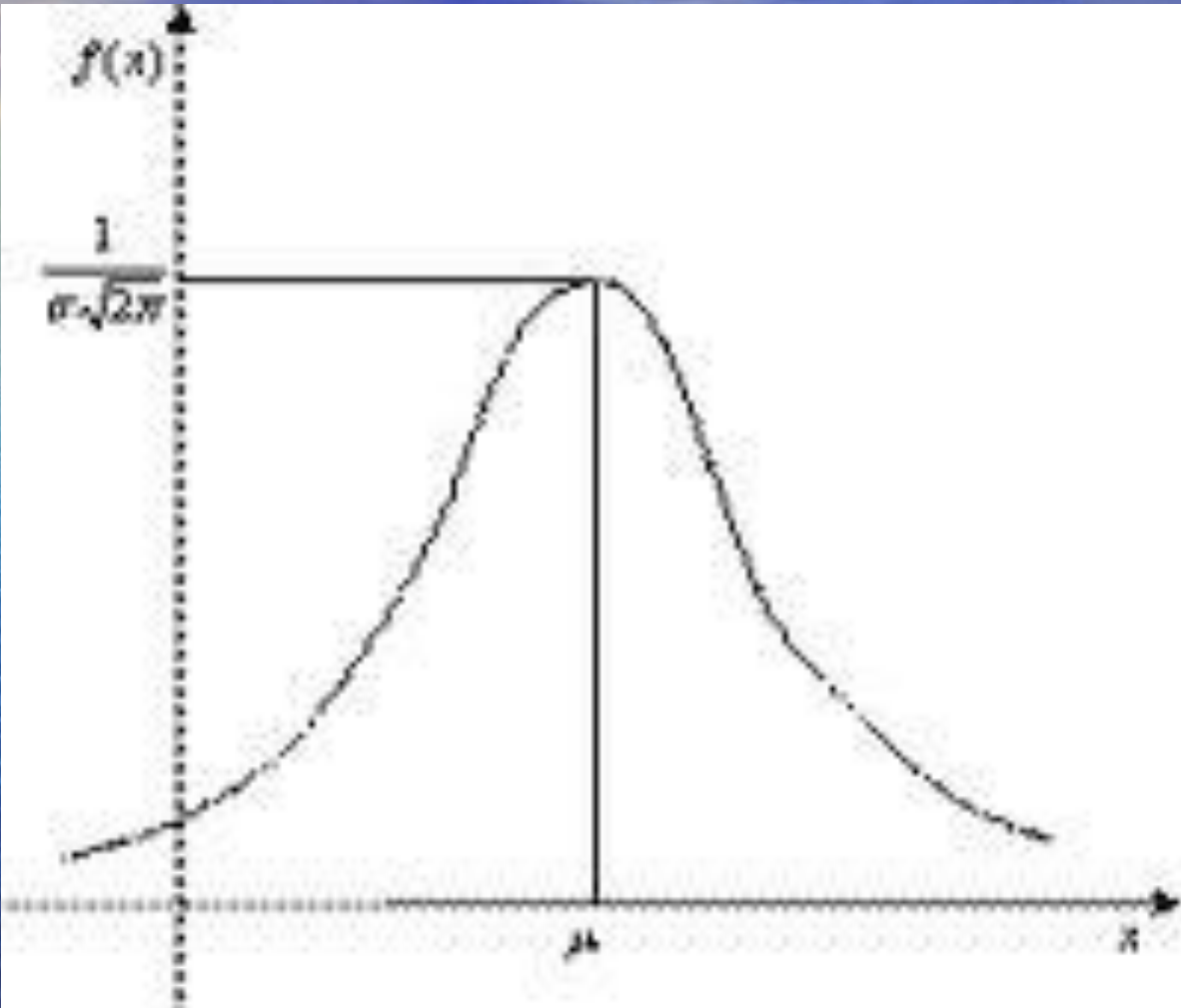
The background of the slide features a serene sunset over a vast ocean. The sky is a deep, vibrant blue, transitioning to a lighter, hazy blue near the horizon. A soft, multi-colored rainbow is visible on the left side, arching over the water. The ocean's surface is covered in gentle, shimmering ripples that catch the light from the setting sun.

