхньої та нижньої довірчої границь знаходиться значення:

a. $\bar{X}_{\text{ген}}$.

b. $\bar{X}_{\text{виб}}$.

c. $t_{\text{гр}}$ генеральної сукупності.

Прикладні задачі:

1. 31 спринтер на сороковому метрі 100-метрової дистанції продемонстрували середню швидкість (м/с). На основі оцінки середньої швидкості на сороковому метрі 100-метрової дистанції створити основу для модельної характеристики спринтерів тієї ж кваліфікації та аналогічного спортивного результату:

11,3; 11,7; 11,3; 11,0; 11,4; 11,8; 11,7; 11,2; 11,2; 11,3; 11,3; 11,3; 11,4; 11,4; 11,8; 11,4; 11,5; 11,8; 11,6; 11,5; 11,3; 11,6; 11,6; 11,7; 11,0; 11,5; 11,4; 11,5; 11,6; 11,2; 11,6.

2. Під час гри у настільний теніс виконується прийом топ-спін: x_i – відсоток використання цього прийому під час гри, n_i – кількість тенісистів, які використовують прийом топ-спін. Підготувати основу для модельної характеристики використання цього прийому для тенісистів аналогічної кваліфікації:

30,5; 30,2; 30,8; 37,2; 35,2; 31,0; 37,8; 33,5; 36,8; 38,1; 35,7.

3. Дзюдоїсти досліджувались на потужність роботи (ум. од). Створити основу для модельної характеристики на основі показників: 165,1; 160,5; 170,3; 167,5; 165,2; 164,5; 170,1; 172,0; 168,2; 168,4; 162,1; 165,2; 166,7; 162,0; 166,3; 165,7; 169,1; 172,0; 162,0; 165,8; 170,1; 170,3; 172,5; 168,3; 164,5; 170,8; 171,2; 172,0; 171,5; 165,2.

4. Під час проведення контрольних випробувань зі спеціальної фізичної підготовки хлопчиків у віці 7 – 9 років вимірювались результати випробовуваних у бігу на 20 метрів (с). Створити основу для модельної характеристики хлопців аналогічного віку.

3,6; 4,2; 4,1; 4,5; 3,7; 3,8; 3,5; 4,1; 4,5; 3,6; 4,2; 4,5; 4,5; 4,4; 4,5; 3,7; 3,5; 3,6; 4,2; 4,8; 4,7; 4,4; 4,9; 4,2; 4,2; 4,6; 3,9; 3,8; 4,5; 4,5.

Завдання для систематизації та узагальнення знань.

1. Закінчити твердження:

1. Найбільша загальна сукупність об'єктів, що об'єднуються за спільними(ою) ознаками (ознакою) називається...

2. Середня арифметична величина генеральної сукупності знаходиться в межах від...

3. Величина, що характеризує різницю між значеннями генерального та вибіркового показників називають..

4. Значення помилки репрезентативності знаходиться за допомогою ...

5. У випадку, коли обсяг вибірки $n \geq 20$, формула для знаходження m має вигляд...

6. Значення критерію надійності t знаходять за допомогою ...

7. Рівнем значущості називається величина, що ...

8. Середнім квадратичним відхиленням генеральної сукупності знаходиться з нерівності...

9. За допомогою вибіркового методу можна ...

10. Модельні характеристики об'єктів, об'єднаних за спільною (ими) ознакою (ами) можна знайти ...

2. Встановити відповідність між назвами статистичних параметрів вибірки та їх позначеннями:

середня арифметична величина вибірки	$\bar{x}_{\text{виб}}$
стандартне відхилення вибірки	P
значення критерію надійності	t
помилка репрезентативності	m
стандартне відхилення генеральної сукупності	$\sigma_{\text{ген}}$
надійність	$\bar{x}_{\text{ген}}$
значення ступеня свободи	$\sigma_{\text{виб}}$

3. Встановити відповідність між назвами та позначеннями статистичних параметрів вибірки:

середня арифметична величина вибірки	$\bar{x}_{\text{виб}}$
помилка репрезентативності для вибірки з обсягом $n < 20$	$\bar{x}_{\text{ген}}$

Середня арифметична
величина генеральної
сукупності

$$m = \frac{\delta_{\text{ген}}}{\sqrt{n-1}}$$

помилка
репрезентативності
для вибірки з обсягом
 $n \geq 20$

$$m = \frac{\delta_{\text{ген}}}{\sqrt{n}}$$

4. Які існують засоби за допомогою яких можна здійснити обчислення числових значень статистичних параметрів вибірки, таких як: середня арифметична величина, мода, медіана, дисперсія, стандартне (середнє квадратичне) відхилення?

5. Знайти онлайн-сервіси, за допомогою яких можна обчислити середню арифметичну величину генеральної сукупності та стандартне (середнє квадратичне) відхилення генеральної сукупності.

Лабораторна робота № 3.

Тема: *параметричні критерії.*

Мета: ознайомитися з основними правилами використання параметричних критеріїв, з алгоритмом застосування критерію Стьюдента.

Сформувати вміння: використовувати критерій Стьюдента для розв'язування практичних задач.

Основні поняття: статистична гіпотеза, нульова гіпотеза, альтернативна гіпотеза, експеримент, помилки першого роду, помилки другого роду, ймовірність помилки першого роду, ймовірність помилки другого роду, критерій, статистичний критерій, вибірка, закон нормального розподілу даних, розрахункове значення критерію, граничне значення критерію.

Теоретичний матеріал.

Щоб перемагати на змаганнях треба вміти аналізувати свої дії, відслідковувати результати діяльності. В тренувальному процесі важливо вміти аналізувати отримані результати.

Зацікавлений у покращенні результатів підопічних, тренер не може постійно використовувати одну й ту ж технологію тренувального процесу. Адже відомо, що не буде нового кращого результату, якщо виконувати одні й ті ж дії. Саме тому тренер часто стає дослідником, якому слід знати, чи стались якісь зміни в результаті проведення експерименту. Існують статистичні методи, що допомагають перевіряти припущення щодо нововведень.

Це припущення називають *статистичною гіпотезою*. Вони перевіряються за допомогою математичних методів. Статистична гіпотеза позначається за допомогою латинської літери H : (твердження).

Перед проведенням будь-якого експерименту формулюють так звані “нульову” та “альтернативну” гіпотези. *Нульовою* гіпотезою називають таку, за якою відсутні відмінності між сукупностями, що порівнюються (позначається H_0). Протилежна гіпотеза, за якою існує різниця між сукупностями, називається *альтернативною* (позначається H_1).

Експеримент може проводитися на одній і тій же групі спортсменів. Наприклад, щоб оцінити ефективність нововведень,

проводяться дослідження над групою спортсменів до застосування нової методики і через якийсь період після.

Часто перед експериментом обирають дві групи об'єктів, що об'єднані за спільними ознаками. В одній групі проводять тренування за старою методикою (контрольна група), у другій групі застосовують нову методику (експериментальна група). Через певний період часу проводять дослідження над групами, аналізують результати і роблять висновки щодо доцільності та ефективності нововведень.

Для з'ясування достовірності відмінностей між досліджуваними сукупностями спочатку формулюють нульову та альтернативну гіпотези, задають рівень значущості. Прийнято вважати, що для спортивних досліджень достатній рівень значущості $\alpha = 0,05$. Однак, у випадку, коли наслідки гіпотез пов'язані з великою відповідальністю, рекомендують обирати $\alpha = 0,01$ або $\alpha = 0,001$.

Лише після формулювання гіпотез та обрання рівня значущості, формують вибірки, за допомогою яких проводять дослідження. Якщо обчислені параметри не суперечать нульовій гіпотезі, вона зберігається. У випадку, коли різницю між результатами не вдається пояснити лише дією випадкових факторів, то нульова гіпотеза відхиляється і застосовується альтернативна.

В процесі аналізу результатів дослідження дослідник може припуститися помилки. Існують два види помилок – помилки першого та другого роду. *Помилки першого роду* виникають у випадку, коли нульова гіпотеза є вірна, однак її відхиляють. *Помилки другого роду* – нульова гіпотеза є невірною (тобто є вірною альтернативна), однак її приймають.

Ймовірність помилки першого роду позначають латинською літерою α . Іншими словами, це відомий нам рівень значущості, це малий відсоток тих випадків, що суперечать нульовій гіпотезі.

Ймовірність помилки другого роду позначають латинською літерою β і залежить вона від альтернативної гіпотези H_1 . В статистичній літературі прийнято визначати помилки першого та другого роду у вигляді:

Гіпотеза	Рішення	
	Схвалити H_0	Схвалити H_1

Справедлива H_0	Правильна з ймовірністю $1-\alpha$	Помилкова з ймовірністю α
Справедлива H_1	Помилкова з ймовірністю β	Правильна з ймовірністю $1-\beta$

Рішення щодо прийняття чи відхилення статистичної гіпотези приймають на основі статистичного критерію.

Статистичний критерій – це правило, за яким перевіряють гіпотези, приймають істинну або відхиляють хибну гіпотезу.

Алгоритм перевірки гіпотез відбувається за схемою: за вибірковими даними обчислюється значення деякої величини, що часто має назву критерій, який в свою чергу має відомий стандартний розподіл (нормальний, t-розподіл Стьюдента тощо). Знайдене значення критерію порівнюється з критичним (граничним) значенням критерію, що знаходиться з відповідної таблиці (для кожного критерію це своя таблиця). Ці значення порівнюються між собою і на основі цього виконуються висновки щодо прийняття чи відхилення гіпотези. У випадку, коли обчислене значення критерію виявилось більше граничного (критичного) значення критерію за певного значення рівня значущості α , різницю між сукупностями пояснити вже випадковостями не можна. В цьому випадку гіпотеза H_0 відхиляється на користь гіпотези H_1 за певного значення рівня значущості α і кажуть, що спостережена різниця є значуща (статистично значуща) на рівні значущості α .

