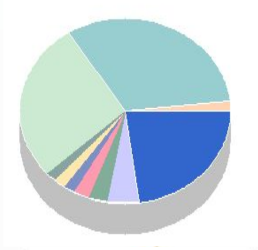
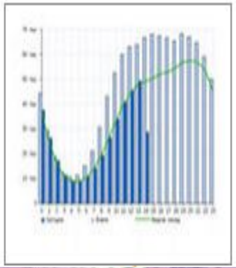
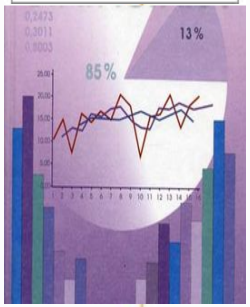






13 %
85 %



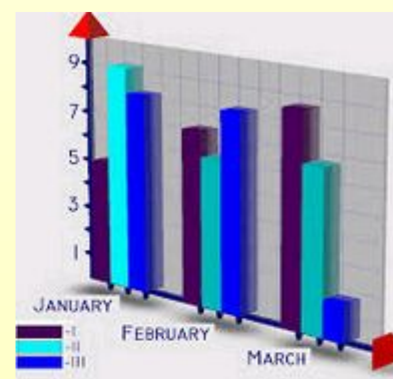

ПИТАННЯ ДЛЯ ОБГОВОРЕННЯ

1. Основні поняття статистики.

2. Оцінка параметрів розподілу за малими вибірками. Довірчі інтервали.

3. Задачі статистичної перевірки гіпотез. Критерії для перевірки статистичних гіпотез.

4. Застосування кореляційного та регресійного аналізу у медицині.

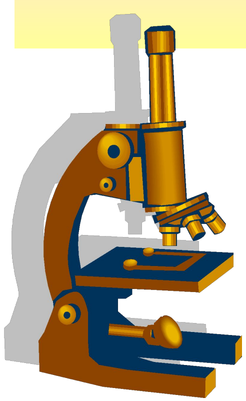


Основні поняття та визначення

Статистика – це наука, яка вивчає статистичні методи збирання, опрацювання, подання, аналізу та інтерпретації даних.

Стан організму пацієнта характеризується сукупністю властивостей (АТ , температура, рівень еритроцитів, лейкоцитів…)

Параметри – властивості, які піддаються оцінці у будь-якій формі якісній або кількісній.



Приклад, рівень глюкози у крові пацієнта становить 5,58 ммоль/л

Основні поняття та визначення

Випадкова величина – величина, яка в результаті експерименту може набути певне значення (X_1, X_2,..,X_n), яке заздалегідь невідоме.

Приклад, кількість пацієнтів, які відвідали поліклініку 16.02.17 (34 пацієнта); артеріальний тиск пацієнта (120/85).

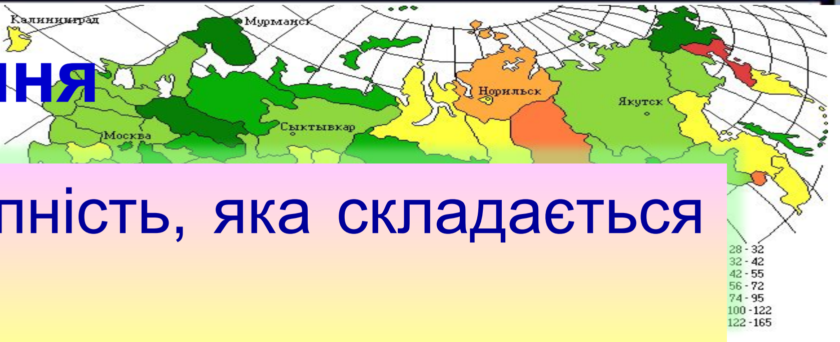
Дискретною випадковою величиною називається величина, яка може набути лише окремі, ізольовані одне від одного значення.

Неперервною випадковою величиною називається величина, яка може набути довільного значення із проміжку (будь-які числа).

Приклад, кількість дітей, які народилися протягом 15.02.17 (7 дітей);

Приклад, 1) рівень тиреотропного гормону (0,0078 mTU/ml).

Основні поняття та визначення



Генеральна сукупність – сукупність, яка складається з усіх одиниць спостереження.

Приклад, кількість усіх хворих на туберкульоз у світі

Статистична сукупність – сукупність, яка складається з певного числа відносно однорідних елементів, взятих разом у певних межах простору і часу.

Приклад, кількість хворих на туберкульоз у 2016 році у Києві.

Вибіркова сукупність (вибірка) – це частина генеральної або статистичної сукупності, яка відображає основні її характеристики.

Основні поняття та визначення

однорідність

Властивості вибірки

репрезентативність

Репрезентативність – здатність вибірки відтворювати генеральну сукупність. Обяг вибірки має бути достатнім для відображення структури генеральної сукупності

Обсяг вибірки (n) – число варіантів, включених у вибіркову сукупність

Із статистичних міркувань рекомендується, щоб число варіантів складало не менше 30-35

Основні поняття та визначення

Варіаційний ряд – сукупність проранжованих по величині значень, які отримані у результаті спостереження певного параметра.

Приклад, 7 пацієнтам виміряли артеріальний тиск

Пацієнт	П1	П2	П3	П4	П5	П6	П7
С.А. Тиск	120	115	120	140	135	120	135

Побудуємо варіаційний ряд:

Пацієнт							
САТ							

Обсяг вибірки:
n=7

Основні поняття та визначення

Варіанти – складові варіаційного ряду.

Пацієнт	П2	П1	П3	П6	П5	П7	П4
САТ	**115**	**120**	**120**	**120**	**135**	**135**	**140**

Приклад, 115, 120, 135, 140

Середні величини – узагальнюючі числові характеристики однорідних величин, які за допомогою одного числа характеризують варіаційний ряд.

Основні поняття та визначення

До середніх величин відносяться:

1. Середньоарифметична величина

Обчислюється за формулою

$$X_{серед} = \frac{\sum_{n}^{i=1} x_i}{n}$$

Приклад: $$X_{серед} = \frac{115+120+120+120+135+135+140}{7} = 126{,}4$$

Excel: функція **СРЗНАЧ**

Варіаційний ряд

Пацієнт	П2	П1	П3	П6	П5	П7	П4
АТ	115	120	120	120	135	135	140

Основні поняття та визначення

До середніх величин відносяться:

2. ***Мода*** – значення, яке найчастіше зустрічається у серії спостережень

Приклад: Мода=120 Excel: функція **Мода**

В.Р.2 (125; 127;130), Мода не існує

3. ***Медіана*** – значення, яке поділяє варіаційний ряд на дві частини

Excel: функція **Медіана**

Приклад: якщо *n* – непарне, то Медіана=120,

Приклад: якщо *n* – парне, то Медіана= $=\frac{120+135}{2}=127,5$

срзнач

Варіаційний ряд

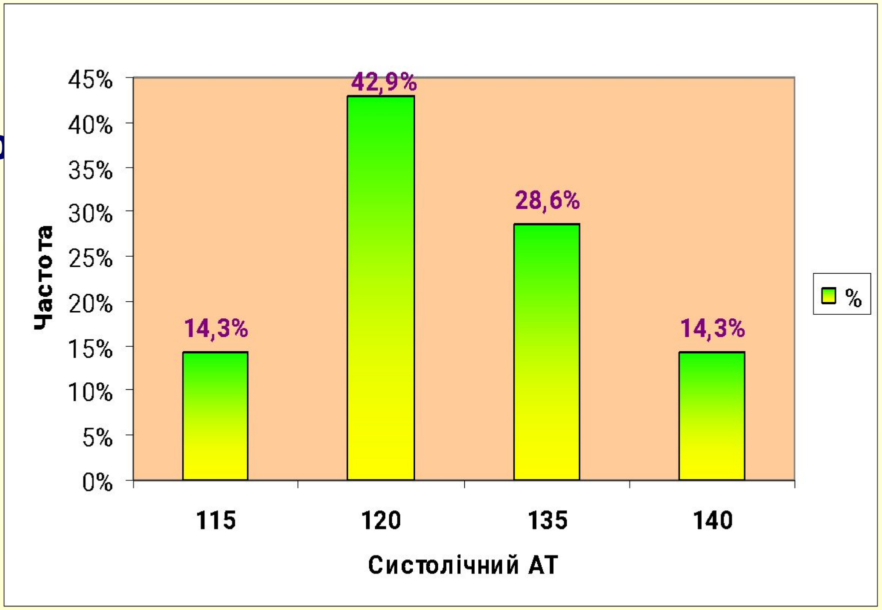
Пацієнт	п2	п1	п3	п6	п5	п7	п4	П8
АТ	115	120	120	120	135	135	140	145

Основні поняття та визначення

Частота – абсолютна чисельність окремих варіант у сукупності, яка вказує на те, скільки разів зустрічається певна варіанта у варіаційному ряді.