Нормальний закон розподілу у сукупностях

- В основі розподілу лежать відповідні математичні закономірності, які для генеральної сукупності (при $n \rightarrow \infty$) характеризуються певним *теоретичним розподілом*. На основі теоретичного розподілу виводяться відповідні статистичні критерії, які використовуються для перевірки гіпотези про досліджувану експериментальну сукупність.

- Значна частина випадкових явищ в природі може бути описана за допомогою нормального закону розподілу (закону Гауса). Це найбільш поширений тип розподілу особин сукупності по класах варіаційного ряду, який можна виразити варіаційною кривою. Показники ознаки (x) знаходяться на осі абсцис, а частоти (f) – на осі ординат.



Нормальний закон розподілу

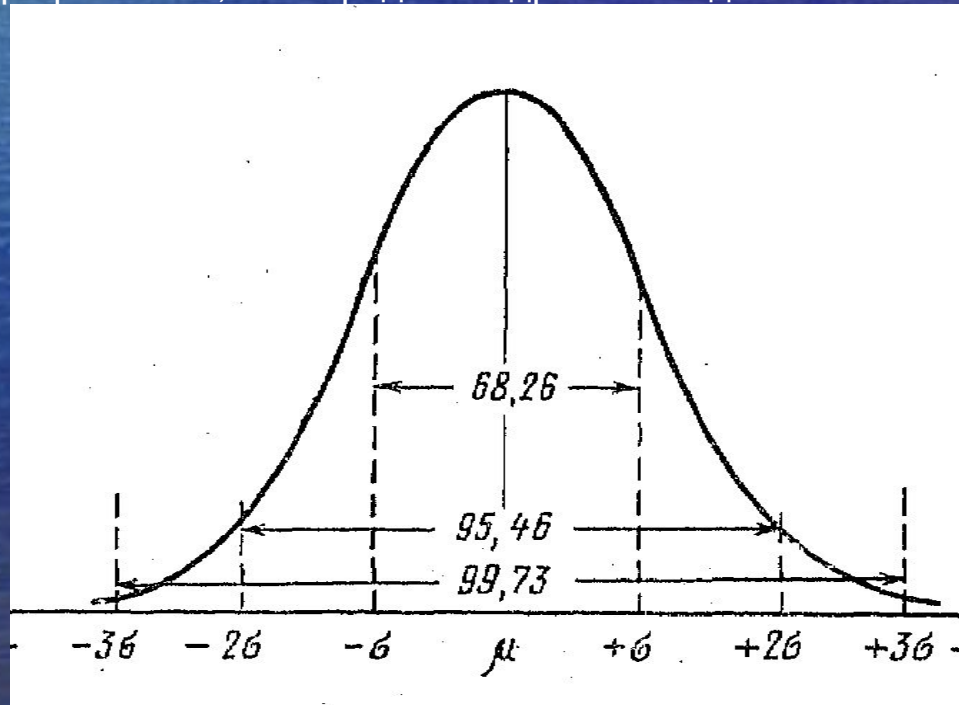
Закон Гауса:

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

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

де x – значення варіант, $f(x)$ – функція нормальної густини ймовірності,

M – середнє арифметичне, σ – середнє квадратичне відхилення.