На практиці за експериментальними даними прийнято обчислювати ймовірність (p) того, що критерій буде за межами значення, розрахованого для вибірок. Тобто p – експериментальний рівень значущості, яке визначається так: у випадку коли обчислене значення критерію не перевищує критичного значення (знайденого за допомогою таблиці) на рівні значущості 0,05, 0,01, 0,001, то відмінності вважають статистично недостовірними і записують: $p > 0,05$, $p > 0,01$, $p > 0,001$; якщо обчислене значення критерію більше критичного значення на рівні значущості 0,05, 0,01, 0,001, то відмінності між вибірками вважають статистично достовірними і записують: $p < 0,05$, $p < 0,01$, $p < 0,001$.

Критерії значущості – це критерії, за допомогою яких перевіряють гіпотези. Критерії значущості діляться на три класи:

1. Критерії значущості, що використовуються для перевірки гіпотез щодо параметрів розподілу генеральної сукупності (частіше для нормального розподілу). Такі критерії називаються *параметричними*.

2. Критерії, що використовуються для перевірки статистичних гіпотез і при цьому не вимагають знань щодо параметрів розподілу сукупностей. Ці критерії отримали назву – *непараметричні критерії*.

3. Критерії, що застосовують для перевірки гіпотез про згоду розподілу генеральної сукупності з раніше прийнятою теоретичною моделлю. Такі критерії називаються *критеріями згоди*.

Існує поняття *однобічного* та *двостороннього критеріїв*.

Двосторонній критерій використовують, якщо до проведення експерименту припускають, що різниця між параметрами, що порівнюються, може бути позитивною (наприклад, результати покращилися і в результаті застосування нововведень відбулися позитивні зміни) чи негативною (наприклад, після нововведень результати стали гірші). *Однобічний критерій* використовується, коли на основі додаткових даних можна припустити, що один з параметрів є більший або менший.

Критерій Стюдента використовується лише коли вибірки мають нормальний розподіл. Критерій визначає достовірність різниці між значеннями обох вибірок, що можуть мати різний обсяг.

Розглянемо на прикладі, як може бути застосований критерій Стюдента для перевірки ефективності нової методики тренувального процесу. Наприклад, маємо дві групи спортсменів одного виду спорту, певного віку та ваги. До початку експерименту обидві групи тестуються, наприклад, виконують якусь важливу вправу. Результати кожного учасника фіксуються. Так утворюються дані двох вибірок. Далі протягом певного періоду часу до однієї групи (контрольної, К) застосовують стару методику тренувального процесу, а в другій групі (експериментальній, Е) проводять заняття за новою методикою. Потім обидві групи (контрольна та експериментальна) тестуються знову, отримують значення двох нових вибірок (K_1 та E_1). Для того, щоб визначити ефективність нововведень, відмінність між вибірками (К, Е, K_1 , E_1) перевіряють на статистичну достовірність.

В нашому випадку для ефективності нової методики тренувального процесу необхідно перевірити на статистичну достовірність відмінність між вибірками E і K , E і E_1 та K і K_1 , E_1 і K_1 . Різниця між контрольною та експериментальною групами на початку експерименту не обов'язкова, має бути не достовірною. Так само відбувається з контрольною групою до і після експерименту (відмінність між показниками може бути, а може ні). Щоб нововведення були ефективні, відмінність між вибірками (E і E_1 та E_1 і K_1) має бути статистично достовірною. Тобто розрахунки повторюються чотири рази.

Алгоритм застосування критерію Стьюдента для незалежних вибірок:

1. Припустити нормальність розподілу генеральних сукупностей X та Y , сформулювати нульову (H_1) та альтернативну гіпотези (H_2).
2. Визначити рівень значущості α .
3. Обрати дві незалежні вибірки (дані цих вибірок мають бути приблизно в однакових умовах і відповідати певним характеристикам), провести вимірювання – отримати значення вибірок.
4. Обчислити значення середньої арифметичної величини, стандартного відхилення, помилки репрезентативності для обох вибірок.
5. Обчислити значення критерію Стьюдента за формулою:

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{m_x^2 + m_y^2}} \quad (11),$$

де \bar{x} , \bar{y} – значення середніх арифметичних величин обох вибірок, m_x та m_y – значення помилок репрезентативності обох вибірок.

6. Обчислити значення ступеня свободи:

$$k = n_1 + n_2 - 2 \quad (12),$$

У таблиці “Критичні точки двобічного розподілу Стьюдента (t_{2p})” за допомогою значення ступеня свободи k та рівня значущості α знайти граничне значення критерію Стьюдента $t_{кр}$ (додаток А).

7. Порівняти значення t та $t_{кр}$.

Якщо $t < t_{кр}$, то приймають нульову гіпотезу, альтернативну гіпотезу відхиляють та вважають різницю між вибірками статистично недостовірною. Тобто різниці між середніми значеннями обох вибірок відрізняються не сильно, саме тому різницю між вибірками вважають не достовірною, тобто сильної різниці немає.

Якщо $t > t_{кр}$, нульову гіпотезу відхиляють, приймають альтернативну гіпотезу і вважають різницю між вибірками статистично достовірною.

Контрольні запитання:

1. Що називається “статистичною гіпотезою”?
2. Яку гіпотезу називають “нульовою”?
3. Яка гіпотеза називається “альтернативною”?
4. Як позначають нульову та альтернативну гіпотези?
5. Коли виникає помилка першого роду?
6. За яких умов виникає помилка другого роду?
7. Як позначають ймовірність помилки першого роду?
8. Яке позначення має ймовірність помилки другого роду?
9. Що називається “статистичним критерієм”?
10. Які критерії називаються “критеріями значущості”?
11. Які критерії називаються “параметричними”, “непараметричними”, “критеріями згоди”?
12. Коли використовують однобічний критерій?
13. За яких умов використовується двосторонній критерій?

Практичні завдання:

Завдання в тестовій формі:

1. Гіпотеза, за якою відсутні відмінності між сукупностями, що порівнюються, називаються:
 - a. Альтернативною.
 - b. Подвійною.
 - c. Нульовою.
2. Гіпотеза, що протилежна альтернативній, має назву:
 - a. Однобічна.
 - b. Альтернативна.
 - c. Подвійна.
3. В основному у спортивних дослідженнях прийнято обирати рівень значущості α рівний:
 - a. 0,05.
 - b. 0,02.
 - c. 0,001.

4. В процесі аналізу результатів дослідження дослідник може припуститися помилки. Розрізняють помилки:

- a. Першого та другого роду.
- b. Першого та третього роду.
- c. Другого та третього роду.

5. Ймовірність помилки першого роду позначають латинською літерою:

- a. α .
- b. β .
- c. γ .

6. Ймовірність помилки другого роду позначають за допомогою латинської літери:

- a. α .
- b. β .
- c. γ .

7. Правило, за яким перевіряють гіпотези, приймають істинну та відхиляють хибну, називається:

- a. Однобічним критерієм.
- b. Рішенням.
- c. Статистичним критерієм.

8. Параметричні критерії, непараметричні критерії та критерії згоди – всі ці критерії є:

- a. Критеріями злагоди.
- b. Критеріями значущості.
- c. Знаковими критеріями.

Прикладні задачі:

1. У двадцятирічних борців вимірювали відносну силу кисті (x_i , Н/кг) (ньютон на 10 кг маси тіла). Аналогічні виміри проводились в іншій групі борців такого ж віку (y_i , Н/кг). Порівняти силу кисті у групах випробовуваних:

x_i : 60,1; 63,4; 65,1; 67,1; 69,0; 68,0; 60,1; 60,2; 64,3; 63,4; 65,7; 66,7; 61,2; 68,3; 65,2; 66,7; 62,5; 63,1; 65,1; 68,7; 65,1.

y_i : 64,7; 65,1; 61,2; 61,5; 65,1; 62,4; 61,3; 68,1; 65,3; 67,1; 65,1; 65,1; 65,0; 67,4; 65,2; 68,0; 65,7; 65,2; 61,5; 67,2; 68,2.

2. Дві групи школярів у віці 11 та 12 років досліджувались на відносну силу привідних м'язів рук, Н/кг. Встановити характер змін відносної сили привідних м'язів рук у цьому віці:

x_i : 0,28; 0,31, 0,35; 0,32; 0,34; 0,31; 0,34; 0,31; 0,32; 0,28; 0,29; 0,28; 0,31, 0,35; 0,32; 0,34; 0,31; 0,32; 0,34; 0,31; 0,35; 0,31; 0,35; 0,34; 0,32; 0,31; 0,34; 0,32; 0,31; 0,29; 0,31.

y_i : 0,32; 0,34; 0,31; 0,34; 0,31; 0,32; 0,28; 0,29; 0,40; 0,36; 0,34; 0,35; 0,31; 0,35; 0,34; 0,32; 0,35; 0,40; 0,38; 0,36; 0,36; 0,34; 0,31; 0,32; 0,35; 0,31; 0,39; 0,31; 0,38; 0,37; 0,37.

3. Порівняти тренувальні процеси контрольної (x_i) та експериментальної груп (x_i). Показником якості тренувань є час максимальної висоти підняття штанги під час виконання підйому штанги на грудну клітину для поштовху (с).

x_i : 0,98; 1,24; 1,25; 1,21; 1,10; 1,15; 0,99; 1,16; 1,15; 1,16; 1,15; 1,25; 0,98; 1,20.

y_i : 0,95; 0,98; 1,10; 1,15; 0,98; 1,15; 1,10; 1,15; 0,99; 0,98; 1,10; 1,15; 1,2; 1,15.