Excel: функція ЧАСТОТА або СЧЁТЕСЛИ

Приклад:
115 – 1 раз, що становить **14,3%**
120 – 3 рази, **42,9%**
135 – 2 рази, **28,6%**
140 – 1 раз, **14,3%**

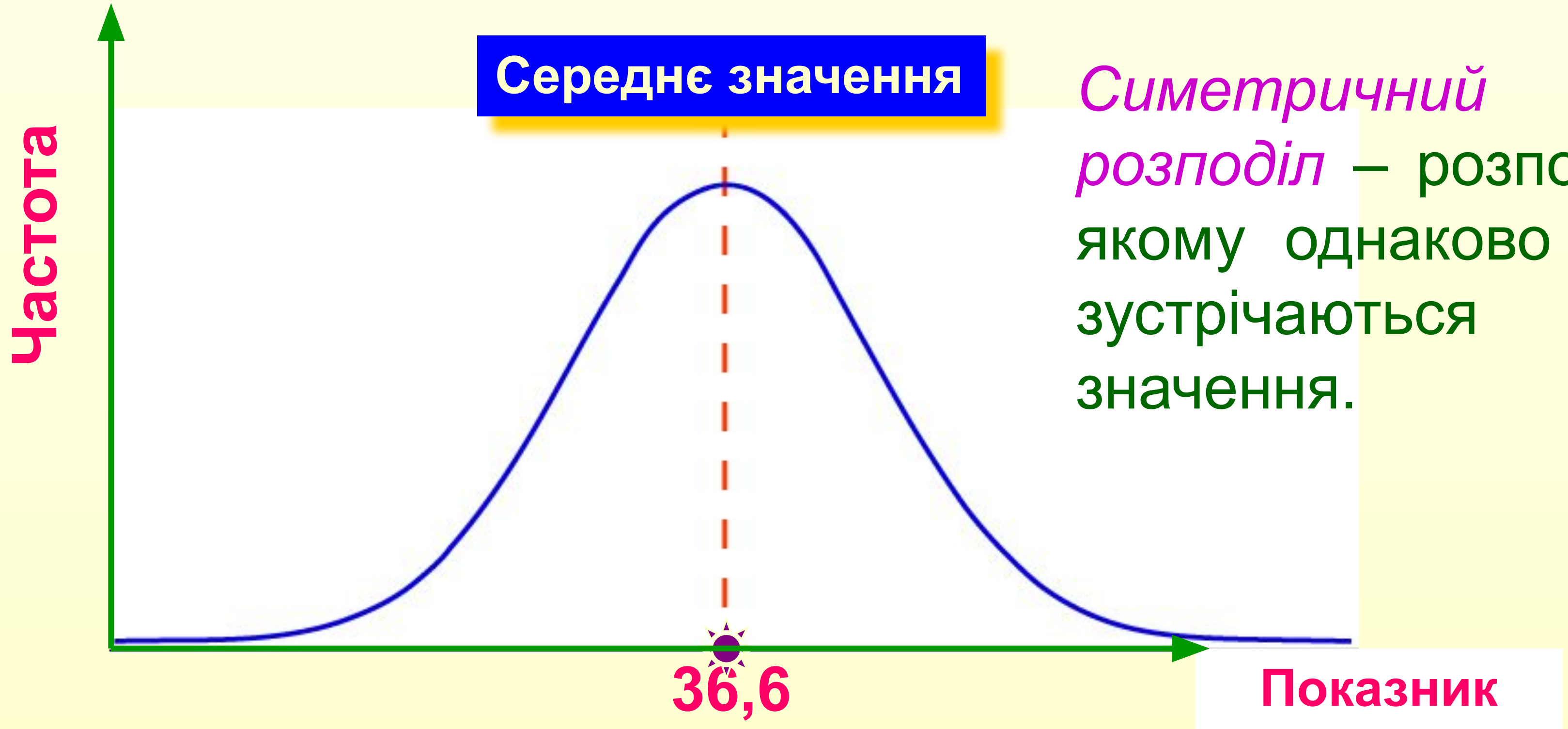


Варіаційний ряд

Пацієнт	П2	П1	П3	П6	П5	П7	П4
АТ	115	120	120	120	135	135	140

Розподіл накопичених частот за значеннями спостережень

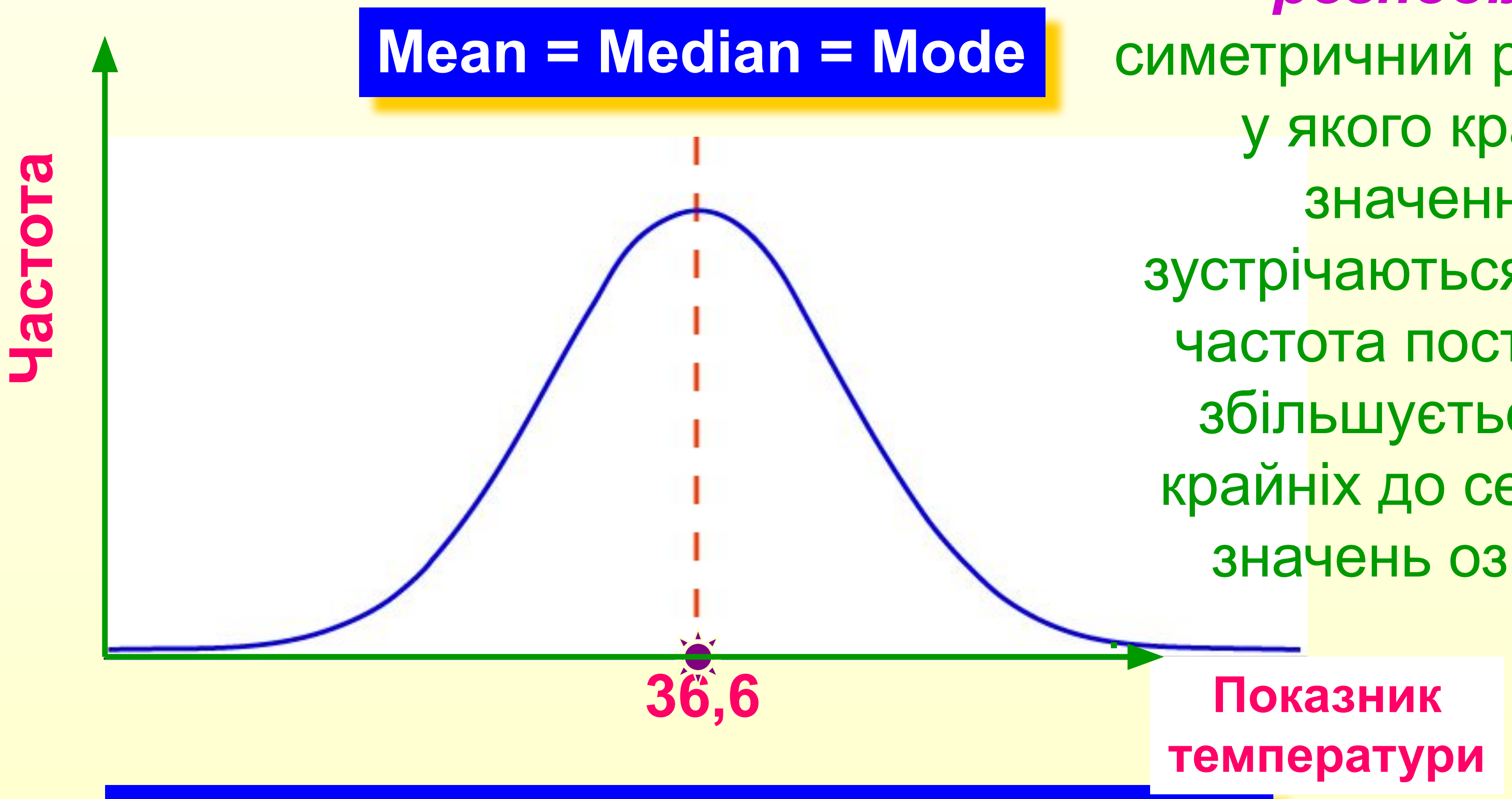
Для емпіричних розподілів використовують їх графічне представлення, найбільш розповсюдженими способами якого є гістограма та полігон частот



Симетричний розподіл – розподіл, у якому однаково часто зустрічаються крайні значення.

Нормальний розподіл симетричний відносно середнього значення

Розподіл накопичених частот за значеннями спостережень



Нормальний розподіл – симетричний розподіл, у якого крайні значення зустрічаються рідко і частота поступово збільшується від крайніх до середніх значень ознаки.

