Національний технічний університет
ДНІПРОВСЬКА ПОЛІТЕХНІКА
1899

Пістунов І.М.

Економічна статистика: навч. наоч. посіб. Дніпро : НТУ «ДП», 2025. 48 с.

В посібнику розглядається сутність, порядок розрахунків та розробку висновків у економічній статистиці.
Подано приклади розрахунків із застосуванням програми Microsoft Excel.
Наведено надано пояснення щодо структури даних,
Призначено для студентів спеціальності 051 «Економіка»

Рецензенти:
Васильєва Н.К., завідувач каф. інформаційних систем ДДАЕУ, проф.
Алексеєв М.О., зав каф. програмних засобів комп’ютерних систем НТУ “ДП”, проф.

ЗМІСТ

РЕСУРСИ ІНТЕРНЕТУ

[Державна служба статистики України](http://www.ukrstat.gov.ua/)

http://www.ukrstat.gov.ua/

[Індекси МінФіну](https://index.minfin.com.ua/ua/reference/coronavirus/geography/)

https://index.minfin.com.ua/ua/reference/coronavirus/geography/

[Міністерство фінансів](https://www.mof.gov.ua/uk/)

https://www.mof.gov.ua/uk/

[Національний банк України](https://bank.gov.ua/)

https://bank.gov.ua/

ВСТУП

Стати́стика — наука, що вивчає методи кількісного охоплення і дослідження масових, зокрема суспільних, явищ і процесів. А також власне кількісний облік масових явищ. Зокрема, облік у будь-якій галузі господарства, суспільного життя, що здійснюється методами цієї науки, а також дані цього обліку. Статистика вивчає кількісний бік масових явищ і процесів у нерозривному зв'язку з їх якісним боком. Статистика поділяється на математичну та прикладну. Прикладну статистику поділяють за галузями на демографічну, економічну, фінансову, соціальну, санітарну, судову, біологічну, технічну тощо.

Статистика розробляє спеціальну методологію дослідження й обробки матеріалів: масові статистичні спостереження, метод угруповань, середніх величин, індексів, балансовий метод, метод графічних зображень й інші методи аналізу статистичних даних.

Статистика поділяється за своїм змістом на демографічну, економічну, фінансову, соціальну, санітарну, судову, біологічну, технічну тощо; математична статистика вивчає математичні методи систематизації, обробки й використання статистичних даних для наукових і практичних висновків.

Статистика складається із трьох розділів:

- збір статистичних відомостей, тобто відомостей, що характеризують окремі одиниці яких-небудь масових сукупностей;
- статистичне дослідження отриманих даних, що полягає в з'ясуванні тих закономірностей, які можуть бути встановлені на основі даних масового спостереження;
- розробка прийомів статистичного спостереження й аналізу статистичних даних. Останній розділ, власне, і становить зміст математичної статистики».

Термін «статистика» вживають ще у двох значеннях. По-перше, у побуті під «статистикою» часто розуміють набір кількісних даних про яке-небудь явище або процес. По-друге, статистикою називають функцію від результатів спостережень, використовувану для оцінки характеристик і параметрів розподілів і перевірки гіпотез.

Основні поняття (категорії) статистики

Статистична сукупність — це маса однорідних в певному відношенні елементів, які мають єдину якісну основу, але різняться між собою певними ознаками і підлягають певному закону розподілу. Статистична сукупність — це певна множина елементів, поєднана умовами існування і розвитку.

Однорідна сукупність — якщо одна чи декілька ознак, що вивчаються, є загальними для всіх одиниць.

Різнорідна сукупність об'єднує явища різного типу.

Одиниця сукупності — це первинний елемент статистичної сукупності, який є носієм ознак, що підлягають реєстрації і є основою обліку.

Ознака — властивість окремої одиниці сукупності.

Якісні ознаки (атрибутивні ознаки) виражаються в вигляді понять, визначень, які характеризують їхню суть, стан або якість. Наприклад, сорт продукції, професія, сімейний статус.

Кількісні ознаки виражають окремі значення якісних ознак у числовому виразі.

Дискретні — ознаки, виражені окремими цілими числами, без проміжних значень.

Неперервні — ознаки, що можуть набувати будь-яких значень у певних чисел.

Прямі — характеризують об'єкт дослідження безпосередньо (вік осіб, кількість присутніх в аудиторії).