4. Контрольна та експериментальна групи досліджувались на час проходження відрізка дистанції 250 м. Порівняти результати, зробити висновки щодо ефективності нововведень.

x_i : 57,8; 58,1; 57,7; 58,1; 58,4; 57,9; 57,9; 58,0; 58,5; 58,6; 58,4; 58,7.

y_i : 57,9; 58,0; 58,5; 58,6; 56,2; 56,2; 58,0; 58,0; 58,1; 56,9; 56,7; 56,4.

Завдання для систематизації та узагальнення знань.

1. Завершити твердження:

1. Припущення під час експерименту, що перевіряють за допомогою критеріїв, називаються...

2. Перед проведенням будь-якого експерименту формулюють гіпотези...

3. Гіпотезу, за якої відсутні відмінності між вибірками, називають...

4. Протилежна гіпотеза до альтернативної має назву...

5. Помилки першого роду виникають якщо...

6. Якщо обчислені параметри не суперечать нульовій гіпотезі, то вона...

7. Ймовірність помилки другого роду залежать від...

8. Критерії значущості ділять на три класи...

9. Параметричні критерії використовують якщо ...

2. Встановити відповідність між назвами статистичних величин та їх позначеннями:

t	обсяг вибірки
n	середня арифметична величина вибірки
\bar{x}	значення критерію Стьюдента
k	помилка репрезентативності
M	ймовірність помилки другого роду
α	ймовірність помилки першого роду
V	значення ступеня свободи

3. Встановити відповідність між позначеннями статистичних величин та формулами, за допомогою яких вони обчислюються :

t	$= n_1 + n_2 - 2$
m	$= \frac{ \bar{x} - \bar{y} }{\sqrt{m_x^2 + m_y^2}}$
\bar{x}	$= \frac{\sum_{i=1}^n x_i \cdot n_i}{n}$
k	$= \frac{\delta_{\text{виб}}}{\sqrt{n}}$

4. Встановити відповідність між залежностями (між розрахунковим та граничним значеннями критерію Стьюдента) та висновками, що обирають за цих умов:

$t > t_{кр}$	різницю між вбірками вважають статистично недостовірною
$t \leq t_{кр}$	різницю між вбірками вважають статистично достовірною

1. Як використовуючи можливості мережі *Інтернет* перевірити вибірку на відповідність нормальному розподілу даних?

2. Чи існують спеціальні сервіси для перевірки вибірки на відповідність нормальному закону розподілу?

3. Чи можна за допомогою хмарних сервісів перевірити вибірки на достовірність відмінностей?

4. Які сервіси є доступними для користувачів з метою перевірки вибірок на достовірність відмінностей за допомогою критерію *Стьюдента*?

5. Чи можна обчислити за допомогою табличного процесора *MS Excel* (програм-аналогів) статистичні параметри вибірки?

6. Знайти сервіси, за допомогою яких можна знайти достовірність відмінностей між вибірками за допомогою критерію *Стьюдента*. Заповнити таблицю веб-адресами цих сервісів.

--	--	--	--	--	--	--	--	--	--

Лабораторна робота № 4.

Тема: непараметричні критерії. Критерій Вілкоксона.

Мета: ознайомитися з основними правилами використання непараметричних критеріїв, з алгоритмом застосування критерію Вілкоксона.

Сформувані вміння: використовувати критерій Вілкоксона для розв'язування практичних задач.

Основні поняття: вибірка, зв'язані вибірки, залежні (зв'язані) вибірки, критерій, закон нормального розподілу даних, порядковий номер, ранг, розрахункове значення критерію, граничне значення критерію.

Теоретичний матеріал.

Критерій Вілкоксона – це своєрідне правило, за допомогою якого аналізують дані, отримані в результаті двох експериментів (або двох замірів) з однією і тією ж групою (піддослідних, об'єктів), що досліджується. Це непараметричний статистичний критерій, що використовується для порівняння двох зв'язаних між собою вибірок, кожна з яких не потребує відповідності закону нормального розподілу даних.

Зв'язані (залежні) вибірки – це вибірки, варіанти яких є замірами одного і того ж параметра на одній і тій самій сукупності об'єктів в різний час або за різних умов. Часто залежні вибірки – це дані вимірів, що проводилися в одній і тій же групі об'єктів в різний час (наприклад, до і після дії певного фактора). Дві залежні вибірки завжди містять одну і ту саму кількість даних.

Щоб використовувати критерій Вілкоксона, кількість досліджуваних об'єктів має бути не менше п'яти і не більше п'ятдесяти.

Перед тим, як використовувати критерій Вілкоксона, прийнято висувати дві гіпотези.

Нульова гіпотеза H_0 – це твердження про відсутність ефекту внаслідок застосування нової методики (нової програми тренування, тактичні прийоми змагальної діяльності тощо). Тобто дві вибірки отримані з однієї генеральної сукупності і ефект внаслідок нововведень відсутній.

Альтернативна гіпотеза H_1 – це твердження про те, що між двома вибірками є відмінності. Вони з'явилися внаслідок застосування нової

методики (програми тренувань, тактичних прийомів змагальної діяльності тощо).

Можна формувати двосторонню (двобічну) та односторонню гіпотези.

Двостороння (двобічна) альтернативна гіпотеза вказує, що відмінності між вибірками можуть мати позитивний ефект (після нововведень результати покращилися) та негативний ефект (внаслідок нововведень результати стали гірші).

Одностороння гіпотеза використовується в одному з цих випадків (або результати покращилися, або погіршилися).

Розглянемо алгоритм застосування критерію Вілкоксона для зв'язаних вибірок. Будемо використовувати двосторонню гіпотезу.

1. Задати рівень значущості α . Для спортивних досліджень прийнято вважати $\alpha = 0,05$.

2. Не ранжуючи (не впорядковуючи значення вибірок в порядку зростання чи спадання), записати дані вибірок у стовпчик (буде два стовпчика). Відкинути пари з однаковим значенням x_i та y_i . В подальшому кількість пар зменшується на число відкинутих пар.

3. Обчислити різниці $d_i = x_i - y_i$ між представниками обох вибірок x_i та y_i .

4. Визначити та записати порядкові номери для всіх значень, що позначають різниці $d_i = x_i - y_i$. Нульові різниці не беруться до уваги.

5. Знайти ранги $|d_i|$ абсолютних значень різниць d_i .

Ранг – це число дорівнює порядковому номеру модуля різниці $(x_i - y_i)$ якщо це числове значення зустрічається один раз; якщо більше одного разу, то ранг дорівнює середній арифметичній величині порядкових номерів всіх рівних за модулем числових значень $(x_i - y_i)$.

6. Заповнити значеннями рангів стовпчики W_+ (якщо різниця $d_i = x_i - y_i$ є додатнім числовим значенням, тобто $d_i > 0$) та W_- (якщо різниця $d_i < 0$).

7. Обчислити суми рангів для чисел, що позначають додатні та від'ємні різниці d_i . Менше значення з цих сум є розрахунковим значенням критерію Вілкоксона.

8. З таблиці “Критичні значення W - критерію Вілкоксона для зв'язаних пар” (додаток В до цього практикуму) для рівня значущості

α та для фіксованої кількості пар n визначається граничне значення критерію Вілкоксона W_α .

9. Проаналізувати значення W та W_α . Якщо $W \leq W_\alpha$, то нульова гіпотеза відхиляється і різницю між вибірками вважають статистично достовірною. У випадку, коли $W > W_\alpha$, різниця між вибірками є статистично недостовірною.

У випадку, коли обсяги вибірок $n \geq 25$, можна використовувати так званий u -критерій. Значення u знаходять з формули:

$$u = \frac{|W - \frac{n \cdot (n+1)}{4}|}{\sqrt{\frac{n \cdot (n+1) \cdot (2n+1)}{24}}} \quad (13)$$

де W – це значення критерію Вілкоксона, що обчислюється за правилами, описаними вище.

Значення u , обчислене за допомогою цієї формули, порівнюється з критичним значенням u_α , що в свою чергу знаходиться з таблиці:

Критичні значення u -критерію

Рівень значущості α	Критичне значення, u_α	
	Двосторонній (двобічний) критерій	Односторонній критерій
0,05	1,96	1,64
0,01	2,58	2,33
0,001	3,28	3,09

Якщо $u \geq u_\alpha$, то нульова гіпотеза відхиляється і різницю між вибірками вважають статистично достовірною. У випадку, коли $u < u_\alpha$, приймають нульову гіпотезу і вважають, що різниця між вибірками є статистично недостовірною.

Практичні завдання.

Розглянемо, як розв'язується задача зі статистики в програмі *Excel* хмарного сервісу *OneDrive*.

На початку (x_i) та в кінці (y_i) річного циклу тренувальних занять у спринтера 20 разів вимірювалась швидкість бігу (м/с). Оцінити характер зміни швидкості бігу цього спортсмена за минулий рік.

	A	B	C	D	E	F	G
1	x_i	y_i	$x_i - y_i$	N	W	W(+)	W(-)

2	2,9	3,0	-0,1	1	7	-	7
3	3,0	2,8	0,2	14	15,5	15,5	-
4	2,9	2,9	0	-	-	-	-
5	3,1	3,2	-0,1	2	7	-	7
6	3,2	3,1	0,1	3	7	7	-
7	2,9	2,8	0,1	4	7	7	-
8	2,8	2,9	-0,1	5	7	-	7
9	3,0	3,0	0	-	-	-	-
10	3,1	3,2	-0,1	6	7	-	7
11	3,2	3,1	0,1	7	7	7	-
12	2,9	2,8	0,1	8	7	7	-
13	2,8	3,0	-0,2	15	15,5	-	15,5
14	3,1	3,2	-0,1	9	7	-	7
15	3,2	3,1	0,1	10	7	7	-
16	2,9	2,8	0,1	11	7	7	-
17	3,0	3,2	-0,2	16	15,5	-	15,5
18	2,8	2,9	-0,1	12	7	-	7
19	2,9	3,0	-0,1	13	7	-	7
20	3,1	3,1	0	-	-	-	-
21	3,2	3,0	0,2	17	15,5	15,5	-
22						73	80

Рисунок 20, Виконання завдання лабораторної роботи в програмі *Excel* сервісу *OneDrive*

1. Ввести назви стовпчиків аналогічно до діапазону *A1:G1* таблиці, що представлена на рисунку 20.