Нормальний розподіл симетричний відносно середнього значення

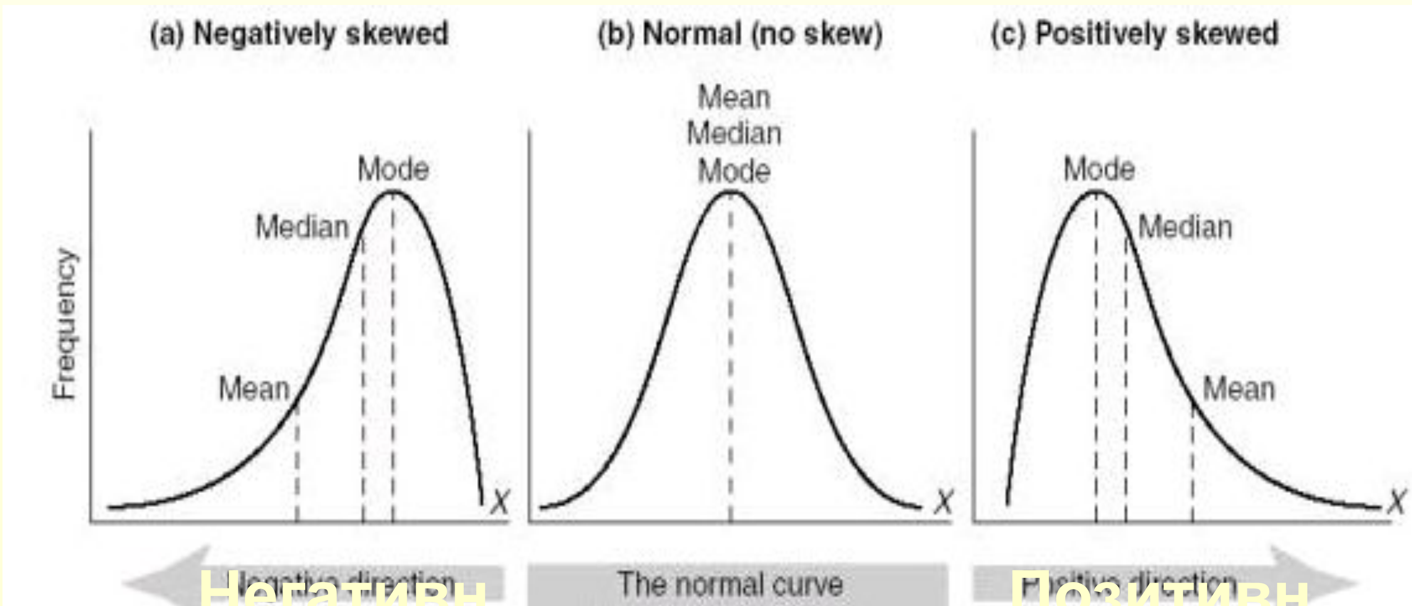
Розподіл накопичених частот за значеннями спостережень

Асиметричний розподіл – *лівосторонній* (переважанням частот малих значень), *правосторонній* (переважанням частот великих значень).

Асиметрія буває **позитивною** і **негативною**.
Негативна асиметрія зсувається **вправо (А<0) – «хвіст спереду»**
Позитивна асиметрія зсувається **вліво (А>0) – "хвіст в кінці"**

$$A_{ss} = \frac{\sum (x_i - \bar{x})^3 / n}{S^3}$$

Розподіл накопичених частот за значеннями спостережень



Негативн
а
асиметрі

A<0

A=0

Позитивн
а
асиметрі

A>0

Закони розподілу випадкових величин

Закон розподілу випадкових величин - функціональна залежність між значеннями випадкових величин та ймовірностями з якими вони приймають ці значення.
Закон розподілу може бути заданий у вигляді *таблиці*, *формули* або *графіка*.

Закони розподілу випадкових величин

Закони розподілу ***дискретних випадкових величин***

Біноміальний розподіл (розподіл Бернуллі)

Розподіл Пуассона

Закони розподілу ***неперервних випадкових величин.***

Нормальний закон розподілу (Гаусса)
Розподіл Х2
Розподіл Ст'юдента *(Госсета)*

Біноміальний розподіл (розподіл Бернуллі)

Дискретна випадкова величина *x*, яка може приймати тільки цілі невід'ємні значення з ймовірностями

$P_n(X=m)=C_n^m p^m q^{n-m}$, *m=0,1,...,n*, де *p* – ймовірність появи події в кожному випробуванні, *m* – кількість сприятливих подій, *n* – загальна кількість випробувань, *q=1–p*,

$C_n^m=\dfrac{n!}{m!(n-m)!}$, називається розподіленою за **біноміальним законом** з математичним сподіванням *np*, та дисперсією – *npq*.

Закон Берніллі використовується тоді, коли необхідно знайти ймовірність появи випадкової події, яка реалізується рівно *m* раз у серії з *n* випробувань.

Біноміальному закону розподілу підпорядковуються випадкові події такі, як число викликів швидкої допомоги за певний проміжок часу, черги до лікаря в поліклініці, епідемії тощо

$$\boxed{P_n(X=m)=C_n^m p^m q^{n-m}}$$

Біноміальний розподіл. Приклад

- Розрахувати ймовірність того, що з 20 дітей, які народяться, рівно 12 дітей будуть дівчатами

- n=20, m=12, p=0.5

$$P_n(X=m)=C_{20}^{12}p^{12}q^{20-12},\qquad C_{20}^{12}=\frac{20!}{12!(20-12)!}$$

Excel: ф-я **БИНОМРАСП (m; n; p)**

Відповідь: БИНОМРАСП (12;20;0,5) = 0,16.
Ймовірність того, що народяться рівно 12 дівчат з 20 дітей становить 0,16

Розподіл Пуассона

Дискретна випадкова величина *X*, яка може приймати тільки цілі невід'ємні значення з ймовірностями

$P_n(X=m)=\frac{\lambda^m e^{-\lambda}}{m!}, m=0,1,\ldots,\lambda>0$, називається розподіленою за **законом Пуассона** з математичним сподіванням λ і дисперсією λ, де $\lambda = np$.

Розподіл Пуассона, як граничний біноміальний використовується при вирішенні задач надійності медичного обладнання та апаратури, розповсюдження епідемії, викликів до хворого дільничих лікарів та в інших задачах масового обслуговування.

$$\boxed{P_n(X=m)=\frac{\lambda\ e^{-\lambda}}{m!}}$$

Розподіл Пуассона. Приклад

Вакцина формує імунітет від деякого захворювання з ймовірністю 0,999. Провакциновано 4000 мешканців міста. Яка ймовірність того, що двоє з них не набули імунітету.

$$\lambda = np = 4000 \times 0{,}001 = 4$$

$$P(x = 2) = \frac{4^2}{2!} e^{-4} = 0{,}147 \; .$$

Excel: ф-я **ПУАССОН** (m; л) (2; 4)

Відповідь: ПУАССОН (2;4)= 0,147, отже, ймовірність того що двоє пацієнтів не набудуть імунітету становить 0,15

Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Середньоквадратичне відхилення – величина, яка показує ступінь розсіювання варіаційного ряду навколо середньої величини.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_{cep})^2}{n-1}}$$

$$\sigma = \sqrt{\frac{(115-126{,}4)^2 + (120-126{,}4)^2 + (120-126{,}4)^2 + (120-126{,}4)^2 + (135-126{,}4)^2 + (135-126{,}4)^2 + (140-126{,}4)^2}{7-1}}$$

$\sigma = 9{,}9$

Excel: функція **Стандотклона**

Варіаційний ряд

Пацієнт	п2	п1	п3	п6	п5	п7	п4
САТ	115	120	120	120	135	135	140

Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Дисперсія (D) – міра відхилення значень випадкової величини від центру розподілу.

$$D=\sigma^2 \qquad D=9{,}9^2=97{,}62$$

Excel: Дисп

$$D=\frac{\sum_{i=1}^{n}(x_i-x_{сер})^2}{n-1}$$

Помилка репрезентативності – найважливіша статистична величина, необхідна для оцінки достовірності результатів дослідження.

$$m=\frac{\sigma}{\sqrt{n}} \qquad m=\frac{9{,}9}{\sqrt{7}}=3{,}73$$

Excel: формула=σ/корень(n)

Варіаційний ряд

Пацієнт	П2	П1	П3	П6	П5	П7	П4
САТ	115	120	120	120	135	135	140

Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Математичним сподіванням будь-якої величини називається сума всіх можливих для неї значень, помножених на їх ймовірності.

Математичне сподівання для *неперервної* випадкової величини:

$$MX = \int_{-\infty}^{+\infty} xf(x)dx$$

Математичне сподівання для *дискретної* випадкової величини:

$$MX = \sum_{i=1}^{n} x_i p(x_i)$$

Дисперсія для *неперервної* випадкової величини:

$$DX = \int_{-\infty}^{+\infty} (x - MX)^2 f(x)dx$$

Дисперсія для *дискретної* випадкової величини:

$$DX = \sum_{i=1}^{n} (x_i - MX)^2 p(x_i)$$

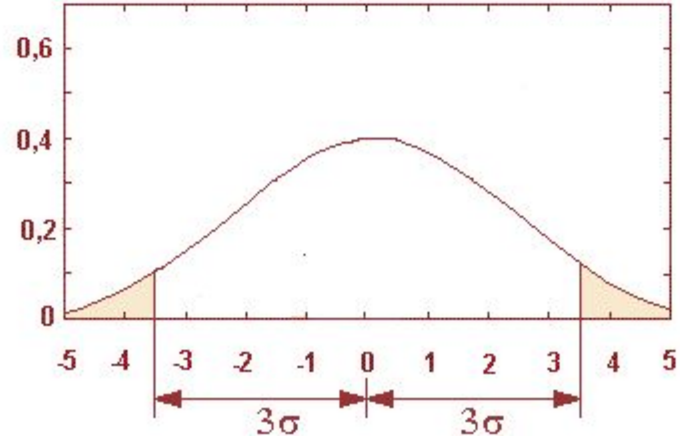
Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

- .В практичних застосування важливим є правило «трьох сігм»

$$P(|X-m| \geq 3\sigma) = 0{,}0027$$

- Тобто ймовірність того, що нормально розподілена випадкова величина відрізняється від свого математичного сподівання більше ніж на три сігма приблизно дорівнює **0,0027**, така подія є практично неможливою.

