

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

Гороховський Є.Ю.

Біометрія

Методичні вказівки
до лабораторних занять
для здобувачів ступеня вищої освіти бакалавра
спеціальності «Біологія» освітньо-професійної програми «Біологія»

Затверджено
вченою радою ЗНУ
Протокол № 7 від 30.01.2018 р.

**Запоріжжя
2018**

УДК : 58.087 (075.8)
Г703

Гороховський Є.Ю. Біометрія : методичні вказівки до лабораторних занять для здобувачів ступеня вищої освіти бакалавра напряму підготовки «Біологія» / Є.Ю. Гороховський – Запоріжжя : ЗНУ, 2018. – 67 с.

Методичні вказівки до лабораторних занять з дисципліни «Біометрія» містять необхідний мінімум теоретичного матеріалу, покрокові рекомендації щодо виконання практичних завдань, питання для самоконтролю. З метою формування необхідних навичок запропоновано завдання для самостійного виконання.

Для здобувачів ступеня вищої освіти бакалавра напряму підготовки «Біологія».

Рецензент *О.В. Маслова*, кандидат фізико-математичних наук, доцент кафедри загальної та прикладної екології і зоології

Відповідальний за випуск *О.Ф. Рильський*, доктор біологічних наук, завідувач кафедри загальної та прикладної екології і зоології

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ.....	4
ВСТУП.....	5
Лабораторне заняття 1. ФОРМУВАННЯ РЕПРЕЗЕНТАТИВНОЇ ВИБІРКИ, ВИЗНАЧЕННЯ РОЗПОДІЛУ ТА ГІСТОГРАМИ.....	7
Лабораторне заняття 2. ВИЗНАЧЕННЯ СТАТИСТИЧНИХ ПОКАЗНИКІВ, ЩО ХАРАКТЕРИЗУЮТЬ СУКУПНІСТЬ.....	20
Лабораторне заняття 3. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ ІЗ ЗАСТОСУВАННЯМ ПАРАМЕТРИЧНИХ ТА НЕПАРАМЕТРИЧНИХ КРИТЕРІЇВ.....	33
Лабораторне заняття 4. ДИСПЕРСІЙНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ.....	43
Лабораторне заняття 5. КОРЕЛЯЦІЙНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ.....	51
Лабораторне заняття 6. РЕГРЕСІЙНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ.....	60
ПЕРЕЛІК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ.....	66

ПЕРЕЛІК СКОРОЧЕНЬ ТА УМОВНИХ ПОЗНАЧЕНЬ

ДІ	–	довірчий інтервал
ККІ	–	кількість класових інтервалів
ЛКМ	–	ліва кнопка миші
ПК	–	персональний комп'ютер
ПКМ	–	права кнопка миші

ВСТУП

Біометрія, або біологічна статистика, – розділ математичної статистики для обробки результатів біологічних експериментів. За допомогою біометрії, в основі якої лежить теорія ймовірностей, надається точна характеристика значення ознаки, яка вивчається, її мінливості; встановлюється вірогідність подібності або відмінності цієї ознаки у двох чи кількох групах особин, встановлюється ступінь залежності зміни одних ознак від зміни інших. Біометрія широко застосовується при вивченні різних теоретичних та практичних питань медицини, біології, рослинництва і тваринництва.

Курс «Біометрія» є необхідною складовою частиною вивчення сучасних методів планування експерименту, подальшого збору та статистичного аналізу отриманих експериментальних даних, а також представлення отриманих результатів у кваліфікаційних роботах, наукових публікаціях та звітах.

Мета курсу – засвоєння теоретичних основ статистичного аналізу біологічних даних, формування знань і умінь, що стосуються організації і проведення наукових і науково-прикладних досліджень, отримання практичних навичок статистичного аналізу з використанням персонального комп'ютера та програмного забезпечення, доступного широкому загалу (Microsoft Excel 2007 - 2013). Завдання курсу полягають у освоєнні методів варіаційної статистики, методів дисперсійного, кореляційного та регресійного аналізу, формуванні практичних навичок планування лабораторного та польового експерименту та використання спеціалізованого програмного забезпечення для аналізу даних, засвоєння правил коректного представлення результатів статистичного аналізу експериментальних даних наукового чи науково-практичного дослідження.

Практична частина курсу «Біометрія» спрямована на формування у студентів комплексу вмінь та навичок роботи із сучасними пакетами прикладних програм, застосуванням методу збору, систематизації та статистичного аналізу інформації у практичній діяльності фахівця-біолога.

Методичні вказівки містять необхідний мінімум теоретичного матеріалу, покрокові рекомендації щодо виконання практичних завдань, контрольні питання.

Видання призначене для здобувачів ступеня вищої освіти бакалавра напряму підготовки «Біологія».

ПРАВИЛА ТЕХНІКИ БЕЗПЕКИ ПРИ ВИКОНАННЯ ЛАБОРАТОРНИХ РОБІТ З КУРСУ «БІОМЕТРИЯ»

Робота з комп'ютером вимагає значної розумової напруги і супроводжується нервово-емоційним навантаженням оператора, високою напругою зорової системи та досить значним навантаженням на м'язи рук під час роботи з клавіатурою ПК. Велике значення має раціональна конструкція і розташування елементів робочого місця, що є важливим для підтримки оптимальної робочої пози в процесі праці.

Перед початком виконання лабораторної роботи потрібно:

- перевірити надійність встановлення апаратури на робочому столі. Повернути монітор так, щоб було зручно дивитися на екран – під прямим кутом (а не збоку) і трохи зверху вниз, при цьому екран має бути трохи нахиленим нижній його край ближче до оператора;
- перевірити загальний стан апаратури, перевірити справність електропроводки, з'єднувальних шнурів, штепсельних вилок, розеток, заземлення тощо;
- відрегулювати освітленість робочого місця;
- відрегулювати та зафіксувати висоту крісла, зручний для користувача нахил його спинки;
- приєднати до системного блоку необхідну апаратуру. Всі кабелі, що з'єднують системний блок з іншими пристроями, слід вставляти та виймати при вимкненому комп'ютері;
- ввімкнути апаратуру комп'ютера вимикачами на корпусах в послідовності: монітор, системний блок, принтер (якщо передбачається друкування); відрегулювати яскравість монітора та контрастність зображення. Не слід робити зображення надто яскравим, щоб не втомлювати очей.

Вимоги безпеки під час виконання роботи:

- необхідно стійко розташовувати клавіатуру на робочому столі, не допускаючи її хитання. Під час набору текстово-цифрових даних сидіти прямо, не напружуватися;
- для запобігання несприятливого впливу на користувача маніпулятору типу «миша» належить забезпечувати вільну велику поверхню столу для його переміщення і зручного упору ліктьового суглоба;
- не дозволяються сторонні розмови, подразнюючі шуми.

ЗАБОРОНЯЄТЬСЯ:

- входити до лабораторії у верхньому одязі;
- ставити на стіл стакани, пляшки, із водою чи будь-якою іншою рідиною, навіть якщо вони закриті кришкою;
- класти будь-які предмети на апаратуру комп'ютера;
- закривати будь-чим вентиляційні отвори апаратури, що може призвести до перегрівання і виходу її з ладу.

ЛАБОРАТОРНЕ ЗАНЯТТЯ 1

ФОРМУВАННЯ РЕПРЕЗЕНТАТИВНОЇ ВИБІРКИ, ВИЗНАЧЕННЯ РОЗПОДІЛУ ТА ГІСТОГРАМИ

Мета: лабораторне заняття передбачає отримання базових навичок зі створення таблиць із експериментальними даними; сортування даних; визначення частоти зустрічальності певної величини ознаки та побудови варіаційного ряду; побудови полігону розподілу та гістограми.

Обладнання та матеріали: ПЕОМ із встановленою ОС Microsoft Windows XP або Windows Seven, процесор таблиць Microsoft Excel 2007 або 2010, Інтернет браузер, методичні вказівки.

План:

1. Створення та сортування таблиць.
2. Визначення частот зустрічальності певної величини ознаки.
3. Складання варіаційного ряду.
4. Побудова графіків розподілу даних.

Теоретичні відомості.



Варіаційний ряд – впорядковане зображення реально існуючого розподілу особин у групі за величиною ознаки. Варіаційний ряд – це подвійний ряд чисел, що складається з позначення класів і відповідних частот. Він показує, як змінюється ознака від мінімальної до максимальної величини (крайні варіанти – *min* і *max*), та яка частота варіант у кожному з класів.

З метою створення візуального відображення статистичної інформації користуються різними графіками. Найпоширеніші види графічного відображення статистичної інформації – це полігони і гістограми. Графічне зображення варіаційних рядів за допомогою полігона чи гістограми допомагає отримати наочне уявлення про закономірності можливих змін спостережуваних значень.

Полігон, як правило, використовують для відображення дискретного варіаційного ряду.

Полігоном частот називають ламану з вершинами у точках (z_i, n_i) , $i = 1, 2, \dots, k$. Тут z_i – значення i -ї варіанти, а n_i – відповідна цій варіанті частота. **Полігоном відносних частот** називають ламану, відрізки якої сполучають точки (z_1, P_{n1}^*) , (z_2, P_{n2}^*) , \dots , (z_k, P_{nk}^*) – тобто це статистичне зображення статистичного розподілу.

Можлива побудова полігона частот не тільки для дискретного варіаційного ряду, а й для інтервального статистичного ряду. В ситуації з інтервальним статистичним рядом при побудові полігонів як абсциси відповідних кінців відрізків обирають точки z_i^* .

Гістограми використовують для зображення винятково інтервальних варіаційних рядів. Для їх побудови в прямокутній системі координат на осі абсцис відкладають відрізки, що є частковими інтервалами спостережень. На

цих відрізках, як на основах, будують прямокутники з висотами, що дорівнюють частотам – абсолютним або відносним. Тобто існують два типи гістограм. Варто знати формальне визначення гістограми.

Гістограмою абсолютних частот називають ступінчасту фігуру, яка побудована з прямокутників, основою яких є інтервали групування, довжини h , а висоти дорівнюють:

$$h_i^* = \frac{n_i}{h}$$

Площа гістограми абсолютних частот дорівнює n .

Гістограмою відносних частот (або просто **гістограмою**) називають ступінчасту фігуру, яка складається з прямокутників, основами яких є часткові інтервали групування довжини h , а висоти дорівнюють:

$$\frac{P_{n_i}^*}{h}$$

Оскільки площа чергового прямокутника становить

$$h \times \frac{P_{n_i}^*}{h} = P_{n_i}^*,$$

то загальна площа гістограми дорівнює одиниці. По суті, гістограма – це графічний статистичний аналог щільності.



У MS Excel складання варіаційних рядів та побудова графіків розподілу ґрунтується на застосуванні лише трьох функцій: МІН(), МАКС(), які визначають найменше та найбільше значення вибірових даних, та функції ЧАСТОТА(), яка визначає частоту зустрічальності кожної з варіант.

Хід виконання роботи.

Завдання 1. Складання варіаційного ряду та побудова полігону розподілу.

У результаті вимірювання довжини правого вуха у 60 особин овець валахської породи були отримані наступні дані (у см):

12	10	14	14	14	14	12	12	12	15
13	11	12	12	14	12	11	13	12	13
14	12	13	12	12	14	12	14	13	13
12	13	12	13	12	11	11	12	13	14
12	14	13	14	13	12	14	15	10	11
10	11	15	11	16	11	14	11	10	11



Вівця домашня
(*Ovis aries*)

За наведеними даними побудувати полігон розподілу.



Побудова полігону розподілу у Microsoft Excel передбачає виконання наступного алгоритму дій:

- Створення таблиці з даними.
- Сортування даних за збільшенням.
- Визначення частоти зустрічальності певної величини ознаки (побудова варіаційного ряду частот).
- Побудова та налаштування графіку.

1.1. Відкрийте додаток Microsoft Excel, та внесіть наведені дані до таблиці. Почніть із комірки A1, позначивши у ній назву змінної (“Довжина вуха, см”). У комірки A2 – A61 внесіть результати вимірювань.

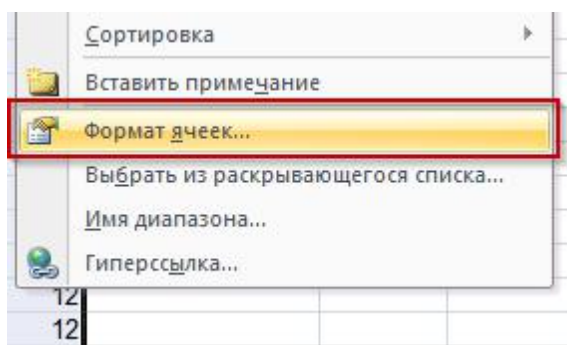
	A	B	C	D	E	F	G
1	Довжина вуха, см						
2	12						
3	10						
4	14						
5	14						



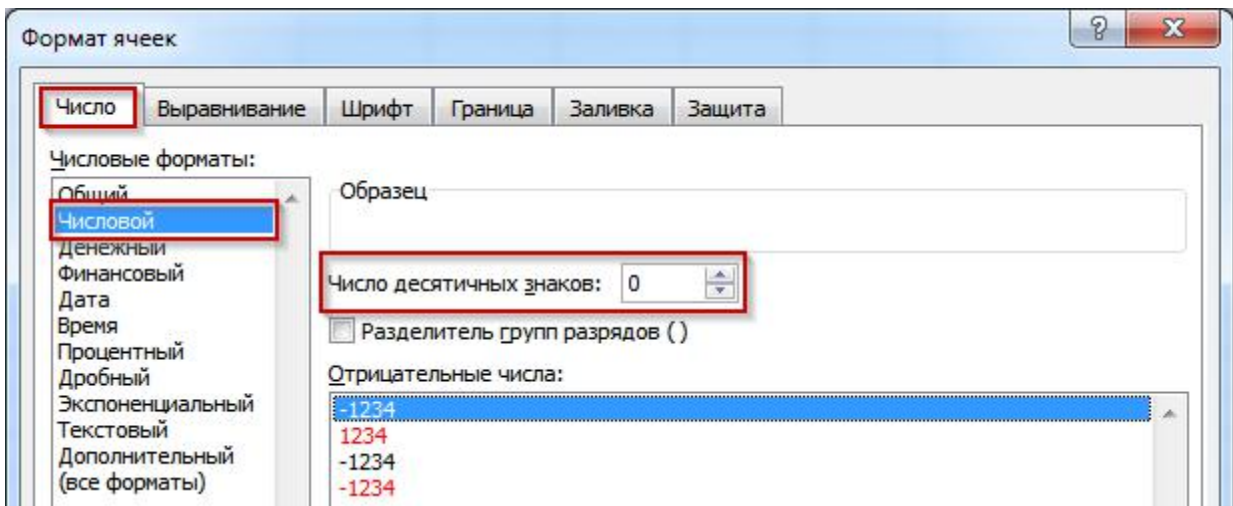
Потрібно пам'ятати, що окрім даних вимірювань чи спостережень, таблиця має містити достатню кількість додаткової інформації (назви змінних, одиниці вимірювання, дослідні групи та ін.), яка може допомогти іншій особі зрозуміти, які саме дані містяться у таблиці.

1.2. Після того, як усі дані будуть внесені до таблиці, виділіть за допомогою миші діапазон комірок, що містять числові дані (A2 – A61).

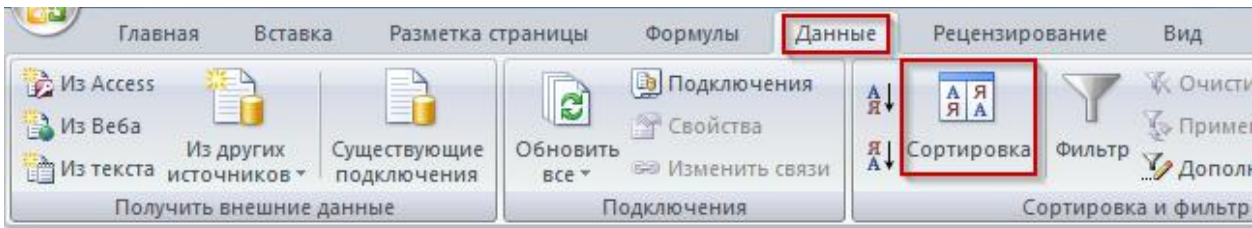
1.3. Клацніть правою кнопкою миші (ПКМ) у виділеній ділянці таблиці для виклику контекстного меню, у якому оберіть пункт **Формат ячеек...**:



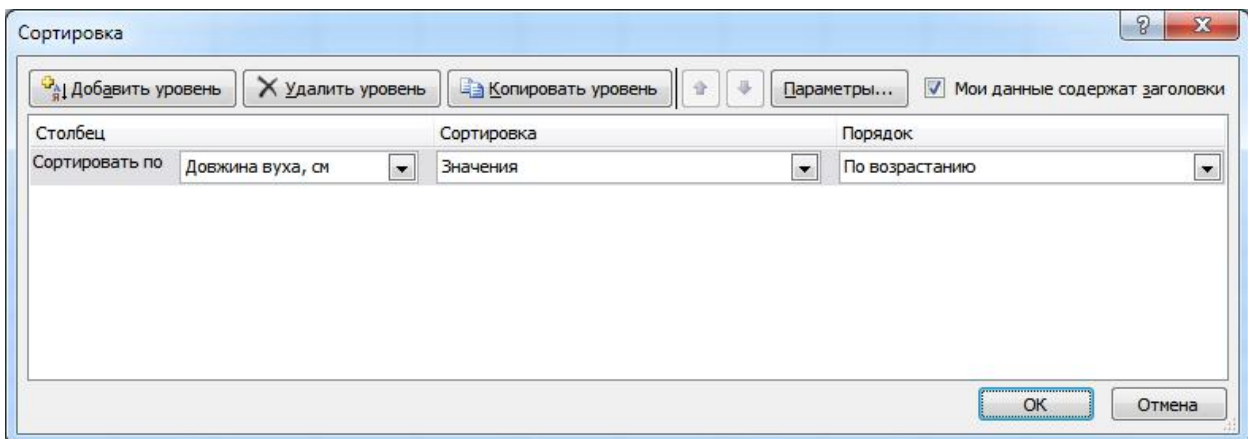
У діалоговому вікні, що з'явиться, на вкладці **Число** змініть числовий формат ваших даних з **Общий** на **Числовой**, та зменшіть значення **Число десятичних знаків** до нуля.



1.4. Виділіть за допомогою миші усі введенні данні (комірки A1 – A61). На стрічці табличного процесору Excel оберіть вкладку **Данніе**, а на ній **Сортировка...**

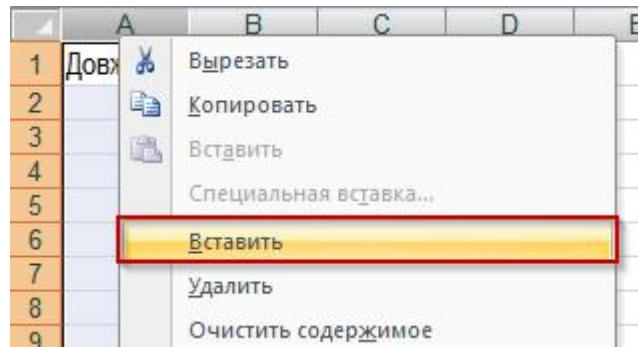


У діалоговому вікні, що з'явиться натисніть кнопку **ОК**. Дані будуть відсортовані за збільшенням (від меншого значення до більшого).

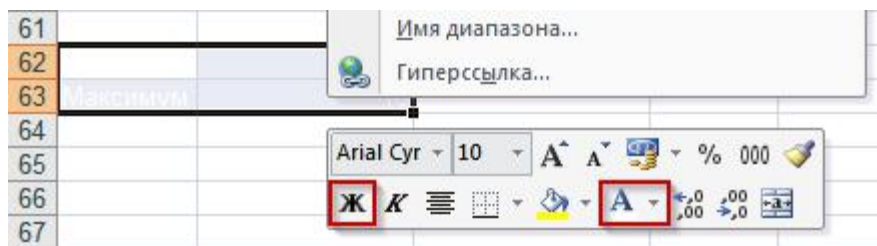


Для того, щоб визначити частоти зустрічальності певних величин ознаки (у даному випадку довжини вуха овець валахської породи) потрібно спочатку встановити межі розсіювання цієї ознаки (тобто мінімальне та максимальне значення).

1.5. Клацніть ПКМ на заголовку стовпчика A, та оберіть у контекстному меню пункт **Вставить**. Буде створено новий стовпчик A, а ваші дані будуть автоматично переміщені у діапазон B1 – B61.



1.6. Виділіть діапазон комірок A62:B63 та виділіть їх шляхом встановлення атрибутів шрифту **Жирный** и кольору **Красный** у контекстному меню Excel.



1.7. Введіть у комірку A62 текст «Мінімум», а у комірку A63 текст «Максимум». У комірку B62 уведіть формулу =МИН(B2:B61), а комірку B63 формулу =МАКС(B2:B61).

