

10(11) ІНФОРМАТИКА

Рівень стандарту

10

Основи статистичного аналізу даних. Ряди даних

За навчальною програмою 2018 року



Урок 11

teach-inf.at.ua



Чи варто відкривати новий супермаркет в даному мікрорайоні?

Чи можна стверджувати, що на двох виробництвах підтримуються однакові технологічні умови?

Як скоригувати план лісозаготівлі?



Відповісти на подібні питання можна за допомогою статистичного аналізу даних.



Для аналізу кількісних характеристик використовують статистичні дослідження.



Статистика — наука про методи збирання, опрацювання, аналізу та інтерпретації даних, що характеризують масові явища і процеси, тобто явища і процеси, що стосуються не окремих об'єктів, а їх сукупностей.





Завданнями статистичного аналізу є:

- ✓ **опрацювання даних, отриманих в ході експерименту,**
- ✓ **виявлення причинних зв'язків у перебігу явищ і процесів,**
- ✓ **оцінка впливу досліджуваних факторів**
- ✓ **формулювання належних висновків, на підставі яких можуть бути прийняті відповідальні рішення.**





Статистичні дані — сукупність чисел, які дають характеристику певних об'єктів та явищ, що нас цікавлять. Усю множину об'єктів, що є предметом статистичного дослідження, називають **генеральною сукупністю** вимірювань.





Статистичне дослідження складається з трьох етапів.

- 1. Збір і групування первинних статистичних даних, отриманих шляхом спостережень або вимірювань (результати соціологічного опитування, спостереження за природними явищами, проведення серії експериментів тощо).**
- 2. Аналіз отриманих статистичних даних (обчислення характеристик вибірок).**
- 2. Отримання висновків за даними спостережень (встановлення закономірностей, властивих генеральній сукупності).**



Генеральна сукупність складається з усіх об'єктів, які підлягають вивченню. Склад генеральної сукупності залежить від цілей дослідження.

Іноді генеральна сукупність — це все населення певного регіону (наприклад, коли вивчається ставлення потенційних виборців до кандидата).



Найчастіше задається кілька критеріїв, що визначають об'єкт дослідження. Наприклад, учні 10 класів, які навчаються за інформаційно-технологічним профілем.





Вибірка, або вибіркова сукупність, — множина об'єктів, випадковим чином вибраних з генеральної сукупності для участі в дослідженні.

Генеральна сукупність у разі перевірки якості виробів — це множина всіх виробів, що підлягають перевірці на відповідність стандартам. Для формування вибіркової сукупності можна випадковим чином відібрати для перевірки 5% від загальної кількості виробів.





Аналітичному відділу фірми необхідно підготувати матеріали про реакцію клієнтів фірми щодо планованої зміни цінових знижок.

Опитати в короткі терміни всіх клієнтів нереально. Потрібно побудувати вибірку з генеральної сукупності всіх клієнтів фірми, внесених в базу даних, і опитати лише клієнтів з вибіркової сукупності.

Таким чином, на основі вивчення частини (реакції тих, хто потрапить у вибірку) можна отримати достовірну уяву про ціле (ймовірну реакцію всіх споживачів).





Вибірка має бути **репрезентативною, тобто її об'єкти повинні добре відображати властивості генеральної сукупності.**

Репрезентативність вибірки — це здатність вибірки коректно відображати генеральну сукупність. Для того, щоб вибірка була репрезентативною, вона повинна генеруватися випадковим чином і мати достатньо великий обсяг.





Вибірка містить набір значень певного параметра вибраних об'єктів, наприклад, набір оцінок учнів з математики. Переглядаючи неупорядковану множину чисел, важко виявити закономірність їх варіювання.

Для вивчення закономірностей варіювання значень величини над експериментальними даними виконують:

ранжування

в порядку зростання або зменшення аналізованої кількісної ознаки

групування

об'єднання в групи однакових значень



Статистичний ряд розподілу — це впорядкований розподіл одиниць сукупності на групи за певною ознакою. Варіаційними рядами називають ряди розподілу, побудовані за кількісною ознакою.