якщо дані підпорядковані нормальному закону розподілу, то майже *вірогідно* (з похибкою до 0,3%) вони повинні бути в межах $\pm 3\sigma$ від середнього значення (математичного сподівання).



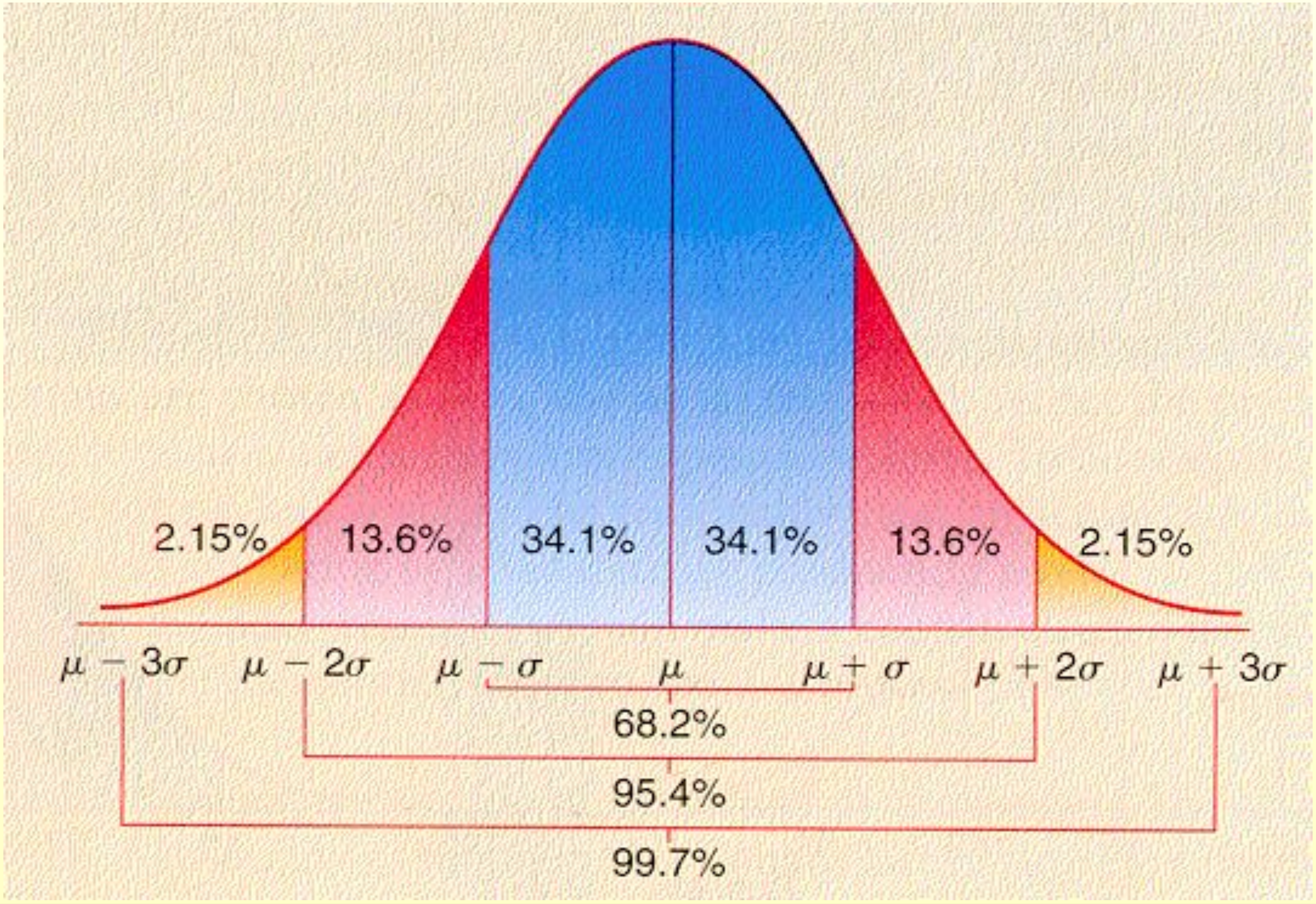
Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Приклад застосування правила “3 сігм”

	Температура				
Пацієнт 1	36,3				
Пацієнт 2	36,4				
Пацієнт 3	36,4				
Пацієнт 4	36,5				
Пацієнт 5	36,5		срз	36,6	
Пацієнт 6	36,5		сигма	0,2	
Пацієнт 7	36,6		3*сигма	0,5	
Пацієнт 8	36,6				
Пацієнт 9	36,6		"СРЗ-3сигма"		"СРЗ +3сигма"
Пацієнт 10	36,6		36,1	ймов. 0,03	37,1
Пацієнт 11	36,6				
Пацієнт 12	36,7				
Пацієнт 13	36,7				
Пацієнт 14	36,7				
Пацієнт 15	36,8				
Пацієнт 16	36,8				
Пацієнт 17	36,9				

Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Приклад, правило “3 сігм”

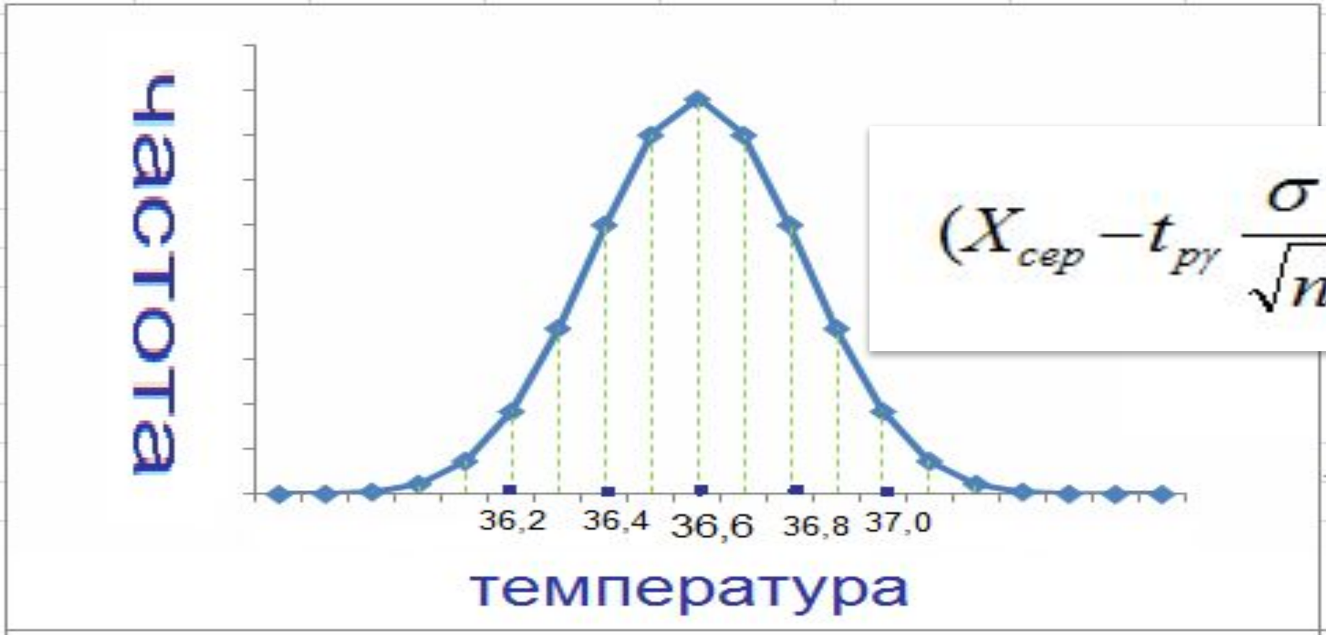


$n < 30$

Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Точкові оцінки параметрів розподілу можна прийняти як орієнтовні дані оцінювання результатів спостережень. Їх недолік – у невідповідності точності оцінювання параметрів, особливо при оцінюванні малих вибірок

Довірчий інтервал – інтервал, у межах якого із заданою довірчою ймовірністю можна очікувати значення оцінюючої випадкової величини.



$$\left(X_{сер} - t_{p\gamma}\frac{\sigma}{\sqrt{n}};\ X_{сер} + t_{p\gamma}\frac{\sigma}{\sqrt{n}}\right)$$

Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Приклад виконання статистичної обробки в середовищі Excel

Пацієнт	Швидкість кровотоку	
x_1	11,2	
x_2	11	
x_3	10,9	
x_4	11,22	
x_5	11,4	
x_6	11	
x_7	11,1	
x_8	11,3	
x_9	11,2	
x_{10}	11,301	
ср значення	11,16	
відхилення (σ)	0,16	
обсяг (n)	10	
помилка (m)	0,05	
Δm	0,10	
11,06	довірчий інтервал	11,26

Ф-я срзнач

Ф-я стандотклона

Ф-я счёт (необов'язково)

=0,16 (відхилення)/корень(10)