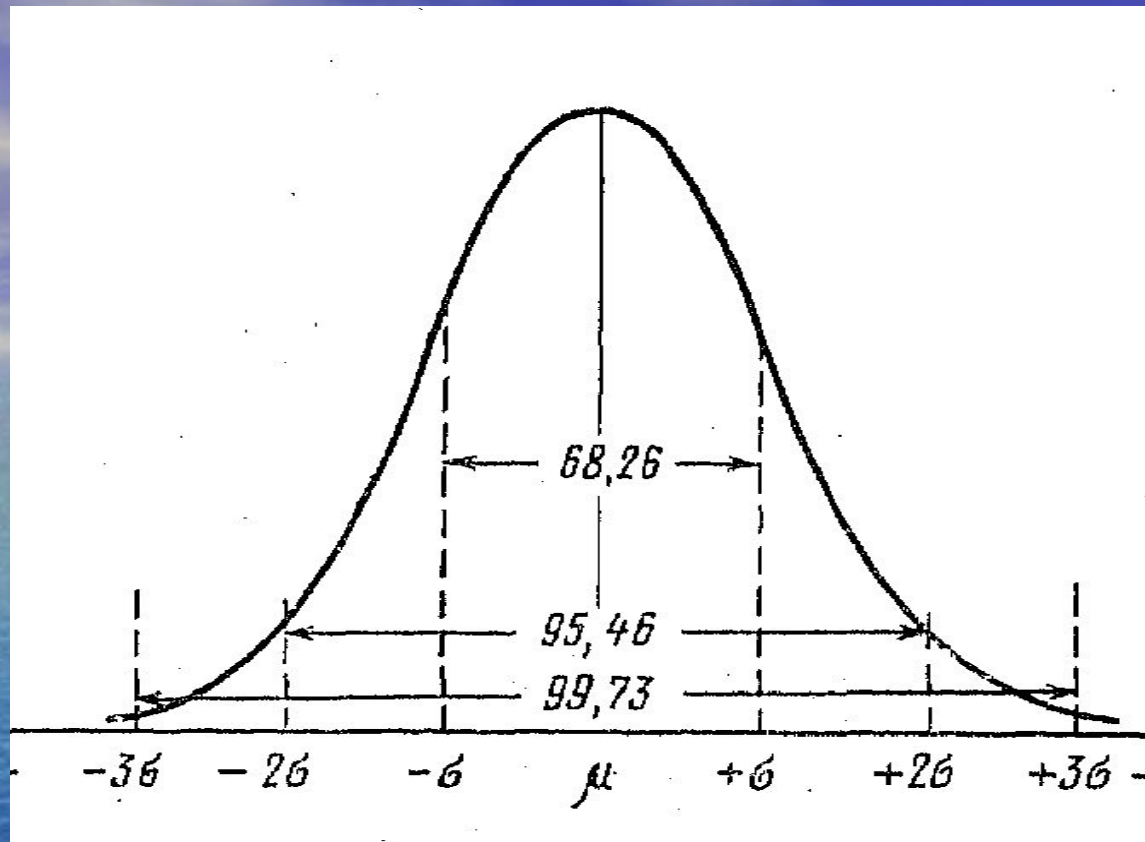
Крива нормального розподілу для різних значень σ .

Для нормального розподілу при $n \rightarrow \infty$
характерні:

- теоретична крива має симетричний вигляд.
- Кінці кривої не зливаються з віссю абсцис, а наближаються до неї в безмежності (асимптотично).
- Вершина кривої нормального розподілу визначається перпендикуляром з точки M (середнє арифметичне значення).
- Максимальне значення f у відповіді найбільшій частоті зустрічі особин, у яких величина ознаки дорівнює середньому арифметичному.
- У нормальному розподілі точка M співпадає з величиною моди та медіани.

Правило трьох сигм

- Криву нормального розподілу характеризує властивість, яку називають правилом трьох сигм: практично весь можливий діапазон відхилень окремих варіант від свого середнього арифметичного значення в сукупностях великого обсягу не виходить за межі $M \pm 3\sigma$. Отже, розподіл ознаки обмежений лімітом $\pm 3\sigma$ від середнього арифметичного значення.



В ці межі входить 99,7% всіх особин сукупності. За межами $\pm 3\sigma$ зустрічається тільки 0,3 % особин, у яких значення ознаки більше за $+3\sigma$ або менше -3σ . У такому випадку значення t для окремих варіант коливаються в межах ± 3 .

- Площа під кривою нормального розподілу варіант у заданому інтервалі t від -3 до $+3$ відображає ймовірність появи поодинокі величини в цьому інтервалі. Також площа під кривою відповідає кількості (в %) особин даної сукупності, що увійшли в даний діапазон. На практиці для визначення ймовірності появи варіант в інтервалі користуються таблицею нормального інтеграла ймовірностей, яку наведено в додатках рекомендованих підручників.