Нехай у результаті проведення деякого експерименту було отримано вибірку

$$x_1, x_2, x_3, \dots, x_n$$

Значення x_1 називають **варіантами**. Розташувавши варіанти x_1 у порядку зростання, отримаємо **варіаційний ряд**.



Частота — це число, яке показує, скільки разів зустрічається кожна варіанта.

Якщо n_i — частоти варіант

$$n_1 (i = 1 \dots k), \text{ то } \sum_{i=1}^k n_i$$

Де n — об'єм вибірки (кількість значень у вибірковій сукупності).



Ще одною характеристикою варіаційного ряду є **відносна частота**, яка визначається, як відношення частоти випадків даного значення до загальної суми частот.

Відносна частота визначається в частках одиниці або відсотках (%). Зрозуміло, що сума відносних частот повинна дорівнювати **1 (100 %)**.





Варіаційні ряди бувають:

дискретні (перервні)

інтервальні (неперервні)

Дискретний варіаційний ряд — це такий ряд розподілу, в якому варіанта як величина кількісної ознаки може набувати тільки певних значень. Варіанти різняться між собою на одну чи кілька одиниць.

Так, кількість вироблених деталей за зміну конкретним робітником може виражатися тільки цілим числом (6, 7, 8 і т. д.).





Упорядкуємо інформацію про кількість деталей, виготовлених за зміну кожним працівником цеху:

Вироблено деталей за зміну, шт.	6	6	...	6	7	...	7	8	...	8	9	...	9	10	...	10
Частота	10			12			15			8			5			

Ряд розподілу працівників за кількістю вироблених деталей:

Вироблено деталей за зміну, шт.	6	7	8	9	10
Кількість робітників (частота)	10	12	15	8	5
Відносна частота	20 %	24 %	30 %	16 %	10 %



Об'єм вибірки n — це кількість варіант у вибірці, тобто кількість джерел інформації.

Для графічного відображення дискретного варіаційного ряду використовують **полігон частот** — ламану з вершинами у точках (z_i, n_i) ,

$$i = 1, 2, \dots, k$$

 z_i

значення i -ї варіанти;

 n_i

відповідна цій варіанті частота.



Для побудови полігона частот на осі абсцис відкладають варіанти z_j , а на осі ординат — відповідні частоти. Точки (z_j, n_j) сполучають відрізками прямих і отримують полігон частот.





Зобразимо полігон частот варіаційного ряду, заданого таблицею.

Вироблено деталей за зміну, шт.	6	7	8	9	10
Кількість робітників (частота)	10	12	15	8	5
Відносна частота	20 %	24 %	30 %	16 %	10 %

Графічне зображення варіаційних рядів за допомогою полігона допомагає отримати наочне уявлення про закономірності значень.