1 спосіб. Ф-я доверит

2 способ. $=t_{py}*m$

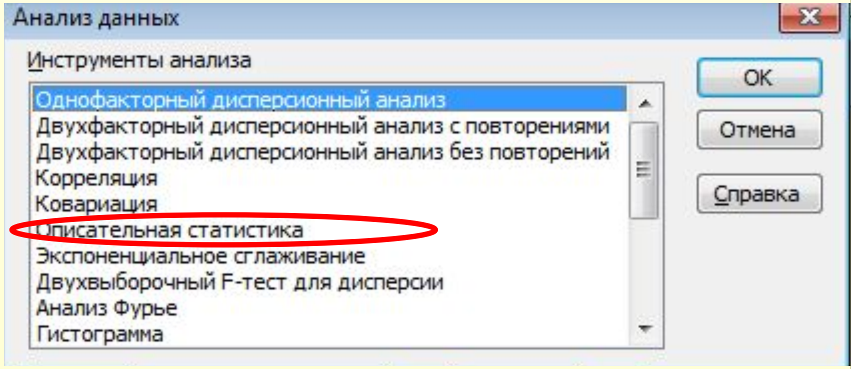
t_{py} - коефіцієнт Стюдента знаходять у спеціальній таблиці (стор. 81)

Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Приклад виконання статистичної обробки в середовищі Excel

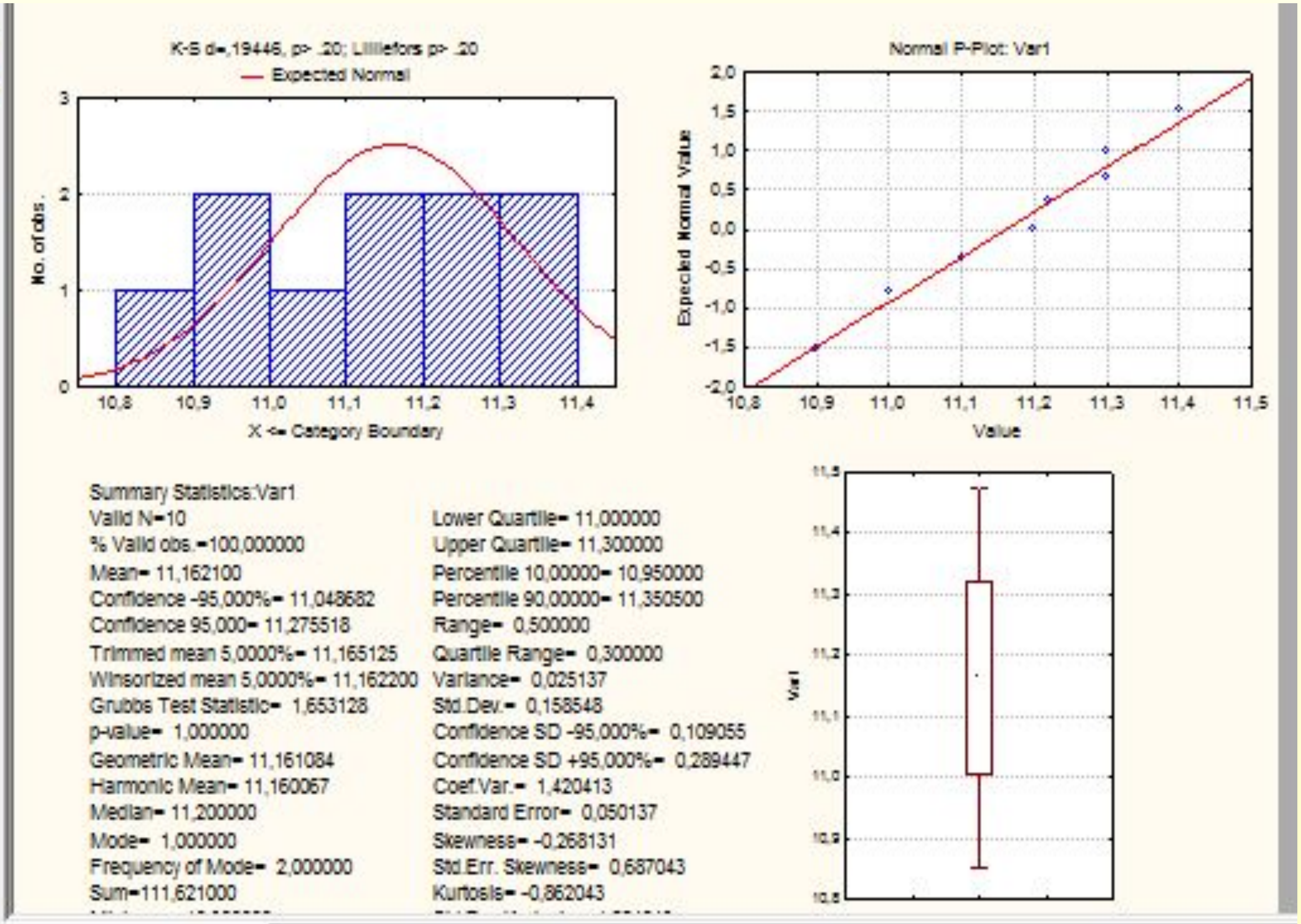
Среднее	11,2
Стандартная ошибка	0,05
Медиана	11,2
Мода	11,2
Стандартное отклонение	0,16
Дисперсия выборки	0,03
Эксцесс	-0,86
Асимметричность	-0,27
Интервал	0,5
Минимум	10,9
Максимум	11,4
Сумма	111,6
Счет	10

Застосування вбудованого пакету аналізу даних. Розділ "Описова статистика"



Оцінка параметрів розподілу за малими вибірками. Довірчий інтервал.

Приклад виконання статистичної обробки в програмі ***STATISTICA***

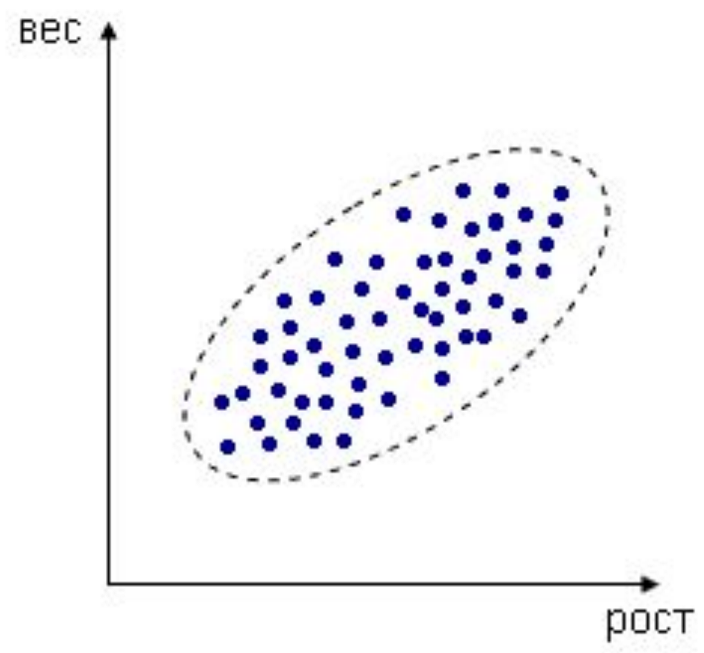


Застосування модуля програми *STATISTICA* *Basic Descriptive Statistics*

Застосування кореляційного аналізу у медицині

Будь - яке явище не може бути усвідомленим і зрозумілим без обґрунтування його зв'язків з іншими явищами. Щоб пізнати сутність явищ, необхідно розкрити їх взаємовплив, кількісно визначити вплив тих або інших факторів. Дослідження зв'язків реалізується за допомогою **кореляційного аналізу**.

Припустимо, проводиться незалежне вимірювання різних параметрів у одного типу об'єктів. З цих даних можна отримати якісно нову інформацію - про взаємозв'язок цих параметрів.



Приклад, вимірювання зросту і ваги людини, кожне вимірювання представлено точкою в двовимірному просторі.

Застосування кореляційного аналізу у медицині

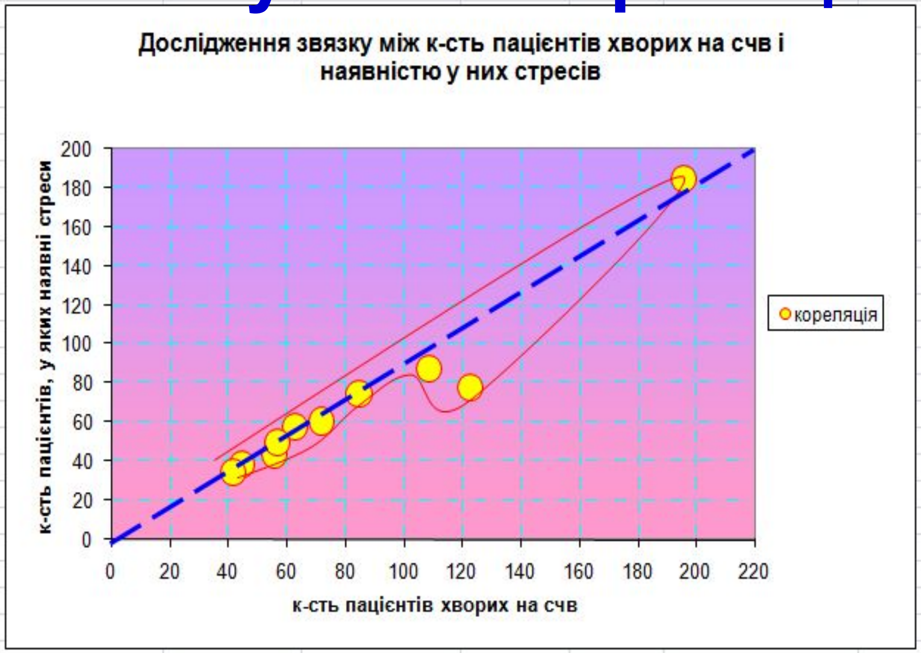
Кореляційний аналіз –- метод, що дозволяє виявити залежність між кількома випадковими величинами

Кореляція (кореляційна залежність) – статистичний взаємозв'язок двох або декількох випадкових величин. При цьому зміни однієї або декількох величин, призводить до зміни іншої.