Непрямі — ознаки, що не належать безпосередньо досліджуваному об'єкту (чи сукупності), а які належать іншій сукупності, що входить в дану.

Багатоваріантні — перш за все характеризуються рангами (шкалою рангів) від більшого до меншого (напр. дуже низький, низький, середній, високий, дуже високий).

Альтернативні — взаємовиключаючі значення: так-ні, позитивне-негативне.

Інтервальні — це ознаки, які характеризують результат процесів.

Моментні — характеризують об'єкт в певний момент часу.

Окремі значення кількісних ознак називаються варіантами.

Первинні варіанти характеризують одиницю сукупності в цілому: абсолютні значення, виміряні, розраховані.

Вторинні варіанти (похідні, розрахункові) — дані, що неможливо перевірити, оскільки вони взяті з певних джерел.

Статистичні показники — це числа в сукупності з набором ознак, що характеризують обставини, до яких вони відносяться, що, де, коли, і яким чином підлягають вимірюванню. Статистичний показник — це кількісна характеристика соціально-економічних явищ і процесів в умовах якісної визначеності.

Статистичні дані — це сукупність показників, отриманих внаслідок статистичного спостереження або обробки даних.

Статистична закономірність — це закономірність, в якій необхідність пов'язана в кожному окремому явищі з випадковістю, і лише в сукупності явищ виявляє себе як закон.

Система статистичних показників — це сукупність статистичних показників, які відображають взаємозв'язки, які об'єктивно існують між явищами.

Загальна мета для статистичного дослідження проекту є вивчення причинно-наслідкові зв'язки, і, зокрема, зробити висновок про вплив зміни значень предикторів або незалежних змінних на залежні змінні або відповідь. Є два основних типи причинних статистичних досліджень: експериментальні дослідження та наглядові дослідження. Різниця між цими двома типами досліджень полягає в тому, як насправді дослідження проводилося. Кожне з цих досліджень може бути дуже ефективним. Експериментальне дослідження включає в себе прийом вимірювання даної системи, маніпулюючи системою, а потім приймати додаткові виміри використовуючи ту ж процедуру, щоб визначити, маніпуляції модифікованих значень вимірювань. На відміну від цього, наглядове дослідження не передбачає експериментальної маніпуляції. Замість цього, дані збираються і кореляції між предикторами і відповідями досліджуються.

Предиктор (от англ. *predictor* «Провісник») - прогностичний параметр; засіб прогнозування.

Абсолютні та відносні величини

В залежності від суті та методики розрахунку розрізняють сім видів відносних величин: планового завдання, виконання плану, динаміки, структури, координації, порівняння та інтенсивності.

$$BB_{пз} = \frac{П_1}{Ф_0} \qquad BB_{пз} = \frac{П_1}{Ф_0} x100$$

Відносна величина динаміки характеризує зміну показника у часі і визначається як відношення значення у наступному періоді до величини у попередньому періоді:

Такі показники називаються темпом росту

$$BB_{\partial} = \frac{\Phi_1}{\Phi_0}$$

А такі – коефіцієнтом росту

$$BB_{\partial} = \frac{\Phi_1}{\Phi_0} x100$$

А такі – коефіцієнтом приросту

$$BB_{\partial} = \left(\frac{\Phi_1}{\Phi_0} - 1 \right) x100\%$$

У статистичній практиці використовують декілька видів середніх:
- середня арифметична;
- середня гармонійна;
- середня квадратична;
- середня геометрична.

Середня арифметична — одна з найбільш поширених видів середньої, застосовується у тих випадках, коли обсяг варіюючої ознаки для всієї сукупності є сумою індивідуальних значень її окремих елементів. Середня арифметична використовується для осереднення прямих значень ознаки шляхом їх підсумування.

$$\overline{X} = \frac{\sum x}{n},$$

де $\overline{x}$ - середнє значення досліджуваної ознаки;

x - окремі значення усереднюваної ознаки (варіанти);

n - число одиниць досліджуваної сукупності;

Середня гармонійна застосовується у тих випадках, коли нам відомі не самі вартості, а їхні обернені числа.

$$\bar{x} = \frac{n}{\sum \frac{1}{x}}$$

Середня квадратична

$$\overline{X} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Середня геометрична

$$\overline{X} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Медіаною називають варіанту, що ділить варіаційний ряд на дві частини з рівною кількістю варіант. Якщо кількість варіант непарна то