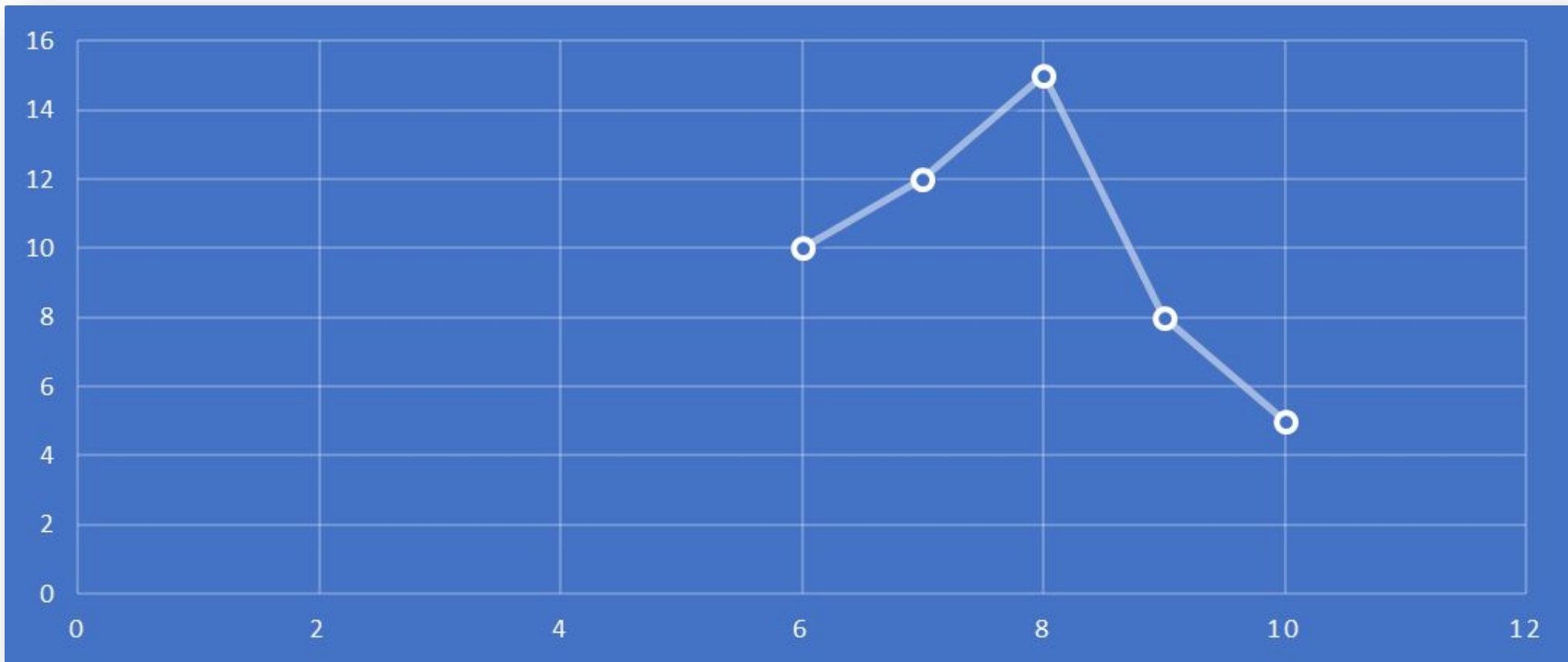
Для цього виділяємо таблицю і вибираємо команду **Вставлення** ⇒ **Діаграми** ⇒ **Точкова** ⇒ **Точкова з прямими лініями та маркерами**.

Зображення інтерфейсу Microsoft Excel, що демонструє процес створення діаграми. На зображенні видно меню "Вставлення" (Insert), групу "Діаграми" (Charts) та підменю "Точкова" (Scatter). Червоні стрілки з номерами 1, 2 та 3 вказують на ці елементи.

	A	B	C	D	E	F	G	H
1	Вироблено деталей за зміну, шт.	6	7	8	9	10		
2	Місяць виробництва (квартал)	10	12	15	8	5		



Частота розподілу працівників за кількістю вироблених деталей





При великому обсязі вибірки доречно побудувати **інтервальний варіаційний ряд** — такий ряд розподілу, в якому значення варіанти подано у вигляді інтервалів, тобто значення ознак, віднесених до однієї варіанти, можуть відрізнятися одне від одного на деяку малу величину.

Інтервали можуть бути:

рівні

і

нерівні

За частоту, що відповідає інтервалу, приймають суму частот, які потрапили в цей інтервал.



Алгоритм побудови інтервального варіаційного ряду

- 1 Виходячи з обсягу вибірки (N), визначте кількість інтервалів (K). Зазвичай кількість інтервалів — у межах 5-20.
- 2 Визначте найбільше (X_{\max}) і найменше (X_{\min}) значення у вибірці.
- 3 Обчисліть розмах ряду: $R = X_{\max} - X_{\min}$
- 4 Визначте ширину інтервалу: $H = R / (K - 1)$.
- 5 Визначте межі кожного інтервалу в порядку зростання.
- 6 Підрахуйте кількість даних, що потрапили в кожний з інтервалів.



Планується випускати шкільну форму. Потрібно з'ясувати, які розміри форми і в якій кількості слід закласти у план. Вимірювання зросту групи школярів дали такі результати (в см):

132, 132, 133, 134, 101, 134, 135, 105, 109, 138, 138, 110, 111, 140, 115, 125, 127, 115, 116, 127, 127, 116, 117, 127, 127, 117, 128, 117, 118, 130, 119, 131, 143, 124, 124, 144, 146, 124, 125, 150, 124, 158, 125, 121, 122, 121.