2. Задати рівень значущості $\alpha = 0,05$. До діапазону *A2:A21* ввести за допомогою клавіатури дані першої вибірки. Далі до діапазону *B2:B21* ввести дані другої вибірки. Ранжувати дані не потрібно.

3. До комірок діапазону *C2:C21* внести числові значення, що утворилися внаслідок різниці відповідного представника першої та другої вибірок ($x_i - y_i$).

Для цього в клітинці з адресою *C2* спочатку обчислити різницю між даними в клітинках *A2* та *B2*. З цією метою обрати клітинку *C2*, ввести з клавіатури знак рівності, після за допомогою лівої клавіші миші обрати клітинку з адресою *A2*, ввести з клавіатури знак "-", обрати клітинку з

адресою $B2$. Натиснути клавішу *Enter*. Поширити формулу з комірки $C2$ у діапазоні $C2:C21$.

4. Проаналізувати обчислені значення різниць в діапазоні $C2:C21$.

Якщо в результаті різниці даних вибірок $(x_i - y_i)$ утворилося число нуль, то надалі така пара чисел не розглядається.

5. До діапазону $D2:D21$ внести з допомогою клавіатури значення порядкових номерів для різниць $(x_i - y_i)$, обчислених за модулем.

Для цього, не звертаючи увагу на знаки “+” та “-” значень різниць “ $(x_i - y_i)$ ” діапазону $C2:C21$, найменшому значенню присвоїти порядковий номер “1” і розмістити це число в комірці з адресою $D2$. Наступному значенню $(x_i - y_i)$, що більше або дорівнює найменшому значенню, є таке, що знаходиться в комірці $C5$. Йому присвоюємо значення порядкового номера “2”. Аналогічно визначаються інші порядкові номери діапазону $C2:C21$ і розміщуються вони у відповідних комірках діапазону $D2:D21$.

6. Стовпчик з назвою W заповнити числовими значеннями, що позначають ранги відповідних чисел, утворених в результаті різниці представників двох вибірок.

В прикладі числа $|0,1|$ та $|-0,1|$ будемо вважати одним числовим значенням, вони зустрічаються тринадцять разів. Щоб обчислити значення рангу для кожного з цих чисел, необхідно знайти середню арифметичну величину порядкових номерів цих різниць $(x_i - y_i)$, тобто $\frac{1+2+3+4+5+6+7+8+9+10+11+12+13}{13} = 7$. Отже, ранг кожної різниці $|\pm 0,1| = 7$.

7. Заповнити значеннями рангів стовпчики W_+ та W_- . До стовпчика з назвою W_+ записати числове значення рангу в тому випадку, коли різниця $(x_i - y_i)$ є додатнім числом. До стовпчика W_- внести значення рангу за умови, коли відповідна різниця $(x_i - y_i)$ є від’ємним числом.

Розглянемо число, що міститься в комірці $C2$. Це від’ємне число, отже, значення рангу, тобто вміст комірки $E2$ потрапить не до комірки $F2$, а до комірки $G2$. Аналогічно заповнити стовпчики W_+ та W_- , тобто діапазони $F2:F21$ та $G2:G21$.

8. Обчислити розрахункове значення критерію Вілкоксона W_p .

Для цього необхідно обчислити суму чисел у стовпчиках W_+ та W_- . Менше значення з цих сум є розрахунковим значенням критерію Вілкоксона W_p .

Для нашого прикладу сума комірок діапазону $F2:F21$, тобто даних стовпчика W_+ виявилася рівною 73. Обчислюючи суму значень діапазону $G2:G21$, тобто даних стовпчика W_- , отримали число 80. Розрахунковим значенням критерію Вілкоксона будемо вважати менше із обчислених сум. Саме тому $W = 73$.

9. Знайти граничне значення критерію Вілкоксона W_α .

Для цього необхідно обчислити кількість пар x_i та y_i . Від цього значення відняти число, що позначає кількість пар, різниця між якими = "0". Відкрити таблицю "Граничні значення критерію Вілкоксона" (див. додаток В до цього лабораторного практикуму), обрати значення W_α , що відповідає числу парних спостережень (кількості пар).

Для нашого прикладу $n = 20$, з таблиці знаходимо граничне значення критерію $W_\alpha = 53$.

10. Визначити, чи достовірна відмінність між даними вибірок (до і після дії якихось факторів).

Для цього необхідно порівняти обчислене значення W і граничне значення критерію W_α .

Якщо $W > W_\alpha$, можна зробити висновок, що різниця між вибірками статистично не достовірна. Якщо $W \leq W_\alpha$, то різницю між вибірками вважають статистично достовірною.

Для нашого прикладу $W=73$, $W_\alpha=53$. Оскільки $W > W_\alpha$, різниця між вибірками статистично не достовірна. Це означає, що показники швидкості бігу спортсмена на початку та в кінці річного циклу тренувальних занять суттєво не змінилися. Отже, можна зробити висновок щодо неефективності технології тренувального процесу.

11. Оформити у хмарному сервісі (на Ваш вибір) завдання свого номера варіанта за вимогами: завдання містить титульну сторінку; у роботі представлений текст задачі номера варіанту; описаний процес розв'язування за допомогою критерію Вілкоксона, відповідно алгоритму, представленою в цій роботі; представлений аналіз результату; є висновки.

Контрольні запитання:

1. Що означає поняття "критерій"?

2. Які непараметричні критерії Вам відомі?
3. Як обчислити значення рангу під час використання критерію Вілкоксона?
4. Чи можна використовувати критерій Вілкоксона для вибірок з різними обсягами?
5. Чи потрібно перевіряти вибірки на відповідність нормальному закону розподілу перед тим, як застосовувати критерій Вілкоксона?
6. Як обчислити розрахункове значення критерію Вілкоксона?
7. Як і за яких умов знаходять граничне значення критерію Вілкоксона?
8. За якої умови різниця між вибірками є статистично недостовірною?
9. Як пояснити вираз “різниця між вибірками статистично недостовірна”?
10. За якої умови різниця між вибірками є статистично достовірною?

Завдання в тестовій формі:

1. Правило, за допомогою якого аналізують дані, отримані в результаті двох експериментів (або двох замірів) з однією і тією ж групою (піддослідних, об’єктів), що досліджується, має назву:
 - a. Гіпотеза.
 - b. Критерій Вілкоксона.
 - c. Припущення.
2. Твердження про відсутність ефекту внаслідок застосування нової методики (нової програми тренування, тактичні прийоми змагальної діяльності тощо) називається:
 - a. Нульовою гіпотезою.
 - b. Альтернативною гіпотезою.
 - c. Критерієм Вілкоксона.
3. Цей вид гіпотези вказує на відмінності між вибірками. Відмінності можуть мати позитивний ефект (після нововведень результати покращилися) та негативний ефект (внаслідок нововведень результати стали гірші). Це:
 - a. Одностороння гіпотеза.
 - b. Нульова гіпотеза.
 - c. Двобічна гіпотеза.

4. Вибірки, варіанти яких є замірами одного і того ж параметра на одній і тій самій сукупності об'єктів в різний час або за різних умов називаються:

- a. Зв'язаними (залежними).
- b. Незв'язаними.
- c. Вибірками.

5. Число дорівнює порядковому номеру модуля різниці $(x_i - y_i)$, якщо це числове значення зустрічається один раз. якщо більше одного разу, то воно дорівнює середній арифметичній величині порядкових номерів всіх рівних за модулем числових значень $(x_i - y_i)$ має назву:

- a. Критерій.
- b. Ранг.
- c. Граничне значення критерію.

6. Обрати правильне твердження. Для використання критерію Вілкоксона:

- a. Спочатку обидві вибірки перевіряють на відповідність нормальному закону розподілу.
- b. Перевіряють гіпотези.
- c. Перевіряти вибірки на відповідність нормальному закону розподілу не обов'язково.

7. Для визначення достовірності відмінності між вибірками (до і після дії якихось факторів), необхідно:

- a. Порівняти обчислене значення W і граничне значення критерію W_α .
- b. Обидві вибірки перевірити на відповідність нормальному закону розподілу;
- c. Обчислити граничне значення критерію W_α .

Прикладні задачі.

1. Для оцінки пригучості 15 гімнасток у віці 14 років виконують стрибок вгору з місця (см) з розмахом рук. Спортсменок досліджують двічі: до початку тренувального процесу x_i і після серії тренувальних занять y_i . Проаналізувати хід тренувального процесу.

x_i : 34, 36, 35, 40, 28, 30, 27, 29, 30, 34, 34, 35, 31, 30, 29;

y_i : 35, 30, 35, 29, 36, 37, 28, 35, 40, 39, 38, 29, 28, 35, 28.

2. Група важкоатлетів до застосування нової технології тренувального процесу (x_i) та після (y_i) досліджувались на

максимальне зусилля на опір, % від ваги штанги, під час виконання підйому штанги на груди для поштовху. Чи ефективна технологія тренувань була використана в тренувальному процесі?

x_i : 151, 160, 162, 157, 169, 164, 155, 153, 170, 158, 157, 160, 162, 158, 167;

y_i : 162, 158, 164, 168, 170, 153, 165, 169, 156, 153, 162, 170, 159, 155, 153.

