

Математичні методи прогнозування рівня попиту та закупівель у компанії

Проф., д.е.н. Ставицький А.В.

Таймінг

- ▶ 09.30 – 11.45 – перша частина учбового дня
- ▶ 11.45– 12.00 - перерва на каву
- ▶ 12.00 – 13.30 - друга частина учбового дня
- ▶ 13.30 – 14.30 - обід
- ▶ 14.30 – 16.00 - третя частина учбового дня
- ▶ 16.00 – 16.15 - перерва на каву
- ▶ 16.15 – 17.30 - четверта частина учбового дня
- ▶ 17.30 – 18.00 – підсумки дня, індивідуальні питання.

Задайте питання!

- ▶ Поставте запитання, на яке Ви хочете отримати відповідь під час тренінгу, запишіть його у блокнот.



Хто знає,
той не прогнозує,
а прогнозує той,
хто не знає

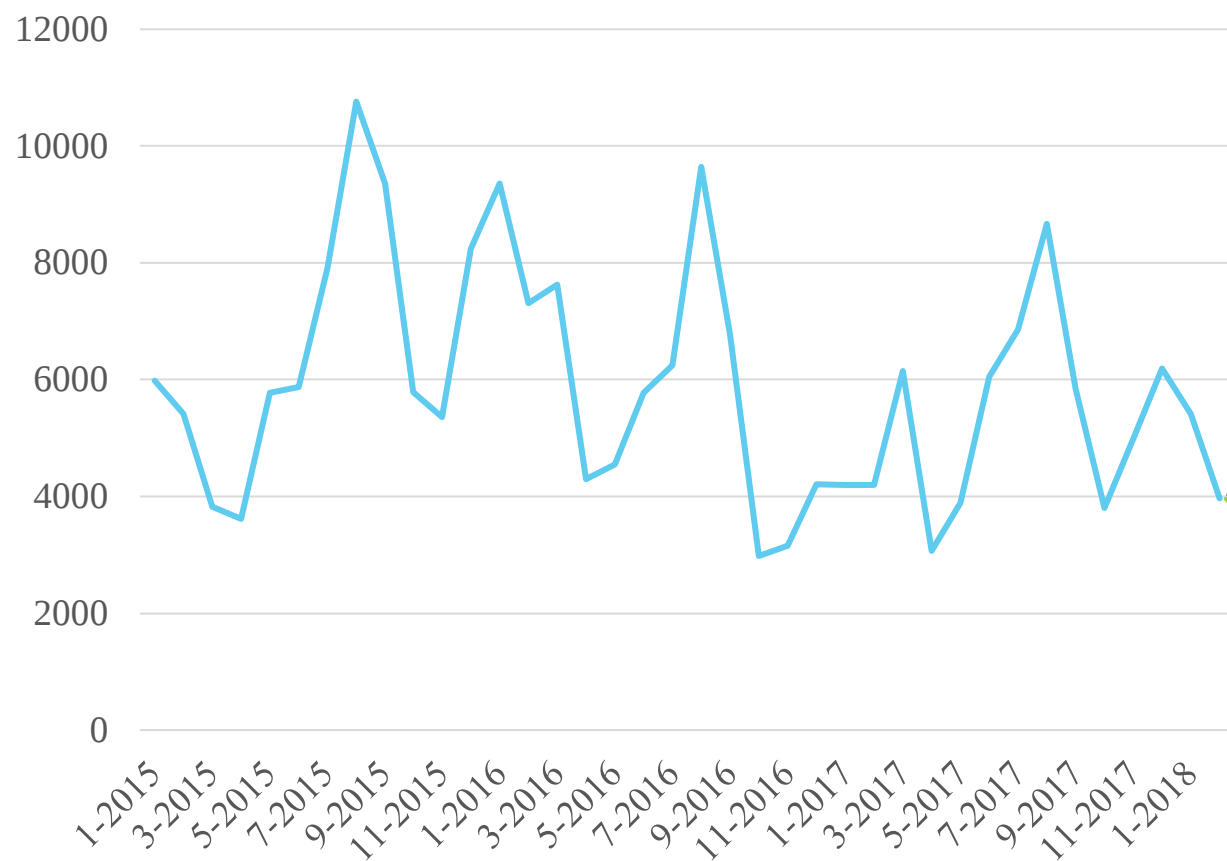
Лві Лао Тсу

Типова ситуація

	A	B
1	Дата	Сумма по полю Количество оборот
14	1-2016	9354
15	2-2016	7306
16	3-2016	7629
17	4-2016	4294
18	5-2016	4547
19	6-2016	5775
20	7-2016	6251
21	8-2016	9639
22	9-2016	6772
23	10-2016	2985
24	11-2016	3153
25	12-2016	4208
26	1-2017	4197
27	2-2017	4195
28	3-2017	6148
29	4-2017	3069
30	5-2017	3889
31	6-2017	6052
32	7-2017	6858
33	8-2017	8668
34	9-2017	5842
35	10-2017	3800
36	11-2017	4987
37	12-2017	6193
38	1-2018	5411
39	2-2018	3962
40		
41		
42		

	A	B	C	D	E
1	Дата календаря	Количество оборот	Сумма продаж, грн.	Количество остаток	
1134	06.02.2018	125	26521	4339	
1135	07.02.2018	123	21723	3753	
1136	08.02.2018	69	18543	3739	
1137	09.02.2018	142	44979	3711	
1138	10.02.2018	215	55852	3669	
1139	11.02.2018	240	60776	3622	
1140	12.02.2018	47	11360	3613	
1141	13.02.2018	80	22286	3597	
1142	14.02.2018	140	35397	3569	
1143	15.02.2018	159	41159	3538	
1144	16.02.2018	145	39619	3509	
1145	17.02.2018	258	69705	3459	
1146	18.02.2018	216	65905	3416	
1147	19.02.2018	114	26732	4089	
1148	20.02.2018	114	34864	4066	
1149	21.02.2018	61	15827	4055	
1150	22.02.2018	145	46016	4026	
1151	23.02.2018	116	32708	4003	
1152	24.02.2018	237	73640	3957	
1153	25.02.2018	173	52281	3922	
1154	26.02.2018	64	18052	3910	
1155	27.02.2018	57	17924	3898	
1156	28.02.2018	100	29368	3875	
1157	01.03.2018				
1158	02.03.2018				
1159	03.03.2018				
1160	04.03.2018				
1161	05.03.2018				
1162	06.03.2018				
1163	07.03.2018				

Або так...



Навіщо?

- ▶ Обґрунтована побудова прогнозів
- ▶ Самостійний вибір моделей для прогнозування
- ▶ Автоматизація розрахунку прогнозів на фірмі

Часовий ряд

Нехай $Y_1, Y_2, \dots, Y_T$ - значення спостережень за економічним процесом протягом T періодів.

Ця послідовність є числовими значеннями, кожне з яких має відповідний індекс, який залежить від номера періоду, коли він спостерігався.

Така послідовність, записана порядку зростання індексу, називається часовим рядом.

Ціль аналізу

Метою прикладного аналізу є побудова математичної моделі ряду, за допомогою якої можна пояснити поведінку ряду та здійснити прогноз на майбутні періоди.

$$\hat{y}_{T+1} - ?$$

Структура часового ряду

- ▶ Будь-який часовий ряд можна представити як суму детермінованого та випадкового компонентів:

$$y_t = d_t + r_t, t = \overline{1, T}.$$

- ▶ У свою чергу, детермінований компонент складається з трьох частин: трендового, сезонного, циклічного компонентів.

$$d_t = tr_t + s_t + c_t, t = \overline{1, T}.$$

Компонети часового ряду - 1

- ▶ Аналіз часового ряду починається із виділення трендового компонента. Його присутність неважко помітити, проаналізувавши графік часового ряду. Як правило, для економічних даних дуже типовим є повільне зростання або падіння протягом тривалого часу.
- ▶ Наявність тренду в економічних часових рядах можна пояснити демографічними змінами, технологічними змінами, змінами у структурі виробництва, попиту тощо. Дія таких факторів є постійною, тому дослідники можуть описувати такі зміни за допомогою кривих, які можна представити в аналітичному вигляді.

Компонети часового ряду - 2

- ▶ **Сезонний компонент** показує коливання довкола трендового компонента. Його наявність пояснюється сезонним характером виробництва, споживання. Наприклад, у четвертому кварталі кожного року перед Новим роком значно зростає споживання товарів.
- ▶ Головна ідея виділення сезонних коливань полягає у порівнянні даних за відповідні періоди, а не за минулі періоди, тобто, наприклад, дані за грудень одного року треба порівнювати із даними грудня минулого року, а не з листопадом.

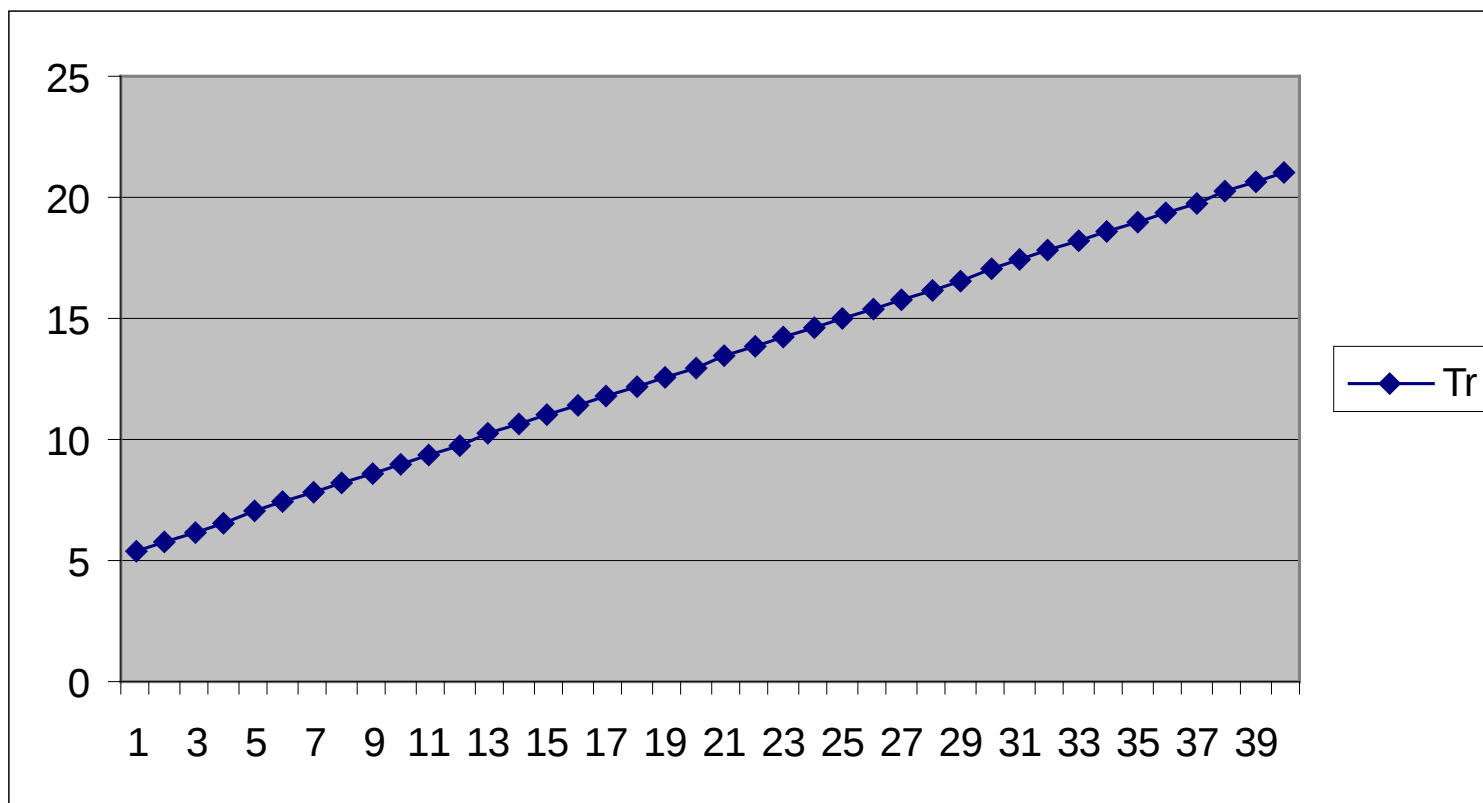
Компонети часового ряду - 3

- ▶ **Циклічний компонент** займає проміжне місце між трендом та сезонним компонентом. Тренд – це гладка зміна, яка проявляється на великому проміжку часу. Сезонний компонент - це періодична функція, що залежить від часу, причому період його значно менше від кількості спостережень.
- ▶ Циклічний компонент розглядається, в основному, як гладка зміна, що залежить від часу, але яка не включається ні до тренду, ні до сезонного компонента.