60		15
61		16
62	Мінімум	=МИН(B2:B61)
63	Максимум	
64		

60		15
61		16
62	Мінімум	10
63	Максимум	=МАКС(B2:B61)
64		



Будь-яка формула у Excel повинна починатись із знаку «=», інакше вона буде розпізнана як текст. Аргументами функцій у дужках може бути як окрема комірка, так і їх діапазон, який потрібно вказувати шляхом його виділення після відкриття лівої дужки формули. Якщо ж вам потрібно увести діапазон комірок вручну, то пам'ятайте, що назви формул набираються у кириличній розкладці клавіатури, а імена комірок – у латинській (англійській) розкладці.

1.8. Наступний крок – визначення частот зустрічальності певної величини ознаки. Для цього у комірку C1 уведіть текст Довжина вуха, см, а у комірки C2 – C8 – можливі величини ознаки, від мінімуму до максимуму із кроком 1 см. Змініть формат комірок C2 – C8. на числовий (см. п. 3). У комірку D1 введіть текст Кількість особин.

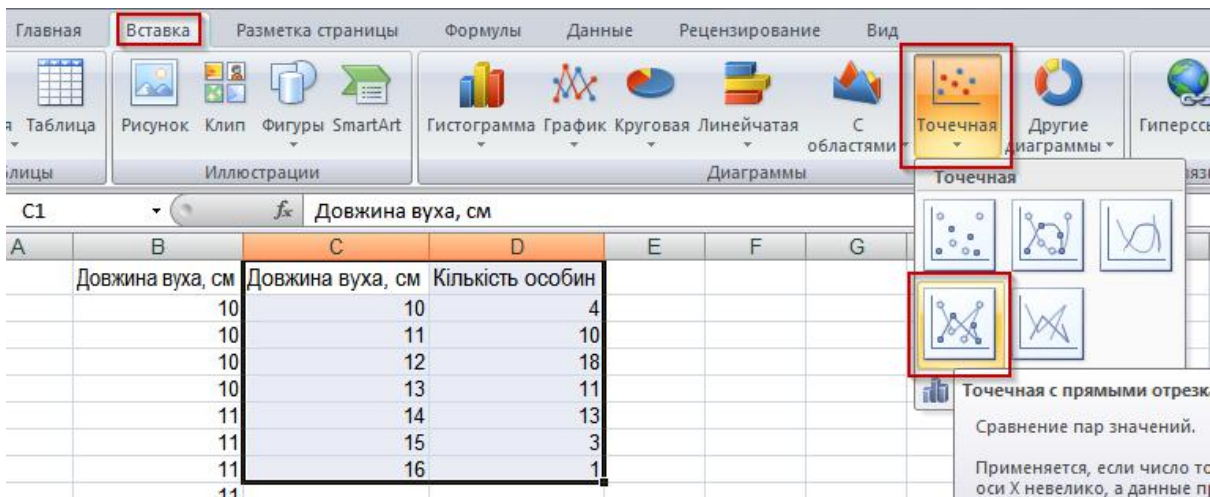
1.9. У комірку D2 введіть функцію ЧАСТОТА, яка має два аргументи – масив_даних та масив_інтервалів, розділених між собою крапкою з комою “;”. У якості першого аргументу виділіть діапазон комірок B2 – B61, натисніть клавішу “;” і виділіть діапазон комірок D2 – D8 (другий аргумент) та натисніть клавішу ENTER. У комірці D2 з’явиться значення частоти, що дорівнює 4.

	A	B	C	D	E
1		Довжина вуха, см	Довжина вуха, см	Кількість особин	
2		10	10	=ЧАСТОТА(B2:B61;C2:C8)	
3		10			
4		10			
5		10			
6		11			
7		11			
8		11			
9		11			

Функція повертає масив частот, тому далі потрібно виділити діапазон комірок D2 – D8, натиснути клавішу <F2>, а після цього – одночасно натиснути комбінацію <CTRL + SHIFT + ENTER>. У комірках D2 – D8 з’являться частоти зустрічальності кожної величини ознаки (довжини вуха), з наявних у вибірці.

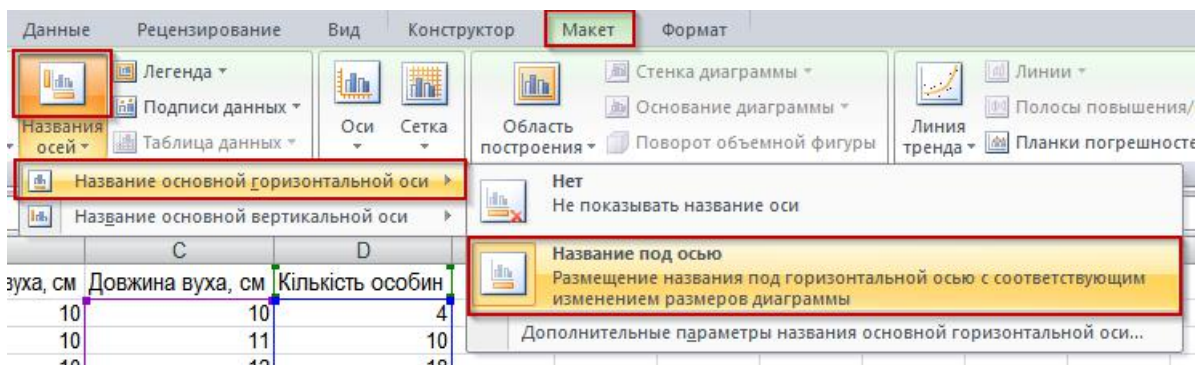
	A	B	C	D
1		Довжина вуха, см	Довжина вуха, см	Кількість особин
2		10	10	4
3		10	11	10
4		10	12	18
5		10	13	11
6		11	14	13
7		11	15	3
8		11	16	1
9		11		

1.10. Для побудови полігону розподілу виділіть діапазон комірок C1 – D8, та на стрічці табличного процесору Excel оберіть вкладку **Вставка** → **Діаграми** → **Точечная** → **Точечная с прямими отрезками**.



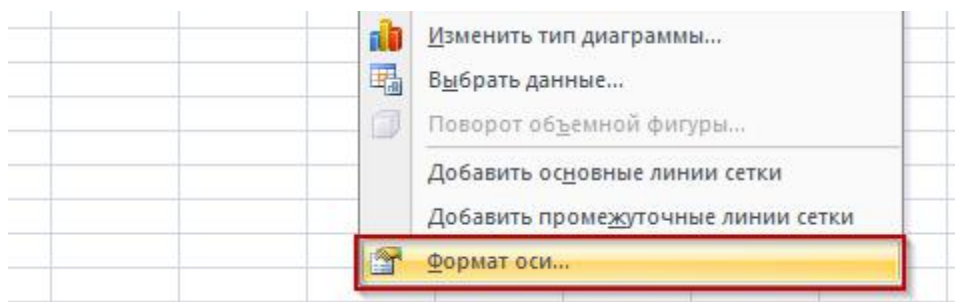
На аркуші Excel з'явиться діаграма. Клацніть ПКМ по заголовку діаграми та оберіть у контекстному меню команду **Удалить**. Те ж саме зробіть і для легенди та ліній сітки.

Після цього у компоненті **Работа с диаграммами**, перейдіть до вкладки **Макет** → **Подписи** → **Название осей** → **Название основной горизонтальной оси** → **Название под осью** та введіть назву горизонтальної осі – «Довжина вуха, см».

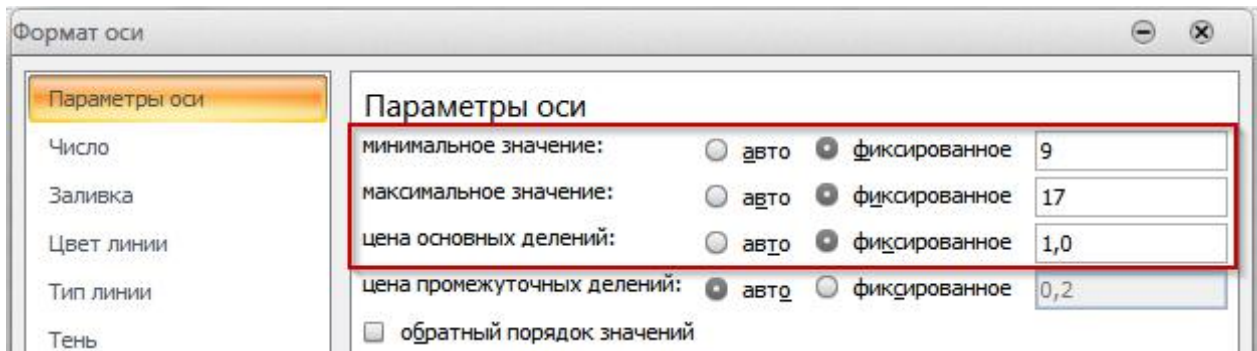


Теж саме зробіть для вертикальної осі та введіть її назву – Кількість особин. Змініть шрифт підписів осей на «Verdana», звичайний, 12 пт.

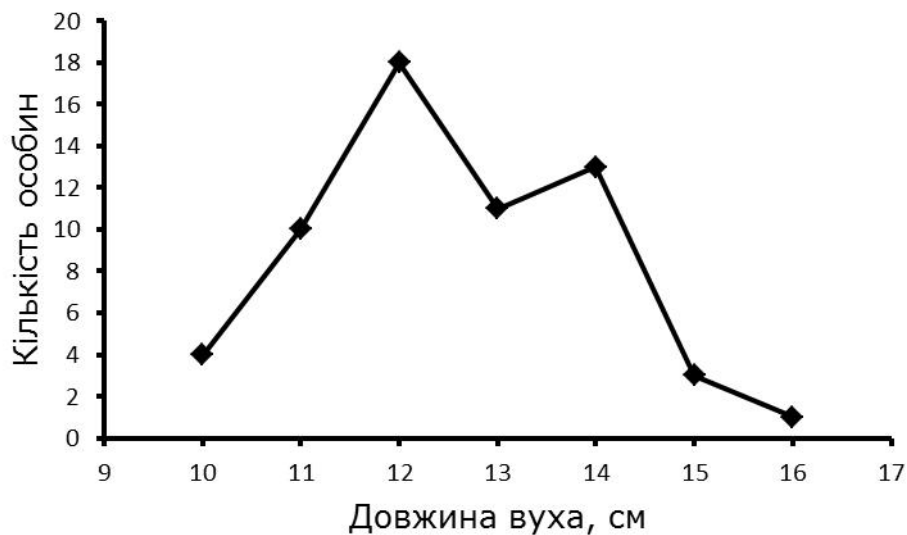
Клацніть ПКМ у ділянці цифр горизонтальної осі, для того щоб її виділити, потім клацніть ЛКМ для виклику контекстного меню.



У діалоговому вікні, що з'явилося змініть мінімальне значення з **авто** на 9, максимальне значення з «авто» на «**фиксированное**» (17) та встановіть фіксоване значення основних поділок, що дорівнює 1. Закрийте діалогове вікно.



У результаті буде побудований полігон розподілу частот, на вертикальній осі якого буде відображена довжина вуха, а на горизонтальній – кількість овець, що мають саме таке значення ознаки.



Завдання 2. Побудова гістограми розподілу.

Була виміряна довжина тіла (см) у 77 екземплярів плітки Дніпровського водосховища, та отримані наступні результати:

143 157 148 153 150 142 164 139 139 140
 143 120 144 130 138 124 127 137 139 129
 128 119 120 138 130 114 126 133 117 132
 130 145 140 153 137 142 145 137 141 125
 148 138 140 135 135 139 125 137 131 120
 127 118 120 124 134 111 132 133 100 132
 143 134 138 130 135 133 134 154 107 110
 94 95 142 148 136 165 172



Плітка (*Rutilus rutilus*)

За наведеними даними побудувати гістограму розподілу довжини тіла плітки Дніпровського водосховища.



Побудова гістограми розподілу у Microsoft Excel передбачає виконання наступного алгоритму дій:

- Створення таблиці з даними.*
- Сортування даних за збільшенням.*
- Визначення частоти зустрічальності певної величини ознаки*
- Визначення величини класових інтервалів.*
- Побудова та налаштування графіку.*

2.1. Використовуючи послідовність дій, що наведені у пп. 1.1 – 1.7, створіть таблицю з даними (комірки A2 : A78), та визначте мінімальне та максимальне значення довжини тіла плитки (комірки C2, C3).

2.2. Наступним етапом у побудові гістограми розподілу є визначення класових інтервалів. Для цього спочатку визначте розмах, віднявши від максимального значення довжини тіла плитки мінімальне значення (у комірці C4, формула =C2-C3). Інтервал класового інтервалу визначають за формулою Стерджеса: $n = 1 + 3,322 * \log n$, де n – кількість усіх значень величини. У комірку C5 введіть формулу для розрахунку величини класового інтервалу (ВКІ) =ОКРУГЛВВЕРХ(1+3,322*LOG10(C4);1). У комірці C6 обчисліть кількість класових інтервалів (ККІ), поділивши розмах значень довжини тіла плитки (комірка C4) на величину класового інтервалу (комірка C5). Отже, формула для обчислення ККІ буде виглядати наступним чином =C4/C5.

2.3. Побудуйте таблицю довжини тіла плитки із урахуванням обраного значення класового інтервалу, для чого у комірку D2 уведіть мінімальне значення довжини тіла плитки, а у комірку D3 – формулу =D2+\$C\$5, відповідно до якої кожна наступна комірка буде містити попереднє значення величини класу із додаванням величини класового інтервалу.



Посилання на комірки (а також на стовпчики та рядки) у Excel можуть бути відносними та абсолютними. Формули, що містять відносні посилання на комірку, змінюються при копіюванні з однієї комірки в іншу, а формули, що містять абсолютні посилання – ні. Змінити тип можна виділивши посилання та натискаючи клавішу F4.

Скопіюйте вміст комірки D3, та вставте у комірки D4 – D13 (або зробіть це за допомогою авто заповнення). Таким чином, у комірках D2 – D13 буде обчислено показники величини довжини тіла плитки із врахуванням величини класового інтервалу.

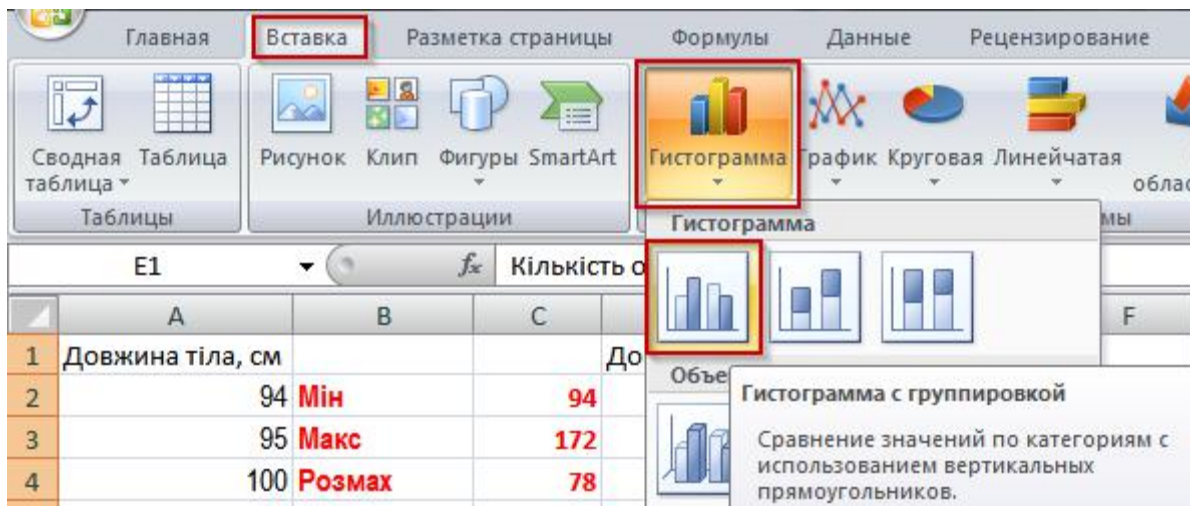
2.4. Обчисліть частоту зустрічальності кожного класового інтервалу. Для цього використайте вже відому формулу ЧАСТОТА. У комірку E2 введіть формулу =ЧАСТОТА(), виділіть діапазон комірок A2:A78 (довжина тіла плитки), натисніть клавішу “;” і виділіть діапазон комірок D2:D13 (класові інтервали). та натисніть клавішу <ENTER>. У комірці E2 з’явиться значення частоти, що

дорівнює 1.

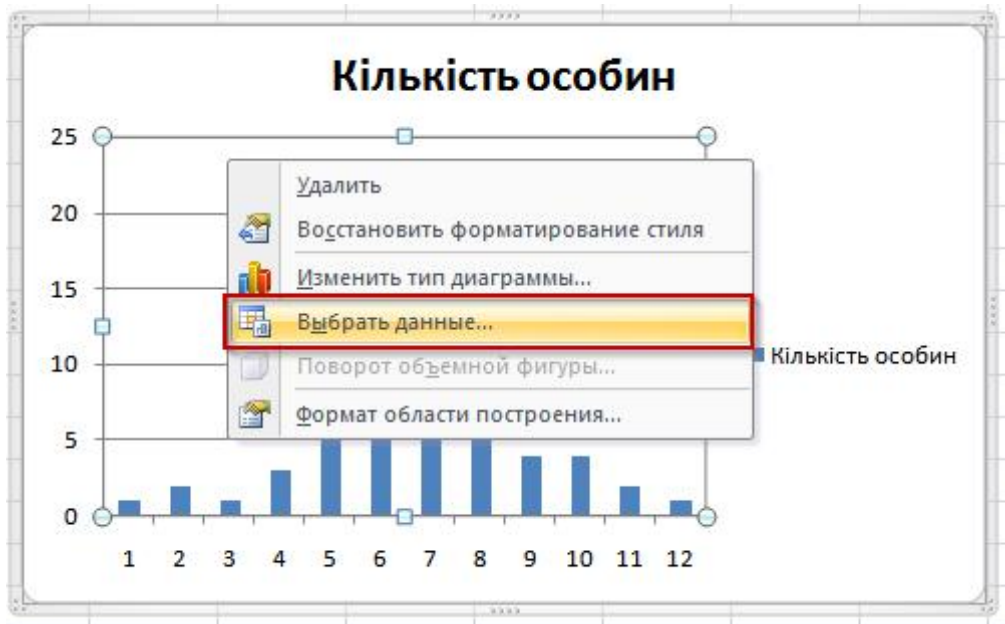
2.5. Далі, виділіть діапазон комірок E2 – E13, натисніть клавішу <F2>, а після цього – комбінацію <CTRL + SHIFT + ENTER>. У комірках E2 – E13 будуть обраховані частоти зустрічальності довжини тіла плітки, із урахуванням класових інтервалів.

	A	B	C	D	E	F
1	Довжина тіла, см			Довжина тіла, см	Кількість особин	
2	94	Мін	94	94	1	
3	95	Макс	172	101,3	2	
4	100	Розмах	78	108,6	1	
5	107	ВКІ	7,3	115,9	3	
6	110	ККІ	11	123,2	7	
7	111			130,5	13	
8	114			137,8	18	
9	117			145,1	21	
10	118			152,4	4	
11	119			159,7	4	
12	120			167,0	2	
13	120			174,3	1	
14	120					

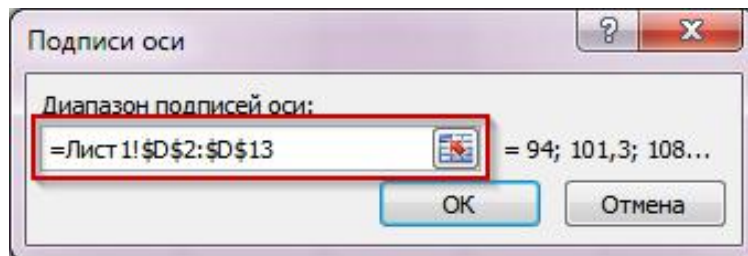
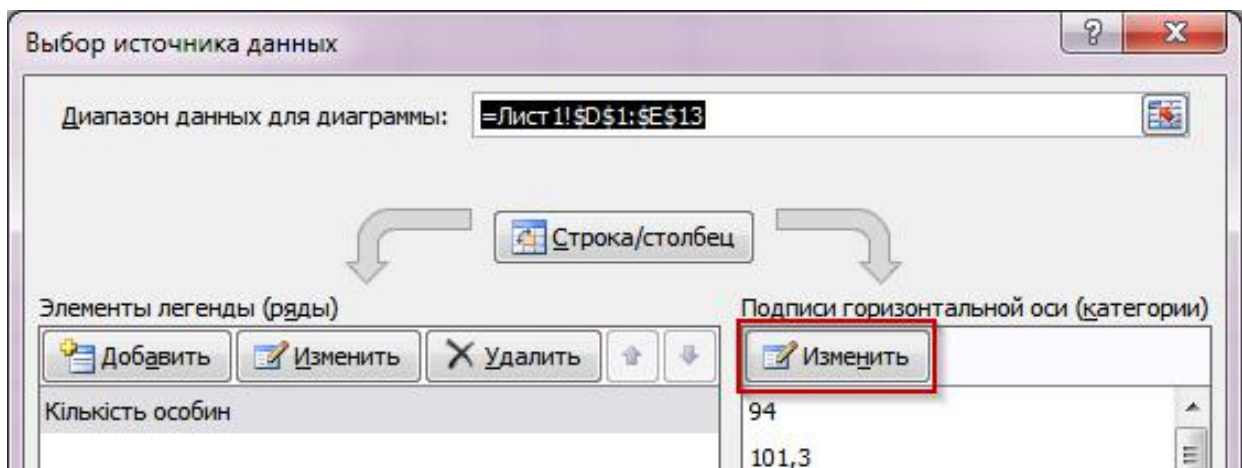
2.6. Побудуйте гістограму, для чого виділіть діапазон комірок E1 – E13; на стрічці табличного процесору Excel оберіть вкладку **Вставка** → **Діаграми** → **Гістограма** → **Гістограма с группировкой**.