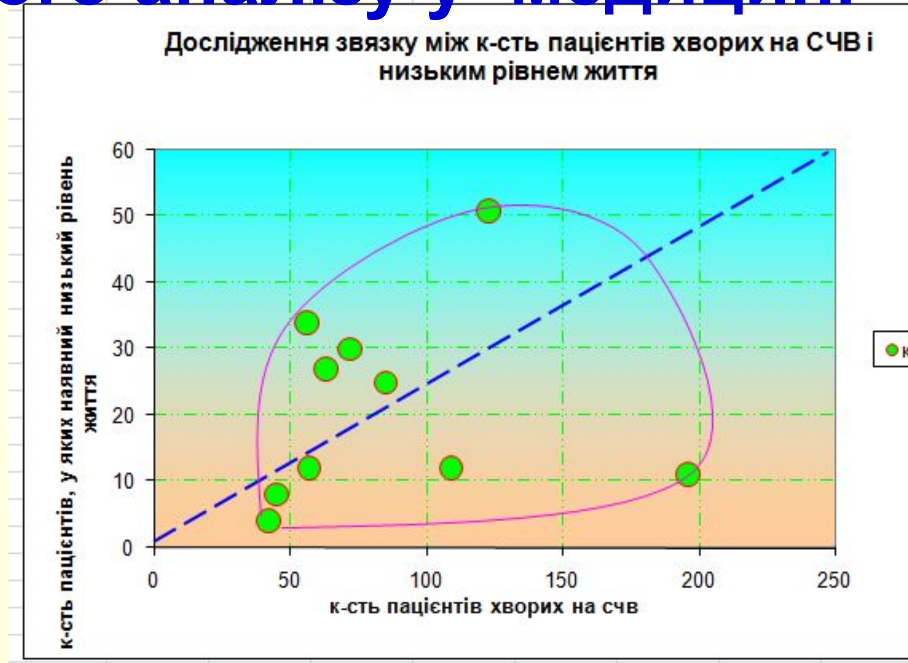
Коефіцієнт кореляції – це число, значення і величина якого характеризує напрямок і силу зв'язку. Значення коефіцієнта кореляції може змінюватися від [-1 до +1]

Excel: Коррел

Застосування кореляційного аналізу у медицині



Висновок: Існує кореляційний зв'язок. Отже, наявність щоденних стресів сприяє розвитку СЧВ. Обрано точкову діаграму для ілюстрації кореляції



Висновок: Не існує кореляційного зв'язку. Отже, низький рівень життя не сприяє

Область	к-сть пацієнтів	наявні щоденні стреси	наявні шкідливі звички	низький рівень життя	Висока частота перебування на сонці
Київська	123	78	5	51	100
Вінницька	109	88	2	12	45
Чернігівська	56	44	17	34	40
Житомирська	85	75	21	25	42
Черкаська	72	61	15	30	50
Харківська	45	39	20	8	36
Донецька	63	58	13	27	54
Волинська	57	50	21	12	40
Львівська	42	35	8	4	22
Крим	196	185	30	11	190

Застосування кореляційного аналізу у медицині

- Знак коефіцієнта кореляції вказує на <u>напрям</u> – прямий чи зворотній взаємозв'язок між двома змінними.
- Абсолютне значення коефіцієнта кореляції $|r|$ характеризує <u>силу</u> та <u>щільність</u> взаємозв'язку, що розглядається.

<u>Позитивна кореляція</u> – **підвищення** рівня однієї змінної супроводжується **підвищенням** рівня іншої.

<u>Негативна кореляція</u> – **зростання** однієї змінної супроводжується **зниженням** значень іншої.

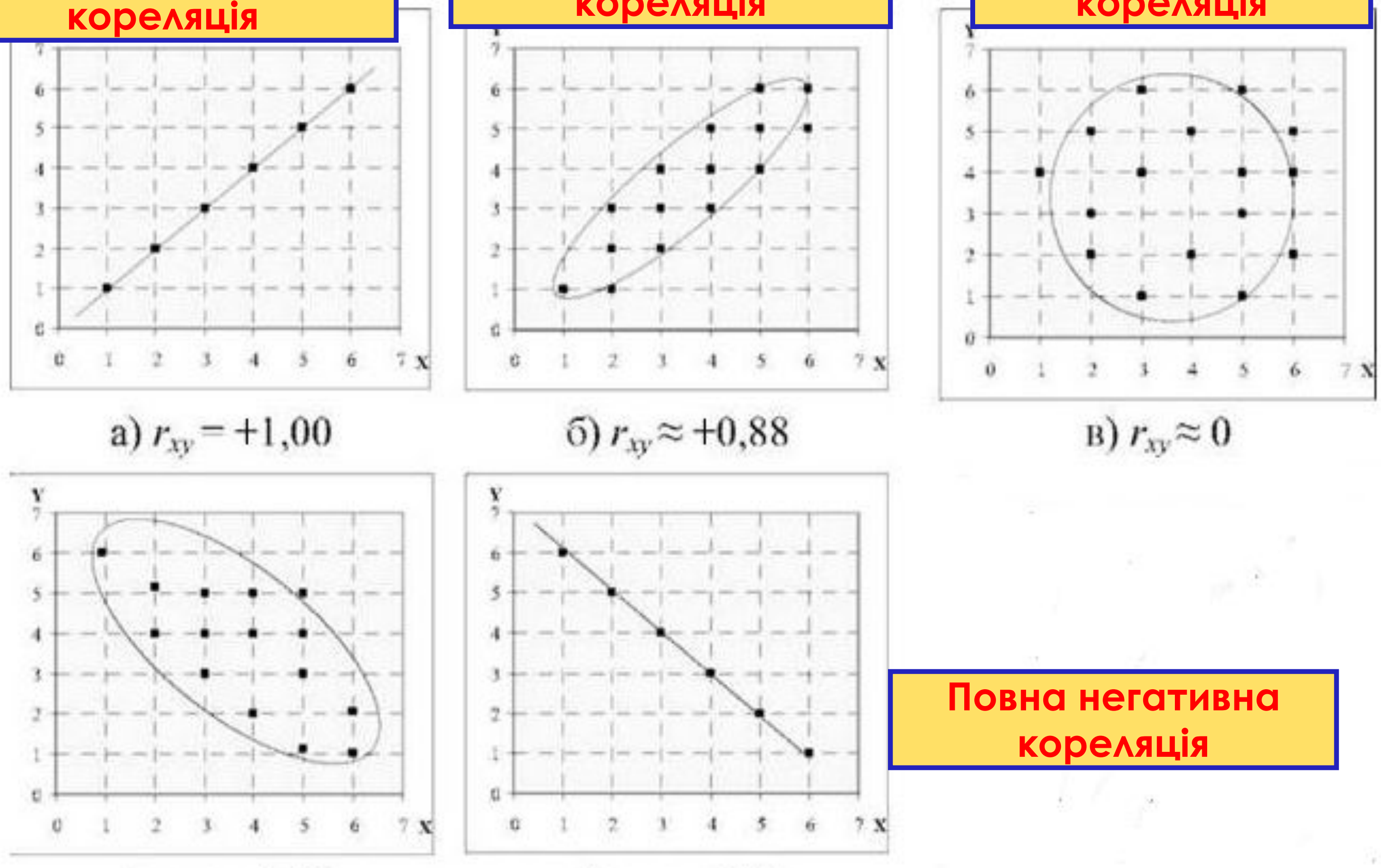
<u>Нульова кореляція</u> – кореляція за відсутності зв'язку змінних.

