

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



Со́я
(*Glycine max* Moench.)

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



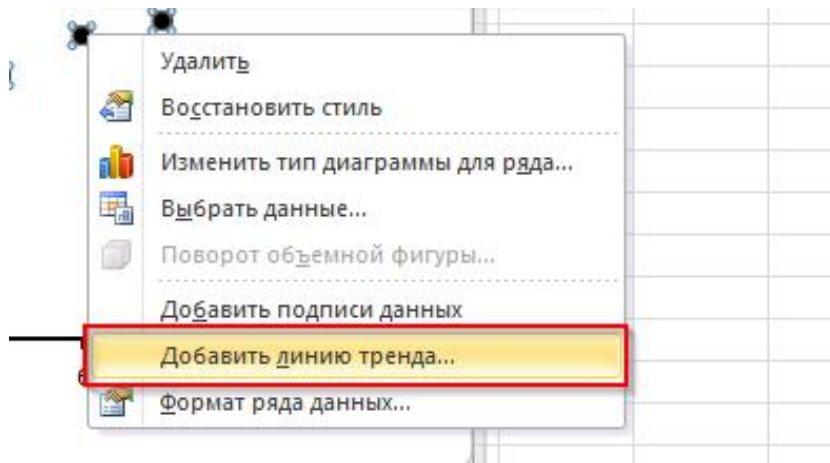
У MS Excel для визначення регресії у графічній формі передбачається наступний алгоритм дій:

- 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 комірки A1 та B1 уведіть назви змінних: «вік, тижнів» та «висота, см», а у комірки A2:B9 – відповідні значення.

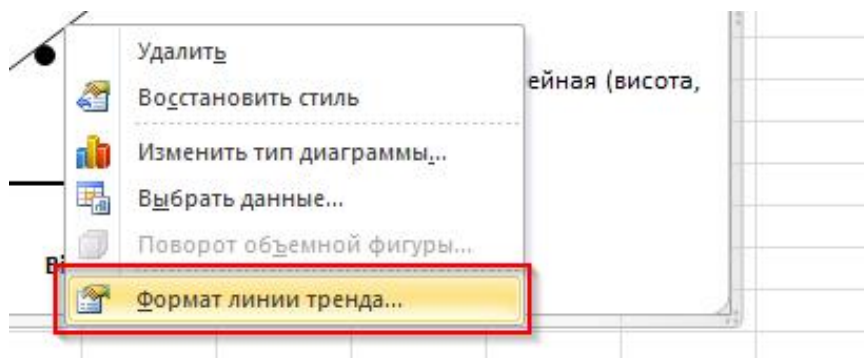
1.2. Побудуйте діаграму розсіювання та відформатуйте її відповідно до рекомендацій завдання 1, лабораторного заняття №5.

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

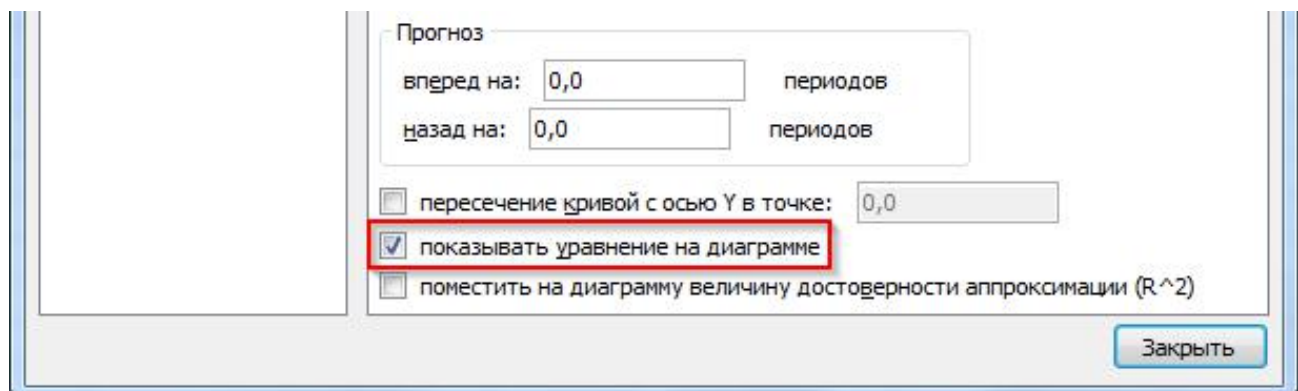


У діалоговому вікні, що з'явиться, оберіть шляхом встановлення перемикача тип лінії тренда (регресії) – «Линейная», за бажанням змініть формат лінії (товщину, колір, вигляд) та натисніть на кнопку «Закреть». На діаграмі з'явиться лінія регресії.