3. Для 10 веслярів 15-річного віку була застосована нова методика тренування. Перед використанням нової методики та після неї проводились дослідження – у спортсменів виміряли наростання зусиль під час гребка. Встановити, чи відбулися суттєві зміни під час гребка, чи ефективні нововведення.

x_i : 331,1; 336,2; 335,2; 337,1; 339,1; 340,1; 346,2; 347,2; 350,1; 351,2;

y_i : 346,5; 348,1; 349,3; 349,5; 349,8; 350,2; 353,2; 355,1; 356,5; 364,1.

4. З метою покращення результатів для групи стрибунів у висоту була застосована нова програма тренування. До нововведень і після група стрибунів у висоту досліджувались на відносну статистичну силу, H . Визначити ефективність експерименту.

x_i : 3,10; 3,15; 3,17; 3,17; 3,17; 3,19; 3,20; 3,21; 3,23; 3,24; 3,25; 3,25; 3,25; 3,27; 3,28; 3,30;

y_i : 3,11; 3,12; 3,15; 3,16; 3,20; 3,20; 3,21; 3,23; 3,23; 3,25; 3,26; 3,26; 3,27; 3,28; 3,29; 3,29.

Завдання для систематизації та узагальнення знань.

1. Закінчити речення:

- a) Непараметричні критерії не вимагають...
- b) Щоб використовувати критерій Вілкоксона кількість досліджуваних об'єктів у вибірках має бути в межах від...
- c) Нульова гіпотеза – це твердження про ...
- d) Альтернативна гіпотеза – це...
- e) Вибірки називаються зв'язаними, якщо...
- f) Критерій Вілкоксона використовують для вибірок, що...
- g) Односторонню гіпотезу обирають у випадку, коли..
- h) Для спортивних досліджень прийнято вважати рівень значущості α рівним ...
- i) Ранг – це число, що...

ж) Для обчислення граничного значення критерію Вілкоксона використовують ...

к) Розрахункове значення критерію Вілкоксона визначають за допомогою ...

л) Щоб знайти граничне значення критерію Вілкоксона необхідно знати кількість пар ...

м) “Різниця між вибірками – статистично недостовірна”. Це означає, що ...

н) Перед тим, як перевіряти достовірність відмінностей між вибірками за допомогою критерію Вілкоксона ці вибірки...

о) Прикладом залежних вибірок можуть бути...

р) Твердження про відсутність відмінностей між вибірками, що досліджуються, називають...

q) Твердження про відмінність між двома вибірками, що досліджуються, має назву...

г) Чи можна за допомогою програмного забезпечення перевірити вибірки на статистичну відмінність?

с) Які існують сервіси для перевірки двох вибірок на статистичну відмінність за допомогою критерію Вілкоксона?

2. Встановити відповідність між залежностями (між розрахунковим та граничним значеннями критерію Вілкоксона) та висновками, що обирають за цих умов:

$$W > W_{\alpha}$$

нульову гіпотезу відхиляють і різницю між вибірками вважають статистично достовірною

$$W \leq W_{\alpha}$$

приймають нульову гіпотезу і різницю між вибірками вважають статистично недостовірною

Лабораторна робота № 5.

Тема: *Кореляційний аналіз. Оцінка та аналіз коефіцієнтів кореляції та детермінації.*

Мета: ознайомити студентів з основними поняттями кореляційного аналізу, навчити обчислювати та аналізувати значення коефіцієнтів кореляції та детермінації.

Сформувані вміння: обчислювати значення коефіцієнтів кореляції та детермінації за допомогою програми *Excel* хмарного сервісу *OneDrive*; аналізувати обчислені значення коефіцієнтів кореляції та детермінації.

Основні поняття: кореляційний аналіз, кореляція, коефіцієнт кореляції, парна кореляція, вибірковий коефіцієнт кореляції, множинна кореляція, прямий (позитивний) зв'язок, негативний (зворотний) зв'язок, функціональний зв'язок, коефіцієнт детермінації.

Теоретичний матеріал.

У різних сферах людської діяльності з метою внесення позитивних змін, для удосконалення існуючих технологій та процесів проводять дослідження, за яких аналізують зв'язки між досліджуваними об'єктами, вивчають закономірності і фактори, що впливають на них.

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

Для функціональних залежностей характерно, коли кожному значенню змінної величини відповідає значення іншої змінної. Такі залежності спостерігаються в математиці, фізиці.

Кореляційні або статистичні зв'язки – зв'язки, за яких чисельному значенню однієї змінної відповідає багато значень іншої змінної.

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

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

рівномірним змінам однієї ознаки відповідають нерівномірні зміни другої ознаки.

Напрямок зв'язку може бути *прямий (позитивний)* або *зворотний (негативний)*. Залежність між ознаками є *прямою (позитивною)*, якщо при збільшенні однієї ознаки друга теж збільшується або при зменшенні однієї ознаки друга теж зменшується. *Зворотною (негативною)* називається така залежність, за якої при збільшенні однієї ознаки друга зменшується або при зменшенні однієї ознаки друга збільшується.

Кореляція – це вид взаємозв'язку між ознаками.

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

Нехай зв'язок між випадковими величинами X та Y є лінійний і генеральна сукупність має двомірний нормальний розподіл. Тоді коефіцієнт кореляції r обчислюється за допомогою формули:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2] \cdot [\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (14),$$

де \bar{x} , \bar{y} – середні арифметичні величини вибірки, n – обсяг вибірки.

Для практичних вимірювань застосовують формулу:

$$r = \frac{n \sum_{i=1}^n (x_i y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{[n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2] \cdot [n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2]}} \quad (14.1).$$

Розрізняють *парну* і *множинну кореляції*. *Парна кореляція* – це зв'язок між двома показниками, один з яких є факторним, а другий – результативним. *Множинна кореляція* виникає під час взаємодії кількох факторів з результативним показником.

Коефіцієнт кореляції обчислюється лише для лінійного зв'язку. Значення коефіцієнта кореляції можуть бути в інтервалі від -1 до $+1$, тобто $-1 \leq r \leq +1$.

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

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

знаходиться в межах від -1 до 0 , то зв'язок між ознаками вважають *негативний (зворотній)*.

Чим сильніший зв'язок між ознаками, тим ближче значення коефіцієнта кореляції наближається до ± 1 . Якщо $r = \pm 1$, зв'язок між ознаками є *функціональним*, тобто кожному значенню ознаки X буде відповідати одне або кілька значень ознаки Y .

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

- якщо $0,90 \leq r \leq 0,99$, між ознаками існує *дуже високий ступінь зв'язку*;
- якщо $0,7 \leq r \leq 0,9$, то між ознаками існує *високий ступінь зв'язку*;
- якщо $0,50 \leq r \leq 0,69$, то між ознаками існує *помірний ступінь зв'язку*;
- якщо $0,2 \leq r \leq 0,49$, між ознаками існує *слабкий ступінь взаємозв'язку*.

Коефіцієнт детермінації – це числове значення, що отримують в результаті піднесення до другого степеня значення коефіцієнта кореляції $d = r^2$.

Значення *коефіцієнта детермінації* вказує на вплив загальних факторів на досліджувані ознаки.

Коефіцієнт детермінації завжди позитивний і перебуває в межах від нуля до одиниці. Він показує долю варіації результативної ознаки Y під впливом факторної ознаки X .

Контрольні запитання:

1. Що означає поняття “кореляція”?
2. Як обчислюють значення коефіцієнта кореляції?
3. За допомогою якої формули обчислюють значення коефіцієнта кореляції під час практичних вимірювань?
4. Яким символом позначають вибірковий коефіцієнт кореляції?
5. В якому числовому інтервалі можуть знаходитися значення коефіцієнта кореляції?
6. За яких умов зв'язок між ознаками є прямий (позитивний)?

7. За якої умови значення коефіцієнта кореляції вважають зворотним (негативним)?
8. В якому діапазоні повинно знаходитись значення коефіцієнта кореляції щоб ступінь зв'язку між ознаками був слабким?
9. За яких умов ступінь зв'язку між ознаками вважають високим і коли помірним?
10. Чи може значення коефіцієнта кореляції бути більшим за “1”?
11. Чи може значення коефіцієнта кореляції бути рівним “0”?
12. Що можна сказати, якщо значення коефіцієнта кореляції $r = 0$?
13. Яку назву має зв'язок між ознаками, якщо значення коефіцієнта кореляції дорівнює ± 1 ?
14. Про який зв'язок між ознаками вказує значення коефіцієнта кореляції, що знаходиться в межах від 0 до + 1 включно?
15. Як побудувати графік кореляційного поля?
16. За допомогою якого програмного забезпечення (звичайне та у хмарі) можна побудувати графік кореляційного поля?
17. Що означає поняття “коефіцієнт детермінації”?
18. Як обчислити значення коефіцієнта детермінації?
19. В якому діапазоні може бути обчислене значення коефіцієнта детермінації?
20. Чи може значення коефіцієнта детермінації бути меншим за нуль?

Практичні завдання:

Розглянемо розв'язання статистичної задачі за допомогою функціональних можливостей програми *Excel* хмарного сервісу *OneDrive*.

Визначити, чи існує зв'язок між результатами бігу школярів на 30 м та 100 м:

x_i : 4,6; 4,6; 4,7; 4,8; 4,8; 4,8; 4,9; 4,9; 4,9; 5,0.

y_i : 12,4; 21,16; 13,0; 13,3; 13,1; 13,2; 13,5; 13,5; 13,6; 13,7.


1. Завантажити програму браузер. Зайти до свого акаунту *Microsoft*.

В рядку адреси у вікні програми ввести “<https://www.microsoft.com/uk-ua/>”. Натиснути кнопку *Вхід*, ввести адресу своєї пошти та пароль.