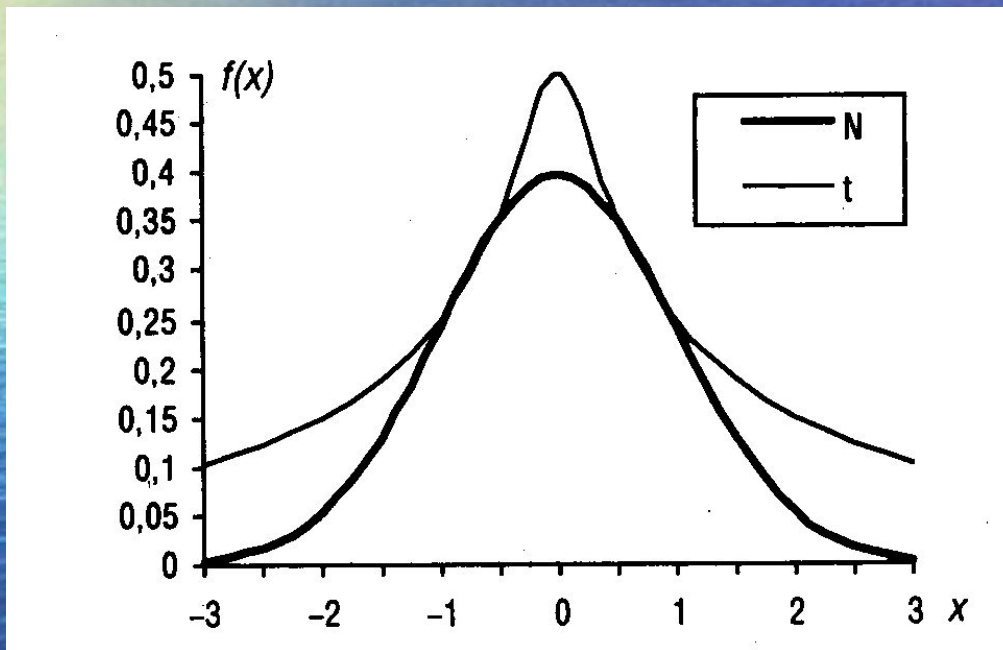
- Встановлено, що ймовірність p появи випадкової величини в інтервалі $M \pm t\sigma$ значень дорівнює:
- у межах $M \pm \sigma$ $p \approx 0,6826$, тобто знаходиться близько 68% усіх даних;
- у межах $M \pm 2\sigma$ $p \approx 0,9545$, тобто знаходиться близько 95% усіх даних;
- у межах $M \pm 3\sigma$ $p \approx 0,9972$, тобто знаходиться близько 99,7% усіх даних.
- Отже, знаючи варіаційну криву розподілу варіант за тою чи іншою ознакою і припускаючи, що розподіл є нормальним, можна передбачити, який відсоток досліджуваних особин (або варіант) укладається: в межах $\pm 1\sigma$ – 68,26%, в межах $\pm 2\sigma$ – 95,46% , в межах $\pm 3\sigma$ – 99,73%.

- У явищах природи діє закон великих чисел, згідно з яким чіткість характеру розподілу варіант у сукупностях залежить від обсягу сукупності: чим він більший, тим яскравіше проявляється закон, якому він підлягає. Характер розподілу варіант ідеально відповідає теоретичному закону розподілу тільки тоді, коли обсяг сукупності нескінченно великий. Ми тоді говоримо про *генеральну сукупність*. Тільки в цьому випадку відносні частоти появи варіант збігаються з їх теоретичними ймовірностями. У реальних обмежених сукупностях ідеального співпадіння кривих розподілу варіант з теоретичною функцією розподілу не спостерігається.

Розподіл Стьюдента

- Закон нормального розподілу проявляється при $n > 20$. Однак експериментатор часто проводить обмежене число досліджень і робить висновки на основі малих вибірок. На початку ХХ ст. в математичній статистиці виник новий напрям, який отримав назву статистики малих вибірок. Найбільше практичне значення для експериментальної роботи мало відкриття в 1908 р. англійським статистиком і хіміком В. Госсетом t -розподілу, який отримав назву розподіл Стьюдента (псевдонім Госсета).

Розподіл Стьюдента



Закон нормального розподілу проявляється при $n > 20$. Однак експериментатор часто проводить обмежену кількість досліджень і робить висновки на основі малих вибірок.

Співвідношення між нормальним (N) і t –розподілом Стьюдента, $f(x)$ – густина ймовірності розподілу

Параметричні критерії перевірки гіпотез

Параметричними називаються такі методи дослідження, у яких усі дані, що входять до вибірок чи генеральних сукупностей попадають під закон нормального розподілу.