Застосування кореляційного аналізу у медицині

А) Повна позитивна кореляція

Б) Висока позитивна кореляція

В) Нульова кореляція



а) $r_{xy} = +1{,}00$

б) $r_{xy} \approx +0{,}88$

в) $r_{xy} \approx 0$

г) $r_{xy} \approx -0{,}60$

г) $r_{xy} = -1{,}00$

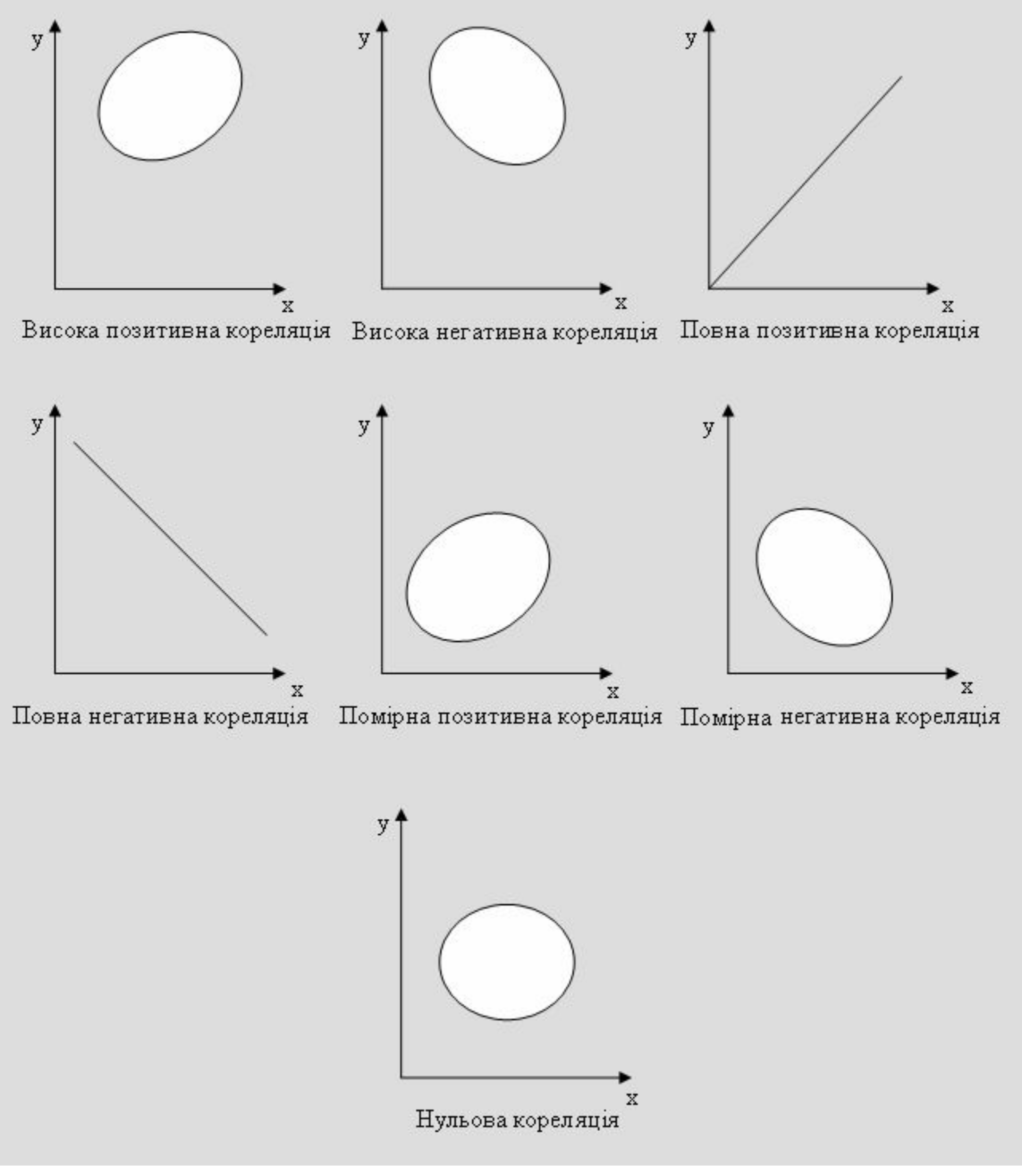
Повна негативна кореляція

Г) Помірна негативна кореляція

Приклади візуальної інтерпретації коефіцієнта кореляції

Застосування кореляційного аналізу у медицині

Зміст концепції кореляції можна з'ясувати за **допомогою так званої діаграми розсіювання**.
При побудові цього графіка по осям координат відкладаються значення відповідних корелюючих характеристик.



Задачі статистичної перевірки гіпотез

- ***Статистичні гіпотези*** – це припущення, котрі відносяться до виду розподілу випадкової величини або окремих його параметрів
- ***Рівень значущості*** – ймовірність, з якою може бути відхилена нульова гіпотеза (для медико-біологічних досліджень достатнім є рівень значущості 0,05).
- ***Довірча ймовірність*** - ймовірність прийняття правильності рішення (для медико-біологічних досліджень достатньою є ймовірність 0,95)
- ***Статистичний критерій*** К – це вирішуюче правило, яке забезпечує прийняття вірності гіпотези
- ***Критична область -*** сукупність значень, при яких основна гіпотеза не приймається

Задачі статистичної перевірки гіпотез

Послідовність операцій при виборі критерію:

1. **Постановка** задачі.
2. Визначити **клас критеріїв**, що використовуються.
3. Визначити **додаткові умови** вибору критерію (багато критеріїв вимагають додаткових умов, без яких їх використання некоректне).
4. **Вибір** конкретного критерію (в багатьох ситуаціях існує декілька рівнозначних критеріїв, придатних для перевірки гіпотези).

Задачі статистичної перевірки гіпотез

Етапи перевірки статистичних гіпотез

1. Визначення статистичної моделі, що буде використовуватися.
2. Формулювання гіпотез **H0** і **H1** (Но – немає відмінностей, Н1 - є).
3. Вибір **критерію**, котрий підходить до висуненої статистичної моделі (наприклад, критерій Стьюдента).
4. Вибір **рівня значущості а** в залежності від надійності висновків, що вимагаються (як правило, а=0,05).
5. *Визначення критичної області* для перевірки гіпотези (**критичне значення критерію – спец.таблиці**).
6. Розрахунок *значення* вибраного статистичного критерію для існуючих даних (**розрахункове значення критерію – обчислюють за формулами**).
7. **Порівняння** розрахованого значення критерію з критичним, а потім вирішують прийняти чи відкинути Н0.

Задачі статистичної перевірки гіпотез

- В медико-біологічних дослідженнях часто виникає задача оцінювання параметрів розподілу за малими вибірками. Для оцінювання параметрів розподілу таких вибірок використовують розподіл Ст'юдента.

Критерій Ст'юдента:

якщо $t_\alpha < t$ – приймається нульова гіпотеза H_0.

якщо $t_\alpha \geq t$ – відхиляється нульова гіпотеза і приймається альтернативна гіпотеза H_1.

Задачі статистичної перевірки гіпотез

Приклад

- Було проведено дослідження дії магнітних полів низької частоти на карциному Герена (результати досліджень в таблиці. В першому стовпчику подані результати інтактної групи, а в другому стовпчику – розмір новоутворення пухлини на яку діяли магнітні поля низької частоти.
- Перевірити ефективність впливу магнітних полів на новоутворення пухлини карциноми Герена

Контрольна група	Дослідна група
0,027	0,075
0,036	0,4
0,1	0,08
0,12	0,105
0,32	0,075
0,45	0,12
0,049	0,06
0,105	0,075

Задачі статистичної перевірки гіпотез

1. Визначення статистичної моделі
2. Формулювання H0 і H1.

Hо – суттєвих відмінностей у групах не виявлено

H1 – наявні суттєвих відмінностей у групах

3. Вибирають *критерій*: *критерій Ст'юдента*.

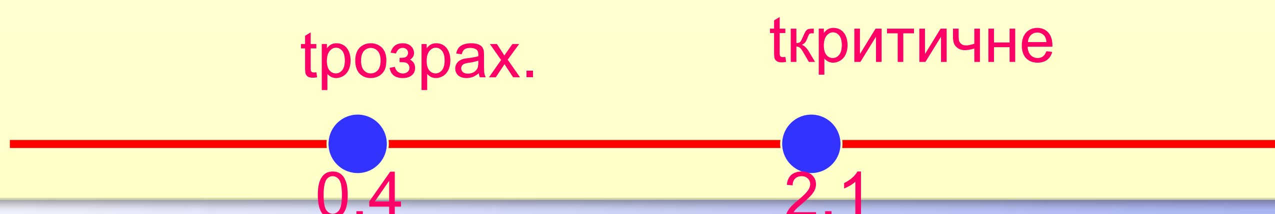
4. Вибирають *рівень* значущості *a*= *0,05*.

5. Визначення *критичної області* для перевірки H0.

спец. таблиці: tкритичне=2,1

6. Розрахунок *значення* вибраного статистичного критерію для існуючих даних. tрозрахункове=0,4

7. Порівняння розрахованого значення критерію з критичним: tкритичне> tрозрахункове (2,1>0,4). Приймається гіпотеза H0 (відмінностей немає)



Задачі статистичної перевірки гіпотез

Приклад

tкритичне=2,1

Знаходження за таблицею кількості ступенів свободи ν значення трьох стандартних критеріїв Ст'юдента (t_{st}), відповідних трьом програмам вірогідності (95%, 99%, 99,9%).