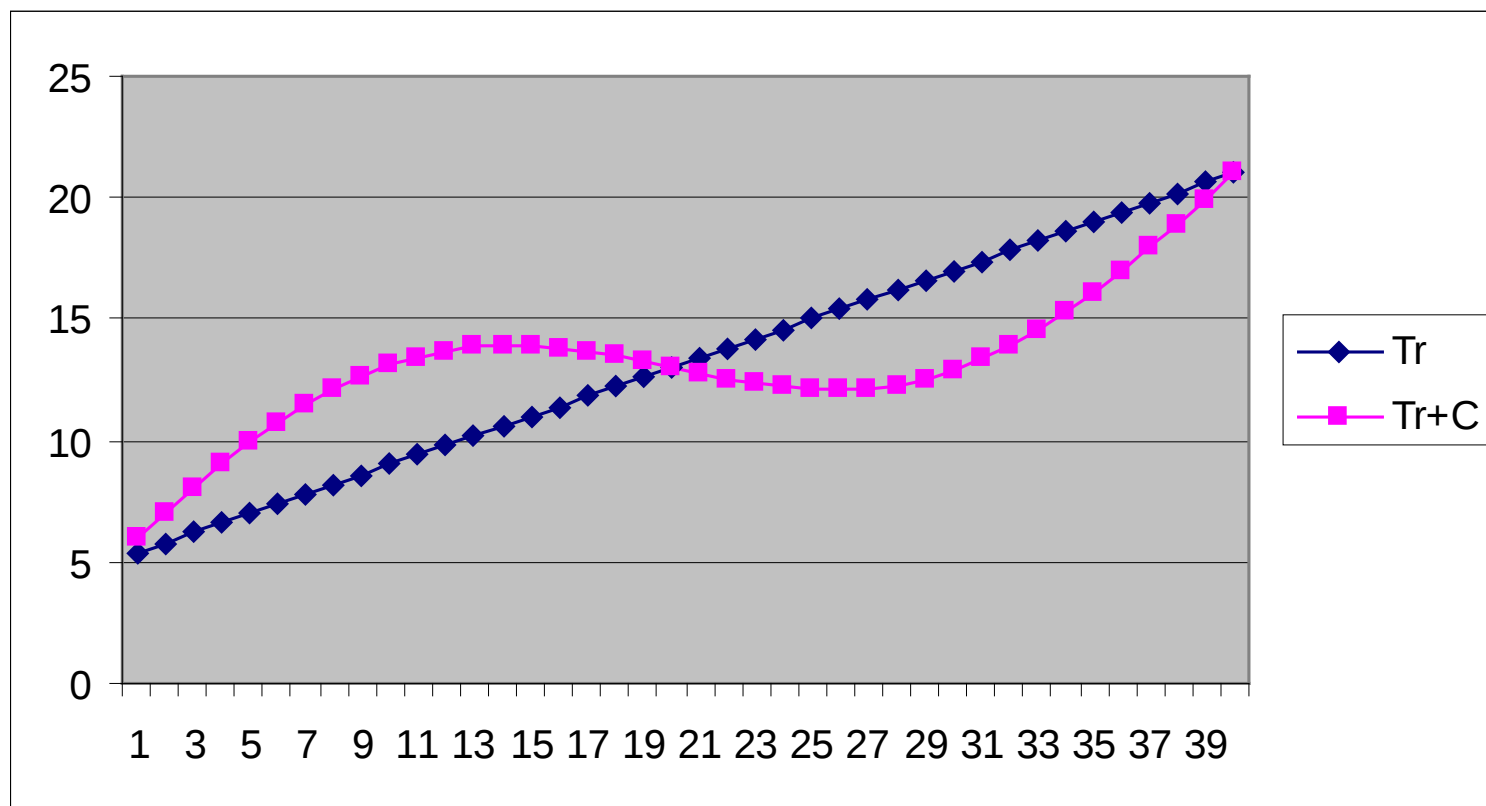
Компонети часового ряду - 4

- ▶ **Випадковий компонент** є тим, що залишилося від часового ряду після виключення тренду, циклічного та сезонного компонентів. Частина таких ефектів може бути віднесена до непередбачених природних катаклізмів (землетруси, пожежі тощо), частина – до випадкових дій людей.
- ▶ За наявності випадкових компонентів неможливо прогнозувати значення часового ряду без помилки. Але будь-який реальний економічний процес включає випадковий компонент.