При параметричних методах дослідження оперують та порівнюють між собою, в основному, показники дисперсій (або середньоквадратичних відхилень) та середньоарифметичні значення (також використовують інші середні величини: геометричні, гармонійні, квадратичні та кубічні).

Вибірки

Залежні

якщо одному показнику з першої вибірки, відповідає один і лише один показник із другої вибірки

Незалежні

якщо взаємозв'язок між вибірками відсутній

- При дослідженні сукупностей, мінливість варіант у яких підлягає нормальному закону, таблицею нормального інтеграла ймовірностей слід користуватися тільки тоді, коли обсяг досліджуваної сукупності більший за 20 варіант. Такий обсяг можна вважати досить великим, так що розподіл у ньому істотно наближається до теоретичних значень нормального розподілу, коли $n \rightarrow \infty$.

- Коли обсяг сукупності менший за 20 варіант, ці відхилення набувають істотного характеру, так що розподіл варіант описується іншим математичним виразом, який називається *розподілом Стьюдента*. В біометрії в основному користуються таблицею розподілу, в якій величина інтеграла ймовірностей (площа під кривою розподілу) в межах $M \pm t\sigma$ представлена залежно від обсягу сукупності n . Таблиця відома під назвою таблиці Стьюдента і наведена в додатку.
- У практичній роботі експериментатора таблиця Стьюдента відіграє дуже важливу роль, бо кількість проведених аналізів, особливо коли вони складні і вимагають багато часу й коштів, не завжди можна і доцільно доводити до великих значень.

МЕТОДИ ПОРІВНЯЛЬНОГО АНАЛІЗУ

- Аналіз достовірності різниці між середніми арифметичними значеннями двох порівнюваних (вбірок) даних
- Аналіз достовірності різниці за мінливістю двох експериментальних груп
- **ПОРІВНЯЛЬНИЙ АНАЛІЗ РОЗПОДІЛІВ ДАНИХ**

Аналіз достовірності різниці між середніми арифметичними значеннями двох порівнюваних (вбірок) даних

- Для порівняння двох експериментальних вибірок, що належать до генеральної сукупності з нормальним розподілом, з метою *встановлення достовірності різниці за середніми арифметичними значеннями* досліджуваного біологічного показника обчислюють абсолютне значення різниці середніх величин $d = |M1 - M2|$ і коефіцієнт Стюдента:

Коефіцієнт Стьюдента

$$t = \frac{d}{m_d}$$

$m_d = \sqrt{m_1^2 + m_2^2}$ стандартна похибка різниці (для вибірок $n > 20$ $n_1 = n_2$).

$$m_d = \sqrt{\frac{\sum (x_1 - M_1)^2 + \sum (x_2 - M_2)^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

Для ($n < 20$) і $n_1 \neq n_2$, то

- Звичайно порівнюють контроль і дослід, експериментальні та літературні дані і т. п. Ймовірність твердження p про статистичну істотність різниці d знаходимо за коефіцієнтом Стьюдента в таблиці залежно від кількості ступенів вільності $v = n_1 + n_2 - 2$. Різниця є статистично істотною, якщо її ймовірність $p \geq 0,95$

Парний критерій Стьюдента

- Для оцінки ефективності лікування, ми обираємо дві групи: одна піддається лікуванню, інша - ні. Далі ми визначаємо середні арифметичні двох груп та встановлюємо статистично істотну різницю між ними. У іншому випадку ми набираємо лише одну групу, вимірюючи значення ознаки *до та після* лікування і визначаємо зміну даної ознаки.

Приклад

- П. Левін досліджував вплив куріння на функцію тромбоцитів, а саме агрегацію тромбоцитів – частку тромбоцитів, які злиплися під впливом аденозиндифосфату – речовини, що стимулює агрегацію.
- Одинадцятьом чоловікам було запропоновано викурити по сигареті. Перед курінням і після нього були взяті зразки крові та визначена агрегація тромбоцитів. Потрібно встановити з певним рівнем ймовірності чи сталися зміни у досліді порівняно з контролем.

Розв'язання.