Кількість ступенів свободи	95%	99%	99,9%	Кількість ступенів свободи	95%	99%	99,9%
1	12,7	63,7	637,0	13	2,2	3,0	4,2
2	4,3	9,9	31,6	14-15	2,1	3,0	4,1
3	3,2	5,8	12,9	16-17	2,1	2,9	4,0
4	2,8	4,6	8,6	18-20	2,1	2,8	3,9
5	2,6	4,0	6,9	21-24	2,1	2,8	3,8
6	2,4	3,7	6,0	25-28	2,1	2,8	3,7
7	2,3	3,5	5,4	29-30	2,1	2,7	3,7
8	2,3	3,4	5,0	31-34	2,0	2,7	3,7
9	2,3	3,3	4,8	35-42	2,0	2,7	3,6
10	2,2	3,2	4,6	43-62	2,0	2,7	3,5
11	2,2	3,1	4,4	63-175	2,0	2,6	3,4
12	2,2	3,1	4,3	176	2,2	2,6	3,3

Задачі статистичної перевірки гіпотез

Приклад

Алгоритм оцінки вірогідності відмінностей дослідження двох незалежних вибірок

1. Знаходження середнього арифметичного значення контрольної та дослідної груп.

2. Знаходження середнього квадратичного відхилення окремих вимірювань у групах.

3. Визначення помилок репрезентативності цих груп.

4. Знаходження абсолютного значення середніх арифметичних дослідної та контрольної груп:

$$d = \left| X_{сер} - Y_{сер} \right|.$$

5. Обчислення середньої похибки різниці.

$$m_d = \sqrt{m_1^2 + m_2^2}$$

tрозрахункове=0,4

6. Визначення критерію вірогідності різниці:

$$t_d = \frac{d}{m_d}.$$

7. Знаходження кількості ступенів свободи:

$\nu = n_1 + n_2 - 2.$

8. Знаходження за таблицею кількості ступенів свободи ν (див. нижче) значення трьох стандартних критеріїв Ст'юдента (t_{st}), відповідних трьом програмам вірогідності (95%, 99%, 99,9%).

9. Порівняння критерію вірогідності t_d знайденими значеннями ($t_{st}95\%$, $t_{st}99\%$, $t_{st}99{,}9\%$).

Якщо $t_d < t_{st}95\%$, то вибіркова різниця ненадійна, тобто відмінності у вибірках випадкові.

Якщо $t_{st}95\% \le t_d \le t_{st}99\%$, то вибіркова різниця надійна з імовірністю 95%.

Якщо $t_d \le t_{st}99{,}9\%$, то вибіркова різниця надійна з імовірністю 99,9%.

Задачі статистичної перевірки гіпотез

Приклад виконання статистичної обробки в середовищі Excel

Microsoft Excel - Статистика

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10

M10

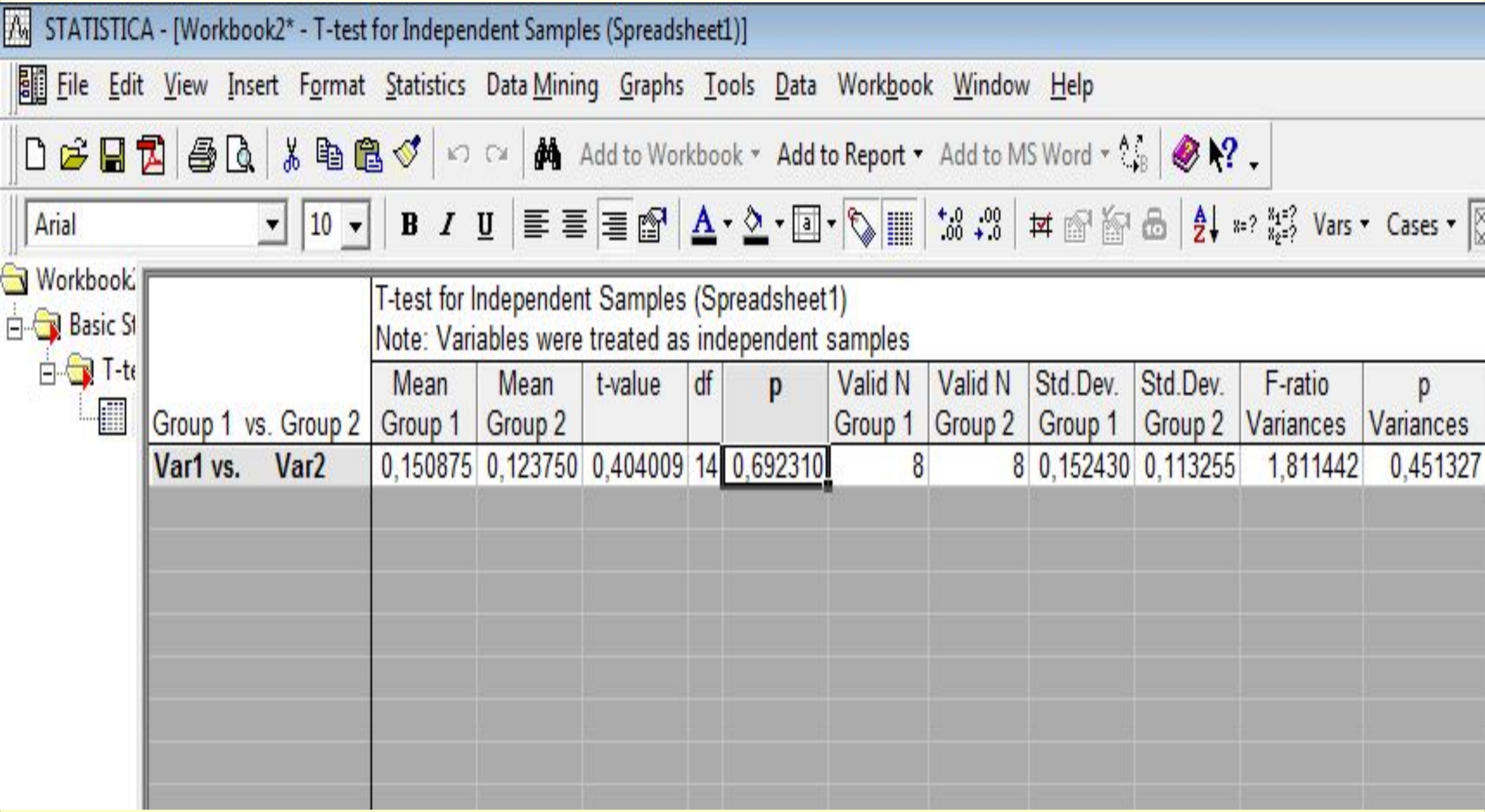
	Контрольна група	Дослідна група
Визначення ефективності методики лікування	0,027	0,075
	0,036	0,4
	0,1	0,08
	0,12	0,105
	0,32	0,075
	0,45	0,12
	0,049	0,06
	0,105	0,075
сер знач	0,15	0,12
відхилення	0,15	0,11
м (помилка)	0,05	0,04
різниця	0,03	
похибка	0,07	
критерій t розрах	0,40	
t критичне	2,1	
Висновок	приймається гіпотеза Н0, про відсутність відмінностей	

Двухвыборочный t-тест с различными дисперсиями

	Переменная	Переменная 2
Среднее	0,15	0,12
Дисперсия	0,02	0,01
Наблюдения	8	8
Гипотетическая разность	0	
df	13	
t-статистика	0,4	
P(T<=t) одностороннее	0,3	
t критическое односторонн	1,8	
P(T<=t) двухстороннее	0,7	
t критическое двухсторон	2,16	

Задачі статистичної перевірки гіпотез

Приклад виконання статистичної обробки в програмі ***STATISTICA***



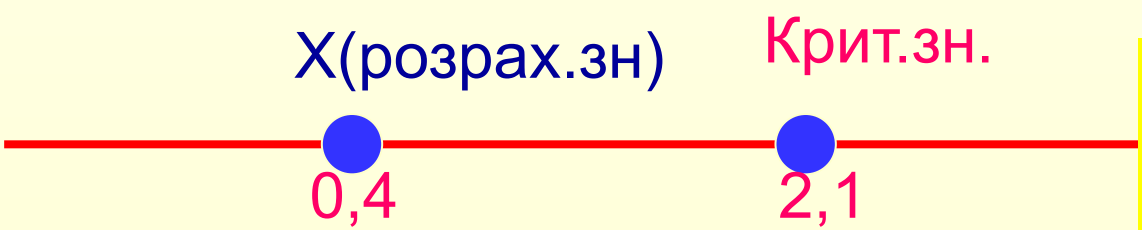
Group 1 vs. Group 2	Mean Group 1	Mean Group 2	t-value	df	p	Valid N Group 1	Valid N Group 2	Std.Dev. Group 1	Std.Dev. Group 2	F-ratio Variances	p Variances
Var1 vs. Var2	0,150875	0,123750	0,404009	14	0,692310	8	8	0,152430	0,113255	1,811442	0,451327

Задачі статистичної перевірки гіпотез

Нехай $\overline{K}$ – значення критерію, що розрахований по вибірці:

– якщо $\overline{K} > x_{\alpha}$, то гіпотеза відхиляється;

– якщо $\overline{K} < x_{\alpha}$, то гіпотеза приймається.



Гіпотеза Н0 приймається – відмінностей немає

Гіпотеза Н1 приймається - відмінності наявні