2.6. Налаштуйте значення підписів горизонтальної осі гістограми, для чого клацніть по ній ПКМ, та у контекстному меню оберіть пункт **Вибрати данные...**

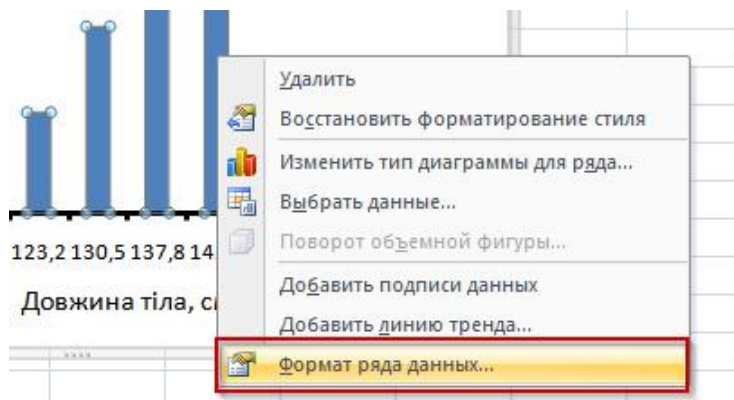


У діалоговому вікні, що з'явиться, вкажіть дані для горизонтальної осі (значення класів), натисніть кнопку **Изменить** та у новому діалоговому вікні вид виділіть діапазон комірок D2 – D13, та натисніть ОК.

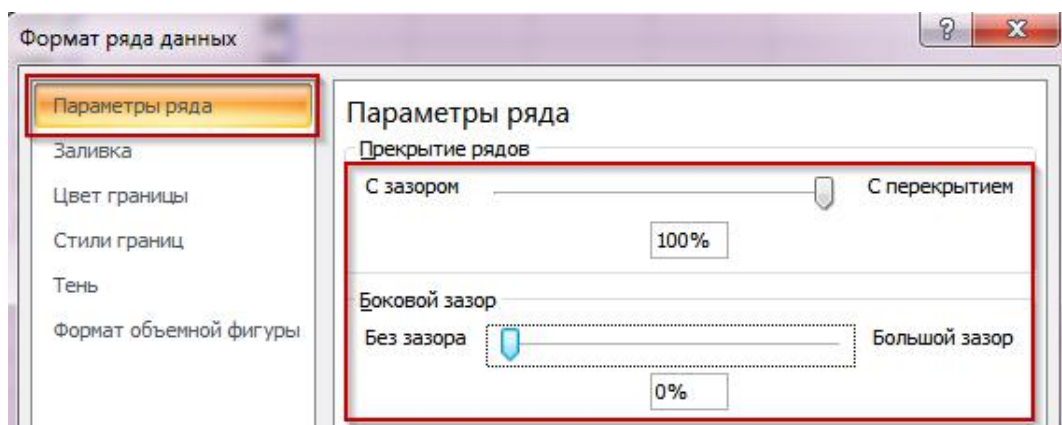


2.7. Відредагуйте гістограму та додайте підписи осей, відповідно до вказівок п. 1.10 даного лабораторного заняття.

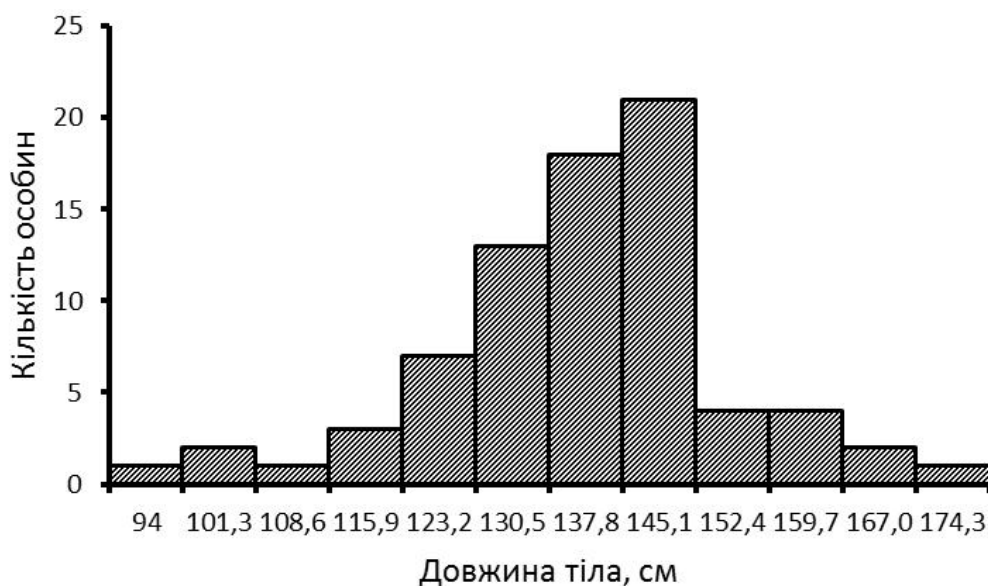
2.8. Клацніть ПКМ по будь-якому стовпчику гістограми, та оберіть у контекстному меню пункт **Формат ряда данных...**



У діалоговому вікні, що з'явиться змінити параметри відображення рядів даних на ті, що наведені на рисунку (**Перекриття рядов – С перекритием, 100%**; **Боковой зазор – Без зазора, 0%**)



У результаті, буде побудовано гістограму розподілу довжини тіла плитки Дніпровського водосховища.



Завдання для самостійного виконання.

1. Була підрахована кількість променів у хвостових плавцях камбали (*Pleuronectes platessa*):

53	51	52	55	56	49	51	52	54	56
54	53	52	53	51	55	53	55	53	51
51	51	56	54	54	53	54	54	55	53
52	55	53	53	56	53	52	56	52	52
56	55	50	54	49	54	54	55	54	55

Складіть варіаційний ряд та побудуйте полігон розподілу.

2. У густери (*Blicca bjoerkna*) озера Біле був вимірний обхват тіла (у мм):

80	75	78	85	78	85	80	77	83	85
88	94	95	88	80	73	78	90	95	90
80	75	83	70	78	83	75	78	86	81
62	77	75	73	80	80	74	73	82	72
80	90	80	78	60	65	75	72	64	67
74	80	68	75	76	65	70	78	75	83
85	70	88	73	56	75	70	73	68	66
65	68	75	78	63	68	68	70	60	56

Складіть варіаційний ряд та побудуйте гістограму.

Контрольні питання.

1. Поясніть, що таке «сукупність»? Наведіть приклади різних сукупностей.
2. Чим відрізняється вибіркова сукупність від загальної?
3. Розкрийте сутність варіанти; випадкової змінної.
4. У чому можуть проявлятися відмінності між окремими варіантами?
5. Охарактеризуйте принципи групування даних при якісній мінливості; при кількісній дискретній.
6. Поясніть, що таке «варіаційний ряд»? У чому полягають особливості розподілу варіант у варіаційному ряді?
7. Порівняйте гістограму та полігон розподілу. Поясніть, у чому полягає різниця між ними?
8. Вкажіть можливі причини багатoverшинності варіаційних кривих.

ЛАБОРАТОРНЕ ЗАНЯТТЯ 2

ВИЗНАЧЕННЯ СТАТИСТИЧНИХ ПОКАЗНИКІВ, ЩО ХАРАКТЕРИЗУЮТЬ СУКУПНІСТЬ

Мета: лабораторне заняття передбачає отримання базових навичок розрахунків точкових (середнього арифметичного, медіани, моди) та інтервальних (стандартного відхилення, помилки середнього арифметичного, довірчого інтервалу) оцінок експериментальних даних; відображення результатів описового статистичного аналізу у графічній формі.

Обладнання та матеріали: ПЕОМ із встановленою ОС Microsoft Windows XP або Windows Seven, процесор таблиць Microsoft Excel 2007 або 2010, Інтернет браузер, методичні вказівки.

План:

1. Обрахування точкових оцінок вибіркової сукупності.
2. Обрахування інтервальних оцінок вибіркової сукупності.
3. Графічне відображення мір центральної тенденції та розсіювання.

Теоретичні відомості.



Для дослідження основних властивостей групового явища, або об'єкта, представленого вибіркою значень, обчислюють точкові та інтервальні оцінки. До основних точкових оцінок у біометрії належать: середнє арифметичне, медіану, моду. До інтервальних – дисперсія, стандартне відхилення, стандартна помилка.

Середнє – дуже інформативна міра «центральної тенденції» («центральної позиції») змінної, яка спостерігається, особливо якщо також повідомляється її довірчий інтервал.

Довірчий інтервал для середнього являє собою інтервал значень поблизу оцінки, де з даним рівнем довіри, знаходиться «істинне» (невідоме) середнє популяції. Наприклад, якщо середнє число вибірки становить 23, а нижня і верхня межі довірчого інтервалу при рівні статистичної значущості дорівнюють 19 і 27 відповідно, то можна зробити висновок, що з вірогідністю 95% інтервал з границями 19 і 27 накриває середнє значення популяції. Якщо встановите більший рівень довіри, то інтервал стане ширше, тому що збільшується вірогідність, з якою він «накриває» невідоме середнє популяції, і навпаки.

Медіана. Якщо упорядкувати дані за величиною, починаючи із самої малої величини і закінчуючи самою великою, то медіана також буде характеристикою усереднення в упорядкованому наборі даних. Медіана поділяє ряд

упорядкованих значень нарівно, з рівним числом цих значень як вище, так і нижче її (ліворуч і праворуч від медіани на чисельній осі).

Мода – це значення, яке найчастіше зустрічається у наборі даних. Деякі набори даних не мають моди, тому що кожне значення зустрічається лише 1 раз, а іноді буває більше однієї моди; це відбувається тоді, коли два або більше значення зустрічаються однаково кількість разів і зустрічальність кожного з цих значень більше, ніж будь-якого іншого значення.

Розмах – це різниця між максимальним та мінімальним значенням змінної в наборі даних.

Дисперсія. Один з способів вимірювання розсіювання даних полягає в тому, щоб визначити ступінь відхилення кожного спостереження від середньої арифметичної. Очевидно, що чим більше відхилення, тим більше змінюється варіабельність спостережень.

Однак ми не можемо використовувати середнє значення цих відхилень як міру розсіювання, тому що позитивні відхилення компенсують негативні відхилення (їх сума рівна нулю). Щоб вирішити цю проблему, у квадрат зводять кожне відхилення і знаходять середнє зведених у квадрат відхилень; ця величина називається **варіацією**, або **дисперсією**.

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

У випадку, якщо справу мають не з загальною сукупністю, а з вибіркою (що у біологічній статистиці є скоріше правилом), то вираховується вибіркова дисперсія:

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

Одиниці виміру (розмірність) варіацій – це квадрат одиниць виміру початкових спостережень. Наприклад, якщо виміри проводили у грамах, то одиниця виміру варіації буде грам у квадраті.

Середнє квадратичне відхилення – це позитивний квадратний корінь з дисперсії, а **стандартне відхилення вибірки** – корінь із вибіркової дисперсії:

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

Стандартне відхилення можна представити як свого роду середнє

відхилення спостережень від середнього значення вибірки. Воно обчислюється у тих же одиницях (розмірностях), що і вихідні дані.

Якщо розділити стандартне відхилення на середню арифметичну і виразити результат у відсотках то буде отриманий **коефіцієнт варіації**, який є мірою розсіювання. Він не залежить від одиниць виміру (безрозмірний), та має деякі теоретичні недоліки і тому не часто використовується.



Обрахування значень описових статистик у табличному процесорі Excel виконують за допомогою ряду статистичних функцій:

<i>Описова статистика</i>	<i>Функції Excel</i>
<i>Середнє арифметичне</i>	<i>СРЗНАЧ(масив_даних)</i>
<i>Стандартна помилка ($n \leq 30$)</i>	<i>СТАНДОТКЛОН.В(масив_даних)/КОРЕНЬ(СЧЕТ(масив_даних) - 1)</i>
<i>Стандартна помилка ($n > 30$)</i>	<i>СТАНДОТКЛОН.В(масив_даних)/КОРЕНЬ(СЧЕТ(масив_даних))</i>
<i>Медіана</i>	<i>МЕДИАНА(масив_даних)</i>
<i>Мода</i>	<i>МОДА(масив_даних)</i>
<i>Стандартне відхилення</i>	<i>СТАНДОТКЛОН.В(масив_даних)</i>
<i>Дисперсія вибірки</i>	<i>ДИСП(масив_даних)</i>
<i>Ексцес</i>	<i>ЭКЦЕСС(масив_даних)</i>
<i>Асиметрія</i>	<i>СКОС(масив_даних)</i>
<i>Інтервал</i>	<i>МАКС(масив_даних) - МИН(масив_даних)</i>
<i>Мінімум</i>	<i>МИН(масив_даних)</i>
<i>Максимум</i>	<i>МАКС(масив_даних)</i>
<i>Сума</i>	<i>СУММ(масив_даних)</i>
<i>Підрахунок</i>	<i>СЧЕТ(масив_даних)</i>
<i>Довірчий інтервал (95%)</i>	<i>ДОВЕРИТ(0,05;СТАНДОТКЛОН(масив_даних);СЧЕТ(масив_даних))</i>
<i>Перцентиль 10%</i>	<i>ПЕРСЕНТИЛЬ(масив_даних; 0,1)</i>

Хід виконання роботи.

Завдання 1. За результатами вимірювань ширини верхнього останнього корінного зуба (мм) у викопного ссавця *Australopithecus rigidus* були отримані наступні дані:

5,8 5,2 6,3 6,1 6,1 5,7 6,5 6,0 6,1 5,8 6,3 6,2
6,2 5,4 5,9 6,0 5,7 5,9 6,1 6,7 6,2 6,5 6,2 6,1
6,2 5,7 6,1 5,7 5,9 6,0 5,7 5,9 6,1 5,9 6,0 6,1



Австралопітек
(*Australopithecus rigidus*)

Обрахувати: середнє арифметичне, медіану, моду; дисперсію, стандартне відхилення, помилку середнього арифметичного, довірчий інтервал (95%); оцінити характер розподілу за коефіцієнтами асиметрії та ексцесу.



Визначення статистичних показників, що характеризують вибіркoву сукупність за допомогою вбудованих формул MS Excel передбачає наступний й алгоритм дій:

- Створення таблиці з даними.
- Обрахування значень точкових оцінок.
- Обрахування значень інтервальних оцінок.
- Визначення характеру розподілу експериментальних даних із використанням вбудованих формул MS Excel.

1.1. Відкрийте додаток Microsoft Excel, та внесіть наведені дані до таблиці. Почніть із комірки В1, позначивши у ній назву змінної («Ширина зуба, мм»). У комірки В2 – В37 внесіть результати вимірювань. Змініть формат комірок з «Общий» на «Числовой».

1.2. У комірки А38:А40 уведіть текстові значення «Середнє арифметичне», «Медіана», «Мода».

Починаючи з комірки В38 уведіть формули для обчислення точкових оцінок вибірки, та вкажіть діапазон даних:

у комірку В38 =СРЗНАЧ(А2:А37)

у комірку В39 =МЕДИАНА(А2:А37)

у комірку B40 =МОДА(A2:A37)

Виділіть вміст комірок A38:B40 червоним кольором.

1.3. У комірки A41:A43 уведіть текстові значення «Дисперсія», «Стандартне відхилення», «Стандартна помилка», відповідно.

1.5 Починаючи з комірки B41 уведіть формули для обчислення точкових оцінок вибірки, та вкажіть діапазон даних:

у комірку B41 =ДИСП(B2:A37)

у комірку B42 =СТАНДОТКЛОН(B2:A37)

у комірку A43 =СТАНДОТКЛОН(B2:A37)/КОРЕНЬ(СЧЕТ(B2:A37))

Виділіть вміст комірок A41:B43 зеленим кольором.

1.6. У комірки A44:A45 уведіть текстові значення «Асиметрія», «Ексцес», а у комірки B44 та B45 – формули для їх обчислення:

у комірку B44 =СКОС(B2:A37)

у комірку B45 =ЭКСЦЕСС(B2:A37)

Виділіть вміст комірок A44:B45 зеленим кольором.

1.6. Зробіть попередній висновок про те, чи підлягають дані у вибірці закону нормального розподілу? Для цього скористуйтеся наступною інформацією. Оскільки асиметрія та ексцес за умов нормального розподілу даних дорівнюють нулю, значення цих показників у випадку, якщо розподіл даних у вибірці також є нормальним, також повинні бути близькі до нуля.

Стандартна помилки асиметрії приблизно дорівнює $\sqrt{6/n}$, де n – розмір вибірки, а стандартна помилка ексцесу – $\sqrt{24/n}$ (також приблизно).

Якщо коефіцієнт асиметрії для вибірових даних перевищує стандартну помилку у два і більше рази, це свідчить, що дані не є симетричними, і, відповідно, їх розподіл не відповідає нормальному. Подібним чином, якщо коефіцієнт ексцесу перевищує стандартну помилку у два і більше рази, дані не є нормально розподіленими.

У комірках C44 та C45 обрахуйте стандартні помилки ексцесу за наступними формулами =КОРЕНЬ(6/СЧЕТ(B2:B37)) та =КОРЕНЬ(24/СЧЕТ(B2:B37)). Порівняйте отримані значення та зробіть висновок, чи підлягають вибіркові дані закону нормального розподілу.



Оцінювання характеру розподілу даних у вибірковій сукупності за допомогою коефіцієнтів асиметрії та ексцесу хоча й має приблизний характер все ж є більш точним, ніж візуальна оцінка гістограми або полігону розподілу, бо він дозволяє зробити математично обґрунтований висновок про те, чи можуть дані у вибірці підлягати закону нормального розподілу.

Завдання 2. У двох племінних заводах була виміряна вага новонароджених телят чорно-строкатої породи, кг:

ПЗ «Маяк-АГРО»:

27 32 32 31 32 28 37 35 26 28
32 39 34 30 37 26 27 40 35 37
28 43 26 35 45 26 35 32 32 35



Теля (*Bos taurus taurus*)

ПЗ «Торосово»:

35 28 32 36 32 36 37 43 28 31
36 33 33 28 23 26 34 32 36 27
32 39 30 30 36 38 24 32 30 31
28 36 36

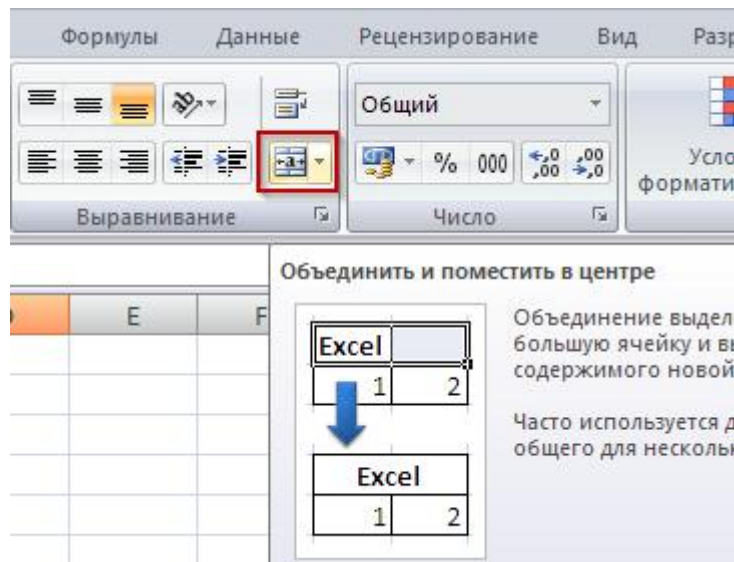
Розрахувати значення середнього арифметичного та стандартного відхилення вибірових сукупностей. Відобразити результати описового статистичного аналізу у графічній формі.



Відображення результатів описового статистичного аналізу у графічній формі при використанні MS Excel передбачає наступний й алгоритм дій:

- Створення таблиці з даними.
- Обрахування значень точкових оцінок.
- Обрахування значень інтервальних оцінок.
- Побудова графіку за обрахованими значеннями точкових та інтервальних оцінок.