№ п/п	Контроль (до куріння)	Дослід (після куріння)	Попарна різниця (x)	$x - M$	$(x - M)^2$
1	25	27	2	-8	64
2	25	29	4	-6	36
3	26	36	10	0	0
4	28	40	12	2	4
5	30	46	16	6	36
6	42	55	13	3	9
7	51	55	4	-6	36
8	51	77	26	16	256
9	56	65	9	-1	1
10	61	60	-1	-11	121
11	65	80	15	5	25
n= 11			$\Sigma x=110$		$\Sigma (x-M)^2 = 588$

- Отже, середня різниця , а її стандартна похибка

$$m = \sqrt{\frac{\sum (x - M)^2}{n(n-1)}} = \sqrt{\frac{588}{110}} = 2,3$$

- Коефіцієнт Стьюдента t у цьому випадку дорівнює

$$t = \frac{M}{m} = \frac{10}{2,3} = 4,3; \quad v = n - 1 = 11 - 1 = 10$$

- Значенню $t = 4,3$ для $v = 10$ відповідає імовірність $p = 0,99$. Отже, можна стверджувати, що куріння приводить до збільшення агрегації тромбоцитів.

Критерій Фішера

- Порівняння двох експериментальних вибірок з метою встановлення достовірності різниці за величиною мінливості досліджуваного біологічного показника здійснюють за допомогою критерію Фішера. Першим кроком цього аналізу – це обчислення дисперсії для кожної з вибірок:

$$\sigma^2 = \frac{\sum(x-M)^2}{n-1}$$

- І визначення на їх основі коефіцієнта Фішера

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

- Обчислене значення коефіцієнта Фішера F порівнюємо з F табл. Для $p \geq 0,95$, яке знаходимо на перетині двох величин кількості ступенів вільності: $v_1 = n_1 - 1$ і $v_2 = n_2 - 1$. Якщо $F \geq F_{\text{табл.}}$, то різниця в мінливості є статистично істотною з заданим рівнем ймовірності.

Порівняльний аналіз з використанням засобів Excel


- Порівняльний аналіз наявного експериментального матеріалу здійснюють за допомогою пакету *Аналізу даних* . Обрати одну з запропонованих програм:
- Двовибірковий t-тест з однаковими дисперсіями;
- Двовибірковий t-тест з різними дисперсіями;
- Парний двовибірковий t-тест для середніх.


- Двовибірковий t-тест перевіряє рівність середніх значень генеральної сукупності для кожної вибірки. Ці три способи допускають наступні умови: рівні дисперсії генерального розподілу, дисперсії генеральної сукупності не рівні, а також представлення двох вибірок до і після спостереження для одного і того ж об'єкта.

Двовибірковий t -тест з різними дисперсіями.

Двухвыборочный t -тест с различными дисперсиями ? X

Входные данные

Интервал переменной 1: 


Интервал переменной 2: 

Гипотетическая средняя разность:

Метки

Альфа:

Параметры вывода

Выходной интервал: 

Новый рабочий лист:

Новая рабочая книга

OK

Отмена

Справка

Microsoft Excel - t-критерій

Файл Правка Вигляд Вставка Формат Сервіс Дані Вікно Довідка

100%

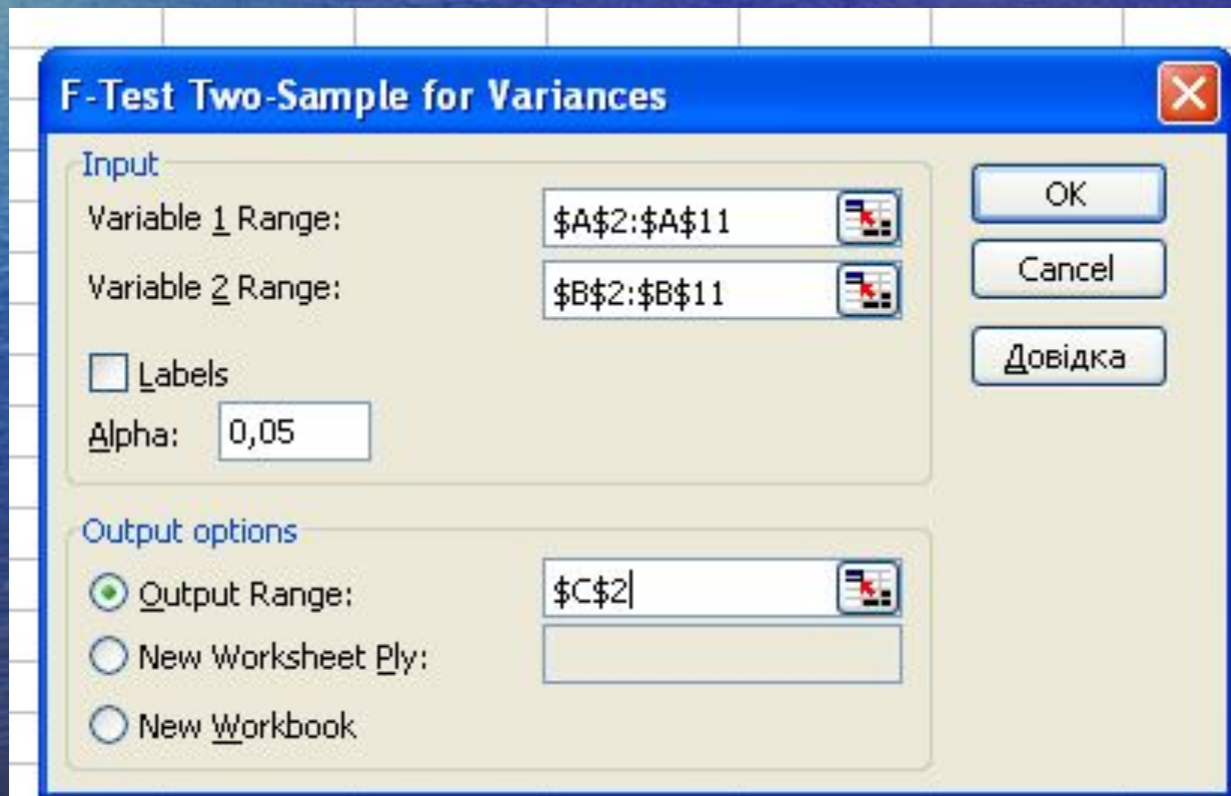
Arial 10 Ж К П

D1 ✗ ✓ fx Двовибірковий t-тест з різними дисперсіями

	A	B	C	D	E	F	G
1				Двовибірковий t-тест з різними дисперсіями			
2	Контрольна	Серія з			Variable 1	Variable 2	
3	серія	опроміненням		Mean	10,788	9,363	
4	12,9	10,3		Variance	1,998	0,900	
5	9	7,9		Observations	8,000	8	
6	11,6	10,1		Hypothesized Mean	0,000		
7	10,5	8,2		df	12,000		
8	9,8	9,9		t Stat	2,368		
9	9,2	10,1		P(T<=t) one-tail	0,018		
10	11,1	9,7		t Critical one-tail	1,782		
11	12,2	8,7		P(T<=t) two-tail	0,036		
12				t Critical two-tail	2,179		
13							

Порівняльний аналіз між дисперсіями двох експериментальних груп

- *Двовибірковий F-тест для дисперсій*



F-Test Two-Sample for Variances

Input

Variable 1 Range:

Variable 2 Range:

Labels

Alpha:

Output options

Output Range:

New Worksheet Ply:

New Workbook

OK

Cancel

Довідка

Оскільки, обчислене значення $F=15,708 > F_{\text{крит.}}=3,179$, то приймається гіпотеза про достовірну різницю між дисперсіями двох вибірок, тобто різницю в їх мінливості.

Microsoft Excel - Книга1

Файл Правка Вигляд Вставка Формат Сервіс Дані Вікно Дов...

Arial 10 Ж К П

	A	B	C	D	E	F
1	Вибірка 1	Вибірка 2				
2	1,45	2,2	F-Test Two-Sample for Variances			
3	1,87	1,95				
4	1	2,09		<i>Variable 1</i>	<i>Variable 2</i>	
5	0,65	2,41	Mean	1,715	2,126	
6	2,3	2,31	Variance	0,361	0,0230	
7	2,52	2,1	Observations	10	10	
8	1,89	2	df	9	9	
9	2,37	2,2	F	15,708		
10	1,5	1,98	P(F<=f) one-tail	0,000174		
11	1,6	2,02	F Critical one-tail	3,179		
12						
13						