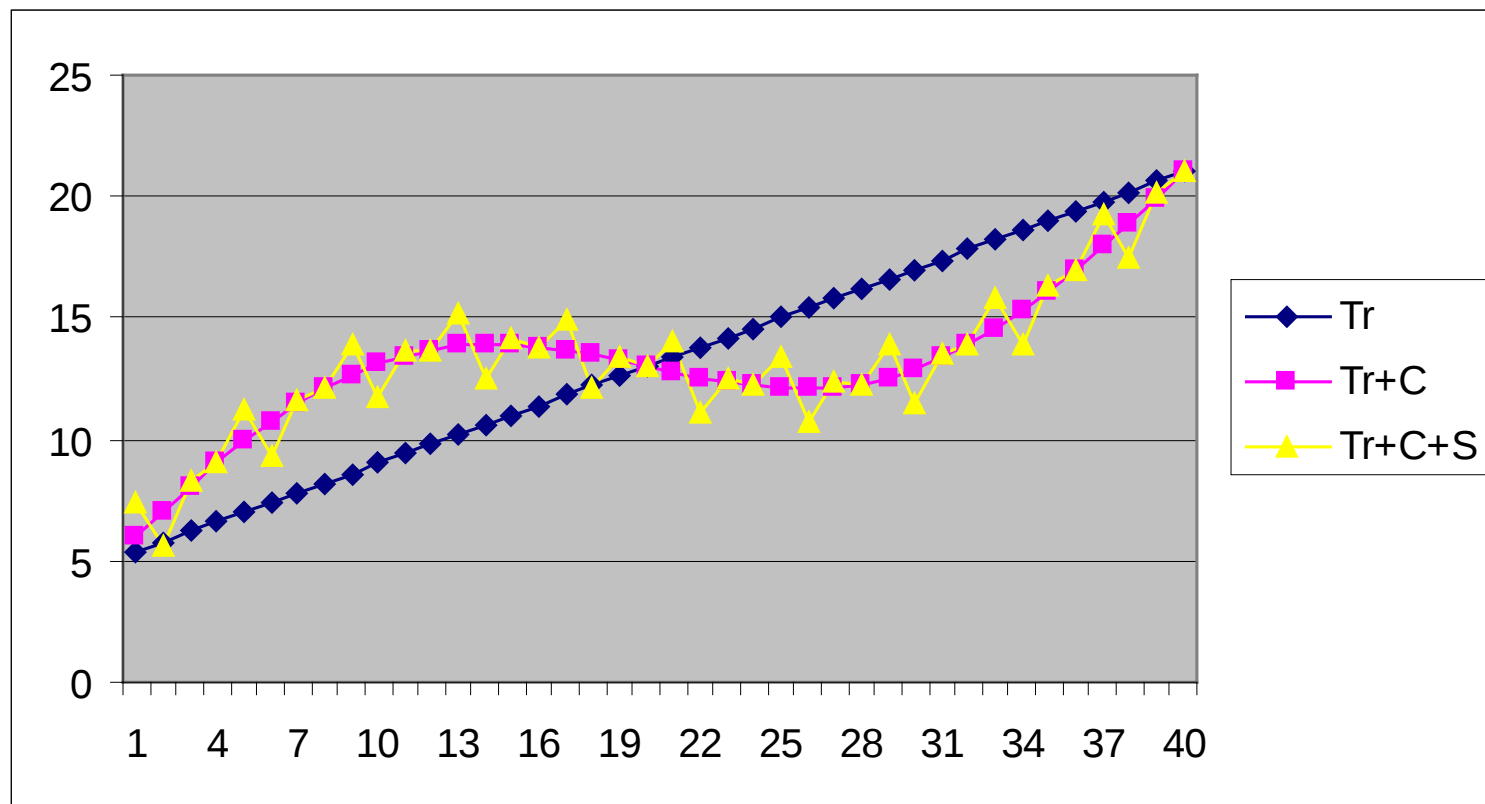
Приклад



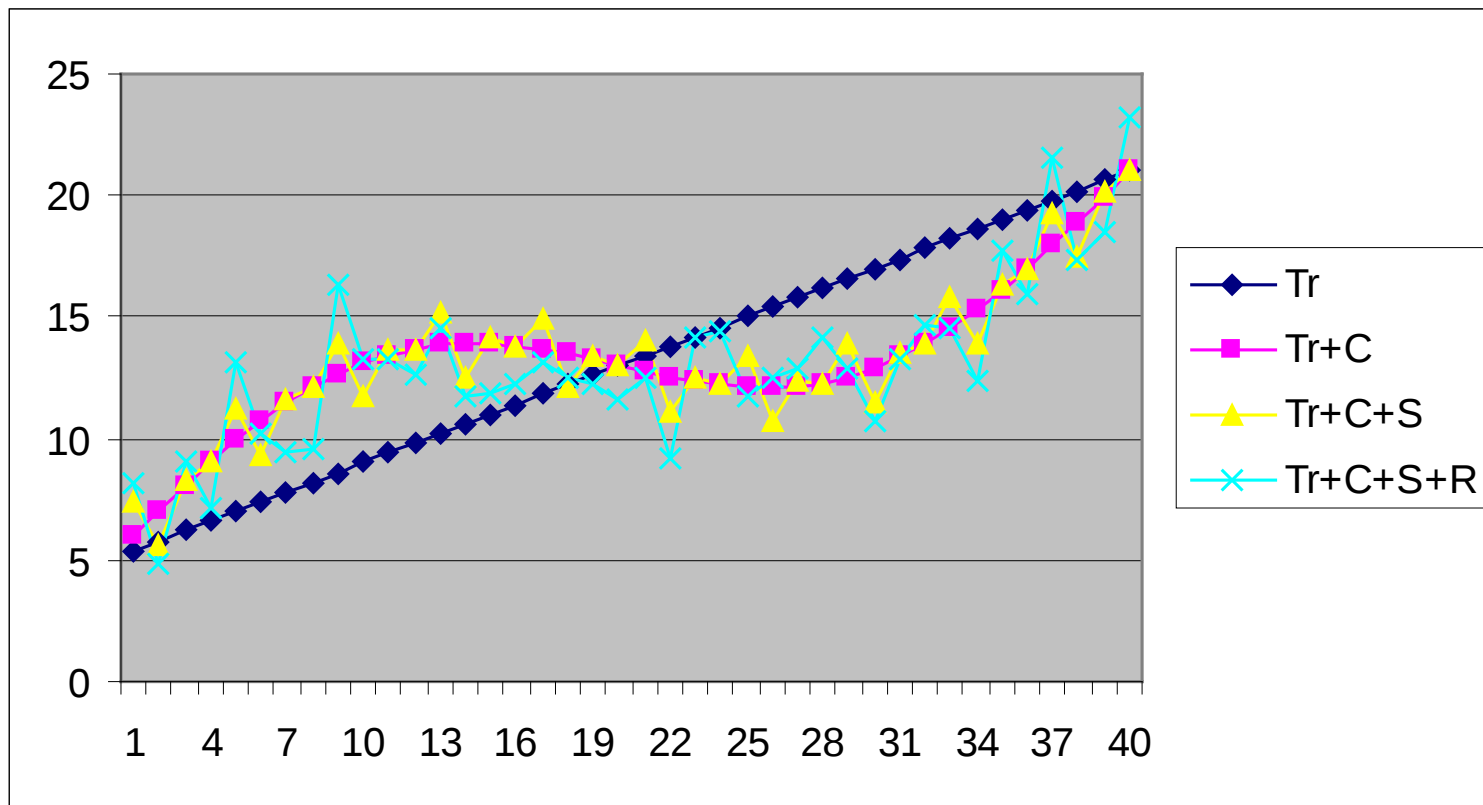
Приклад



Приклад



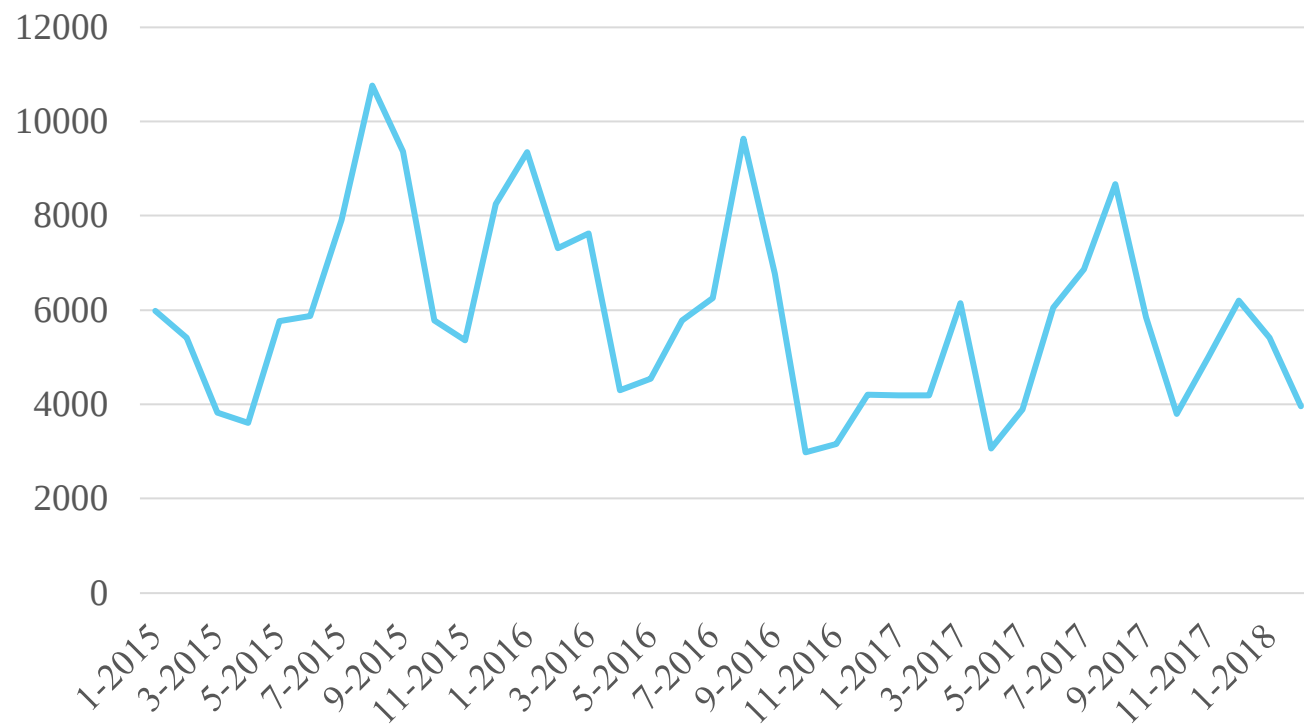
Приклад



Порядок аналізу

1. Побудова та вивчення графіка ряду
2. Вибір моделі
3. Прогнозування

Побудова та вивчення графіка



На графіку визначають

- ▶ наявність тренду та його характер;
- ▶ наявність сезонних та циклічних компонентів;
- ▶ ступінь повільності чи перервності змін послідовних значень ряду після усунення тренду.

Дані для тренінгу

- ▶ http://www.andriystav.cc.ua/T_Train.html

Методи прогнозування обсягу продажу

- ▶ методи експертних оцінок;
- ▶ методи аналізу та прогнозування часових рядів;
- ▶ каузальні (причинно-наслідкові) методи (спроба визначити фактори, що визначають поведінку прогнозованого показника).

Експертні методи

Їх багато, але вони вимагають:

- ▶ Експертів
- ▶ Грошей
- ▶ Часу
- ▶ Знань у конкретній області

Методи прогнозування часових рядів

- ▶ Метод ковзкого середнього
- ▶ Метод експоненціального згладжування
- ▶ Метод подвійного експоненціального згладжування
- ▶ Метод потрійного експоненціального згладжування
- ▶ Метод Холта-Вінтерса
- ▶ Метод виділення сезонності

Причинно-наслідкові методи

- ▶ Трендові моделі
- ▶ Регресійний аналіз
- ▶ Моделі з фіктивними змінними

Коли використати?

Ситуація	Метод
Нічого не відомо, немає даних	Експертний метод
Є ряд даних продажів, немає тренду	Експоненціальне згладжування Ковзке середнє
Є ряд даних продажів, є лінійний тренд	Подвійне експоненціальне згладжування Метод Холта-Вінтерса
Є ряд даних продажів, є квадратичний тренд	Потрійне експоненціальне згладжування
Є ряд даних продажів, є сезонність	Метод виділення сезонності
Є ряди даних продажів та інших факторів	Регресійний аналіз

Точність прогнозів

- ▶ Про точність прогнозу прийнято судити за розміром помилки прогнозу - різниці між прогнозним та фактичним значенням змінної, що досліджується. Однак такий підхід до оцінки точності можливий лише у двох випадках.
- ▶ По-перше, коли період прогнозування вже закінчився, і дослідник знає фактичні значення змінної. За короткострокового прогнозування це цілком реально.
- ▶ По-друге, коли прогноз розробляється, тобто прогнозування здійснюється на деякий час у минулому, вже є фактичні дані.

Важливо!

- ▶ Перевірка точності одного прогнозу мало може сказати досліднику. Хороший одиничний прогноз то, можливо отримано і з неадекватної моделі, і навпаки, тому про якість прогнозів застосовуваних методик і моделей можна судити лише з сукупності зіставлень прогнозів та його реалізації.

Критерії точності прогнозів

- ▶ Найбільш простим критерієм якості прогнозів за умови, що є дані про їх реалізацію, може стати відносна кількість випадків, коли фактична реалізація потрапляла в довірчий інтервал прогнозу, до загального числа прогнозів, тобто

$$\eta = \frac{m}{m + n}$$

де

- ▶ m – кількість прогнозів, підтверджених фактичними даними;
- ▶ n – кількість прогнозів, не підтверджених фактичними даними.

Абсолютні критерії точності прогнозів

$$MSE = \frac{1}{n} \sum_t (y_t - \hat{y}_t)^2$$

Середньоквадратична похибка прогнозу за n періодів.

$$RMSE = \sqrt{\frac{1}{n} \sum_t (y_t - \hat{y}_t)^2}$$

Корінь із середньоквадратичної похибки прогнозу за n періодів.

$$MAD = \frac{1}{n} \sum_t |y_t - \hat{y}_t|$$

Середня абсолютна похибка за n періодів.

Відносні критерії точності прогнозів

$$RMSPE = 100 \sqrt{\frac{1}{n} \sum_t \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2}$$

корінь із
середньоквадратичної
похибки у відсотках від
фактичних значень за n
періодів.

$$MAPE = \frac{100}{n} \sum_t \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

середня абсолютна
похибка у відсотках за n
періодів.

Оцінка точності мікроекономічного прогнозу

MAPE, RMSPE	Точність прогнозу
менше 5%	Висока
5% - 10%	Гарна
10% - 15%	Задовільна
15% - 20%	Погана
більше 20%	Незадовільна

Ефективність прогнозування

- ▶ кількість зусиль, що витрачаються на побудову моделі та наявність готових машинних програм;
- ▶ існування серійної кореляції у помилках;
- ▶ незмінність первинних даних;
- ▶ повний обсяг роботи у деяких сферах діяльності - тисячі рядів щомісяця вимагають оновлення, невеликі витрати та швидкість мають першорядне значення;
- ▶ терміновість прогнозування.