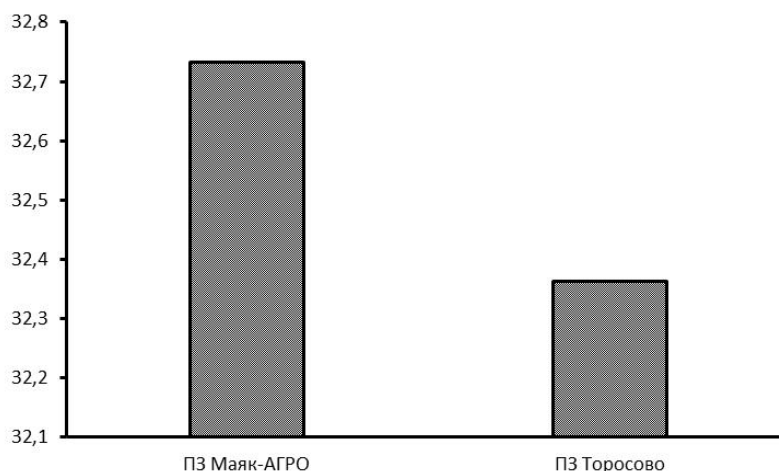
2.1. У комірку A1 уведіть назву вимірюваної величини «Вага новонароджених телят, кг». У комірки A1 та A2 – назви племінних заводів, у яких були проведені вимірювання ваги («ПЗ Маяк-АГРО» та «ПЗ Торосово» відповідно.). Виділіть комірки A1 та A2 та об'єднайте їх (вкладка на стрічці «Главная», панель інструментів «Выравнивание»):



Починаючи з комірок В1 та В2 уведіть дані про вагу новонароджених телят на обох племінних заводах.

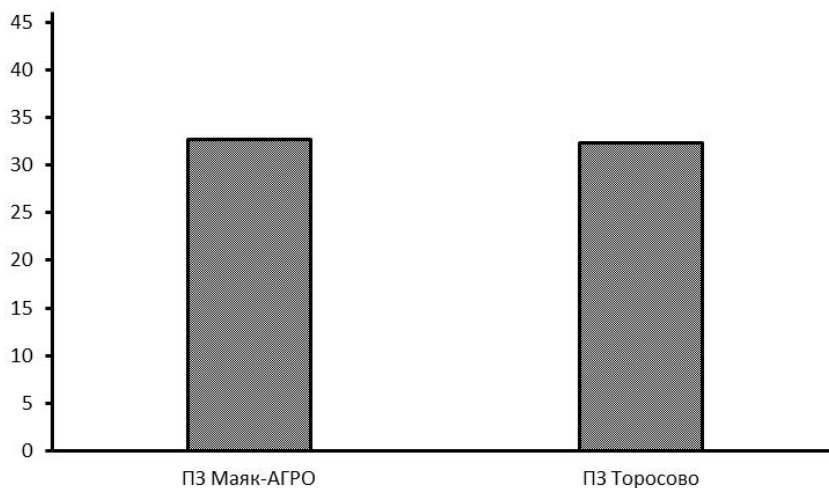
2.2. Додайте додатковий стовпчик шляхом клацання ЛКМ на заголовку стовпчика А. Після додавання ваші дані буде автоматично зміщено у діапазон комірок В1:С35. У комірки А36 та А37 уведіть текстові «Середня вага телят, кг» та «Стандартне відхилення». У комірках В36, В37 та С36, С37 обрахуйте значення відповідних показників.

2.3. Після обрахування середнього арифметичного та стандартного відхилення виділіть комірки В2:С2, та тримаючи натиснутою клавішу CTRL, виділіть комірки В36:С36. На стрічці оберіть вкладку «**Вставка**», та на панелі інструментів «**Діаграма**» оберіть вид графіку «**Гістограма**» та далі – «**Гістограма с группировкой**». Клацніть ЛКМ у будь-якій вільній ділянці робочого аркуша для вставки графіку.

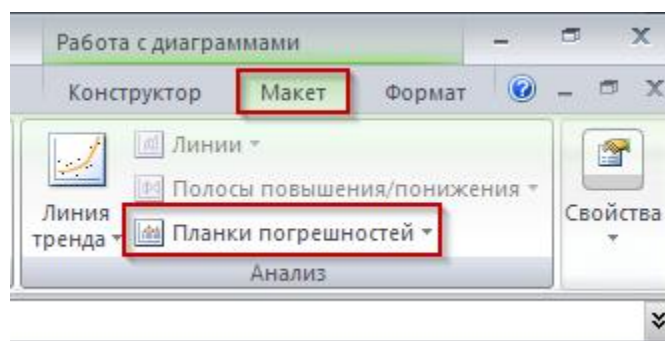


Зверніть увагу на те, що хоча різниця у вазі телят на двох племінних заводах складає лише 0,367 кг, на графіку вона виглядає суттєвою, що

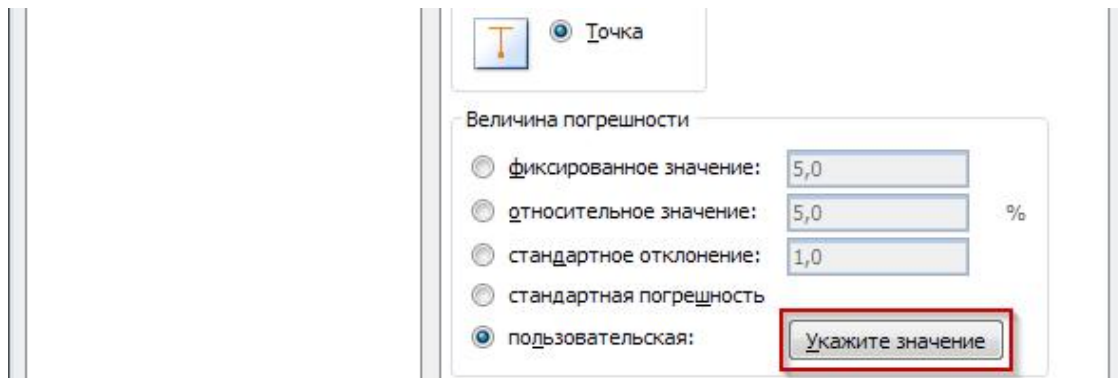
обумовлено невірним вибором масштабування. У переважній більшості графіків, на яких відображено біологічну або медичну інформацію, відлік прийнято вести від нуля. Для того, щоб змінити графік, клацніть ЛКМ по його вертикальній осі та оберіть у контекстному меню пункт «**Формат осі...**» та у діалоговому вікні, що з'явилось, змініть такі показники як «**Минимум**» на 0 та «**Максимум**» – на 46. Із такими налаштуваннями діаграма буде виглядати наступним чином:



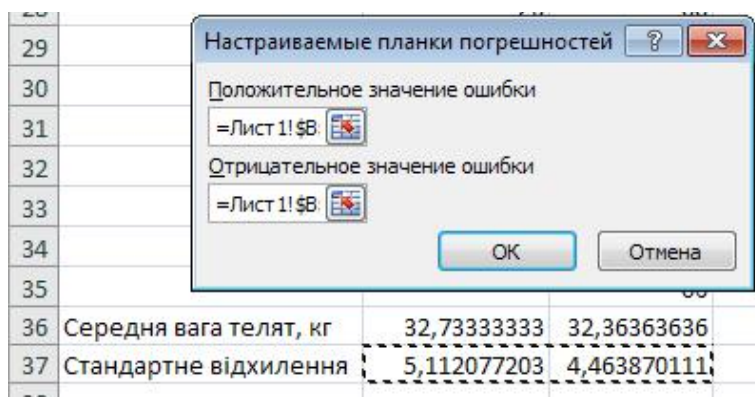
2.4. Наступним кроком є додавання до графіку показнику варіації ознаки, яким, у нашому випадку, є стандартне відхилення. Для цього потрібно виділити будь-який стовпчик на графіку, клацнувши по ньому ЛКМ. При цьому на стрічці з'являться три додаткові вкладки для роботи з діаграмами (**Конструктор**, **Макет** та **Формат**). Оберіть вкладку **Макет**, на ній панель **Аналіз** і далі: **Панки погрешностей** → **Дополнительные параметры планок погрешностей...**



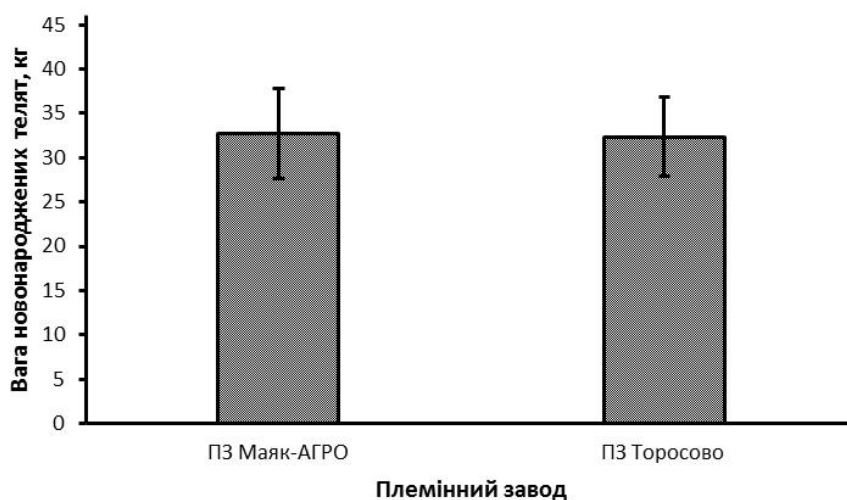
У діалоговому вікні, що з'явиться змініть значення групи «**Величина погрешности**» на «**Пользовательская**» та натисніть на кнопку «**Укажите значение**».



У наступному діалоговому вікні у якості позитивного та негативного значень погрешності вкажіть діапазон комірок B37:C37, які містять розраховані значення стандартного відхилення для обох вибірових сукупностей.



2.5. Відредагуйте отриманий графік: додайте підписи до осей, видаліть лінії сітки та заголовков. У результаті отриманий графік має виглядати наступним чином:



Отже, на отриманому графіку представлені не тільки середні значення ваги новонароджених телят, а й показники варіювання ознаки – стандартне відхилення.

Завдання 3. Була виміряна довжина правого вуха у 70 особин кроликів породи Рекс та породи Різен:

Порода Рекс:

12	12	12	11	11	12	14
11	12	11	10	12	11	12
13	14	13	15	13	14	14
14	13	10	12	11	12	11
10	13	12	11	12	14	10
13	14	13	11	12	11	16
13	13	12	13	14	13	11
13	14	11	13	16	12	12
14	12	12	12	12	14	12
12	15	14	15	14	11	12



Кроль
(*Oryctolagus cuniculus*)
породи Рекс

Порода Різен:

16	14	16	14	15	13	16
16	14	14	15	18	12	12
17	17	13	16	15	20	16
15	14	16	18	16	13	17
16	15	16	19	14	16	15
14	14	15	15	17	14	17
17	17	19	18	15	14	14
16	15	15	15	18	16	14
13	12	16	14	15	15	15
15	18	17	16	16	16	16



Кроль
(*Oryctolagus cuniculus*)
породи Різен

Розрахуйте значення середнього арифметичного та 95% довірчій інтервал для даних вибіркової сукупності. Відобразіть результати описового статистичного аналізу у графічній формі.

3.1. Створіть таблицю із даними, де у комірці A1 уведіть текст «Довжина вуха кролів, см», у комірці A2 та B2 назви порід «Рекс» та «Різен» відповідно. Об'єднайте комірці A1 та B1.

3.2. Діапазон комірок A3:B72 заповніть даними, наведеними у таблиці.

3.3. У комірках A73 та B73 розрахуйте середнє арифметичне значення довжини вуха кролів обох порід. У комірках A74 та B74 розрахуйте 95% довірчій інтервал для середнього арифметичного за допомогою формули: $\text{DOVERIT.NORM}(0,05;\text{СТАНДОТКЛОН.В}(\text{масив_даних});\text{СЧЕТ}(\text{масив_даних}))$

70	12	14
71	12	15
72	12	16
73	12,5	15,5
74	0,33	=ДОВЕРИТ.НОРМ(0,05;СТАНДОТКЛОН.В(B3:B72);СЧЁТ(B3:B72))
75		

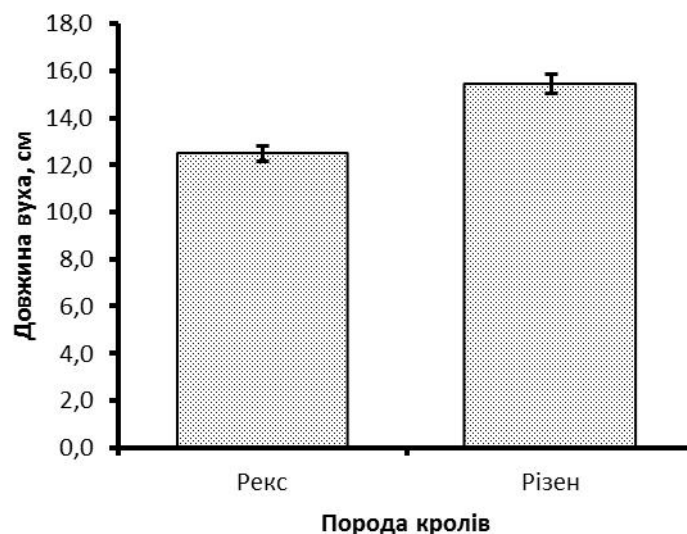
Значення 0,05 використовується для того, щоб отримати 95% (0,95) довірчій інтервал, та можуть використовуватись й інші значення, наприклад 0,01 для 99% довірчого інтервалу.



Довірчій інтервал (ДІ) – це 95% вірогідність того, що справжнє середнє знаходиться у межах \pm ДІ від визначеного, а верхнє та нижнє значення цього діапазону являють собою довірчі межі. Серед усіх інтервальних оцінок 95% довірчій інтервал є найбільш корисним показником розсіювання даних навколо середнього значення, та також є найбільш простим для розуміння.

Довірчій інтервал – найкраща статистика для порівняння різних наборів даних, а також при позначенні меж похибки на графіку. ДІ слід розраховувати кожного разу, коли обчислюють середнє значення, і звертатися до нього кожного разу, при оцінці даних. Якщо ДІ невеликий порівняно з середнім значенням, то середнє значення є надійним, але якщо ДІ значно відрізняється від середнього, то середнє значення є ненадійним.

3.4. За обрахованими показниками побудуйте графік, на якому відобразить середнє значення довжини вуха кролів обох порід, та позначте 95% довірчі інтервали:



Завдання для самостійного виконання.

1. Під час експерименту по відгодівлі баранів, 15 тварин отримували щодня в якості підгодівлі по 5 г фосфорного борошна (група А), а 15 інших тварин приблизно того ж віку, ваги і походження були контрольними (група Б). Добовий приріст був наступним (в г):

Груп а А	23 4	27 7	21 4	20 1	17 4	16 7	18 4	15 7	19 6	17 3	19 0	19 1	14 1	15 0	19 1
Груп а Б	18 3	15 4	17 5	15 9	15 7	18 9	19 8	16 5	17 6	12 4	17 3	18 2	20 4	15 1	14 7

Яким методом можна встановити, чи достовірна різниця між дослідною і контрольною групами за добовим приростом маси тіла? Визначте цю різницю і з'ясуйте, чи достовірна вона.

2. Отримані наступні дані про довжину крилів у представників 3 видів шпаків (*Sturnus*) з Індії (в мм):

<i>Sturnus contra</i>	120	120	121	122	125	126	126	125	122	123	122				
<i>Sturnus ginginiamus</i>	129	123	128	125	126	127	129	125	124	129	128	127	127		
<i>Sturnus fuscus</i>	122	127	125	128	127	122	129	127							

Обрахуйте середні арифметичні довжини крил шпаків та 95% довірчі інтервали до них. Зобразіть результати у графічній формі. Дайте відповідь, чи правдива різниця між середніми довжин крил трьох видів?

Контрольні питання.

1. Поясніть, що таке «точкові» та «інтервальні» оцінки вибіркової сукупності; у чому полягають відмінності між ними?
2. Розкрийте сутність середнього арифметичного вибіркової сукупності.
3. Визначте значення довірчого інтервалу для оцінки середнього арифметичного.
4. Порівняйте закономірності випадкової варіації вибіркової і генеральної сукупностей?
5. Якою мірою середня арифметична вибіркової сукупності характеризує середню арифметичну генеральної сукупності?
6. Поясніть, що характеризує середня помилка? Яка її формула?

7. Вкажіть межі, у яких може перебувати середня арифметична генеральної сукупності (μ) по відношенню до \bar{X} вибіркової сукупності. З якою вірогідністю?

ЛАБОРАТОРНЕ ЗАНЯТТЯ 3

ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ ІЗ ЗАСТОСУВАННЯМ ПАРАМЕТРИЧНИХ ТА НЕПАРАМЕТРИЧНИХ КРИТЕРІЇВ


Мета: ознайомитись із загальними принципами перевірки статистичних гіпотез; опанувати базові методи порівняння груп із застосуванням параметричного (Т-критерій) та непараметричного (U-критерій) критеріїв.

Обладнання та матеріали: ПЕОМ із встановленою ОС Microsoft Windows XP або Windows Seven, процесор таблиць Microsoft Excel 2007 або 2010, Інтернет браузер, методичні вказівки.

План:

1. Формулювання нульової та альтернативної гіпотез.
2. Перевірка статистичних гіпотез із використанням параметричних критеріїв.
 - 2.1. Т-критерій для незалежних вибірок.
 - 2.2. Т-критерій для залежних вибірок.
3. Непараметричний критерій Мана-Уїтні.

Теоретичні відомості.

 Загальні принципи порівняння відмінностей між групами ґрунтуються на аналізі так званої **нульової гіпотези**. Відповідно до цієї гіпотези, спочатку приймається припущення, що між даними показниками (або групами, на основі яких вони отримані) достовірної відмінності немає, тобто **обидві ці групи разом складають один і той же однорідний матеріал, одну сукупність**. Статистичний аналіз повинен привести або до відхилення нульової гіпотези, якщо доведена достовірність отриманих відмінностей, або до її збереження, якщо достовірність відмінностей не доведена, тобто виявлені відмінності визнаються **випадковими**. Але, так як всі статистичні показники та відмінності між ними характеризуються певними рівнями значущості, то відкидання нульової гіпотези має бути пов'язане з прийняттям певного рівня значущості (P). Так, якщо необхідним визнаний рівень значущості 0,01 і якщо ймовірність достовірності (довірча імовірність) даного статистичного показника, або різниці між показниками, не задовольняє цій умові, тобто вона нижче 0,99 (наприклад, 0,97, 0,91 або 0,88), то немає підстав для відкидання нульової гіпотези. Її треба вважати правильною принаймні до тих пір, поки нові дані не дадуть можливості її спростувати, довівши, що існуючі відмінності не є чисто випадковими.

Звичайно, і у тому випадку, коли нульова гіпотеза вважається спростованою, якийсь шанс, що вона в дійсності вірна, залишається. При рівні значущості 0,01 цей шанс становить 1 на 100, тобто у 1% випадків відкидання нульової гіпотези буде помилкою. Якщо досягнуто рівня значимості не 0,01, а 0,001, то впевненість у тому, що нульова гіпотеза дійсно відкинута правильно, різко зростає (лише 1 шанс на 1000 випадків, що вона все ж вірна). При $P = 0,05$

впевненість у правильності виведення становить лише 95 випадків з 100, а у 5 випадках можливий неправильний висновок.

Таким чином, якщо отримані дані характеризуються рівнем значущості $P > 0,05$, то немає підстав відхиляти нульову гіпотезу. Якщо $P < 0,01$, то для відкидання нульової гіпотези підстави достатні. Але значно невизначений стан речей, якщо результати аналізу або порівняння задовольняють рівню значущості 0,05, але не задовольняють рівню значущості 0,01. Надійне судження виявляється неможливим. Очевидно, що у таких випадках повинні бути проведені додаткові дослідження, щоб вирішити, чи слід відкинути нульову гіпотезу. Взагалі треба мати на увазі, що збереження нульової гіпотези ще не означає її правильності. Може виявитися все ж, що вона неправильна. Збереження ж нульової гіпотези залишає питання відкритим.

Якщо була отримана різниця між середніми арифметичними двох генеральних сукупностей, то, очевидно, не може стояти питання про статистичну помилку цієї різниці. Ця різниця завжди достовірна, навіть якщо вона і дуже мала. Інша річ, якщо порівнюються дві вибіркові сукупності, наприклад: дві групи морських свинок, яких піддавали впливу хімічних речовин або фізичних факторів, дві групи корів, порівнювані за масою і т.п. У цих випадках різниця між середніми має свою статистичну помилку, з якою її можна порівняти і встановити, чи достовірна ця різниця чи ні. Нульова гіпотеза в даному випадку буде зводитися до того, що дві досліджувані вибіркові сукупності походять з однієї і тієї ж генеральної сукупності і що різниця між їх середніми арифметичними випадкова, тобто лежить у межах помилки вибірковості. **Щоб мати право відкинути нульову гіпотезу, треба довести, що різниця між середніми арифметичними достовірна, тобто задовольняє необхідному рівню значущості.**

Хід виконання роботи.

Завдання 1. Досліджували вплив двох видів комплексних добрив («Гуміноорм» та «Фітомакс») на врожайність картоплі (кг/м²) на 20 дослідних ділянках. Були отримані наступні результати:

Гуміноорм	27	20	16	18	22	19	23	21	17	19
Фітомакс	28	19	18	21	24	20	25	27	29	21



Картопля
(*Solanum tuberosum*)

Визначити, чи якийсь з цих двох добрив збільшує врожайність картоплі у більшій мірі, ніж інше?