2. Створити новий документ “Книга *Excel*”.

3. **За даними вибірок побудувати графік кореляційного поля.**

1) Для цього ввести дані обох вибірок у стовпчик (спочатку даними першої вибірки заповнити перший стовпчик, після даними другої вибірки заповнити другий стовпчик);

2) обрати діапазон значень обох вибірок; на вкладці *Вставлення* в групі *Діаграми* обрати *Точкова діаграма*. У списку, що з'явиться, обрати . В результаті програма побудує графік кореляційного поля.

3) Для нашого прикладу ми отримаємо наступний графік (рис. 20):

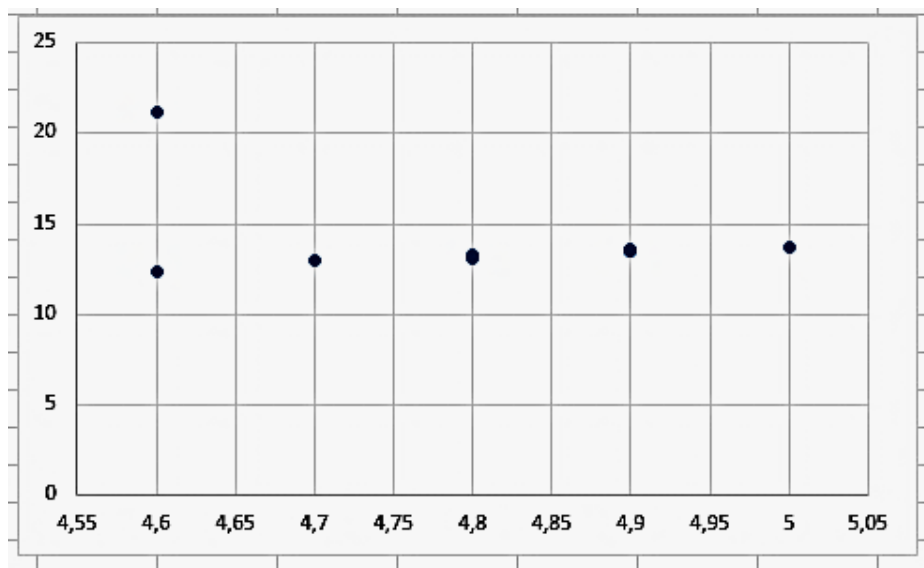


Рисунок 20. Графік кореляційного поля, побудований у програмі *Excel* сервісу *OneDrive*

4. Аналіз графіка дозволяє задати гіпотезу щодо лінійності зв'язку між результатами в бігу на 30 м та на 100 м. Силу зв'язку можна оцінити за допомогою коефіцієнта r .

5. Обчислити значення коефіцієнта кореляції r за допомогою формули (14.1):

$$r = \frac{n \sum_{i=1}^n (x_i - y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{\left[n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 \right] \cdot \left[n \sum_{i=1}^n (y_i)^2 - (\sum_{i=1}^n y_i)^2 \right]}}$$

Для цього необхідно:

1. Побудувати таблицю (рис. 21).
2. У комірки діапазону *A1:F1* ввести за допомогою клавіатури назви стовпчиків.

3. ввести числові значення ознак між якими планується знайти вид зв'язку. Для нашого прикладу введемо результати забігу на 30 м (x_i) та на 100 м (y_i).

4. в комірці з адресою $D2$ створити формулу “ $=x_i * x_i$ ”. Для цього в комірці з цією адресою ввести з клавіатури “ $=B2*B2$ ”;

5. у комірці з адресою $F2$ створити формулу “ $=x_i * y_i$ ”. Тобто формула матиме вигляд: “ $=B2*C2$ ”;

6. поширити формулу у діапазоні від комірки з адресою $D2$ до комірки з адресою $D11$. З цією метою слід піднести курсор миші до правого нижнього кута комірки з адресою $D2$. Як тільки курсор миші змінить свій вигляд на зображення хрестика, натиснути лівою клавішею миші в правому нижньому куті комірки і, не відпускаючи її, перемістити курсор до комірки з адресою $D11$;

7. аналогічно до попереднього пункту, поширити формулу від комірки $E2$ до комірки $E11$, з комірки $F2$ до комірки $F11$;

8. в комірці з адресою $B2$ обчислити значення суми за даними стовпчика B таблиці (рис. 21). Для цього слід обрати комірку $B12$, натиснути кнопку *Автосума* в групі *Редагування* вкладки *Основне*;

9. обчислити суму значень ($\sum x_i$) стовпця B таблиці, суму значень ($\sum y_i$) стовпця C таблиці, суму значень ($\sum x_i * x_i$) стовпця D таблиці, суму значень ($\sum y_i * y_i$) стовпця E таблиці, суму значень ($\sum x_i * y_i$) стовпця F таблиці.

Отже, побудувавши відповідну таблицю, ми обчислили значення проміжних сум, що входять до складу формули (2): $\sum x_i = 48,0$; $\sum y_i = 132,0$; $\sum x_i^2 = 230,56$; $\sum y_i^2 = 1743,94$; $\sum x_i y_i = 634,08$.

Підставити обчислені значення сум у формулу (14.1). Отримаємо:

$$r = \frac{n \sum_{i=1}^n (x_i - y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{[n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2] \cdot [n \sum_{i=1}^n (y_i)^2 - (\sum_{i=1}^n y_i)^2]}}$$

$$= \frac{10 \cdot 634,08 - 48,0 \cdot 132,0}{\sqrt{[10 \cdot 230,56 - 48,0^2] \cdot [10 \cdot 1743,94 - 132,0^2]}} = 0,967.$$

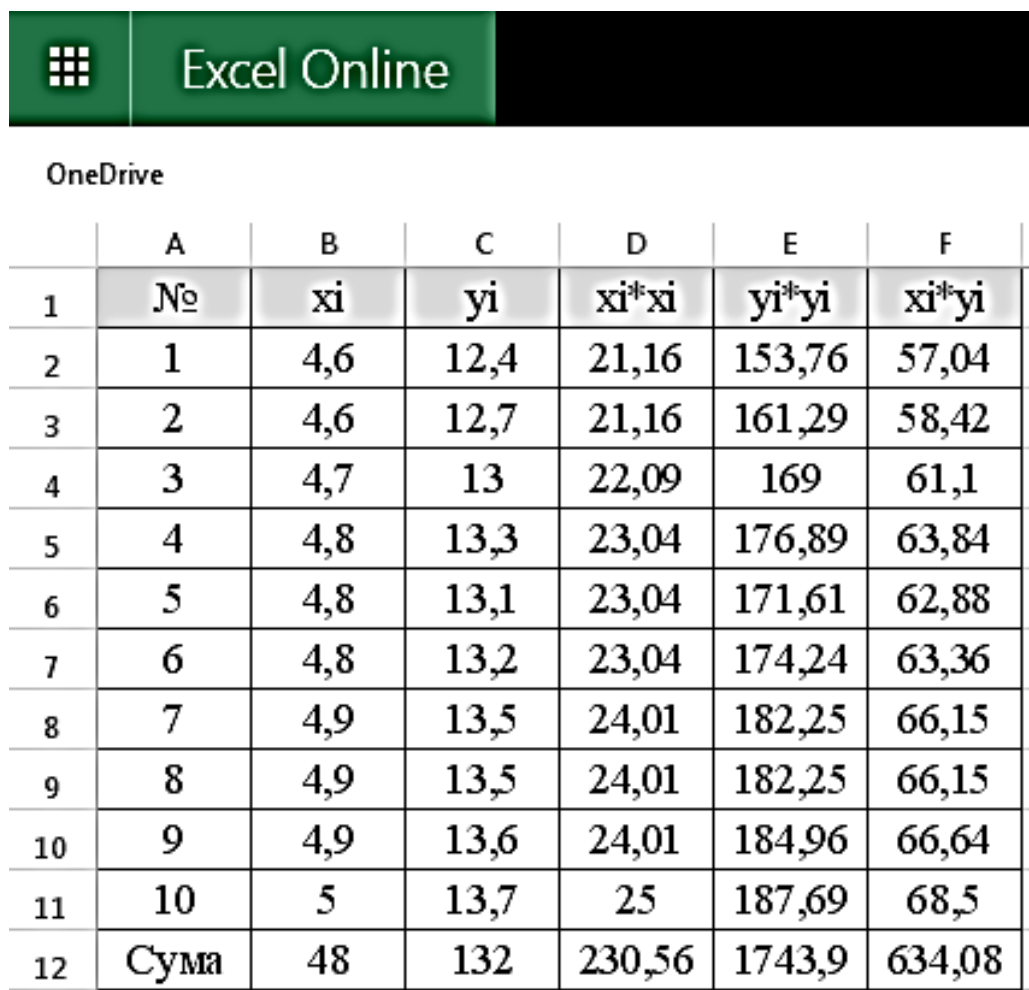
Обчислене значення коефіцієнта кореляції свідчить про сильну додатну кореляцію між результатами в бігу на 30 м та 100 м.

10. Знайти значення коефіцієнту детермінації.

Для нашого прикладу $d = 0,967^2 = 0,935$.

11. Проаналізувати обчислене значення коефіцієнта детермінації.

93,5% розсіювання (розкид) результатів у бігу на 100 м може бути пояснено зміною результатів бігу на 30 м. Іншими словами, на досліджувані ознаки (результати бігу на 30 м та на 100 м) впливають загальні фактори, що викликають варіювання цих ознак. Доля загальних факторів складає 93,5%. Інші 6,5% факторів впливають на досліджувані ознаки вибірково.