Поліпшення прогнозування за допомогою правильної підготовки даних

- ▶ Піки в даних
- ▶ Кризові значення
- ▶ Інфляція

Методи згладжування

Метод ковзкого середнього

- ▶ Цей метод є одним із найпростіших, який дозволяє виділити тренд. Для цього методу дослідник повинен мати довгий ряд спостережень. Формально метод описується виразом:

$$\tilde{y}_t = \frac{1}{k} \sum_{j=-k_1}^{k_2} y_{t+j}, \quad k = k_1 + k_2 + 1.$$

- ▶ Для квартальних даних часового ряду формула набуває вигляду

$$\tilde{y}_t = \frac{1}{4} \sum_{j=-3}^0 y_{t+j} = \frac{1}{4} (y_{t-3} + y_{t-2} + y_{t-1} + y_t).$$

Звичайне експоненціальне згладжування

- ▶ Найкраще цей метод зарекомендував себе, коли дані мають дуже гладкий або навіть горизонтальний тренд. Новий ряд будується за правилом:

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad 0 < \alpha < 1.$$

- ▶ Вибір початкового значення :

$$S_1 = y_1$$

Вибір константи

- ▶ Єдиний параметр α може вибиратися кількома шляхами.
- ▶ По-перше, якщо вибирається значення близько до 1, то будуть більш важливими при прогнозуванні останні дані часового ряду, при виборі близьким до 0 більш впливовими будуть попередні значення.
- ▶ По-друге, можна покласти

$$\alpha = \frac{2}{T + 1}.$$

Прогноз

- ▶ Прогноз значень часового ряду дорівнює останньому члену послідовності

$$\hat{y}_{T+p} = S_T, \quad p = 1, 2, \dots$$

Подвійне експоненціальне згладжування Брауна

- ▶ Цей метод будується аналогічно до попереднього, тільки процес згладжування робиться двічі:

$$S'_t = \alpha y_t + (1 - \alpha)S'_{t-1},$$

$$S''_t = \alpha S'_t + (1 - \alpha)S''_{t-1}, \quad 0 < \alpha < 1.$$

- ▶ Метод використовується, коли дані часового ряду мають тренд

Прогноз

- ▶ Прогноз будується як останнє значення другої послідовності:

$$\hat{y}_{T+p} = S_T'', \quad p = 1, 2, \dots$$

Потрійне експоненціальне згладжування Брауна

- Цей метод аналогічний двом попереднім, лише згладжування проводиться тричі. Це дозволяє прогнозувати нестаціонарні часові ряди з великими перепадами мінімального та максимального значень. Нові послідовності будуються за правилом:

$$S_t' = \alpha y_t + (1 - \alpha)S_{t-1}',$$

$$S_t'' = \alpha S_t' + (1 - \alpha)S_{t-1}'',$$

$$S_t''' = \alpha S_t'' + (1 - \alpha)S_{t-1}''', \quad 0 < \alpha < 1.$$

Прогноз

- ▶ Прогноз на наступні періоди має вигляд:

$$\hat{y}_{T+p} = S_T''', \quad p = 1, 2, \dots$$

Несезонна модель Холта-Вінтерса

- ▶ Ця модель схожа на подвійне експоненціальне згладжування, але дозволяє виділяти трендовий компонент за допомогою другої послідовності:

$$S_2' = y_2, \quad S_2'' = y_2 - y_1,$$

$$S_t' = \alpha y_t + (1 - \alpha)(S_{t-1}' + S_{t-1}''), \quad 0 < \alpha < 1,$$

$$S_t'' = \beta(S_t' - S_{t-1}') + (1 - \beta)S_{t-1}'', \quad 0 < \beta < 1.$$

Прогноз

- ▶ Прогноз на наступні періоди :

$$\hat{y}_{T+p} = S'_T + pS''_T, \quad p = 1, 2, \dots$$

- ▶ Слід зазначити, що прогнози, зроблені за цим методом, зазвичай або сильно завищені, або занижені внаслідок того, що додається постійний трендовий компонент, який на практиці змінюється протягом року.

Завдання

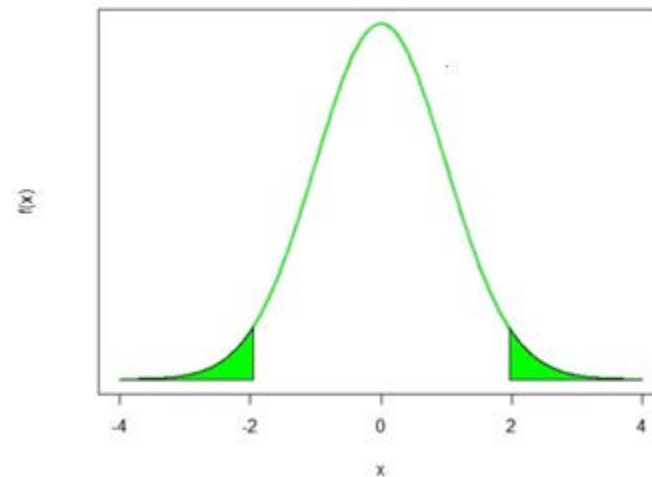
- ▶ Використати методи прогнозування:
 - ▶ Ковзкого середнього
 - ▶ Експоненціального згладжування
 - ▶ Подвійного експоненціального згладжування
 - ▶ Потрійного експоненціального згладжування
 - ▶ Холта-Вінтерса
- ▶ Оптимізувати параметри методів для вибірки 31-39.
- ▶ Зробити прогноз з усіх методів на дні 47-50 за даними 1-46. Розрахувати RMSPE.
- ▶ Вибрати найкращий спосіб прогнозування для даних.

Довірчі інтервали

- ▶ Довірчий інтервал - це допустиме відхилення значень від істинних. Розмір цього припущення визначається дослідником з урахуванням вимог щодо точності інформації.

Інтервальна оцінка

- ▶ Як правило, довірчий рівень позначають таким чином:
 $(1 - \alpha) \times 100\%$, де величина α являє собою площу, обмежену хвостом розподілу, що виходить за межі довірчого інтервалу.
- ▶ Величину α називають рівнем значущості довірчого інтервалу.
- ▶ Площі, обмежені як лівим, і правим хвостами розподілу, які виходять межі довірчого інтервалу, рівні $\alpha / 2$.



Довірчий інтервал

$$\hat{y}_{T+n} - t_{\alpha} \sigma_p \leq y_{T+n} \leq \hat{y}_{T+n} + t_{\alpha} \sigma_p$$

- ▶ T – довжина часового ряду;
- ▶ n -період попередження;
- ▶ $\hat{y}_{T+n}$ -точковий прогноз на момент $T + n$;
- ▶ t_{α} – значення t-статистики Стьюдента;
- ▶ σ_p - середня квадратична помилка прогнозу.

Сезонність у часових рядах

- ▶ Отримання коефіцієнта сезонності: знаходження середнього відношення для кожного кварталу (місяця, тижня, дня) до середньорічного (середньоквартальне, середньомісячне, середньотижнєве) значення.

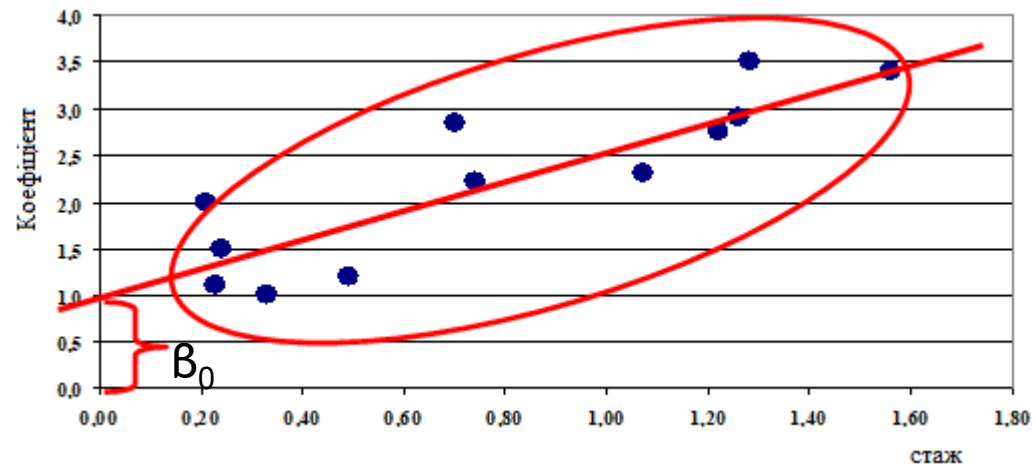
Регресійний аналіз

- ▶ Дає нам правила, що визначають лінію регресії, яка краще за інших передбачає одну змінну на підставі іншої.
- ▶ По осі Y відкладають змінну, яку ми хочемо передбачити, а по осі X – змінну, на основі якої будемо прогнозувати.
- ▶ Передбачене значення Y зазвичай позначають як $\hat{Y}$

Проста лінійна регресія

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

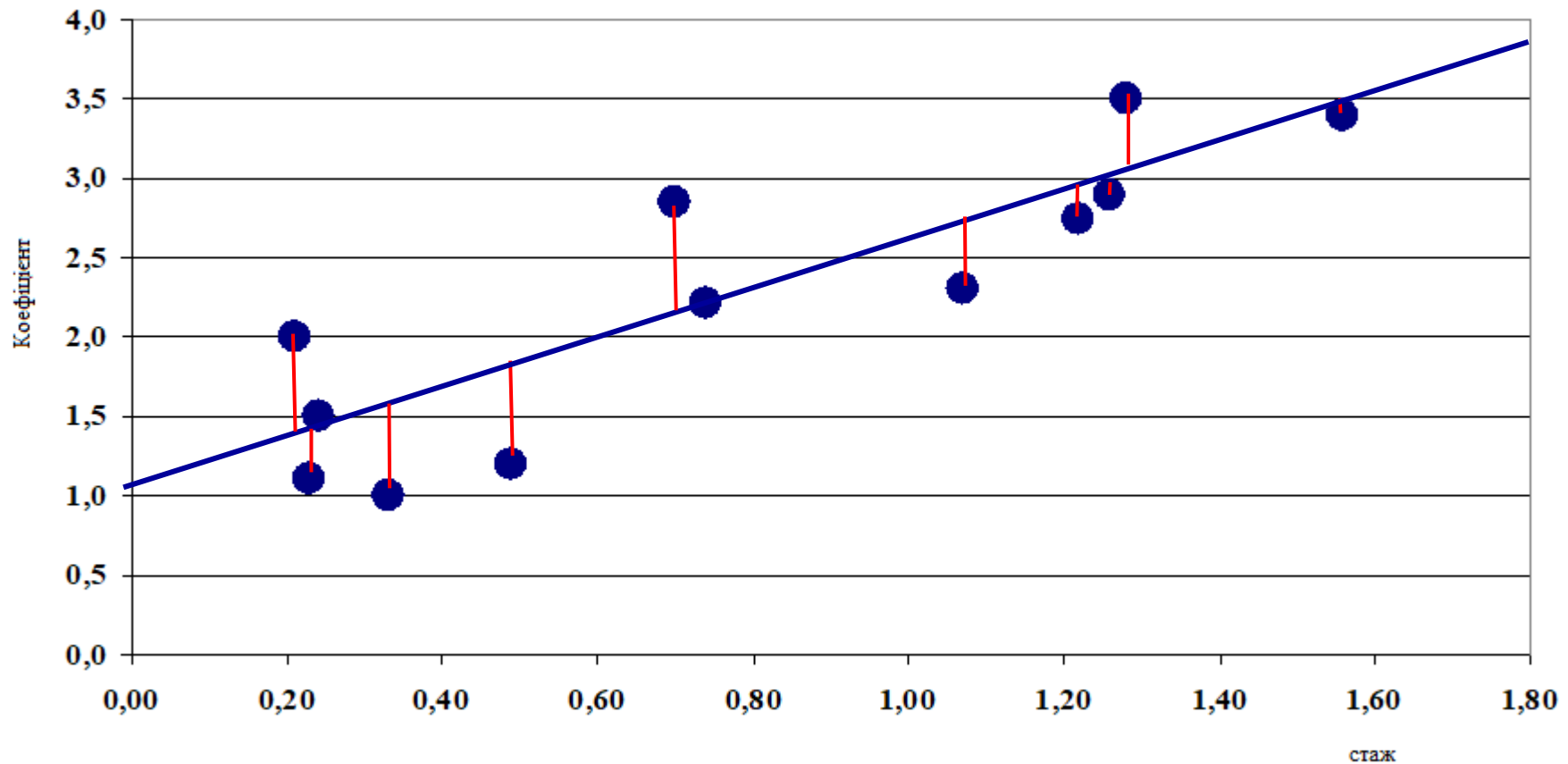
- ▶ Y – залежна змінна
- ▶ X – незалежна змінна
- ▶ β_0 та β_1 - коефіцієнти регресії
- ▶ β_1 – характеризує НАКЛОН прямої;
- ▶ β_0 – визначає точку перетину прямої з віссю OY .