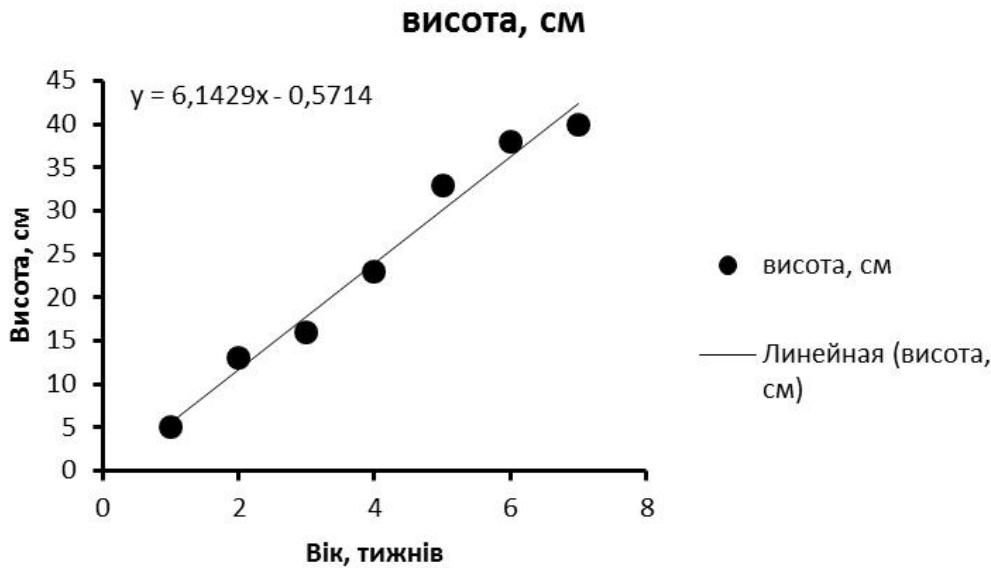
1.4. Додайте до діаграми рівняння регресії. Для цього клацніть ПКМ по лінії тренду (регресії) та оберіть пункт «Формат линии тренда...».



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

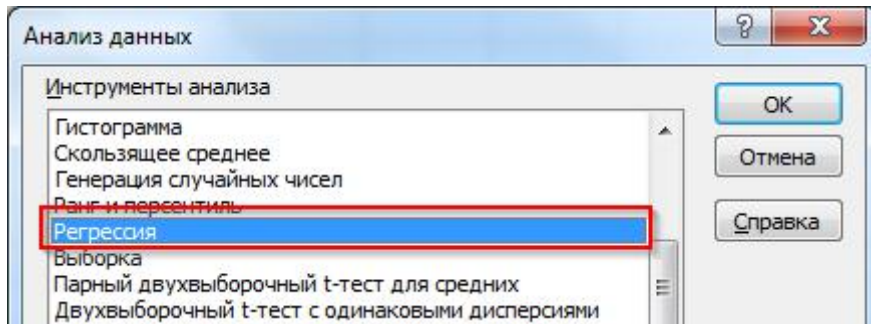
Визначте коефіцієнт регресії та його довірчу імовірність при $\alpha = 0,05$.



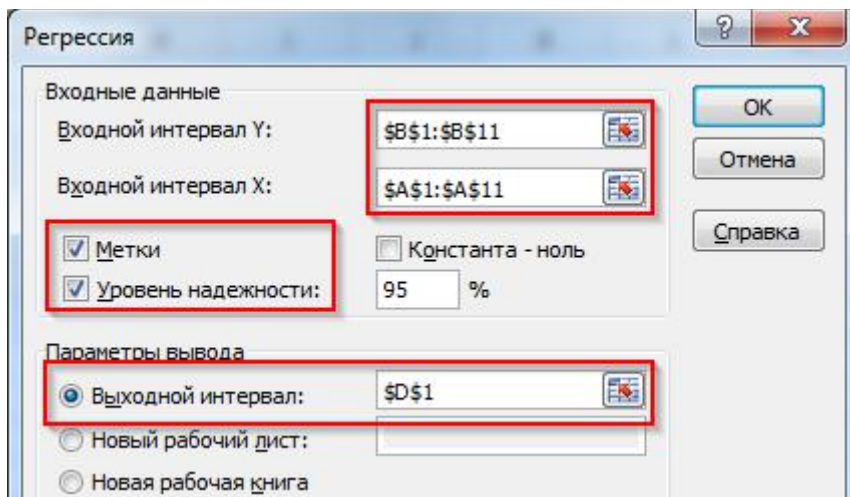
У MS Excel для визначення коефіцієнту регресії використовується процедура «Регрессия» надбудови «Анализ данных» за наступним алгоритмом дій:

- 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 «**OK**».



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 ?