	A	B	C	D	E	F
1	№	xi	yi	xi*xi	yi*yi	xi*yi
2	1	4,6	12,4	21,16	153,76	57,04
3	2	4,6	12,7	21,16	161,29	58,42
4	3	4,7	13	22,09	169	61,1
5	4	4,8	13,3	23,04	176,89	63,84
6	5	4,8	13,1	23,04	171,61	62,88
7	6	4,8	13,2	23,04	174,24	63,36
8	7	4,9	13,5	24,01	182,25	66,15
9	8	4,9	13,5	24,01	182,25	66,15
10	9	4,9	13,6	24,01	184,96	66,64
11	10	5	13,7	25	187,69	68,5
12	Сума	48	132	230,56	1743,9	634,08

Рисунок 21. Обчислення значення коефіцієнта кореляції у програмі *Excel* хмарного сервісу *OneDrive*

12. Оформити завдання за такими вимогами:

- створити титульну сторінку;

- b. розмістити завдання свого номера варіанта (варіант прикладної задачі визначає викладач);
- c. описати процес розв'язування за алгоритмом, представленим у цій роботі;
- d. проаналізувати результат, зробити висновки (наявність висновків в кінці роботи обов'язкова).

Завдання в тестовій формі:

1. В якому числовому діапазоні може бути значення коефіцієнта кореляції (оберіть правильний варіант)?
 - a. Від -2 до $+1$.
 - b. Від 0 до $+1$.
 - c. Від -1 до $+1$.
2. Якщо обчислене значення коефіцієнта кореляції знаходиться в межах від $0,2$ до $0,49$, то ступінь зв'язку між досліджуваними ознаками вважають:
 - a. Сильним.
 - b. Слабким.
 - c. Помірним.
3. Аналіз цього графіка дозволяє задати гіпотезу щодо лінійності зв'язку між показниками, що досліджуються:
 - a. Кореляційного поля.
 - b. Кореляція.
 - c. КОРЕЛ.
4. Чи може бути значення коефіцієнта детермінації меншим за нуль?
 - a. Так, але не завжди.
 - b. Так.
 - c. Ні.
5. Зв'язок між показниками вважають *функціональним*, якщо значення коефіцієнта кореляції дорівнює:
 - a. ± 1 .
 - b. 0 .
 - c. $\pm 0,5$.
6. Зв'язок між величинами вважають відсутнім, якщо значення коефіцієнта кореляції дорівнює:
 - a. 0 .

- b. ± 1 .
- c. $- 1$.

Прикладні задачі:

1. Визначити, чи існує зв'язок між ростом важкоатлетів x_i (см) та максимальною швидкістю руху штанги у фазі тяги під час виконання ривка y_i (м/с) за інших аналогічних умов.

x_i : 180; 165; 181; 167; 169; 180; 167; 168; 182; 167; 166; 180; 181; 182; 169.

y_i : 1,32; 1,33; 1,38; 1,37; 1,38; 1,37; 1,36; 1,37; 1,40; 1,33; 1,35; 1,36; 1,33; 1,37; 1,38.

2. Перевірити, чи існує зв'язок між кількістю кидків за 10 хвилин з різними суперниками (x_i) та частотою серцевих скорочень у спортсменів (y_i , ударів за хвилину):

x_i : 5; 5; 7; 9; 8; 7; 8; 9; 8; 5; 6; 7; 9; 5; 7.

y_i : 160; 155; 160; 180; 185; 170; 170; 180; 185; 155; 160; 160; 185; 155; 160.

3. Для оцінки техніки захисту 10 боксерів застосували тест: під удар знизу лівою під тулуб підставляється зігнута в лікті права рука. Експерт проводить оцінку виконання цього тесту у балах: 5; 4; 3; 2 у порядку зниження якості виконуваного тесту. В якості критерію тестування береться кількість виграних сутичок y_i , коли у відповідь на удар вниз лівою по тулубу для захисту підставляють зігнуту в лікті праву руку. Встановити інформативність тесту.

x_i : 5; 4; 5; 4; 4; 5; 3; 3; 3; 5; 5; 5; 5; 4; 3; 5; 5;

y_i : 7; 3; 4; 2; 3; 4; 2; 3; 2; 5; 7; 7; 5; 2; 2; 1; 4.

4. Плавці вільним стилем на дистанції 100 метрів тестувались на максимальну інтенсивність плавання за 30 секунд (x_i , м/с). Ці результати необхідно порівняти в плаванні вільним стилем на дистанції 100 м (y_i , м/с). Перевірити на інформативність тестове завдання.

x_i : 1,0; 1,1; 1,2; 1,5; 1,1; 1,4; 1,2; 1,1; 1,0; 1,5;

y_i : 58,0; 59,0; 62,0; 68,0; 65,0; 64,0; 59,0; 58,0; 60,0; 65,0.

5. Вам необхідно встановити взаємозв'язок між результатами в бігу п'яти велосипедистів на дистанції 60 м (9,1; 9,4; 9,2; 9,5; 9,5) та швидкістю їх їзди на велосипеді, км/год (18,2; 18,4; 17,6; 17,5; 16,5). Який метод статистики Ви використаєте? Чи будете Ви

використовувати програмне забезпечення? Якщо так, то конкретизувати та обґрунтувати Ваш вибір.

6. Уявіть ситуацію: Ви вже тренер. Для групи спортсменів необхідно підібрати якомога ефективнішу програму тренувань. Ви розумієте, що якщо використовувати стару програму, нового результату не буде. Тому Ви вирішуєте оновити програму тренувань, що містила б вправи для одночасного тренування кількох показників, пов'язаних між собою і посилення одного підвищувало показники другого. Який статистичний метод та яке програмне забезпечення Ви будете використовувати для створення необхідного комплексу вправ?

Завдання для систематизації та узагальнення знань.

1. Чи існують хмарні сервіси за допомогою яких можна обчислити значення коефіцієнта кореляції? За допомогою яких хмарних сервісів це можна зробити?

2. За допомогою якого програмного забезпечення можна обчислити значення коефіцієнта кореляції?

3. Чи можна обчислити значення коефіцієнта конкордації онлайн? Перелічити хмарні сервіси, якщо такі існують?

4. За допомогою яких онлайн-сервісів можна обчислити значення коефіцієнта кореляції двох вибірок.

5. Навести приклади сервісів, що обчислюють значення коефіцієнта кореляції та аналізують отримане значення.

6. Записати формулу для обчислення значення коефіцієнта кореляції та описати з яких компонентів складається формула.

7. Описати основні кроки користувача для обчислення значення коефіцієнта кореляції за допомогою програми *Excel* хмарного сервісу *Диск*.

8. Як за обчисленим значенням коефіцієнта кореляції можна зробити висновок щодо виду зв'язку між ознаками або його відсутності?

9. Що вказує на відсутність зв'язку між досліджуваними явищами?

10. Закінчити речення:

а) Розрізняють такі види кореляції...

б) Значення коефіцієнта кореляції можуть знаходитися в інтервалі...

- с) Зв'язок між досліджуваними ознаками вважають додатнім (прямим) за умови ...
- д) Від'ємним (оберненим) вважають зв'язок між ознаками якщо...
- е) Значення коефіцієнта кореляції вказує на...
- ф) Значення коефіцієнта детермінації завжди...
- г) Зв'язок між ознаками називають функціональним якщо...
- h) Зв'язок між досліджуваними явищами вважають середнім якщо значення коефіцієнта кореляції знаходиться в інтервалі...
- і) Якщо значення коефіцієнта кореляції знаходиться в діапазоні $[0,5; 0,69]$, то зв'язок між ознаками вважають...
- j) Якщо значення коефіцієнта кореляції знаходиться в числових межах від 0,7 до 0,9, то зв'язок між ознаками є...
- к) Про відсутність зв'язку між ознаками вказує значення...
- l) Чим більшим є числове значення коефіцієнта кореляції, тим зв'язок між ознаками вважають...
- т) Якщо числове значення коефіцієнта кореляції наближається до нуля, зв'язок між ознаками стає...

11. З'єднати відповідності:

$0,2 \leq r \leq 0,49$	<i>зв'язок між ознаками середній</i>
$0,7 \leq r \leq 0,9$	<i>зв'язок між ознаками сильний</i>
$0,5 \leq r \leq 0,69$	<i>зв'язок між ознаками слабкий</i>

12. Пояснити чи може бути значення коефіцієнта детермінації:

- а) дорівнювати нулю;
 б) бути меншим за нуль?

За яких умов це може відбутися?