«Найкраща» лінія регресії

$$e_t = Y_t - \hat{Y}_t$$

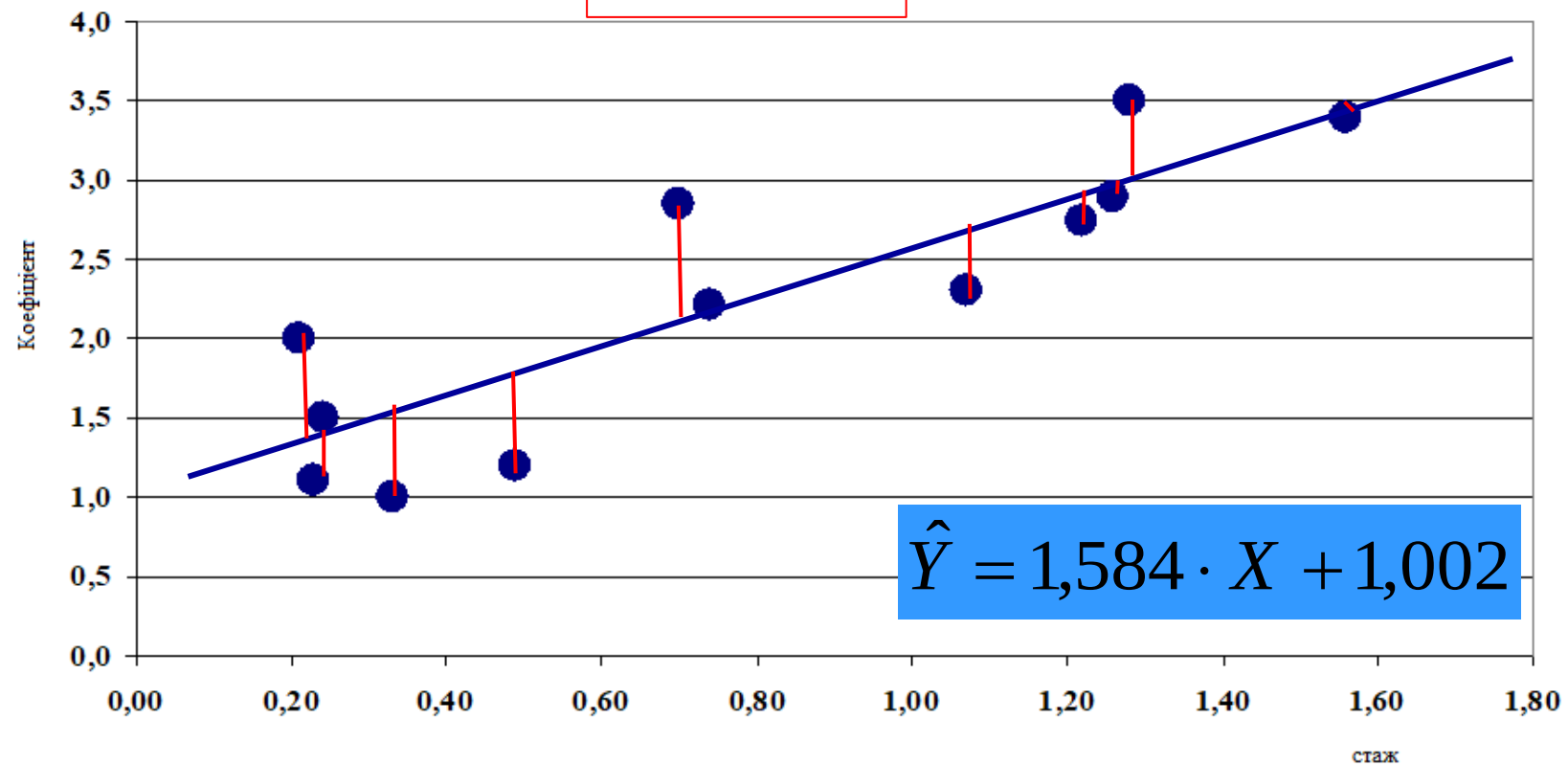
$$\sum e_t = 0$$



Метод найменших квадратів

- ▶ лінію регресії підбирають таку, щоб загальна сума квадратів помилок (residuals) була найменшою.

$$\sum e_t^2 \rightarrow \min$$



Лінійна регресія у MS Excel

- ▶ **Спосіб 1.** Лінія тренду на графіку
- ▶ Побудувати точковий графік за наявними даними виду (y_i, x_i)
- ▶ Клацнути правою кнопкою миші на серії точок та вибрати «додати лінію тренду»
- ▶ Наочний, але не проводяться F-тест та оцінка довірчих інтервалів коефіцієнтів регресії
- ▶ **Спосіб 2.** Використання пакета аналізу даних
- ▶ Вибрати вкладку дані, натиснути на пункт меню «аналіз даних»
- ▶ Із запропонованих опцій вибрати регресію
- ▶ Вказати вхідні дані та вивчити результат
- ▶ Наочний, містить F-тест та оцінку довірчих інтервалів коефіцієнтів регресії

Опис регресії

ВЫВОД ИТОГОВ									
<i>Регрессионная статистика</i>									
Множественный R	0,996								
R-квадрат	0,992								
Нормированный R-квадрат	0,991								
Стандартная ошибка	121,674								
Наблюдения	25								
<i>Дисперсионный анализ</i>									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>				
Регрессия	1	39785221,53	39785221,53	2687,367	2,5E-25				
Остаток	23	340504,3072	14804,5351						
Итого	24	40125725,84							
<i>Коэффициенты</i>									
	<i>Коэффициент</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	-141,41	67,96105028	-2,080753151	0,0488	-281,998	-0,82203	-281,998	-0,82203	
Средний доход людей, грн.	0,77	0,014880843	51,83982221	0,0000	0,74064	0,8022	0,74064	0,8022	

Тренд та його види

Види тренду

Лінійний

$$f(t) = a_0 + a_1 t$$

Квадратичний

$$f(t) = a_0 + a_1 t + a_2 t^2$$

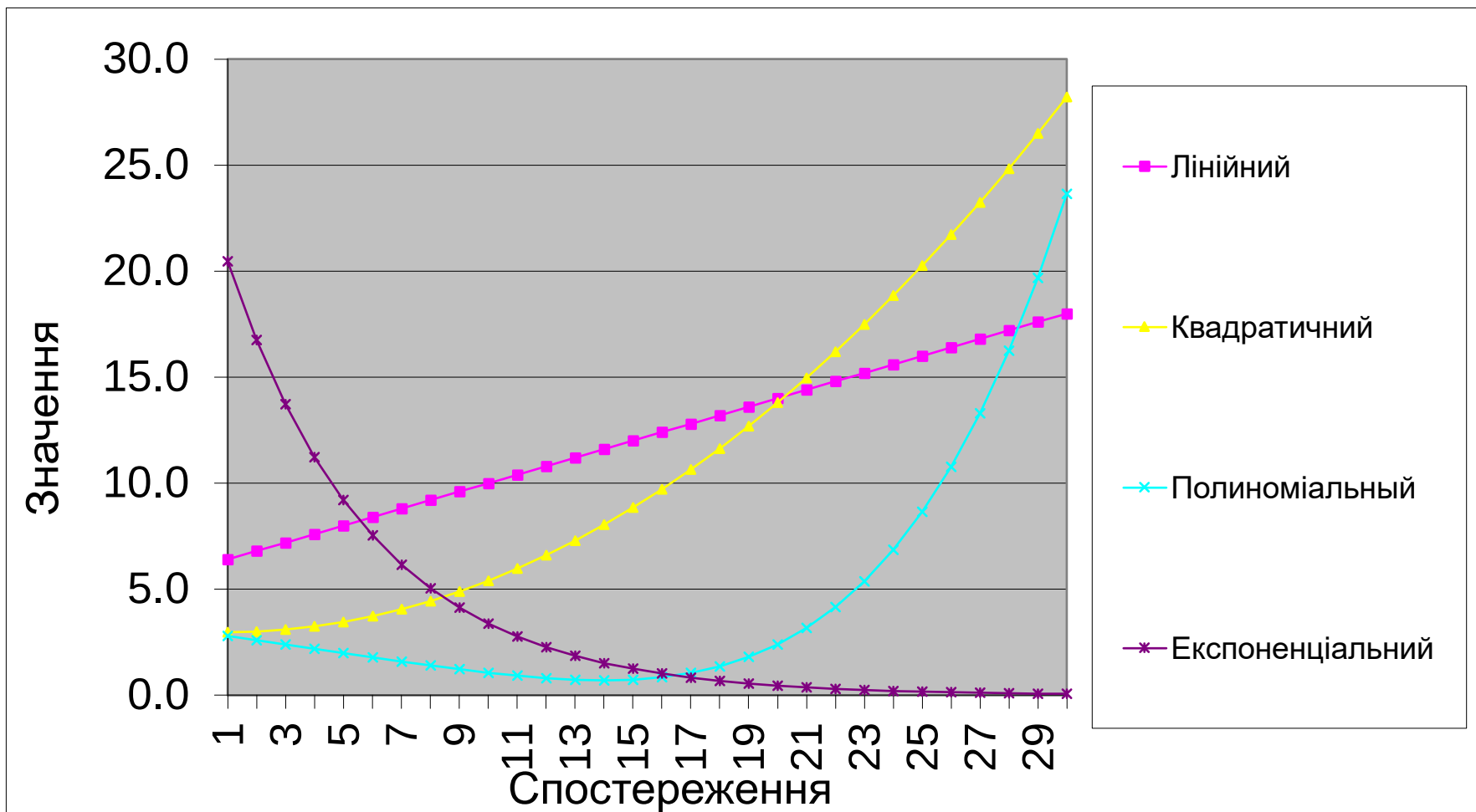
Поліноміальний

$$f(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n$$

Експоненціальний

$$f(t) = a_0 e^{a_1 t}$$

Приклад



Множинна регресія

Вивчення:

- ▶ проблем попиту,
- ▶ прибутковості акцій,
- ▶ вивчення функції витрат виробництва, у мікроекономічних розрахунках.

Врахування впливу кількох факторів

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1, n}$$

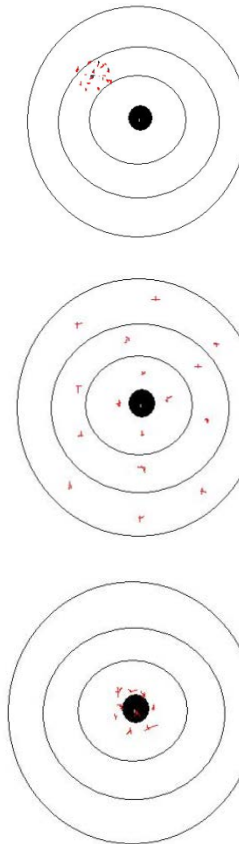
y_t - Залежна змінна ;

$x_{1t}, x_{2t}, \dots, x_{k-1t}$ - незалежні змінні ;

ε_t - залишки .

Бажані характеристики

- ▶ **Відсутність зміщення**
- ▶ **Ефективність**
 - ▶ Стандартні помилки будуть мінімальними
- ▶ **Консистентність**
 - ▶ При збільшенні кількості спостережень стандартна помилка зменшується.



Основна мета множинної регресії

- ▶ побудувати модель з великою кількістю факторів, визначивши при цьому вплив кожного з них окремо, а також сукупний їх вплив на показник, що моделюється.

Специфікація

- ▶ Побудова рівняння множинної регресії починається з вирішення питання специфікації моделі.