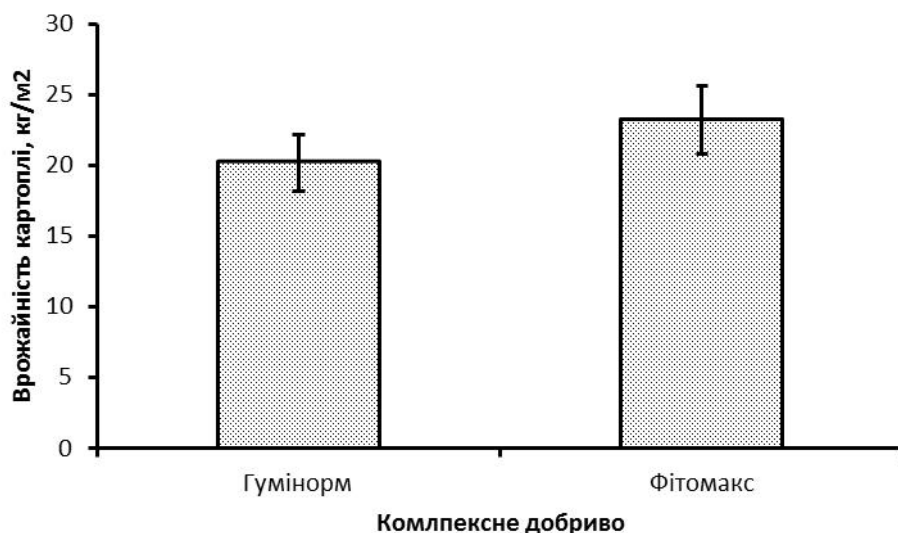
Порівняння відмінностей між вибірковими сукупностями у MS Excel передбачає наступний загальний алгоритм дій:

- Визначення типу вибірок (незалежні або залежні).
- Визначення типу даних.
- Формулювання нульової та альтернативної гіпотез.
- Вибір належного статистичного критерію.
- Перевірка нульової гіпотези на прийнятому рівні статистичної значимості за допомогою обраного статистичного критерію
- Інтерпретація результатів та формулювання висновку.

1.1. Створіть таблицю з даними. У комірку B1 уведіть текст «Врожайність картоплі, кг/м²» та поєднайте її з коміркою C1, у комірки B2 та C2 уведіть назви добрив: «Гуміноорм» та «Фітомакс» відповідно. У діапазон комірок уведіть дані про врожайність картоплі B3:C12.

1.2. У комірки A13 та A14 уведіть текст «Середнє значення» та «95% ДІ», та обрахуйте відповідні показники у комірках B13:C13 та B14:C14. Виділіть вміст комірок A13:C14 червоним кольором.

1.3. Побудуйте графік, на якому відображено середнє значення врожайності картоплі за умов внесення зазначених добрив та 95% довірчі інтервали за наступним зразком:



Середня врожайність картоплі за умов внесення комплексного добрива «Фітомакс» відповідно до графіку є майже на 15% більшою, ніж за умов використання добрива «Гуміноорм». Але чи дійсно «Фітомакс» збільшує врожайність картоплі?

1.4. Визначте, чи існують **статистично значимі** відмінності у врожайності картоплі за умов внесення комплексних добрив на прийнятому рівні

статистичної значущості. Сформулюємо нульову гіпотезу (H_0): «**Вплив кожного з добрив на врожайність картоплі однаковий**». Альтернативна гіпотеза (H_1) у такому випадку буде сформульована наступним чином: «**За умов внесення одного з добрив врожайність картоплі більша, ніж за умов внесення іншого**». Рівень статистичної значущості прийемо на рівні 5% ($p = 0,05$), що є достатнім для досліджень подібного роду. Отже, дуже спрощуючи: якщо обраховане значення статистичного критерію буде меншим за 0,05, то нульова гіпотеза буде відхилена і підтверджена альтернативна, якщо більшим за 0,05 – то вірною залишиться нульова гіпотеза.

Для перевірки висунутих гіпотез застосуємо параметричний Т-критерій (відомий як Т-критерій Ст'юдента). Для розрахунку даного критерію у Excel є формула ТТЕСТ, яка має наступні аргументи ТТЕСТ(масив1; масив2; хвости; тип), де масив1 та масив2 – дані вибірок сукупностей для порівняння; хвости – може приймати значення 1 (для одностороннього розподілу) та 2 (для двостороннього розподілу); тип – може приймати значення 1 (для залежних вибірок), 2 (для незалежних вибірок) та 3 (для незалежних вибірок із нерівними дисперсіями).

Отже, у комірку A15 уведіть текст « $p =$ », а у комірку B15 – формулу для обрахування Т-критерію (=ТТЕСТ(B3:B12;C3:C12;2;3)):

11		17	29
12		19	21
13	M	20,2	23,2
14	95% ДІ	2,00	2,44
15	$p =$	=ТТЕСТ(B3:B12;C3:C12;2;3)	
16			

Отриманий результат ($p = 0,0794$) є більшим за прийнятий рівень статистичної значущості (0,05 або 5%). Інакше кажучи, вірогідність, що дані у вибірках не є дійсно різними становить майже $(0,0794 * 100) = 8\%$, що більше за обраний нами рівень статистичної значущості.

Отже, ми не можемо підтвердити альтернативну гіпотезу та відхилити нульову. Таким чином висновок, підтверджений результатом статистичного аналізу, буде наступним: «**Вплив кожного з добрив на врожайність картоплі однаковий**», або, інакше кажучи, ми не можемо стверджувати що будь-яке з добрив суттєво збільшує врожайність картоплі, ніж інше; виявлені варіації врожайності – випадкові.

Завдання 2. У вісьмох осіб двічі була виміряна частота пульсу: до та після прийому великої кількості їжі:

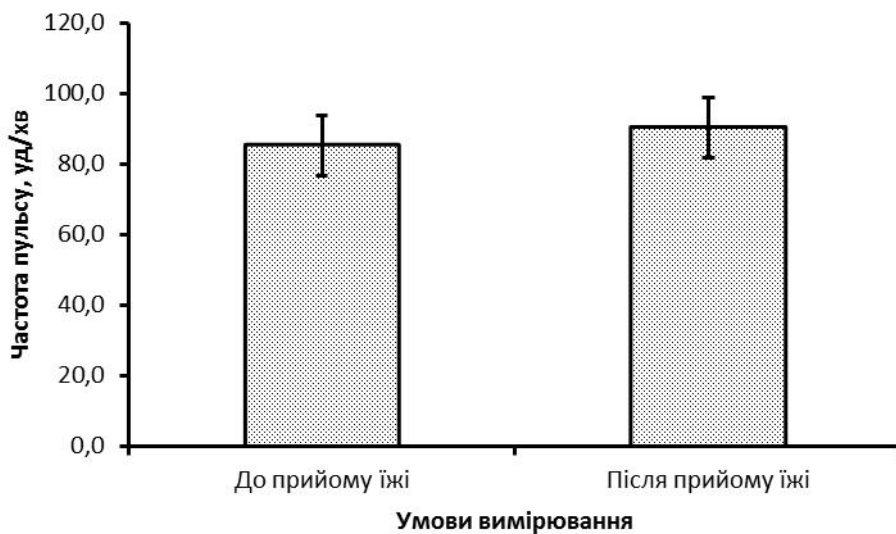
Умови вимірювання	Частота пульсу, уд/хв.							
	До прийому їжі	105	79	79	103	87	74	73
Після прийому їжі	109	83	86	109	90	78	78	89

Встановити, чи впливає прийом великої кількості їжі на частоту пульсу?

2.1. Створіть таблицю з даними. У комірку B1 уведіть текст «Частота пульсу, уд/хв» та поєднайте її з коміркою C1, у комірки B2 та C2 уведіть умови визначення частоти пульсу: «До прийому їжі» та «Після прийому їжі» відповідно. У діапазон комірок B3:C10 уведіть дані про частоту пульсу.

2.2. У комірки A11 та A12 уведіть текст «Середнє значення» та «95% ДІ», та обрахуйте відповідні показники у комірках B11:C11 та B12:C12. Виділіть вміст комірок A13:C14 червоним кольором.

2.3. Побудуйте графік, на якому відображено середнє значення частоти пульсу у осіб до і після прийому їжі:



Середнє значення частоти пульсу у осіб після прийому їжі збільшено лише на 5,9%. То ж чи дійсно прийом їжі впливає на частоту пульсу? Чи навпаки, не впливає?

Сформулюємо нульову гіпотезу: «**Прийом їжі не змінює частоту пульсу**», та альтернативну – «**Прийом їжі змінює частоту пульсу (збільшує)**». Рівень статистичної значущості (P) прийемо на рівні 0,05 ($P = 0,05$).

Для перевірки висунутих гіпотез знову застосуємо параметричний Т- критерій. Для цього у комірку A13 уведіть текст «р =», а у комірку B13 – формулу для обрахування Т-критерію =ТТЕСТ(B3:B10;C3:C10;2;1). **Зверніть увагу**, що тип тесту (останній аргумент функції) змінився на «1», тому що у даному випадку ми маємо справу із парною (синоніми – пов'язаною, залежною) вибіркою.

Оцініть результат тесту. $P = 0,00003$, отже вірогідність, що дані у вибірках не є дійсно різними, складає лише $0,00003 * 100 = 0,003\%$, що значно нижче, за обраний нами рівень статистичної значимості ($p = 0,05$, або 5%). Таким чином ми можемо відхилити нульову гіпотезу і прийняти альтернативну «**Прийом їжі змінює частоту пульсу (збільшує її)**».

Завдання 3. Визначали врожайність двох сортів пшениці (*Triticum turgidum* та *Triticum polonicum*) при культивуванні у засушливих умовах.

Сорт пшениці	Врожайність, ц/г						
<i>Triticum turgidum</i>	15	19	20	18	21	25	31
<i>Triticum polonicum</i>	16	17	22	21	24	23	28



Пшениця
(*Triticum spp*)

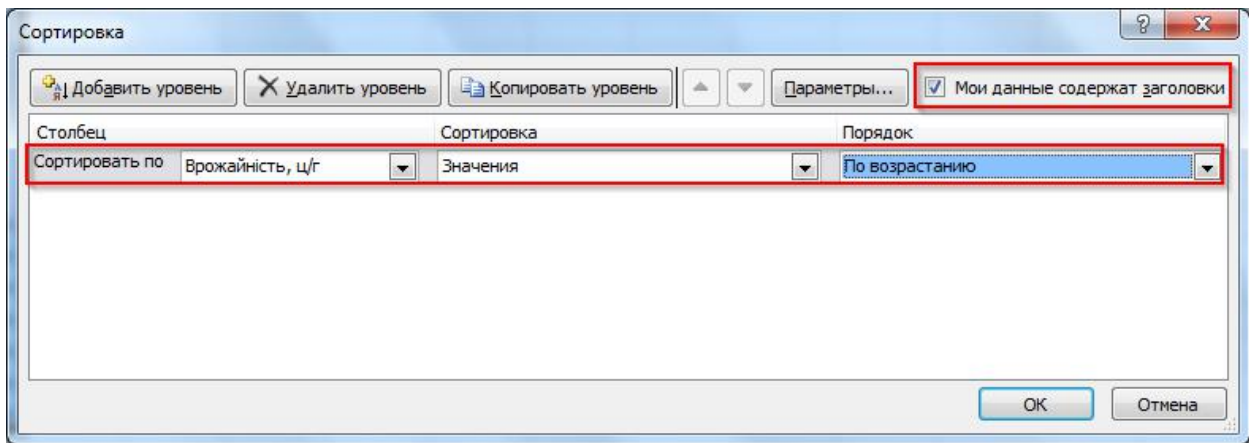
Визначите за допомогою U-критерію Мана-Уїтні, чи відрізняється врожайність вищезазначених сортів пшениці.

3.1 Створіть робочу таблицю із даними. У комірку A1 уведіть текст «Врожайність пшениці», у комірку B2 уведіть текст «Група». Отже, комірки у стовпчику будуть містити значення окремих варіант, а комірки стовпчику B – так звану **групувальну змінну**.

У комірки A2:A8 уведіть дані про врожайність сорту *Triticum turgidum*, а у комірки B2:B8 – назву сорту «*Triticum turgidum*». Відповідні дії виконайте для іншого сорту пшениці («*Triticum polonicum*»), увівши дані у діапазон комірок A9:A15 та B9:B15 відповідно.

	A	B
1	Врожайність, ц/г	Група
2	15	<i>Triticum turgidum</i>
3	19	<i>Triticum turgidum</i>
4	20	<i>Triticum turgidum</i>
5	18	<i>Triticum turgidum</i>
6	21	<i>Triticum turgidum</i>
7	25	<i>Triticum turgidum</i>
8	31	<i>Triticum turgidum</i>
9	16	<i>Triticum polonicum</i>
10	17	<i>Triticum polonicum</i>
11	22	<i>Triticum polonicum</i>

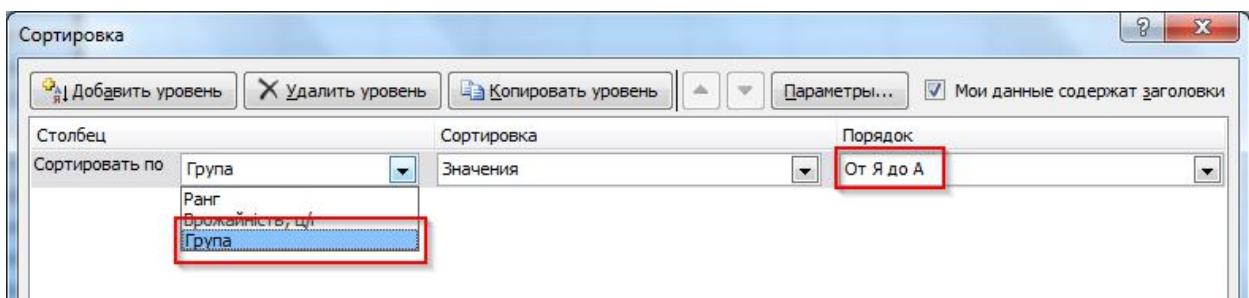
3.2. Наступним кроком є ранжування даних. Для цього спочатку відсортуйте дані про врожайність обох сортів пшениці за зростанням. Виділіть діапазон комірок A1:B15 та на стрічці табличного процесору Excel оберіть вкладку «Данные», панель інструментів «Сортировка и фильтр» кнопка «Сортировка»:



Після того, як ваші дані будуть відсортовані, додайте новий стовпчик А до таблиці; у комірку А1 уведіть текст «Ранг», та за допомогою автозаповнення надайте відсортованим значенням ранг починаючи з 1 та з кроком, який також дорівнює 1. Отже, дані у таблиці повинні виглядати наступним чином:

	A	B	C
1	Ранг	Врожайність, ц/г	Група
2	1	15	Triticum turgidum
3	2	16	Triticum polonicum
4	3	17	Triticum polonicum
5	4	18	Triticum turgidum
6	5	19	Triticum turgidum
7	6	20	Triticum turgidum
8	7	21	Triticum turgidum
9	8	21	Triticum polonicum
10	9	22	Triticum polonicum
11	10	23	Triticum polonicum
12	11	24	Triticum polonicum
13	12	25	Triticum turgidum
14	13	28	Triticum polonicum
15	14	31	Triticum turgidum

3.3. Далі, розподіліть ранжовані дані у дві групи. Для цього знову відсортуйте дані, але за значенням групувальної змінної «Група» (стовпчик С):



Отже, таблиця із даними прийме наступний вид:

	A	B	C
1	Ранг	Врожайність, ц/г	Група
2	1	15	Triticum turgidum
3	4	18	Triticum turgidum
4	5	19	Triticum turgidum
5	6	20	Triticum turgidum
6	7	21	Triticum turgidum
7	12	25	Triticum turgidum
8	14	31	Triticum turgidum
9	2	16	Triticum polonicum
10	3	17	Triticum polonicum
11	8	21	Triticum polonicum
12	9	22	Triticum polonicum
13	10	23	Triticum polonicum
14	11	24	Triticum polonicum
15	13	28	Triticum polonicum

3.4. Наступний крок – підрахунок суми рангів у кожній групі. Для цього, додайте новий рядок до таблиці із даними, клацнувши ПКМ по номеру рядка 9, та оберіть у контекстному меню пункт «Вставити». У комірці A9 обчисліть суму рангів (R1) = СУММ(A2:A8), а у комірці B9 – кількість варіант у групі (N1) = СЧЕТ(A2:A8). Відповідно у комірках A17 та B17 обчисліть суму рангів для іншого сорту пшениці.

3.5. Далі, у комірку A18 уведіть позначення критерію – літеру «U» а у комірці B18 обчисліть значення критерію Мана-Уїтні за відомою формулою:

$$U = N1 * N2 + ((N1 * (N1 + 1)) / 2) - R1$$

15	11	24	Triticum polonicum		
16	13	28	Triticum polonicum		
17	56	7			
18	U=	=B9*B17+((B9*(B9+1)/2))-A9			
19					

3.7 Наступний етап – обчислення у комірці A19 середнього арифметичного (μU) та у комірці A20 – стандартного відхилення (σU), за наступними формулами:

$$\mu U = N1 * N2 / 2$$

$$\sigma U = \text{КОРЕНЬ} ((N1 * N2 * (N1 + N2 + 1)) / 12)$$

3.8. У комірці обчисліть Z оцінку за наступною формулою:

$$Z \text{ оцінка} = (U - \mu U) / \sigma U$$

15	11	24	Triticum polonicum		
16	13	28	Triticum polonicum		
17	56	7			
18	U=	28			
19	$\mu U=$	24,5			
20	$\sigma U=$	7,826			
21	Z=	0,447			
22					

3.9. Останній крок при порівнянні груп за допомогою критерію Мана-Уїтні полягає у порівнянні обрахованої Z оцінки з Z критичним для заданого рівня статистичної значимості α .

Для тесту із однобічним розподілом Z критичне =НОРМСТОБР (α), а для тесту із двобічним розподілом Z критичне =НОРМСТОБР ($\alpha/2$).

Якщо потрібен 95% рівень довірчої імовірності, тоді рівень статистичної значимості = 0,05; тоді для тесту з однобічним розподілом, Z критичне буде обчислюватись за формулою: =НОРМСТОБР (α) = НОРМСТОБР (0,05), а для тесту з двобічним розподілом, Z критичне обчислюється за формулою: =НОРМСТОБР ($\alpha/2$) = НОРМСТОБР (0,025).

18	U=	28		
19	$\mu U=$	24,5		
20	$\sigma U=$	7,826		
21	Z=	0,447		
22	Zкр.1бічн.=	-1,644853627		
23	Zкр.2бічн.=	-1,959963985		
24				



Результат статистичного тесту є значущим у випадку якщо

$$| Z \text{ оцінка} | > | Z \text{ критичного} |$$

У нашому випадку $| Z \text{ оцінка} | < | Z \text{ критичного} |$ (1,96 для двобічного розподілу та 1,64 – для однобічного). Отже, результат не є статистично значимим, і ми не можемо стверджувати із 95% ступенем довірчої вірогідності, що є відмінності між цими двома групами. Таким чином остаточний висновок такий: «**При культивуванні у засушливих умовах врожайність сортів пшениці *Triticum turgidum* та *Triticum polonicum* не відрізняється**».

Завдання для самостійного виконання.

1. Були отримані наступні дані про масу тіла тушканчиків (*Dipus aegyptius*):

Самці	186	190	165	182	182	182	180	173	157	179	164	146	173
	144	156	156	165	160	160	161	144	153	152	151	173	
Самиці	162	163	190	188	147	146	145	157	162	186	175	147	164
	145	145	155	174	180	148	175	145	144	153	165	141	

Чи відрізняються за вагою самці від самок? Сформулюйте нульову гіпотезу та зробіть статистично обґрунтований висновок.

2. Були отримані наступні дані про температуру тіла тушканчиків (*Dipus aegyptius*):

Самці	37,4	37,5	37,9	37,8	36,8	37,8	37,5
Самиці	37,0	37,8	38,1	37,5	37,7	37,8	37,6

Чи відрізняються за температурою тіла самці від самок? Сформулюйте нульову гіпотезу та зробіть статистично обґрунтований висновок.

Контрольні питання.

1. Поясніть сутність нульової гіпотези та наведіть приклади.
2. Яким чином оцінюється достовірність різниці між середніми арифметичними? Чи однакові способи оцінки при малих і великих n ?
3. Наведіть приклад формулювання нульової гіпотези при порівнянні двох середніх арифметичних.
4. Поясніть, у чому перевага попарного порівняння даних? Наведіть приклади з біології.
5. Вкажіть відмінності між графічним та математичним методами порівняння середніх арифметичних.
6. У чому полягає нульова гіпотеза при порівнянні дисперсій?
7. Чи можна вважати достовірним відмінність між середніми арифметичними, якщо фактичне значення T більше критичного? Якщо воно менше критичного?

ЛАБОРАТОРНЕ ЗАНЯТТЯ 4

ДИСПЕРСІЙНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

Мета: лабораторне заняття передбачає засвоєння методів проведення однофакторного та двофакторного дисперсійного аналізу із використанням табличного процесору MS Excel.