Розіб'ємо множину значень на 7 інтервалів:

$[101, 109)$, $[109, 117)$, $[117, 125)$, $[125, 133)$, $[133, 141)$, $[141, 149)$, $[149, 158]$.



Для кожного інтервалу визначимо частоту, тобто число значень, що потрапили до інтервалу. Обчислимо відносні частоти в інтервалах шляхом ділення частоти в інтервалі на суму всіх частот, яка в даному випадку дорівнює 46.

Відносна частота в інтервалі дає відсоток потрапляння до інтервалу даних від їхньої загальної кількості.





Заповнимо таблицю частот інтервального варіаційного ряду:

Інтервал	[101, 109)	[109, 117)	[117, 125)	[125, 133)	[133, 141)	[141, 149)	[149, 158]
Частота	2	7	12	13	7	3	2
Відносна частота	0,0435	0,152	0,261	0,283	0,152	0,065	0,0435

Відносна частота кожної варіанти показує, яку частку кожного розміру слід закласти в план.



Обчислення статистичних характеристик здійснюється за достатньо складними формулами.

В наступному параграфі ви дізнаєтесь, як обчислення характеристик вибірок реалізоване у відповідних функціях **MS Excel.**





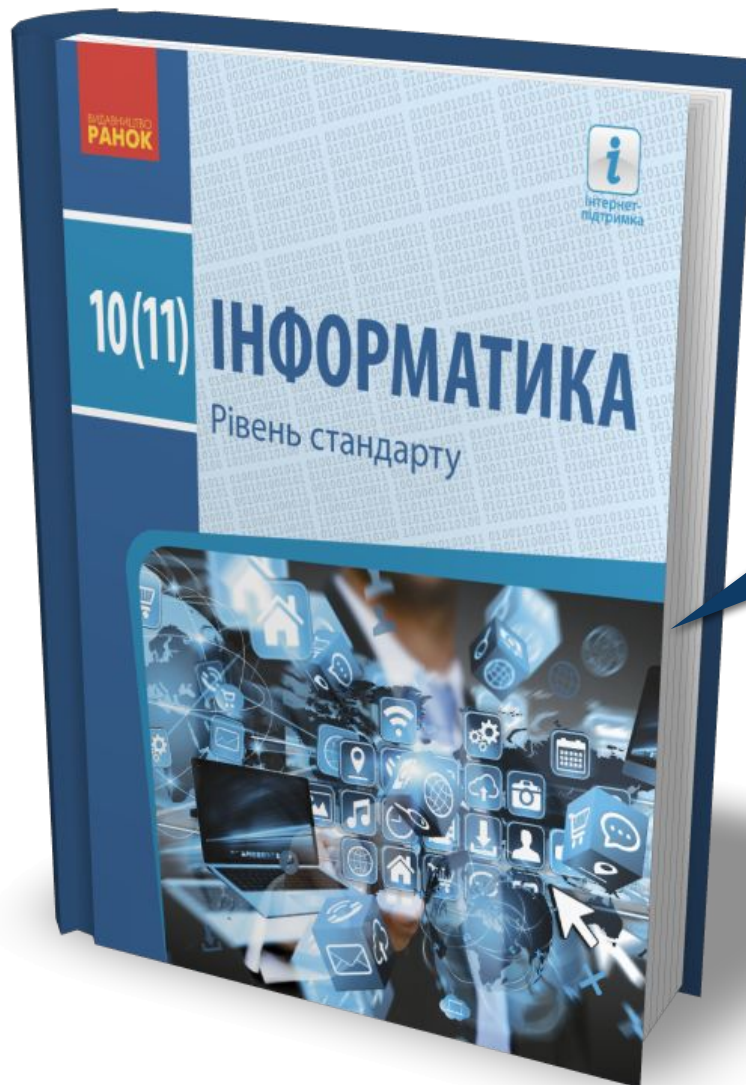
Статистика





- 1. Якою може бути вибіркова сукупність для опитування аудиторії телевізійного каналу серед мешканців міста, в якому проживає 100 тис. чоловік, 60 % з яких чоловіки і 40 % — жінки?**
- 2. Чи залежить репрезентативність вибірки від її розміру?**
- 3. Для яких вибірок будують дискретний варіаційний ряд?**
- 4. Поясніть відмінності між поняттями варіанта і частота.**