Умови включення факторів при побудові множинної регресії

- ▶ Вони мають бути кількісно вимірні. Якщо необхідно включити модель якісний фактор, що не має кількісного виміру, йому потрібно надати кількісну визначеність.
- ▶ Чинники не повинні бути інтеркорельовані. Якщо між факторами існує висока кореляція, то не можна визначити їхній ізольований вплив на результативний показник і параметри рівняння регресії виявляються неінтерпретованими.

Відбір факторів при побудові множинної регресії

- ▶ У першій стадії підбираються фактори, виходячи з сутності проблеми;
- ▶ У другій – з урахуванням матриці показників кореляції визначають суттєвість включення у рівняння регресії кожного з чинників. Якщо фактори явно залежні, то вони дублюють один одного, і один із них рекомендується виключити з регресії.

Приклад

	Y	X1	X2	X3
Y	1			
X1	0,8	1		
X2	0,7	0,8	1	
X 3	0,6	0,5	0,2	1

Очевидно, що фактори X1 та X2 дублюють один одного.

Мультиколінеарність – 1

- ▶ Для оцінки мультиколінеарності факторів може використовуватись визначник матриці парних коефіцієнтів кореляції між факторами.
- ▶ Якби фактори не корелювали між собою, то матриця парних коефіцієнтів кореляції було б одиничною матрицею тобто:

$$\text{Det}|R| = \begin{vmatrix} r_{x_1x_1} & r_{x_2x_1} & r_{x_3x_1} \\ r_{x_1x_2} & r_{x_2x_2} & r_{x_3x_2} \\ r_{x_1x_3} & r_{x_2x_3} & r_{x_3x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1,$$

Мультиколінеарність – 2

- ▶ Якщо ж, навпаки, між факторами існує повна лінійна залежність і всі коефіцієнти кореляції дорівнюють одиниці, то визначник такої матриці дорівнює нулю:

$$\text{Det}|R| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Завдання:

- ▶ На основі даних (закладка «Морозиво») побудувати регресійну модель
- ▶ Виділити значущі чинники (пояснити чому)
- ▶ Побудувати функцію попиту з урахуванням лише значущих параметрів та на її основі визначити рівень цін, за яких дохід фірми буде максимальним
- ▶ Розрахувати значення еластичності за ціною для рівня ціни = 25

Еластичність

$$E_p = \frac{\partial Q}{\partial P} \cdot \frac{P}{Q}$$

Визначення ціни

$$\Pi = Q \cdot P = (b_0 + b_1 P)P \rightarrow \max$$

$$\frac{\partial \Pi}{\partial P} = b_0 + 2b_1 P = 0$$

$$P = -\frac{b_0}{2b_1}$$

Виділення сезонності за допомогою фіктивних змінних

- ▶ Фіктивна змінна – величина, яка набуває значення 1 за наявності деякої ознаки, 0 – за його відсутності.

Перспекії

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \varepsilon_t$$

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \beta_4 t + \varepsilon_t$$

ФІКТИВНІ ЗМІННІ

$$q_1 = (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \dots)'$$

$$q_2 = (0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \dots)'$$

$$q_3 = (0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \dots)'$$

1 квартал

2 квартал

3 квартал

4 квартал

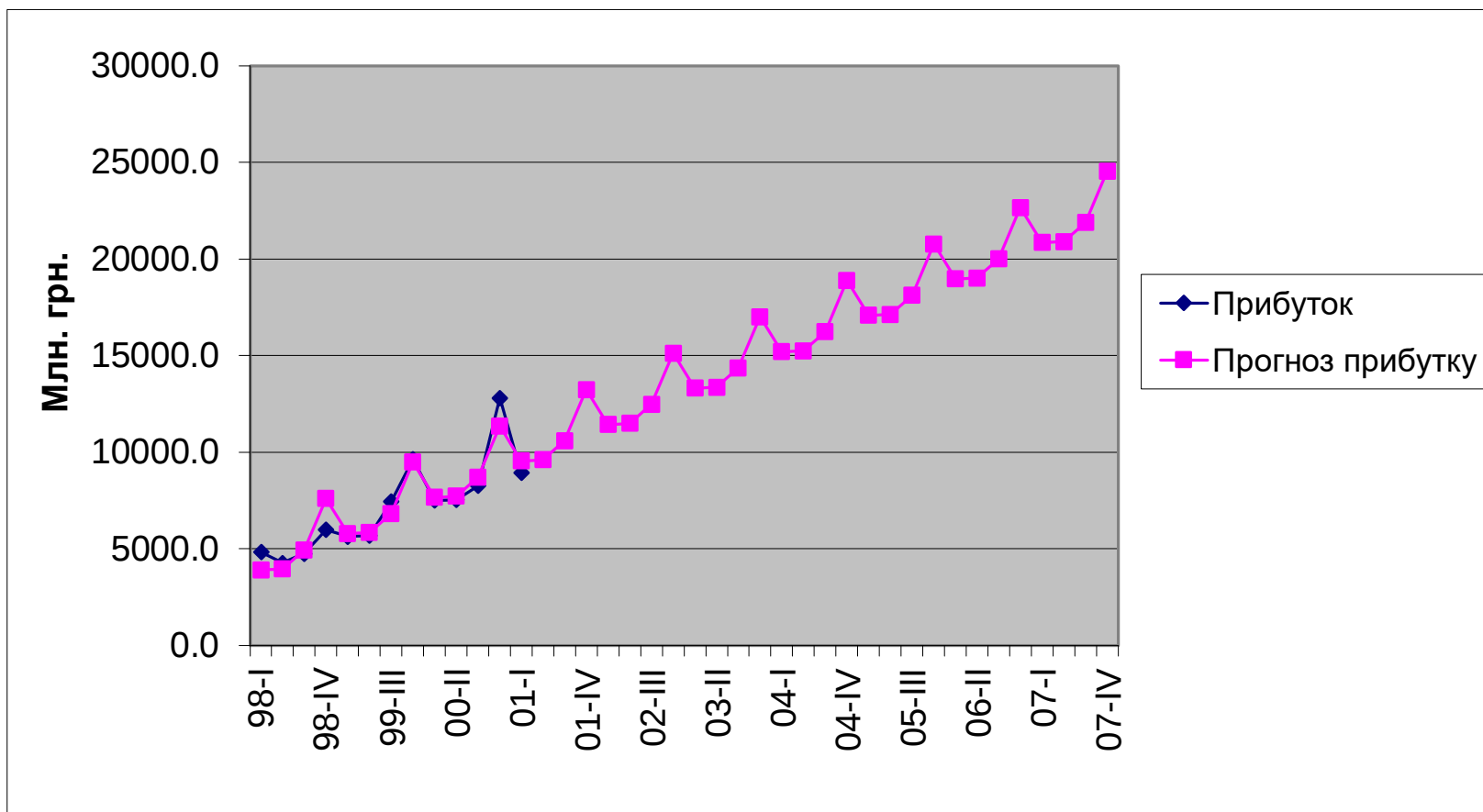
$$y_t = \beta_0 + \beta_1 + \varepsilon_t$$

$$y_t = \beta_0 + \beta_2 + \varepsilon_t$$

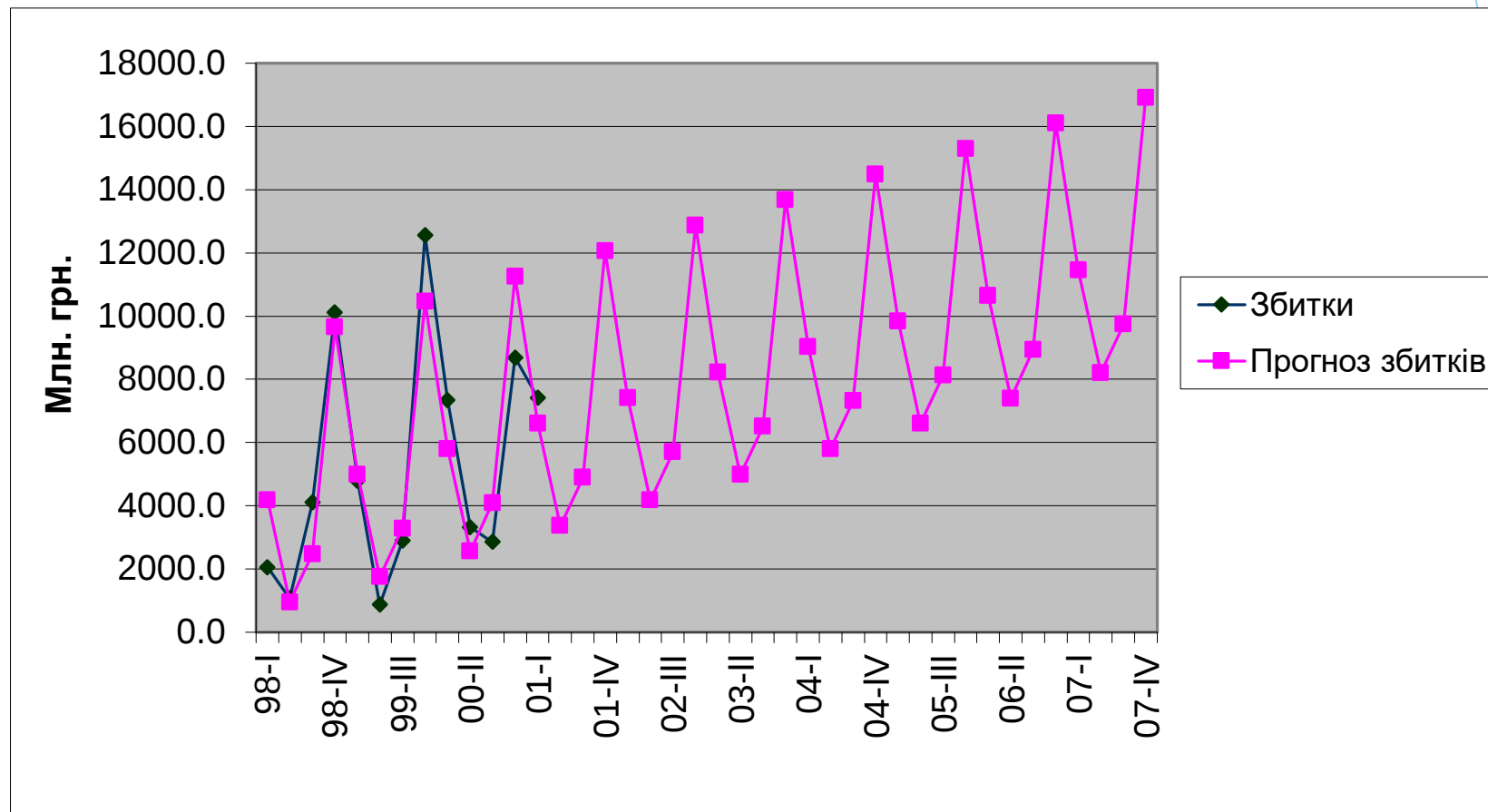
$$y_t = \beta_0 + \beta_3 + \varepsilon_t$$

$$y_t = \beta_0 + \varepsilon_t$$

Приклад – 1



Приклад - 2



Застосування фіктивних змінних

- ▶ Виділення сезонних коливань
- ▶ Нівелювання кризових явищ
- ▶ Моделювання якісних показників
- ▶ Нівелювання пропущених значень