13. Заповнити схему “Статистичні функції, доступні у програмах”:

Табличний процесор <i>MS Excel</i>	Програма <i>Google Документи</i> хмарного сервісу <i>Диск</i>	Програма <i>Excel Документи</i> хмарного сервісу <i>OneDrive</i>

ДОДАТКИ

Додаток А

Критичні точки двобічного розподілу Стьюдента (t_{α})

N	Рівень значущості α				N	Рівень значущості α			
	0,1	0,05	0,01	0,001		0,1	0,05	0,01	0,001
1	6,314	12,706	63,657	636,61	21	1,721	2,080	2,831	3,819
2	2,920	4,308	9,925	31,599	22	1,717	2,074	2,819	3,792
3	2,353	3,182	5,841	12,924	23	1,714	2,069	2,807	3,768
4	2,132	2,776	4,604	8,610	24	1,711	2,064	2,797	3,745
5	1,943	2,447	3,707	5,959	25	1,708	2,060	2,787	3,725
6	1,943	2,447	3,707	5,959	26	1,706	2,056	2,779	3,707
7	1,895	2,365	3,499	5,408	27	1,703	2,052	2,771	3,690
8	1,860	2,306	3,355	5,041	28	1,701	2,048	2,763	3,674
9	1,833	2,262	3,250	4,781	29	1,699	2,045	2,756	3,659
10	1,812	2,228	3,169	4,587	30	1,697	2,042	2,750	3,646
11	1,795	2,201	3,106	4,437	40	1,684	2,021	2,704	3,551
12	1,782	2,179	3,055	4,318	50	1,676	2,009	2,678	3,505
13	1,771	2,160	3,012	4,221	60	1,664	2,000	2,660	3,505
14	1,761	2,145	2,977	4,140	80	1,664	1,990	2,639	3,416
15	1,753	2,131	2,947	4,073	100	1,660	1,984	2,626	3,391
16	1,746	2,120	2,921	4,015	120	1,658	1,980	2,617	3,373
17	1,740	2,110	2,898	3,965	200	1,653	1,972	2,601	3,340
18	1,734	2,101	2,878	3,922	500	1,648	1,965	2,586	3,310
19	1,729	2,093	2,861	3,883	∞	1,645	1,960	2,580	3,291
20	1,725	2,086	2,845	3,850					

Додаток Б

Значення q для визначення довірчих меж 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

Додаток В

Граничні значення критерію Вілкоксона ($W_{гр}$) для зв'язаних пар (n – кількість парних спостережень)

n	α		n	α		n	α	
	0,05	0,01		0,05	0,01		0,05	0,01
6	1		13	18	11	20	53	39
7	3		14	22	14	21	60	44
8	5	1	15	26	17	22	67	50
9	7	3	16	31	21	23	74	56
10	9	4	17	36	24	24	82	62
11	12	6	18	41	29	25	90	69
12	15	8	19	47	33			

Список літературних джерел:

1. Англо-русский словарь по информационным технологиям / [С. Б. Орлов]. – М. : РадиоСофт, 2011. – 640 с.
2. Білак Ю.Ю. Інформатика та інформаційні технології: практикум. Ч.1 / Ю.Ю. Білак, В.О. Лавер, Ю.В. Андрашко, І.М. Лях. – Ужгород: ПП «АУТДОР - ШАРК», 2015. – 96 с.
3. Буйницька О.П. Інформаційні технології та технічні засоби навчання / О. П. Буйницька. – К. : Центр учбової літератури, 2012. – 240 с.
4. Воскобойников Ю. Е. Теория вероятностей и математическая статистика (с примерами в Excel) / Ю. Е. Воскобойников, Т. Т. Баланчук. – Новосибирск, 2013. – 200 с.
5. Гришин В. Н. Информационные технологии в профессиональной деятельности : учебник / В. Н. Гришин, Панфилова Е. Е. – М. : Форум, Инфра-М, 2007. – 416 с.
6. Громов Ю.Ю. Информационные технологи / Ю. Ю. Громов, В. Е. Дидрих [и др.]. – Тамбов :ТГТУ, 2011. – 152 с.
7. Грошев А.С. Информатика : лабораторный практикум / А. С. Грошев. – Архангельск : Арханг.гос.техн.ун-т, 2012. – 145 с.
8. Гуда А. Н. Информатика. Общий курс : учебник / А. Н. Гуда, М. А. Бутакова, Н. М. Нечитайло, А. В. Чернов ; под общ. ред. В. И. Колесникова. – [4-е изд.]. – М. : Издательско-торговая корпорация Дашков и К, 2011. – 399 с.
9. Денисова Л. В. Измерения и методы математической статистики в физическом воспитании и спорте : учеб. пособие для вузов / Л. В. Денисова, И. В. Хмельницкая, Л. А. Харченко. – К. : Олимп. лит., 2008. – 127 с.
10. Дибкова Л. М. Информатика та комп'ютерна техніка : навч. посіб. / Л. М. Дибкова. – [3-тє вид., доп.]. – К. : Академ-видав, 2011. – 464 с.
11. Ершова Е. Е. Лабораторный практикум по современным компьютерным технологиям : учеб. пособие. ч.2 Excel / Е. Е. Ершова. – Новосибирск : НГАСУ (Сибстрин), 2007. – 80 с.
12. Ефимова О. Курс компьютерной технологии / О. Ефимова, В. Морозов, Ю. Шафрин. – М. : АБФ, 2010. – 656 с.

13. Жалдак М.І. Теорія ймовірностей і математична статистика: Підручник для студентів фізико-математичних спеціальностей педагогічних університетів. – Вид. 2, перероб. і доп. / М.І. Жалдак, Н.М. Кузьміна, Г.О. Михалін. – Полтава: «Довкілля-К», 2009. – 500 с.
14. Ильина Л. С. Основы обработки экономической информации в Excel 2007: метод. указания и задания для выполнения лаб. работ для студ. направлений Прикладная информатика и экономика / Л. С. Ильина, Н. Ю. Прокопенко. – Н. Новгород : ННГАСУ, 2010. – 135 с.
15. Іванов В. Г. Основи інформатики та обчислювальної техніки : підручник / В. Г. Іванов, В. В. Карасюк, М. В. Гвозденко ; за заг. ред. В. Г. Іванова. – Х. : Право, 2012. – 312 с.
16. Інформаційні системи і технології : навч. посіб. по базовій підготовці студ. рівня бакалавр і спеціаліст ден. і заоч. форм навч. / М-во освіти і науки, молоді та спорту України; Донец. нац. ун-т економіки і торгівлі ім. М. Туган-Барановського; Кафедра інформ. систем і технологій упр. / [Н. М. Спіцина, Т. В. Шабельник, С. В. Бондаренко]. – Донецьк : ДонНУЕТ, 2011. – 290 с.
17. Кашуба В.О. Комп'ютерна техніка та методи математичної статистики: навч. посібник / В.О. Кашуба, Л.В. Денисова, В.В. Усиченко, Л.А. Харченко, Ю.Л. Хлевна, В.П. Вишневецька, М.В. Караватська, А.М. Бойко – 2014. – 213 с.
18. Косинський В. І. Сучасні інформаційні технології : навч. посіб. для студ. ВНЗ / В. І. Косинський, О. Ф. Швець. – [2-ге вид., випр.]. – К. : Знання, 2012. – 318 с.
19. Костенко А. Б. Конспект лекцій по курсам Информатика, Информатика и компьютерная техника / А. Б. Костенко, Н. О. Манакова, Е. В. Кузьмичева. – Х. : Харьк. нац. акад. гор. хоз-ва, 2011. – 94 с.
20. Кузнецов В. М. Теорія ймовірностей [Текст]: методичні вказівки до виконання модульної роботи № 7: у 2 ч. / уклад.: В. М. Кузнецов, Т. М. Бусарова, О. В. Звонарьова, Т. А. Агошкова; Дніпропетр. нац. ун-т залізн. трансп. ім. акад. В. Лазаряна. – Д.: Вид-во Дніпропетр. нац. ун-ту залізн. трасп. ім. акад. В. Лазаряна, 2013. – Ч. 2. Випадкові величини. 49 с.

21. Кукушкина О. И. Использование информационных технологий в различных областях специального образования : дис. доктора пед. Наук : 13.00.03 / Кукушкина Ольга Ильинична. – М., 2005. – 381 с.
22. Макарова Н. В. Информатика : учеб. пособие / Н. В. Макарова, В. Б. Волков. – СПб.: Питер, 2011. – 576 с.
23. Немцова Т. И. Практикум по информатике : учеб. пособие / Т. И. Немцова, Ю. В. Назарова. – М. : ИНФРА-М, 2012. – Т. 1. – 320 с.
24. Новожилов О. П. Информатика : учеб. пособие / О. П. Новожилов. – М. : Юрайт, 2012. – 564 с.
25. Підгорна Т.В. Особливості формування системи інформатичних компетентностей майбутніх вчителів природничо-математичних дисциплін / Т.В. Підгорна // Науковий часопис НПУ імені М.П. Драгоманова. Серія 2: Комп'ютерно-орієнтовані системи навчання. – 2014. – № 14. – С. 65-70.
26. Подгорная Т.В. Формирование практических умений учащихся поиска химических сведений в сети Интернет / ТВ Подгорная // Комп'ютер у школі та сім'ї. – 2016. – № 7. С. 25 – 27.
27. Статистика. Обработка спортивных данных на компьютере : учеб. пособие для студ. выс. учеб. заведений физ. культуры / [под ред. М. П. Шестакова и Г. И. Попова] – М. : СпортАкадемПресс, 2002. – 278 с.
28. Степанов А. Н. Информатика. Базовый курс : для студ. гуманит. спец. вузов / А. Н. Степанов. – [6-е изд.]. – СПб. : Питер, 2012. – 719 с.
29. Федотова Е. Л. Информатика : курс лекций / Е. Л. Федотова, А. А. Федотов. – М. : Форум, 2012. – 479 с.
30. Швабе Р. В. Microsoft Word 2007 – это просто! / Р. В. Швабе. – М. : НТ Пресс, 2014. – 400 с.
31. Freund Rudolf J., Wilson William J., Sa Ping. Regression Analysis : Statistical Modeling of a Response Variable 2nd Edition. – Elsevier Inc., 2006. – 480 pages.
32. Milton M. Head First Excel / Michael Milton. – O'Reilly, 2010. – 440 pages.

У лабораторному практикумі подано теоретичні основи математичної статистики, на прикладі деяких хмаро орієнтованих технологій показаний аналіз даних експериментів для визначення ефективності тренувального процесу в сфері фізичної культури та спорту.

Укладачі:

Вишневецька В.П. – старший викладач кафедри інноваційних та інформаційних технологій у фізичній культурі і спорті Національного університету фізичного виховання і спорту України.

Рецензенти:

Струтинська О.В. – кандидат педагогічних наук, доцент кафедри інформаційних технологій і програмування, факультету інформатики, заступник декана з наукової роботи та міжнародної діяльності НПУ імені М.П. Драгоманова.

Глушак О. М. – кандидат педагогічних наук, доцент кафедри інноваційних технологій і математичних дисциплін Київського університету імені Б. Грінченко.