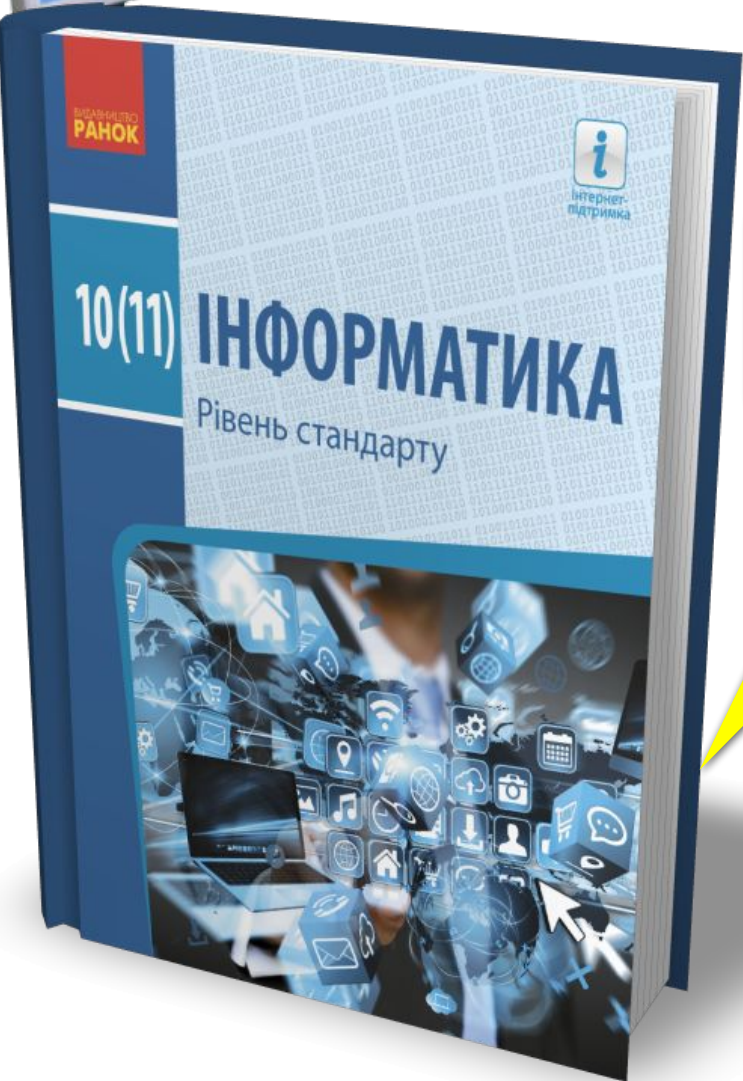




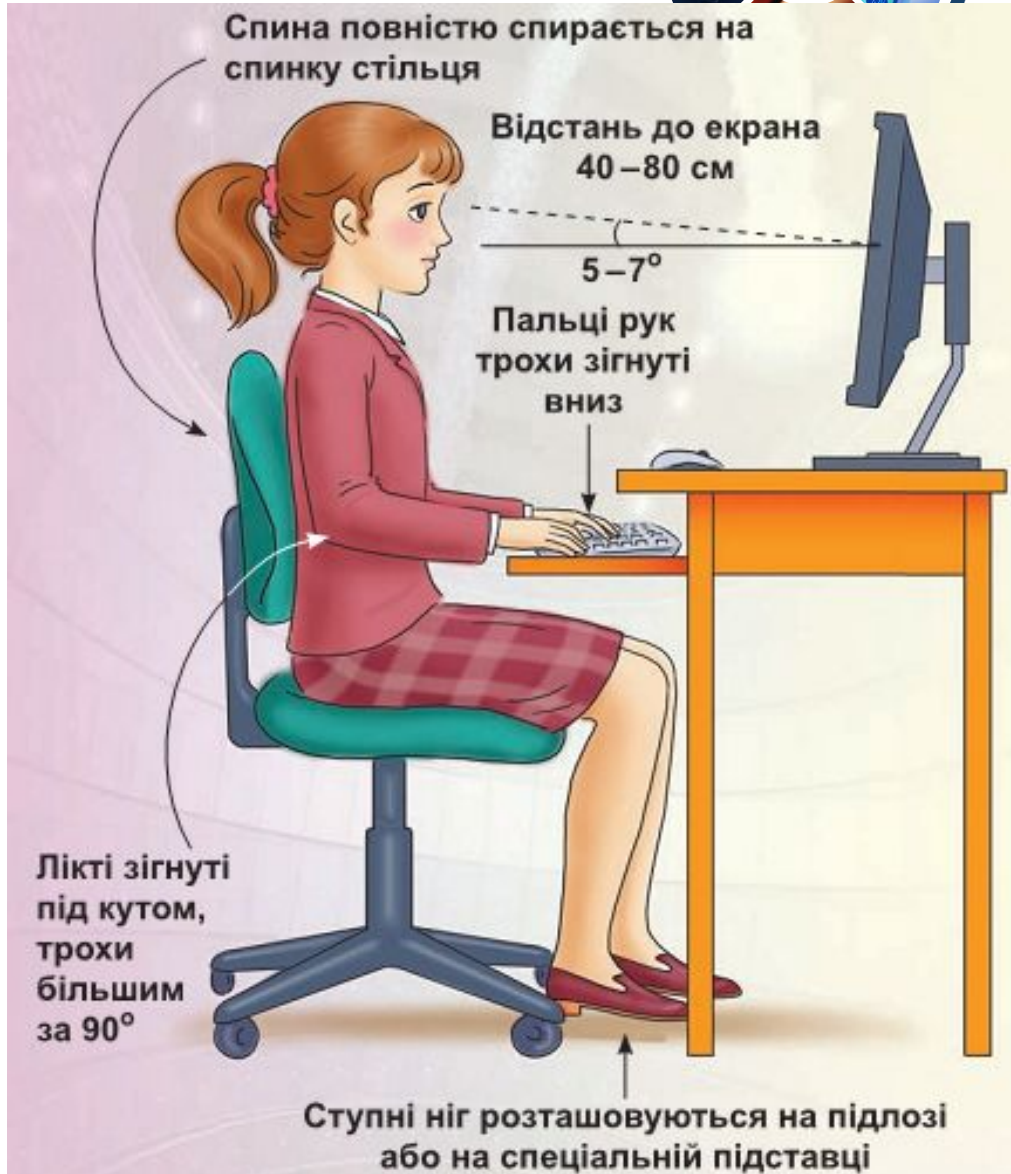
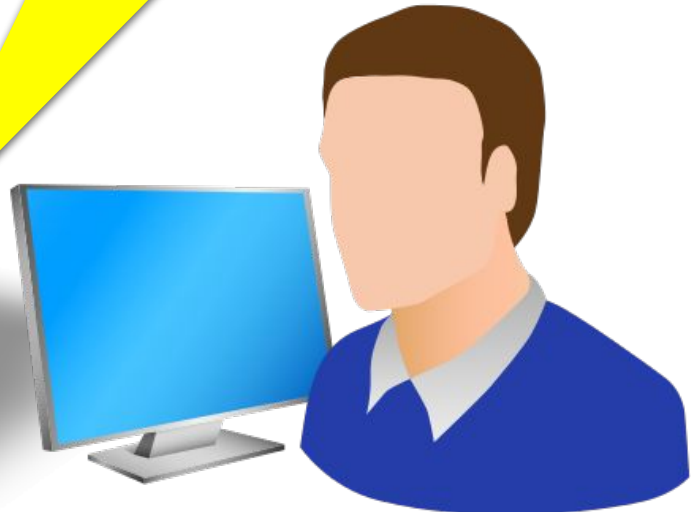
Проаналізувати
§ 11, ст. 61-67

Працюємо за комп'ютером

Розділ 2
§ 11



**Сторінка
66-67**



10(11) ІНФОРМАТИКА

Рівень стандарту

10

Дякую за увагу!

За навчальною програмою 2018 року



Урок 11

teach-inf.at.ua