Створення фіктивних сезонних змінних

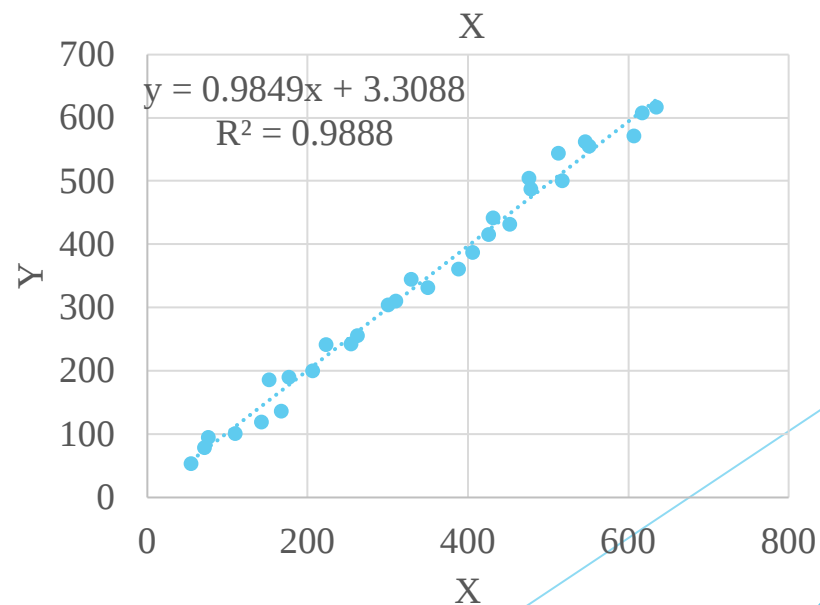
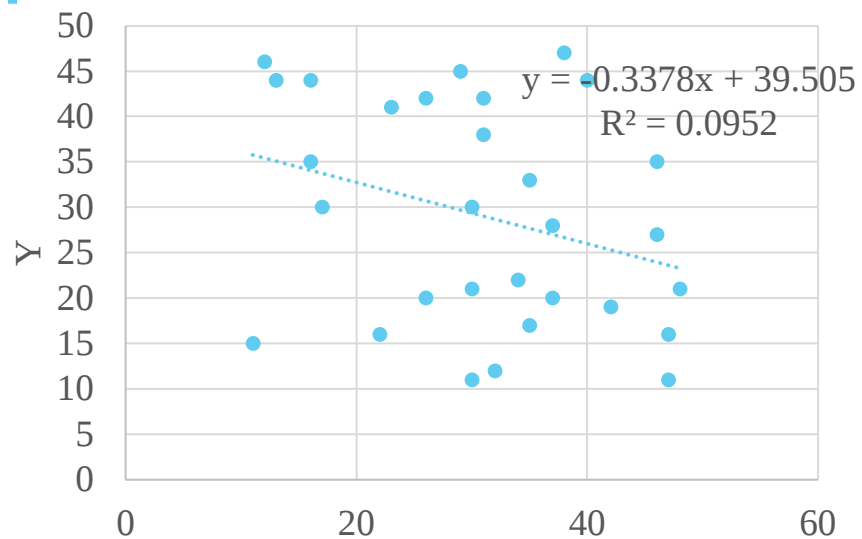
- ▶ Функції Month / Місяць, WeekDay / День Тижня
- ▶ Вручну

Стаціонарний процес

- ▶ Випадковий процес $X(t)$ називається стаціонарним, якщо всі його ймовірні характеристики не змінюються з часом t .

Чому стаціонарність важлива?

№	X	Y	T	X*	Y*
1	35	33	20	55	53
2	31	38	40	71	78
3	16	35	60	76	95
4	30	21	80	110	101
5	42	19	100	142	119
6	47	16	120	167	136
7	12	46	140	152	186
8	17	30	160	177	190
9	26	20	180	206	200
10	23	41	200	223	241
11	34	22	220	254	242
12	22	16	240	262	256
13	40	44	260	300	304
14	30	30	280	310	310
15	29	45	300	329	345
16	30	11	320	350	331
...					
30	35	17	600	635	617



Можливі проблеми

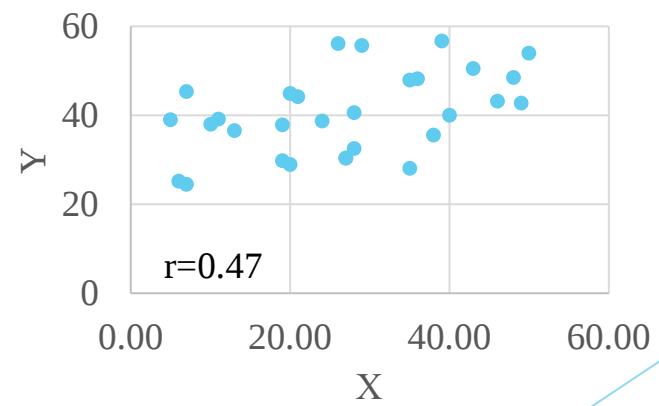
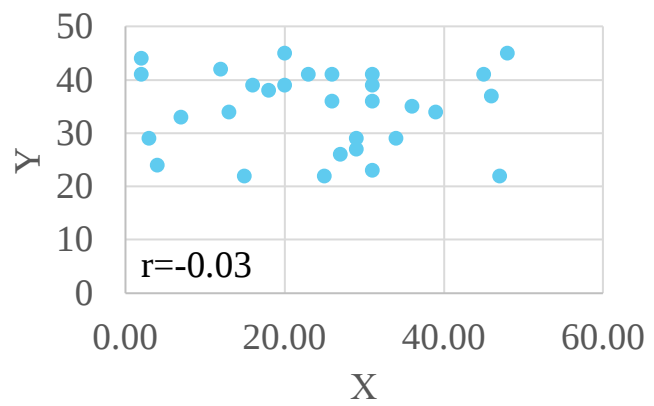
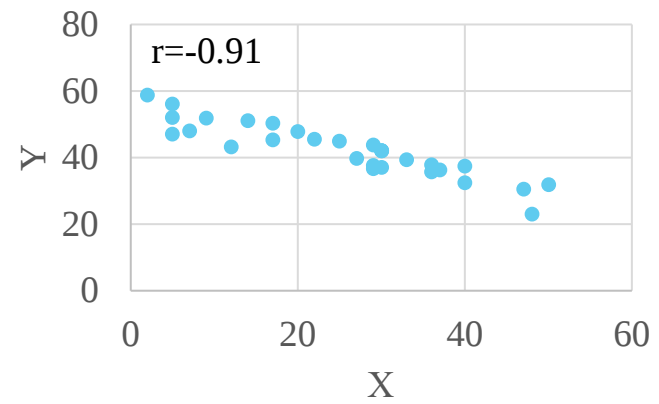
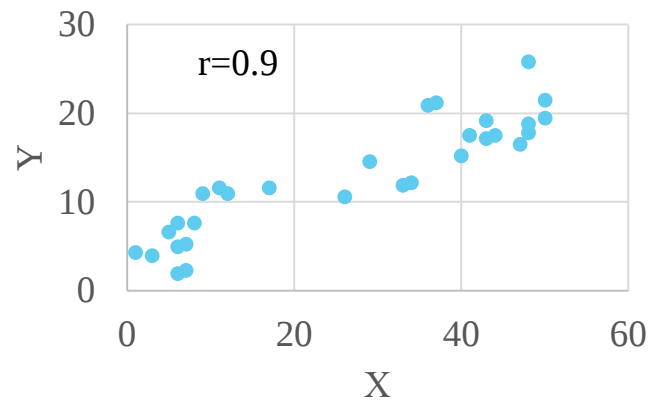
- ▶ Стійкість моделі
- ▶ Розподіл залишків
- ▶ Використання лагів
- ▶ Пропуски даних

Порівняння моделей та вибір оптимальної

- ▶ В основному порівнюємо скоригований коефіцієнт детермінації.
- ▶ Він дорівнює квадрату коефіцієнта кореляції.

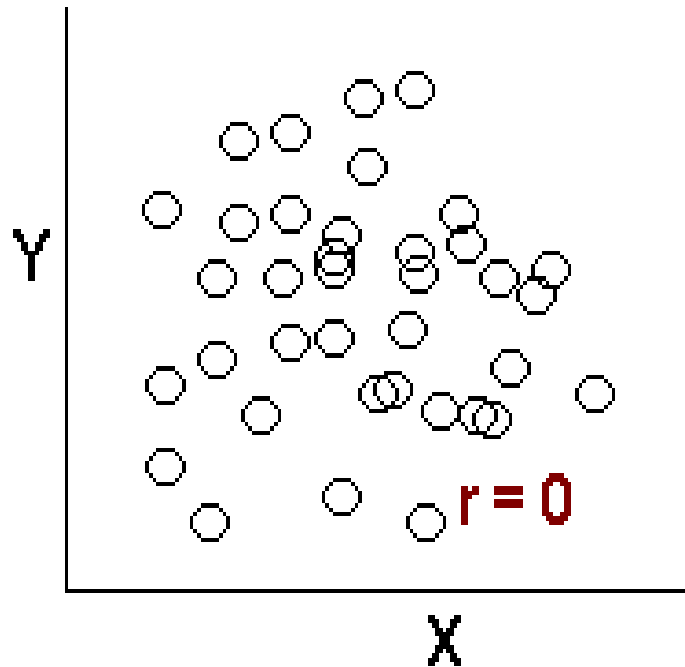
Коефіцієнт кореляції

- ▶ Може набувати значень від -1 до +1
- ▶ Знак коефіцієнта показує напрямок зв'язку (прямий або зворотний)
- ▶ Абсолютна величина показує силу зв'язку
- ▶ Завжди заснований на парах чисел



Але!

- Складається враження, що близький до нуля коефіцієнт кореляції свідчить, що зв'язку між змінними немає чи майже немає.

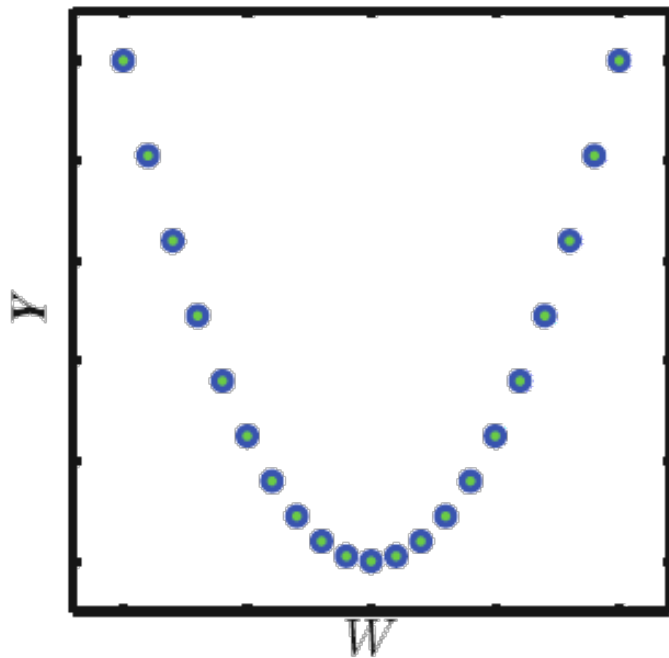


Тут її немає

Але це не завжди так, є винятки!