Обладнання та матеріали: ПЕОМ із встановленою ОС Microsoft Windows XP або Windows Seven, процесор таблиць Microsoft Excel 2007 або 2010, Інтернет браузер, методичні вказівки.

План:

1. Однофакторний дисперсійний аналіз.
2. Двофакторний дисперсійний аналіз.

Теоретичні відомості.



Аналізуючи біологічні особливості тварин і рослин в природі або при розведенні їх в культурі, легко переконатися в їх постійній залежності від багатьох зовнішніх і внутрішніх факторів. Так, миші варіюють в своєму зростанні і розвитку в залежності від розмірів їхніх батьків, від часу народження, від числа дитинчат у пометах, в яких вони народилися, від погодних умов, від кількості одержуваної ними їжі і т.д. Вага телят при народженні і в наступні періоди їхнього життя визначається спадковими особливостями їх матерів і батьків, умовами ембріонального розвитку, рівнем годівлі та утримання. Все це створює величезне різноманіття тварин і рослин за їх морфологічними і фізіологічними властивостями не тільки в природних умовах, але навіть при розведенні їх людиною у відносно однорідних умовах.

Така складна обумовленість біологічних властивостей організмів привела до необхідності розробки відповідних математичних методів, за допомогою яких можна було б виділити вплив окремих факторів і оцінити їх відносну роль в загальній мінливості цих властивостей. До їх числа відноситься дисперсійний аналіз, який на даний час являє собою самостійну і дуже важливу галузь біологічної статистики. Сутність його полягає у встановленні ролі окремих факторів в мінливості тієї чи іншої ознаки.

Справа в тому, що вплив тих чи інших факторів на досліджувану ознаку (або ознаки) ніколи не може бути виділено в чистому вигляді. Хоча при проведенні дослідів і намагаються зберегти умови максимально однорідними, все ж різні досліді дають дещо неоднакові результати. Це пояснюється тим, що на них впливають численні випадкові обставини, багато інших чинників, які змінюються від досліді до досліді і не піддаються контролю. Тим більше є великою роль таких додаткових неконтрольованих факторів при проведенні аналізу не в експериментальних умовах, а безпосередньо в природних умовах.

Тому виникає важливе завдання розкладання загальної мінливості ознаки

на складові частини, які з одного боку визначаються досліджуваними конкретними факторами, а з іншого – викликаються випадковими, неконтрольованими причинами. Дисперсійний аналіз дозволяє оцінювати значимість впливу окремих факторів, а також їх відносну роль у загальній мінливості.

Хід виконання роботи.

Завдання 1. Для перевірки впливу гучності сигналу на швидкість реакції на швидкість реакції випадковим чином відібрали 3 групи досліджуваних. Першій групі (5 осіб) пред'являли звукових сигнал у 10 дБ, другій (6 осіб) – 30 дБ, третьої (4 особи) – 50 дБ. У досліджуваних кожної групи фіксували час реакції у мілісекундах.

Номер групи	1 (10дБ)	2 (30дБ)	3 (50дБ)
	304	272	223
	268	264	184
Результати	272	256	209
вимірювань, мс	262	269	183
	283	285	
		247	

Сформулюйте статистичні гіпотези (H_0 та H_A) за даними умовами та перевірте їх.



Для проведення дисперсійного аналізу у MS Excel використовується процедури «Однофакторный дисперсионный анализ» та «Двухфакторный дисперсионный анализ» застосування яких передбачає наступний й алгоритм дій:

- Створення таблиці з даними.
- Формулювання H_0 та H_A гіпотез (H_0^A та H_0^B – у випадку двофакторного дисперсійного аналізу).
- Виклик надбудови «Анализ данных».
- Встановлення необхідних умов проведення статистичного тесту.
- Інтерпретація результатів статистичного тесту.
- Відхилення чи прийняття H_0 гіпотези на прийнятому рівні статистичною значимості за результатами статистичного тесту.

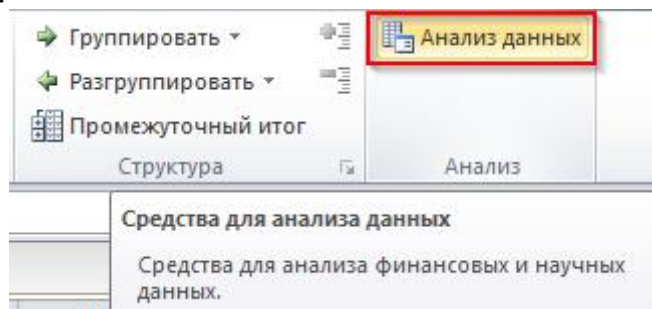
1.1. Сформулюємо нульову (H_0) та альтернативну (H_A) гіпотези:

H_0 – Гучність сигналу впливає на швидкість реакції.

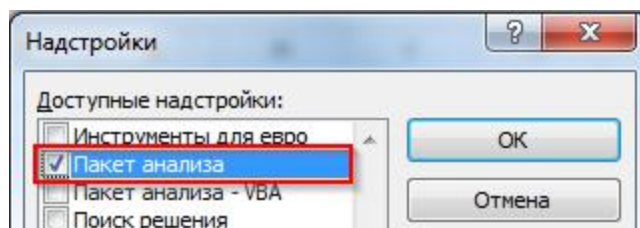
N_A – Гучність сигналу не впливає на швидкість реакції.

Для перевірки вищезазначених припущень використаємо однофакторний дисперсійний аналіз. Для цього уведіть дані у відповідну таблицю (починаючи із комірки A1), таким чином, щоб у одному стовпчику опинились дані, які відповідають одному значенню досліджуваного фактора, а стовпчики розташовувались у порядку зростання (зменшення) фактора, що досліджується.

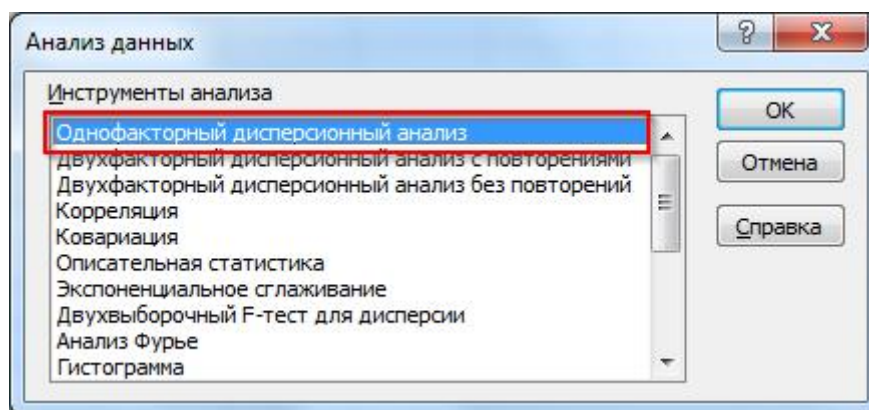
1.2. Далі, на стрічці MS Excel активуйте вкладку «Данные» та натисніть на кнопку «Анализ».



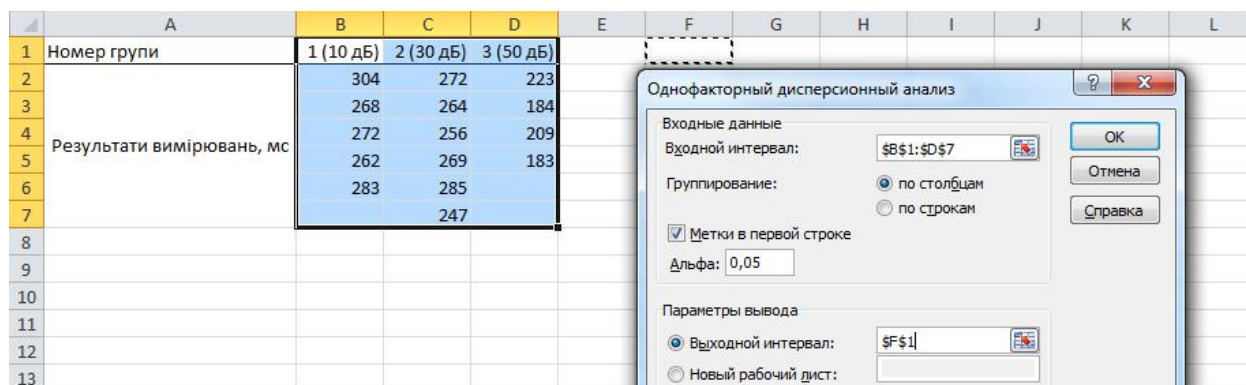
Якщо ця вкладка відсутня, зробіть наступні кроки: вкладка «Файл» на стрічці, пункт «Параметры» та у діалоговому вікні, що з'явиться оберіть пункт «Надстройки» і далі – «Пакет анализа» та натисніть кнопку «Перейти». У діалоговому вікні, що з'явиться зробіть відмітку на пункті «Пакет анализа» та натисніть кнопку «ОК»



1.3. У списку «Инструменты анализа» діалогового вікна «Анализ данных» оберіть процедуру «Однофакторный дисперсионный анализ», після чого натисніть кнопку «ОК»;



1.4. У діалоговому вікні, що з'явилося задайте вхідний інтервал (комірки B1:D7). У розділі «Группирование» встановити перемикач у положення «**по столбцам**». Встановіть позначку «**Метки в первой строке**» та вкажіть вихідний діапазон, тобто те місце в якому будуть виводитись результати аналізу, для чого встановіть перемикач у положення «**Выходной интервал**», клацніть ЛКМ по кнопці із зображенням таблиці у діалоговому вікні та оберіть комірку F1. Натисніть кнопку «**ОК**».



У результаті цих дій вихідний діапазон буде включати в себе результати дисперсійного аналізу: середні значення, дисперсії, критерій Фішера та інші показники.

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
1 (10 дБ)	5	1389	277,8	273,2		
2 (30 дБ)	6	1593	265,5	173,9		
3 (50 дБ)	4	799	199,75	384,9166667		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	15341,88	2	7670,941667	29,53154425	0,0000232	3,885293835
Внутри групп	3117,05	12	259,7541667			
Итого	18458,93	14				

1.5. **Інтерпретація результатів.** Вплив фактора, що досліджується, визначається за величиною значимості критерію Фішера, яка знаходиться у таблиці «Дисперсионный анализ» на перетині рядка «Между группами» и столбца «P-Значение». У випадках, коли P-Значение < 0,05, критерій Фішера є значимим, і вплив досліджуваного фактору можна вважати доведеним на визначеному рівні довірчої вірогідності (у нашому випадку – 0,95).

Завдання 2. Досліджували процес росту бактерій у рідкому культуральному середовищі при трьох різних значеннях рН та чотирьох різних значеннях температури. Усі дослідні зразки знаходились у 12 окремих

пробірках, ступінь росту визначалась шляхом вимірювання оптичної щільності культури (табл. 4.1).

Таблиця 4.1 – Оптична щільність культури бактерій за умов культивування при різних значеннях температури та рН

Температура, °С	рН 5,5	рН 6,5	рН 7,5
25	10	19	40
30	15	25	45
35	20	30	55
40	15	22	40

Визначити, чи впливає температура культивування та рН культурального середовища на інтенсивність росту бактерій?

Визначена задача полягає у дослідженні впливу двох факторів (температури та рН) на зміну значень випадкової величини x (інтенсивність росту бактеріальної культури). Різні значення факторів називають рівнями факторів. Для кожного набору рівнів факторів проводиться одне вимірювання, отже необхідно застосовувати двофакторний дисперсійний аналіз без повторень (у іншому випадку потрібна процедура двофакторного дисперсійного аналізу з повтореннями).



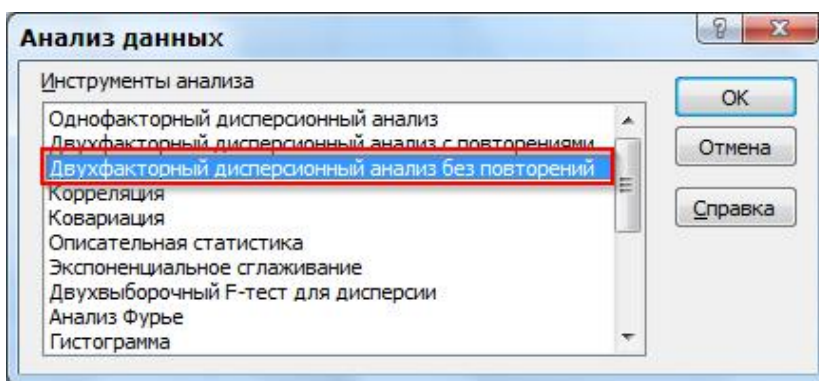
Дані для задач однофакторного дисперсійного аналізу зазвичай формуються у прямокутну таблицю, кожний рядок якої – дані про зміну випадкової величини за рівнями першого фактора (A), стовпчики – дані про зміну випадкової величини за рівнями другого фактора (B). Задача двофакторного дисперсійного аналізу у тому, щоб встановити, чи є фактори (кожний окремо або обидва разом) причиною мінливості випадкової величини. Далі формулюються дві нульових гіпотези, для кожного фактора своя, про те, що математичні очікування випадкових величин за дії різних рівнів фактора однакові. Якщо нульова гіпотеза не відхиляється, то відповідний фактор не впливає на мінливість, а у протилежному випадку – впливає.

2.1. Сформулюйте нульові гіпотези для вищезазначеної задачі. Останні можуть бути сформульовані наступним чином H_0^A : «Температура не впливає на інтенсивність росту бактеріальної культури», H_0^B «рН не впливає на інтенсивність росту бактеріальної культури»; альтернативні гіпотези H_0^A – «Температура впливає на інтенсивність росту бактеріальної культури», H_0^B – «рН впливає на інтенсивність росту бактеріальної культури».

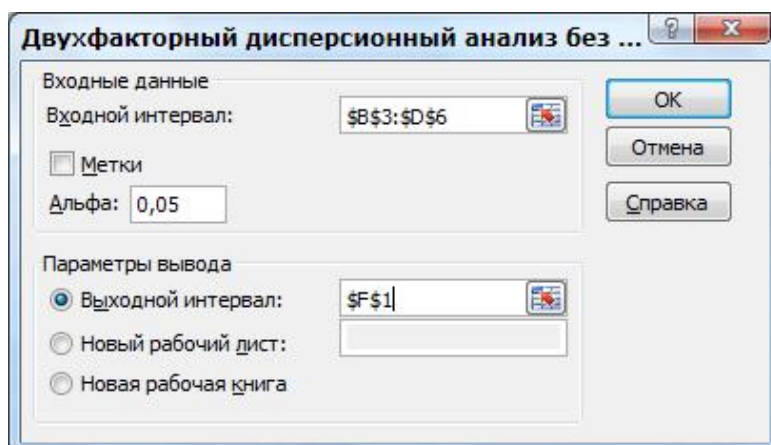
2.2. Створіть таблицю із даними.

	A	B	C	D	E
1		Оптическая щільність, ум.од.			
2	Температура, °C	pH 5,5	pH 6,5	pH 7,5	
3	25	10	19	40	
4	30	15	25	45	
5	35	20	30	55	
6	40	15	22	40	

2.3. На вкладці «Данные» стрічки MS Excel оберіть панель «Анализ» та натисніть на кнопку «Анализ данных». У діалоговому вікні «Анализ данных» оберіть пункт «Двухфакторный дисперсионный анализ данных без повторений» та натисніть на кнопку «ОК».



2.4. У діалоговому вікні, що з'явилося вкажіть вхідний інтервал (комірки B3:D6), вихідний інтервал (комірку F1) та натисніть на кнопку «ОК».



Результати аналізу будуть виведені у вказаний вами діапазон:

	F	G	H	I	J	K	L	M
1	Двухфакторный дисперсионный анализ без повторений							
2								
3	ИТОГИ							
		<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>			
4	Строка 1	3	69	23	237			
5	Строка 2	3	85	28,33333	233,3333333			
6	Строка 3	3	105	35	325			
7	Строка 4	3	77	25,66667	166,3333333			
8								
9	Столбец 1	4	60	15	16,6666667			
10	Столбец 2	4	96	24	22			
11	Столбец 3	4	180	45	50			
12								
13								
14	Дисперсионный анализ							
15		<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>	
16	Строки	238,6666667	3	79,55556	17,46341463	0,002280032	4,757062663	
17	Столбцы	1896	2	948	208,097561	0,000002870	5,14325285	
18	Погрешность	27,33333333	6	4,555556				
19								
20	Итого	2162	11					

2.5. Інтерпретація результатів, полягає у оцінці отриманих Р-значень. У нашому випадку воно $< 0,05$, як для дії фактора «Температура» (рядки) так і для фактора «рН» (стовпчики). Окрім цього, обраховане значення F-критерію (17,46341463 та 208,097561) більше, ніж F-критичне (4,757062663 та 5,14325285 відповідно). Отже із 95% довірчою імовірністю ми можемо прийняти гіпотези H_A^A та H_A^B і стверджувати, що **«Як температура так і рН впливають на інтенсивність росту бактеріальної культури»**.

Завдання для самостійного виконання.

1. Досліджували відмінність продуктивності відтворення одного і того ж навчального матеріалу у трьох групах випробовуваних (по 5 осіб), які розрізнялися умовами пред'явлення цього матеріалу для запам'ятовування. Результати обстеження наведені в таблиці.

Обстежувана особа, № з/п	Умо ва 1	Умо ва 2	Умо ва 3
1	5	8	11
2	4	7	9
3	3	6	7
4	6	9	10
5	7	5	8

Визначити, чи впливають умови пред'явлення навчального матеріалу на продуктивність його відтворення?

2. Досліджували активність ензиму манозо-6-фосфат ізомерази (MPI) та MPI генотипи у ракоподібних – бокоплавів *Platorchestia platensis*. Визначити, чи впливає генотип та стать бокоплавів на активність манозо-6-фосфат ізомерази.

Генотип	Самиці	Самці
FF	2,838	1,884
	4,216	2,283
	2,889	4,939
	4,198	3,486
FS	3,550	2,396
	4,556	2,956
	3,087	3,105
	1,943	2,649
SS	3,620	2,801
	3,079	3,421
	3,586	4,275
	2,669	3,110

Визначити, чи впливає а) генотип та б) стать бокоплавів на активність манозо-6-фосфат ізомерази?

Контрольні питання.

1. Сформулюйте завдання дисперсійного аналізу.
2. Наведіть приклади задач однофакторного дисперсійного аналізу.
3. Поясніть, що таке рівні фактора?
4. Яку інформацію про мінливість випадкової величини містить коефіцієнт детермінації?
5. Сформулюйте нульові гіпотези двофакторного дисперсійного аналізу.
6. Які твердження можна формулювати в завданні двофакторного дисперсійного аналізу? Наведіть приклади.
7. Опишіть алгоритм виконання однофакторного дисперсійного аналізу.

ЛАБОРАТОРНЕ ЗАНЯТТЯ 5

КОРЕЛЯЦІЙНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

Мета: лабораторне заняття передбачає засвоєння методів графічного (діаграми розсіювання) та математичного (коефіцієнт кореляції Пірсона та ранговий коефіцієнт кореляції Спірмена) визначення кореляційного зв'язку із застосуванням табличного процесору MS Excel 2010.

Обладнання та матеріали: ПЕОМ із встановленою ОС Microsoft Windows XP або Windows Seven, процесор таблиць Microsoft Excel 2007 або 2010, Інтернет браузер, методичні вказівки.

План:

1. Представлення кореляції у графічному вигляді та визначення та її типу.
2. Визначення параметричного коефіцієнту кореляції Пірсона.
 - 2.1. Обчислення довірчої імовірності коефіцієнту кореляції Пірсона.
3. Визначення непараметричного коефіцієнту рангової кореляції Спірмена.
 - 3.1. Обчислення довірчої імовірності коефіцієнту кореляції Спірмена.

Теоретичні відомості.



При вирішенні багатьох наукових та практичних завдань досліднику часто доводиться проводити статистичний аналіз зв'язків між факторними та результативними ознаками статистичної сукупності (причинно-наслідковий зв'язок) або визначати залежності паралельних змін декількох ознак цієї сукупності від якої-небудь третьої величини (від їх загальної причини). Для вивчення особливостей зв'язків подібного роду, визначення їх розмірів та напрямків, а також оцінки їх достовірності використовують методи визначення кореляційного зв'язку. Є два види проявів кількісних зв'язків між ознаками: а) функціональний зв'язок; б) кореляційний зв'язок.

Функціональний зв'язок – це така форма співвідношень між двома ознаками, коли кожному значенню однієї з них строго відповідає лише певне значення іншої (наприклад площа кола залежить від його радіусу). Функціональний зв'язок є характерним для фізико-математичних процесів.

Кореляційний зв'язок – такий зв'язок, при якому кожному визначеному значенню однієї ознаки відповідає кілька значень іншої, взаємозв'язаної з нею ознаки. Інакше кажучи – це зв'язок, де вплив окремих факторів проявляється лише як тенденція (у середньому) при масовому спостереженні фактичних даних. Прикладами кореляції є: зв'язок між ростом і масою тіла людини, зв'язок між температурою тіла і частотою пульсу та ін. Кореляційний зв'язок характерний для медико-біологічних процесів.