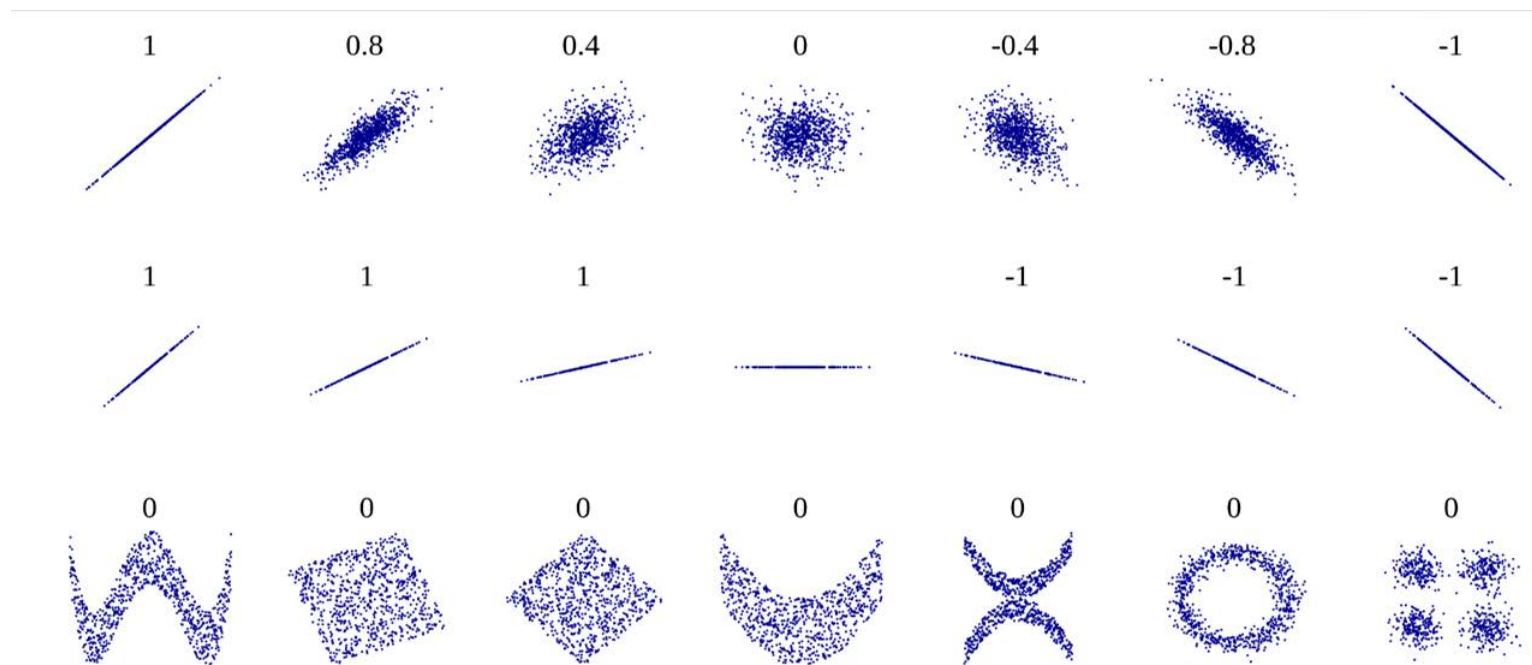
Недолік коефіцієнта кореляції

- ▶ Коефіцієнт кореляції оцінює лише лінійний зв'язок змінних!
- ▶ Він не вказує на наявність нелінійного зв'язку!



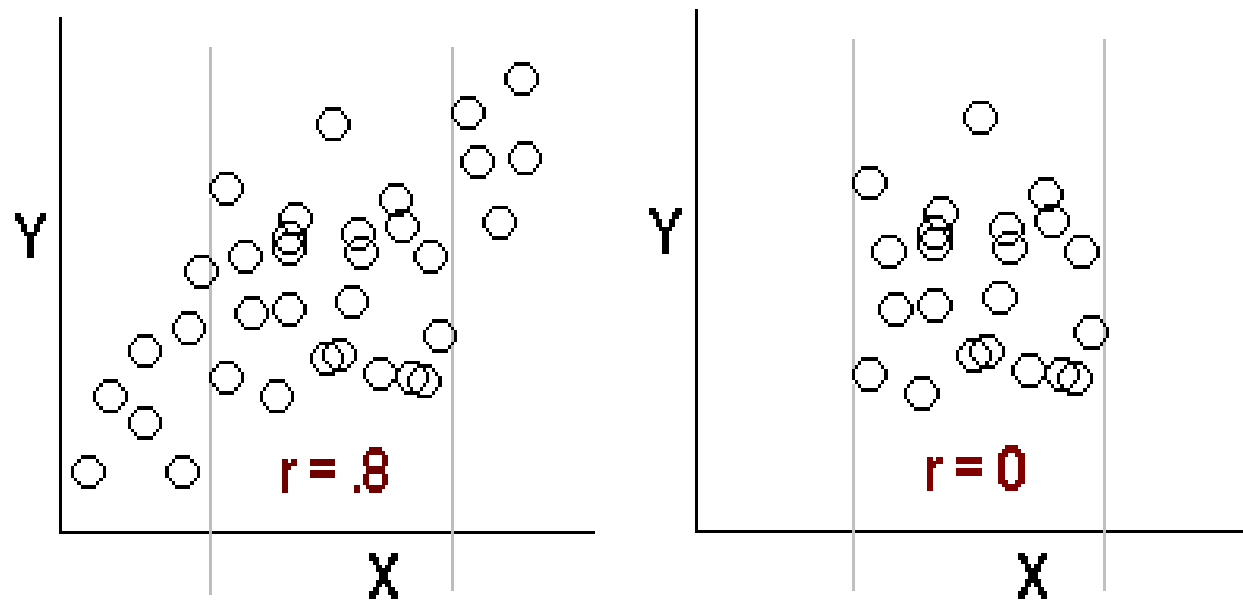
Тут зв'язок змінних дуже сильний, але $r = 0.00$

Приклади

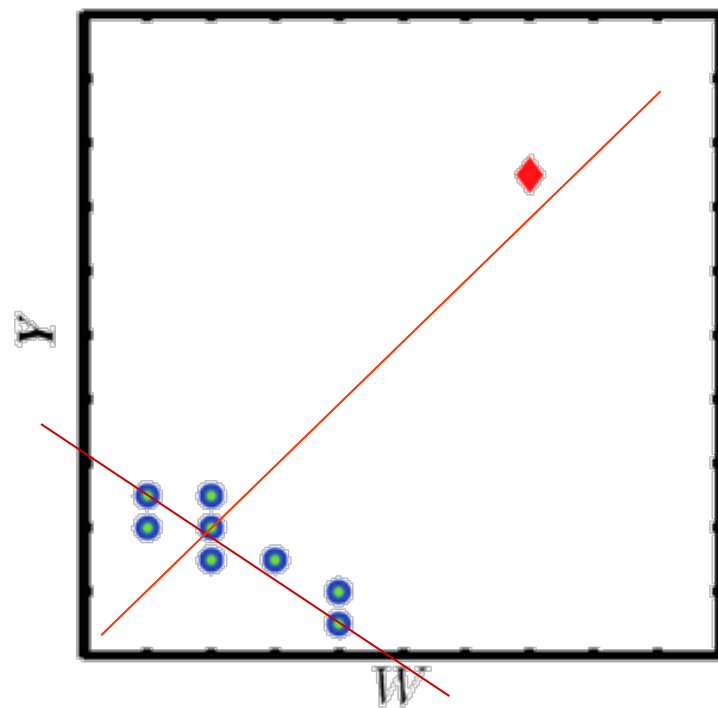


Пам'ятаємо!

- ▶ Потрібно, щоб у змінних була значна мінливість! Якщо сформувати вибірку з однотипних осіб, сподіватися виявити кореляцію там не варто.



Викид

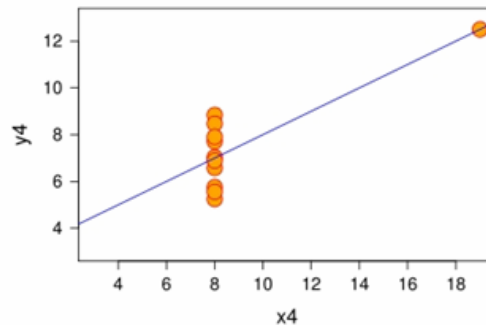
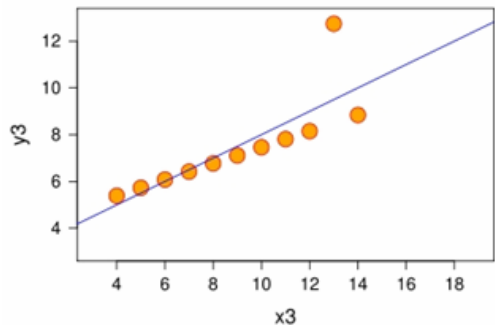
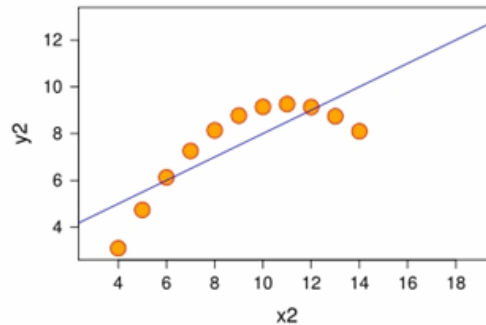
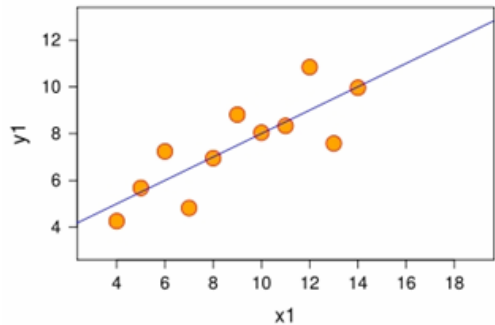


Важливо!

- ▶ Кореляція зовсім не має на увазі наявність причинно-наслідкового зв'язку!
- ▶ Вона взагалі нічого про нього не говорить (навіть дуже великий r)

Рішення про залежність

- ▶ Коефіцієнт кореляції – параметр вибірки. Чи можна з його основі судити про населення?
- ▶ Просто дивлячись на коефіцієнт – Ні.



Кореляція між
змінними
скрізь
 $r = 0.816$

Використання ринкових тенденцій для прогнозування

- ▶ Індекси настрої споживачів
- ▶ Індекс очікувань ділової активності
- ▶ Облікова ставка НБУ
- ▶ Зростання грошової маси
- ▶ Зміна внутрішнього боргу
- ▶ Зміна платіжного балансу
- ▶ Baltic Dry Index (BADI)
- ▶ Oil WTI Futures
- ▶ US Dollar Index Futures

Прогнозування попиту кількох товарів

- ▶ Написання макросів (VBA)
- ▶ Формування аркуша формул

Демонстрація інформаційної системи

Що ми зробили?

- ▶ Прогнозування – важливий момент формування закупівель.

Розглянуто методи:

- ▶ методи аналізу та прогнозування часових рядів;
- ▶ каузальні (причинно-наслідкові) методи (спроба знайти фактори, що визначають поведінку прогнозованого показника);
- ▶ Аналіз якісних показників.

Анкета

- ▶ andriystav.cc.ua/T_Train.html

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the slide, creating a modern, dynamic feel.

Дякую за увагу!

Проф., д.е.н. Ставицький А.В.

Email: a.stavytskyy@gmail.com