Величина, що характеризує напрямок і силу зв'язку між ознаками це коефіцієнт кореляції, який одним числом дає уявлення про напрямок та силу зв'язків між ознаками (явищами); межі його коливань знаходяться в діапазоні

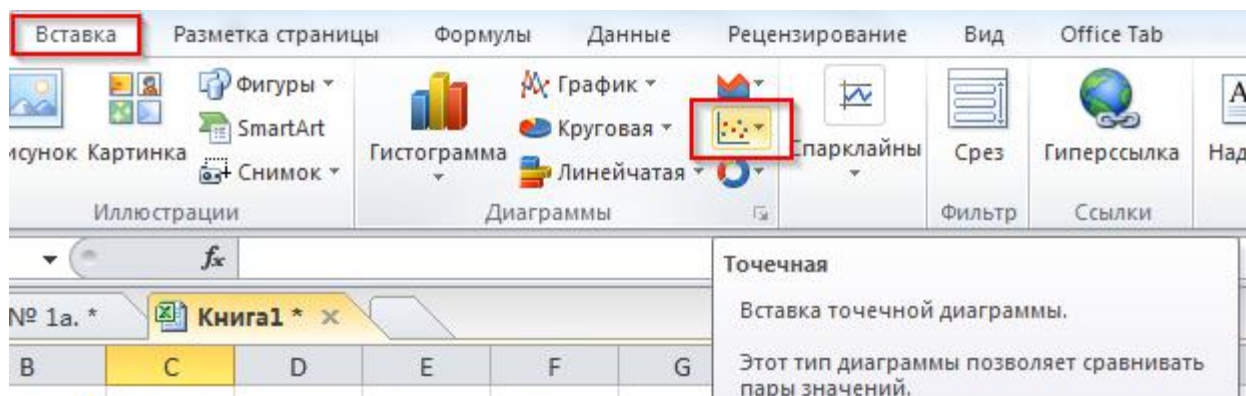


У MS Excel для визначення типу та ступені кореляційного зв'язку в графічній та математичній формі потрібно виконати наступний й алгоритм дій:

- Створення таблиці з даними.
- Побудова діаграми розсіювання
- Попередня оцінка ступеня та спрямованості кореляції за виглядом діаграми
- Формулювання H_0 та H_A гіпотез.
- Обрахування обраного коефіцієнту кореляції.
- Обрахування довірчої ймовірності обрахованого коефіцієнту кореляції.
- Відхилення чи прийняття H_0 гіпотези на прийнятому рівні статистичної значущості за результатами інтерпретації результатів статистичного тесту.

1.1. Для визначення кореляції у графічному вигляді використовують діаграми розсіювання. Для цього створіть таблицю із даними. У комірці A1 та B1 уведіть назви змінних «Довжина голови, см» та «Довжина грудного плавця, см», а у комірці A2:B17 – значення відповідних змінних.

1.2. Виділіть діапазон комірок A1:B17, та оберіть на стрічці Excel вкладку «**Вставка**», панель інструментів «**Діаграми**», тип діаграми «**Точечная**».



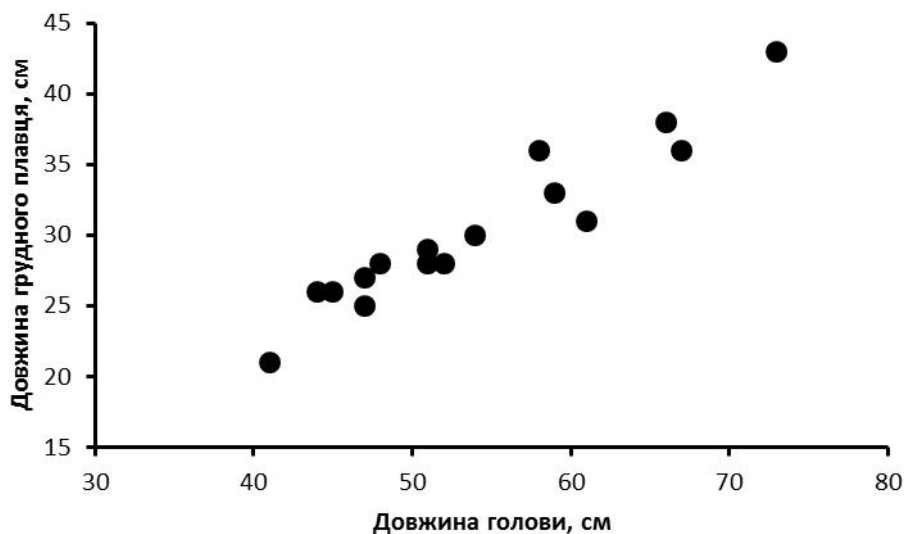
Відредагуйте отриманий графік, змінивши мінімальне значення вертикальної осі з «авто» на «**фиксированное**» із значенням, що дорівнює 15, та мінімальне значення горизонтальної осі з «авто» на «**фиксированное**» із значенням, що дорівнює 30.

1.3. Видаліть з графіку додаткові лінії, заголовок та легенду, додайте підписи до осей, змініть форму маркерів на круглу а колір – на чорний.



Щоб змінити параметри будь-якої осі графіку клацніть по ньому ПКМ та у контекстному меню оберіть пункт «Формат осі...».
Щоб змінити вигляд маркерів, клацніть по будь-якому з них ПКМ та у контекстному меню оберіть пункт «Формат ряда даних...»

У результаті буде побудована точкова діаграма (діаграма розсіювання), на якій у графічній формі буде відображено наявність або відсутність кореляційного зв'язку.



1.4. Інтерпретація результатів. У даному випадку дійсно, спостерігається позитивна лінійна кореляція між довжиною голови та довжиною грудного плавця у окуня озера Біле, оскільки точки вишукались зліва направо та знизу догори. Отже, можна зробити висновок, що чим більша довжина голови у окуня, тим більша довжина його грудного плавця.



Тип кореляції можна визначити у залежності від виду діаграми розсіювання. Якщо точки вишиковуються в уявну лінію знизу догори (починаючи з лівого боку діаграми), то наявна позитивна лінійна кореляція. У випадку коли точки вишиковуються з гори донизу – то це свідчить про наявність негативної лінійної кореляції. Якщо ж кореляції між показниками немає, точки майже рівномірно розсіюються по усій площі діаграми.

1.5. Визначте коефіцієнт кореляції для даної вибіркової сукупності. Для цього використайте функцію PEARSON(масив1;масив2), де масив1 – набір даних незалежної змінної, а масив2 – набір даних залежної змінної. У комірку A19 уведіть загальноприйняте позначення коефіцієнту кореляції (малу латинську літеру r) та знак «дорівнює», а у комірку B19 – формулу для обчислення коефіцієнту кореляції =PEARSON(A2:A18;B2:B18).

17	45	26		
18	73	43		
19	r=	0,96498775		
20				

У результаті буде обрахований коефіцієнт кореляції, який у нашому

випадку буде дорівнювати 0,96. Отже, дані графічної оцінки кореляції підтверджено, і остаточний висновок проведеного кореляційного аналізу наступний: **«Між довжиною голови та довжиною грудного плавця окуня озера Біла наявна сильна позитивна кореляція».**



Коефіцієнти кореляції, як міри зв'язку між випадковими величинами, також є величинами випадковими та носять імовірнісний характер. Статистичні висновки про кореляційний зв'язок між величинами роблять не з генерального коефіцієнта кореляції ρ (значення якого є зазвичай невідомим), а за його вибірковою аналогом – r . Оскільки коефіцієнти кореляції r розраховується за значеннями змінних, які випадково потрапили у вибірку з генеральної сукупності, то й статистика r є величиною випадковою, яка потребує статистичної оцінки. Як правило, перевіряють нульову гіпотезу про відсутність кореляційного зв'язку між змінними в генеральній сукупності, тобто $H_0: \rho = 0$. Достовірність (довірча ймовірність) коефіцієнтів кореляції залежить від прийнятого рівня значущості α і обсягу вибірки n .

1.6. Для визначення довірчої ймовірності коефіцієнта кореляції Пірсона спочатку потрібно впевнитись, що кореляція має лінійний характер. З побудованої у завданні 1 діаграми розсіювання добре видно, що кореляція лінійна. Далі, у комірці B20 розрахуйте емпіричний критерій t_{cm} за допомогою формули $=B19*КОРЕНЬ((СЧЁТ(A2:A18)-2)/(1-B19^2))$ та отримайте значення t_{cm} . У комірку B21 уведіть рівень статистичної значущості (α), що дорівнює 0,05. Отримайте двобічне критичне значення t-критерію Стьюдента ($t_{кр}$) за допомогою функції $=СТЮДРАСПОБР()$. Для цього у комірку B22 уведіть вираз $=СТЮДРАСПОБР(B21;СЧЁТ(B2:B18)-2)$.

19	r=	0,96498775			
20	t_{cm} =	14,2487741			
21	α =	0,05			
22	$t_{кр}$ =	2,13144955			

1.7. Інтерпретація результатів та прийняття рішення. Оскільки $t_{cm} > t_{кр}$ ($14,25 > 2,13$), нульова гіпотеза відхиляється, тобто із **95% довірчою ймовірністю** ми можемо стверджувати що **«Між довжиною голови та довжиною грудного плавця окуня озера Біле наявна сильна позитивна кореляція»**

Завдання 2. Під час моніторингу для проекту MoorLIFE еколог зібрав дані про два види рослин – верес (*Calluna vulgaris*) та чорницю (*Vaccinium myrtillus* L.) з району вересової пустки, яку було відновлено у 2003 році.

Ділянка	Чорниця, % покриву	Верес, % покриву
1	5	0
2	40	0
3	50	5
4	5	0
5	10	0
6	25	0
7	0	1
8	4	0
9	0	0
10	0	1
11	10	6
12	2	0,5

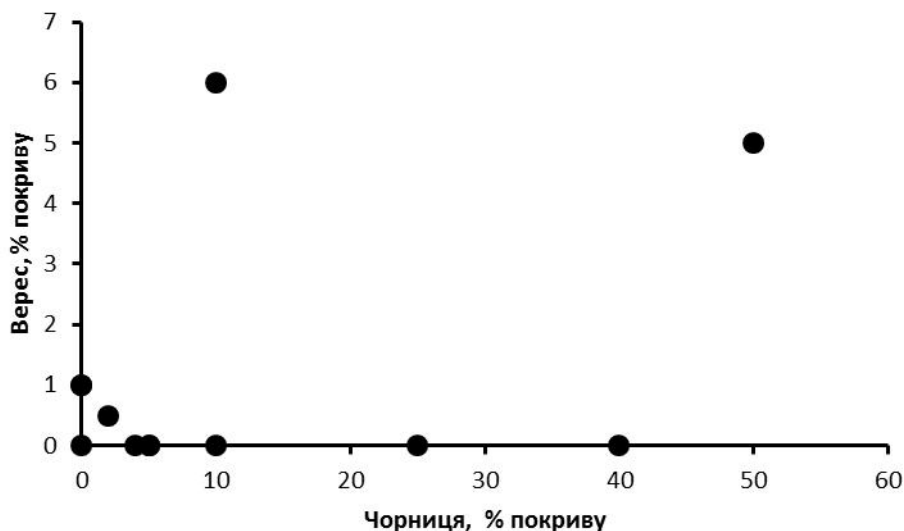


Чорниця
(*Vaccinium
myrtillus*)

Верес
(*Calluna
vulgaris*)

Оцінити наявність та значущість кореляції (із 99% довірчою ймовірністю, $\alpha = 0,01$) між зростанням цих двох видів рослин.

2.1. Перенесіть дані до електронної таблиці. У комірці A1, B1 та C1 уведіть назви змінних, а у комірці A2:C13 – варіанти. Для того, щоб зробити попередню оцінку наявності та характеру кореляційного зв'язку між змінними побудуйте діаграму розсіювання.

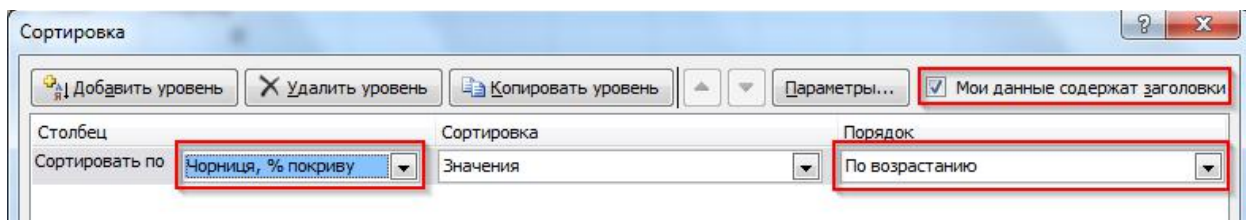


З діаграми видно, деякі точки вишиковуються у пряму (ліворуч знизу, а праворуч догори), що може вказувати на наявність слабкої негативної лінійної кореляції між двома видами рослин.

2.2. Наступний крок – розрахунок коефіцієнту кореляції. Оскільки дані виражені у відсотках, то розумним вибором буде використання рангового коефіцієнту кореляції Спірмена. Для цього спочатку потрібно проранжувати усі дані спостережень **окремо для кожного виду рослин**. На новому аркуші створіть таблицю за наступним зразком:

	A	B	C	D	E	F	G
			Чорниця, % покриву	Ділянка	Ранг	Верес, % покриву	
1	Ділянка	Ранг					
2		1		5	1		0
3		2		40	2		0
4		3		50	3		5
5		4		5	4		0
6		5		10	5		0

2.3. Відсортуйте дані у таблиці за збільшенням за змінною «Чорниця, % покриву». Виділіть комірки A1:C13 та оберіть на стрічці вкладку «**Данные**», панель інструментів «**Сортировка и фильтр**», кнопка «**Сортировать**». У діалоговому вікні встановіть необхідні параметри.



Після того, як ваші дані будуть відсортовані, у комірках B2:B13 за допомогою автозаповнення проранжуйте варіанти, починаючи із 1 із кроком 1. Знову відсортуйте дані (комірки A1:C13) але за змінною «Ділянка».

Подібну послідовність дій виконайте для частини таблиці, що містить дані про площу покриття вересової пустки для вересу. Таким чином ви отримаєте ранжовані дані для кожного виду рослин, окремо на кожній ділянці.

2.4. Створіть таблицю зі значеннями рангів за наведеним зразком:

	A	B	C	D	E
1	Ранг				
2	чорниця	верес			
3	6	1			
4	11	2			
5	12	11			
6	7	3			

Коефіцієнт рангів r_s розраховується за формулою:

$$r_s = 1 - \frac{6 \times \sum_{i=1}^n (x_i - y_i)^2}{n \times (n^2 - 1)}$$

де: n – обсяг сукупності об'єктів; $(x_i - y_i)$ – різниця рангів i -го об'єкта. Коефіцієнт r_s приймає значення в інтервалі від -1 до +1.

2.5. Виконайте наступну послідовність дій. У комірки C1 та D1 уведіть текст «(x-y)» і «(x-y)²» відповідно, а у комірках C2:C14 та D2:D14 розрахуйте відповідні значення. Для цього у комірку C2 уведіть формулу =A2-B2, а у комірку D2 – формулу =C2^2.

Додайте новий стовпчик A, та у комірку A15 уведіть текст «Суми», а у комірках B15:E15 обчисліть суму усіх значень у відповідних стовпчиках за допомогою формули СУММ().

У комірку A16 уведіть текст «r_s=», а у комірку B16 – формулу для обчислення коефіцієнта кореляції Спірмена: =1-((6*E15)/СЧЁТ(B3:B14)*((СЧЁТ(B3:B14)^2)-1)). Отримане значення r_s дорівнює приблизно -0,15, що свідчить про наявність слабкої негативної кореляції між обома видами рослин.

2.6. Далі обчисліть довірчу ймовірність для визначеного коефіцієнту кореляції. Для цього у комірку B17 уведіть формулу для обчислення t-критерію =B15/КОРЕНЬ(ABS(1-B15^2)/(СЧЁТ(B3:B14)-2)), а у комірку B18 – формулу для обчислення критичного значення t-критерію =СТЮДРАСПОБР(0,01/2;СЧЁТ(B3:B14)-2).

15	Суми	78	78	0	328
16	r _s =	-0,14685			
17	t _{емп} =	3,162538			
18	t _{кр} =	3,581406			

2.7. Отриманні значення t_{емп} = 3,16 та t_{кр} = 3,58 не дозволяють нам відхилити нульову гіпотезу на прийнятому рівні статистичної значущості (α = 0,01), оскільки t_{емп} < t_{кр}. Отже, остаточний висновок буде наступним: **«За результатами кореляційного аналізу площі покриття вересової пустки рослинами верес (*Calluna vulgaris*) та чорниця (*Vaccinium myrtillus* L.) зв'язку між зростанням цих рослин на визначений ділянці не виявлено».**

Завдання для самостійного виконання.

1. Для 10 півників леггорнів 15-денного віку були отримані наступні дані про масу їх тіла x (в г) і масу гребеня y (в мг):

x	83	72	69	90	90	95	95	91	75	70
y	56	42	18	84	56	107	90	68	31	48

Дайте відповідь на питання, чи наявний кореляційний зв'язок між масою тіла та масою гребеня півників, для чого побудуйте діаграму розсіювання та розрахуйте коефіцієнт кореляції Пірсона та його довірчу ймовірність при рівні статистичної значущості α = 0,05.

2. При відновленні деяких земельних ділянок, переданих до заповідного фонду, екологами проводився моніторинг процесів сукцесії. У таблиці подано інформацію про домінуючі види рослин на одній із ділянок, умовно поділену на 26 квадратів.

Ідентифікаційний код квадрату	Deschampsia flexuosa, % покриття	Ранг 1	Agrostis spp, % покриття	Ранг 2
JP001	30		15	
JP003	35		1	
JP004	15		5	
JP005	15		10	
JP006	15		2	
JP007	10		10	
JP007b	3		3	
JP008	5		1	
JP009	10		5	
JP010	25		0,25	
JP011	35		30	
JP012	50		20	
JP013	50		10	
JP014	30		15	
JP015	40		15	
JP016	50		5	
JP017	30		4	
JP018	15		10	
JP019	40		7	
JP020b	35		15	
JP021	30		30	
JP022	50		5	
JP023	15		10	
JP024	40		10	
JP025	40		12	
JP026	10		4	

Побудуйте діаграму розсіювання. Проаналізуйте припущення про наявність або відсутність кореляційного зв'язку між зростанням цих двох видів рослин на визначеній ділянці. Визначте довірчу ймовірність для коефіцієнта кореляції ($\alpha = 0,05$).

Контрольні питання.

1. Дайте визначення поняттю «кореляція».
2. Поясніть, у чому полягає різниця між кореляційною і функціональною залежностями?
3. Розкрийте сутність позитивної і негативної кореляції. Які між ними відмінності?
4. Чому дорівнює коефіцієнт кореляції при повному кореляційному зв'язку?
5. Проаналізуйте, чим відрізняється r від ρ ?

ЛАБОРАТОРНЕ ЗАНЯТТЯ 6 РЕГРЕСІЙНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

Мета: лабораторне заняття передбачає засвоєння методів графічного (побудова лінії регресії) та математичного (розрахунок рівняння регресії та обчислення коефіцієнту регресії) проведення регресійного аналізу даних із застосуванням табличного процесору MS Excel 2010.

Обладнання та матеріали: ПЕОМ із встановленою ОС Microsoft Windows XP або Windows Seven, процесор таблиць Microsoft Excel 2007 або 2010, Інтернет браузер, методичні вказівки.

План:

1. Графічне представлення регресії.
2. Розрахунок рівняння регресії.
3. Регресійний аналіз за допомогою надбудови «Аналіз даних».
4. Інтерпретація результатів регресійного аналізу.

Теоретичні відомості.

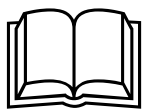


Відомо, що різні залежності широко поширені як в органічній, так і в неорганічній природі. Їх вивчення привело до розробки великої кількості методів їх математичного охарактеризування. Першим з них був розглянутий вище кореляційний метод. Однак коефіцієнт кореляції вказує лише на ступінь зв'язку в варіації двох змінних величин, або, як іноді кажуть, на міру тісноти зв'язку між ними, але не дає можливості судити про те, як кількісно змінюється одна величина по мірі зміни іншої. На це питання дозволяє відповісти інший метод визначення зв'язку між ознаками, що варіюють, – метод регресії.

Аналіз взаємовідносин двох мінливих величин за допомогою методу регресії часто може дати дуже цінні у практичному відношенні результати. У деяких випадках для висвітлення різних сторін питання, яке вирішується, треба застосовувати як кореляційний, так і регресійний методи аналізу.

При простому кореляційному аналізі вивчається залежність між мінливістю двох ознак – x і y . За допомогою регресійного аналізу ставиться додаткове завдання встановити, як кількісно змінюється одна величина при зміні іншої на одиницю. Так як мінливих величин дві, то регресія, вочевидь, може бути двосторонньою: визначення зміни y при зміні x та навпаки, визначення зміни x при зміні y . У цьому й полягає головна відмінність регресійного аналізу від аналізу кореляційного.

Регресія може бути виражена декількома способами; графічно – шляхом побудови емпіричних ліній регресії, шляхом складання рівнянь регресії і побудови теоретичних ліній регресії і, нарешті, математично – за допомогою обчислення коефіцієнта регресії.



Регресія може бути:

- лінійною ($y = a + bx$);
- параболічною ($y = a + bx + cx^2$);
- експоненціальною ($y = a * \exp(bx)$);
- ступеневою ($y = a * x^b$);
- гіперболічною ($y = b/x + a$);
- логарифмічною ($y = b * \ln(x) + a$);
- показовою ($y = a * b^x$).

Хід виконання роботи.

Завдання 1. Шляхом щотижневого взяття проб з поля було вивчено зміна висоти рослин сої (*Glycine max* Moench.) у (в см) з віком x (в тижнях):

x	y
1	5
2	13
3	16
4	23
5	33
6	38
7	40



Соя
(*Glycine max* Moench.)

Викладіть ці дані на графіку і побудуйте емпіричну лінію регресії у по x . Складіть рівняння регресії.



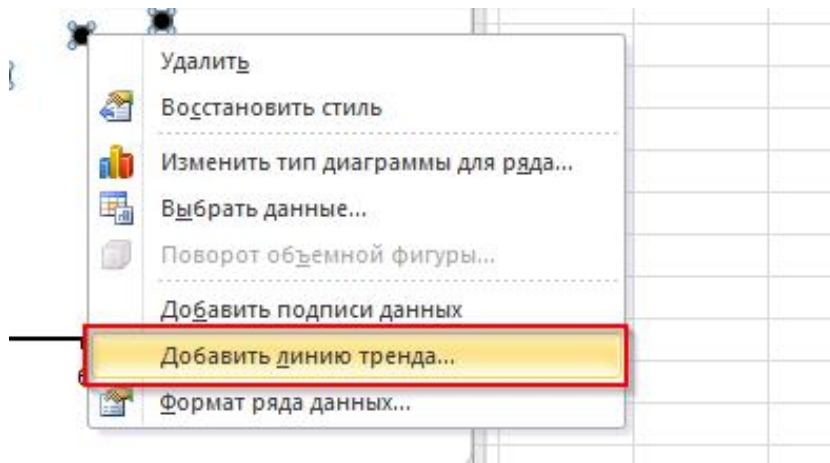
У MS Excel для визначення регресії у графічній формі передбачається наступний алгоритм дій:

- Створення таблиці з даними.
- Побудова діаграми розсіювання.
- Додавання до діаграми лінії тренду.
- Визначення типу регресії та додавання до діаграми її рівняння.

1.1. Створіть таблицю із даними. У комірці A1 та B1 уведіть назви змінних: «вік, тижнів» та «висота, см», а у комірці A2:B9 – відповідні значення.

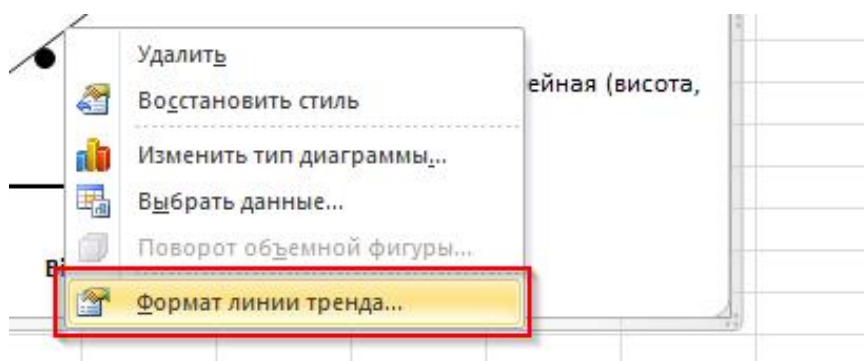
1.2. Побудуйте діаграму розсіювання та відформатуйте її відповідно до рекомендацій завдання 1, лабораторного заняття №5.

1.3. Додайте лінію тренду. Для цього клацніть ПКМ по будь-якому маркеру та оберіть у контекстному меню пункт «Добавить линию тренда...»

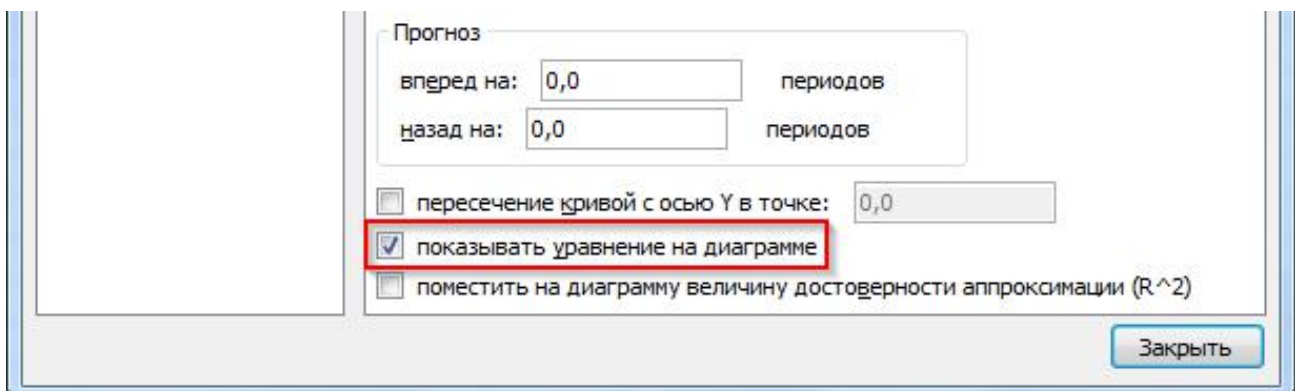


У діалоговому вікні, що з'явиться, оберіть шляхом встановлення перемикача тип лінії тренда (регресії) – «Линейная», за бажанням змініть формат лінії (товщину, колір, вигляд) та натисніть на кнопку «Закреть». На діаграмі з'явиться лінія регресії.

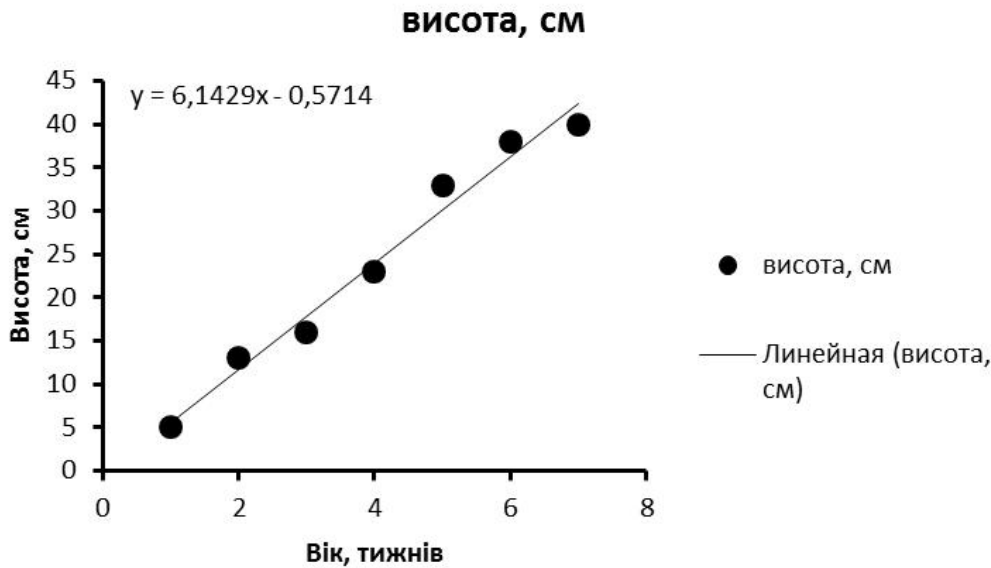
1.4. Додайте до діаграми рівняння регресії. Для цього клацніть ПКМ по лінії тренду (регресії) та оберіть пункт «Формат линии тренда...».



У нижній частині діалогового вікна, що з'явиться, встановіть прапорець на пункт «показывать уравнение на диаграмме», та натисніть на кнопку «Закреть».



Клацніть ЛКМ по ділянці, на якій розташовано рівняння регресії та перемістіть його у ліву верхню ділянку діаграми.



Завдання 2. У 20 дорослих чоловіків була виміряна довжина тіла, см (x) та маса тіла (y):

x	165	176	175	168	167	172	175	180	179	173
y	56	75	70	61	61	63	72	80	76	68

x	166	178	169	169	170	176	180	169	177	176
y	58	76	60	64	63	71	78	63	75	71

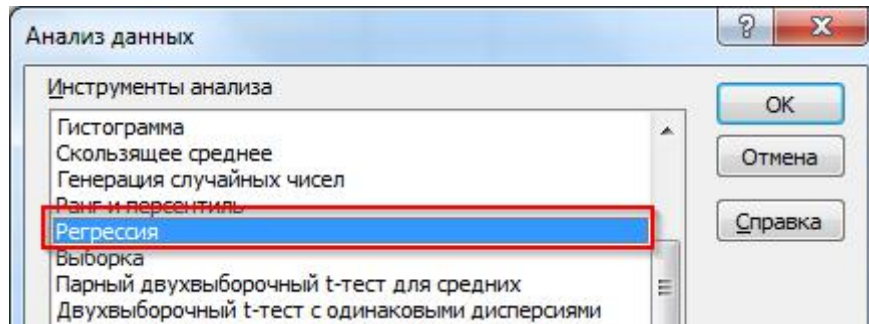
Визначте коефіцієнт регресії та його довірчу імовірність при $\alpha = 0,05$.



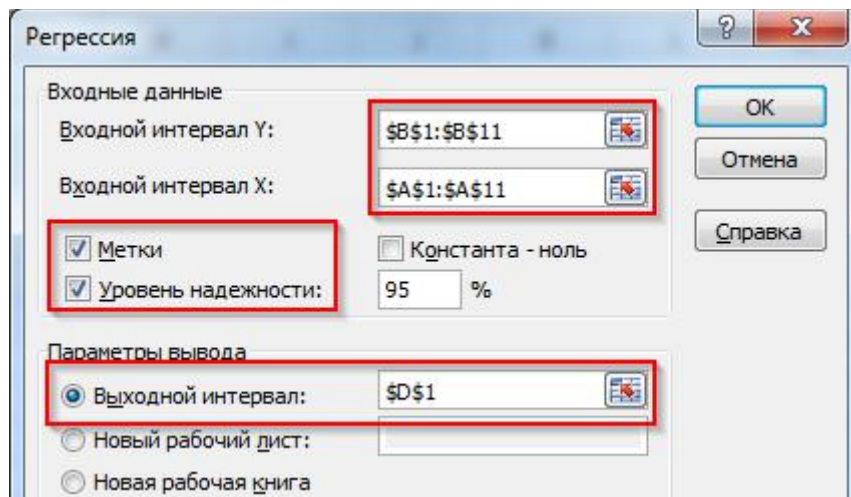
У MS Excel для визначення коефіцієнту регресії використовується процедура «Регрессия» надбудови «Анализ данных» за наступним алгоритмом дій:

- Створення таблиці з даними.
- Визначення H_0 та H_A гіпотез.
- Проведення регресійного аналізу.
- Визначення типу регресії та додавання до діаграми її рівняння.
- Відхилення чи прийняття H_0 гіпотези на прийнятому рівні статистичною значимості за результатами інтерпретації результатів статистичного тесту.

2.1. Перенесіть дані у таблицю Excel. Для визначення коефіцієнту регресії застосуйте надбудову Excel «Анализ данных». У діалоговому вікні, що з'явилося, виберіть зі списку пункт «Регрессия».



У діалоговому вікні вкажіть: вхідні інтервали змінної x та y, шляхом встановлення прапорців вкажіть, що ваші дані містять мітки, а також зазначте рівень надійності (довірчу вірогідність, що дорівнює 95%). У якості вихідного інтервалу вкажіть комірку D1, та натисніть кнопку «**ОК**».



2.2. Оцінка та інтерпретація результатів. У першу чергу зверніть увагу на R-квадрат та коефіцієнти.

	D	E	F	G	H	I	J	K	L	
Вывод итогов										
<i>Регрессионная статистика</i>										
Множественный R		0,978282385								
R-квадрат		0,957036425								
Нормированный R-квадрат		0,95464956								
Стандартная ошибка		1,54876733								
Наблюдения		20								
<i>Дисперсионный анализ</i>										
		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>				
Регрессия		1	961,7737557	961,7737557	400,9595523	9,43506E-14				
Остаток		18	43,17624434	2,398680241						
Итого		19	1004,95							
		<i>Коэффициенты</i>		<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение		-187,1445701	12,74915593	-14,67897727	1,85026E-11	-213,9295528	-160,3595875	-213,9295528	-160,3595875	
x		1,475113122	0,073667349	20,02397444	9,43506E-14	1,320343764	1,62988248	1,320343764	1,62988248	

R-квадрат – коефіцієнт детермінації. У нашому прикладі він дорівнює 0,957, або 95,7%. Це означає, що розрахункові параметри моделі на 95,7% пояснюють залежність між досліджуваними параметрами. Чим вище коефіцієнт

детермінації, тим якісніше модель. Добре – вище 0,8. Погано – менше 0,5 (такий аналіз навряд чи можна вважати резонним). У нашому прикладі – «дуже добре».

Коефіцієнт -187,14 показує, яким буде Y , якщо всі змінні в даній моделі будуть рівні 0. Тобто на значення аналізованого параметра впливають і інші фактори, які не описані в моделі.

Коефіцієнт 1,47 показує вагомість змінної X на Y . Тобто довжина тіла у межах даної моделі впливає на масу тіла з вагою 1,48 (це достатньо великий ступінь впливу). Знак «+» вказує на позитивний вплив: чим більше довжина тіла, тим більша маса.

Щодо довірчої імовірності видно, що P -значення суттєво менше рівня статистичної значущості ($9,44 \cdot 10^{-14} < 0,05$).

Отже, ми справедливо можемо ствержувати **із 95% довірчою ймовірністю**, що **«Маса тіла обстежених чоловіків залежить від довжини їх тіла: чим більшим є зріст, тим більшою є вага тіла»**

Завдання для самостійного виконання.

1. Передбачається, що між кількістю настриженої вовни (y) і живою вагою овець (x) є залежність. Для 10 овець були отримані наступні дані (в кг):

x	50	55	60	50	65	60	50	55	50	65
y	4,3	4,1	4,4	4,0	4,2	4,0	4,2	4,1	4,2	4,5

Викладіть ці дані на графіку і побудуйте емпіричну лінію регресії y по x . Складіть рівняння регресії. Визначте коефіцієнт регресії при рівні статистичної значущості 0,05.

2. Для встановлення зв'язку між вмістом фосфору в ґрунті (x) і вмістом фосфору в злакових рослинах (y) було проведено 9 аналізів з наступними результатами:

x	1	4	5	9	13	11	23	23	28
y	64	71	54	81	93	76	77	95	109

Викладіть ці дані на графіку і побудуйте емпіричну лінію регресії y по x . Складіть рівняння регресії. Визначте коефіцієнт регресії при рівні статистичної значущості 0,01.

Контрольні питання.

1. Дайте визначення поняттю «регресія»
2. Поясніть, у чому полягає перевага регресії порівняно з кореляцією.
3. Які способи вираження регресії вам відомі?
4. Наведіть приклади формулювання нульової гіпотези стосовно регресії.
5. У чому полягають відмінності між рівнянням регресії x по y та рівнянням регресії y по x ?

ПЕРЕЛІК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Основна

1. Лакин Г.Ф. Биометрия : учебное пособие для биол. спец. вузов / Г.Ф. Лакин. – М.: Высшая школа, 1990. – 352 с.
2. Атраментова Л.О. Біометрія : підруч. для студ. вищ. навч. закл. / Л.О. Атраментова, О. М. Утевська. – Х. : Ранок, 2007. – 176 с.
3. Дёрффель К. Статистика в аналитической химии / К. Дёрффель. – М. : Мир, 1969. – 248 с.
4. Животовский Л.А. Популяционная биометрия / Л.А. Животовский. –М. : Наука, 1991. – 276 с.
5. Зайцев Г.Н. Методика биометрических расчётов. Математическая статистика в экспериментальной ботанике / Г.Н. Зайцев. – М. : Наука, 1973. – 256 с.
6. Ивантер Э.В., Коросов А.В. Элементарная биометрия / Э.В. Ивантер, А.В. Коросов. – Петрозаводськ : Изд-во ПетрГУ, 2010. – 104 с.
7. Акімова О.В. Статистика в малюнках та схемах : навчальний посібник / О.В. Акімова, О.С. Дубинська. – К.: Центр учбової літератури, 2007. – 168 с.
8. Плохинский Н.А. Биометрия / Н.А. Плохинский. – М. : Изд-во Моск. ун-та, 1970. – 370 с.
9. Рокицкий П.Ф. Биологическая статистика / П.Ф. Рокицкий. – Минск, Вышэйш. школа, 1973. – 320 с.

Додаткова

1. Общая теория статистики : учебник для ВУЗов / [Под ред. М.Г. Назарова]. – М. : Омега-Л, 2010. – 410 с.
2. Громько Г.Л. Теория статистики / Г.Л. Громько. – М. : ИНФРА-М, 2005. – 476 с.
3. Бабак В.П. Теорія ймовірюностей, випадкові процеси та математична статистика : підручник / В.П. Бабак, Б.Г. Марченко, М.Є. Фриз. – К. : Техніка, 2004. – 288 с.
4. Бродский Я.С. Статистика. Вероятность. Комбинаторика / Я.С. Бродский. – М. : ООО «Издательство Оникс», 2008. – 544 с.
5. Лагутин М.Б. Наглядная математическая статистика / М.Б. Лагутин. – М. : БИНОМ. Лаборатория знаний, 2009. – 472 с.
6. Рунион Р. Справочник по непараметрической статистике: Современный подход / Р. Рунион [пер. с англ. Е.З. Демиденко; предисл. Ю. Н. Тюрина]. – М.: Финансы и статистика, 1982. – 198 с.
7. Мармоза А.Т. Практикум з теорії статистики : навчальний посібник / А.Т. Мармоза. – К : Ельга, Ніка – Центр, 2007. – 348 с.
8. Лялин В.С. Статистика. Теория и практика в Excel / В.С. Лялин, И.Г. Зверева, Н.Г. Никифорова. – М: Финансы и статистика: Инфра-М, 2010. – 448 с.

9. Макарова Н.В. Статистика в Excel : Учеб. пособие / Н.В. Макарова, В.Я. Трофимец. – М.: Финансы и статистика, 2002. – 368 с.

Інформаційні ресурси

1. БИОМЕТРИКА [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу: <http://www.biometrica.tomsk.ru> (дата звернення 24.03.2018). – Назва з екрана.
2. Академия Анализа Данных StatSoft: курсы лекций по статистическому анализу [Електронний ресурс] : [Веб-сайт] – Електронні дані. – [Москва : StatSoft Russia, 1999-2018]. – Режим доступу до ресурсу: <http://statsoft.ru/academy/lections.php> (дата звернення 24.03.2018). – Назва з екрана.
3. Статистический анализ в MS Excel: статьи о статистических методах анализа, формулах (функциях) и диаграммах табличного процессора Excel [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу до ресурсу: <http://stataliz.info> (дата звернення 24.03.2018). – Назва з екрана.
4. Методы интеллектуального анализа данных [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу до ресурсу: <https://www.ibm.com/developerworks/ru/library/ba-data-mining-techniques/index.html> (дата звернення 24.03.2018). – Назва з екрана.
5. Санкт-Петербургский институт информатики и автоматизации РАН [Електронний ресурс] : [Веб-сайт] – Електронні дані. – [СПб : Санкт-Петербургский институт информатики и автоматизации РАН, 2001-2018]. – Режим доступу до ресурсу: <http://inftech.webservis.ru/it/database/datamining> (дата звернення 24.03.2018). – Назва з екрана.
6. Сайт американської статистичної асоціації [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу до ресурсу: <http://www.amstat.org> (дата звернення 24.03.2018). – Назва з екрана.
7. NAG's Statistical software [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу до ресурсу: <http://www.nag.co.uk> (дата звернення 24.03.2018). – Назва з екрана.
8. With MaxStat you don't need to be a statistician to analyze your data! [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу до ресурсу: <http://www.maxstat.de/> (дата звернення 24.03.2018). – Назва з екрана.
9. Free statistical software [Електронний ресурс] : [Веб-сайт] – Електронні дані. – Режим доступу до ресурсу: <http://freestatistics.altervista.org/> (дата звернення 24.03.2018). – Назва з екрана.

Навчально-методичне видання
(українською мовою)

Гороховський Єгор Юрійович

БИОМЕТРИЯ

Методичні вказівки
до лабораторних занять
для здобувачів ступеня вищої освіти бакалавра
спеціальності «Біологія»

Рецензент *О.В. Маслова*
Відповідальний за випуск *О.Ф. Рильський*
Коректор *Є.Ю. Гороховський*

Підписано до друку
Формат 60 x 90/16
Замовлення № . Наклад 20 прим.

Запорізький національний університет
69600, м. Запоріжжя, МСП-41
вул. Жуковського, 66

Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру видавців, виготівників
і розповсюджувачів видавничої продукції
ДК № 2952 від 30.08.2007