у випадку парної кількості варіант медіана дорівнює

Поняття варіації та її основні показники

Розмах варіації: $R = x_{\max} - x_{\min}$

Просте лінійне відхилення:

$$d = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n}$$

Зважене лінійне відхилення

$$d = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right| \cdot f_i}{\sum_{i=1}^{n} f_i}$$

Формула зваженої середньої дисперсії

$$\sigma_x^2 = \frac{\sum_{i=1}^{n} (x - \bar{x})^2 f_i}{\sum_{i=1}^{n} f_i}$$

Зважене середнє квадратичне відхилення

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2 f_i}{\sum f_i}}$$

Відносні показники варіації

Коефіцієнт осциляції, який відображує відносну коливність крайніх значень ознаки навколо середньої:

$$V_R = \frac{R}{\bar{x}} 100\%$$

Лінійний коефіцієнт варіації:

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} 100\%$$

Коефіцієнт варіації:

$$V_\sigma = \frac{\sigma_x}{\bar{x}} 100\%$$

Коефіцієнт варіації є певною мірою критерієм типовості середньої. Якщо коефіцієнт дуже великий - це означає, що середня характеризує сукупність за ознакою, яка суттєво змінюється в окремих одиниць. Типовість такої середньої сумнівна, тобто невелика.

Вибіркове спостереження є такий вид спостереження, при якому обстеженню піддається частина одиниць досліджуваної сукупності, що дозволяє по ній отримати дані для характеристики всієї сукупності.

Вся вивчається сукупність явищ називається генеральної (позначається N). Та частина одиниць, яка відібрана з генеральної сукупності називається вибіркової сукупністю (позначається n).

Розрізняють середню і граничну помилки вибірки. Ці два види помилки пов'язані наступним співвідношенням:

$$\Delta = t \cdot \mu$$

Середня помилка вибірки

$$\mu = \sqrt{\frac{\sigma^2}{n}}$$

Формула середньої помилки при випадковій бесповторній вибірці

$$\mu = \sqrt{\left[\frac{\sigma^2}{n} \cdot \left(1 - \frac{n}{N}\right)\right]}$$

Діапазон існування середнього

$$\tilde{X} - \Delta_{\overline{X}} \leq \overline{X} \leq \tilde{X} + \Delta_{\overline{X}}$$

Індекси

Економічний індекс - це відносна величина, яка характеризує зміну досліджуваного явища в часі, в просторі або в порівнянні з певним еталоном

кількість даного виду продукції визначається буквою - *q* (фізичний обсяг); ціна одиниці виробу - *p*; собівартість одиниці виробу - *z*; трудомісткість одиниці виробу – *t*, P_0 - значення показника в базисному періоді; P_1 - значення показника в поточному періоді.

Індивідуальний індекс:

$$i_p = \frac{P_1}{P_0}$$

Індекс фізичного обсягу реалізації:

$$J_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$$

Загальний індекс ціни

$$J_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}$$

Загальний індекс товарообігу (реалізації продукції)

$$J_{qp} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0}$$

Індивідуальний індекс собівартості

$$i_z = \frac{z_1}{z_0}$$

Загальний індекс собівартості

$$J_z = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1}$$

Загальний індекс фізичного обсягу продукції, зважений за собівартістю

$$J_q = \frac{\sum q_1 \cdot z_0}{\sum q_0 \cdot z_0}$$

Загальний індекс витрат на виробництво

$$J_{zq} = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_0}$$

Середні рівні собівартості

$$\bar{z}_0 = \frac{\sum z_0 \cdot q_0}{\sum q_0},$$

$$\bar{z}_1 = \frac{\sum z_1 \cdot q_1}{\sum q_1}.$$

Індекс змінного складу

$$J_{\text{п.с.}} = \frac{\bar{z}_1}{\bar{z}_0}$$

Індекс фіксованого складу (або постійного складу)

$$J_{фикс.сост.} = \frac{\sum z_1 \cdot q_1}{\sum q_1} : \frac{\sum z_0 \cdot q_1}{\sum q_1} = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1}$$

Індекс структурних зрушень

$$\mathrm{J}_{\text{стр.сдв.}} = \frac{\sum z_0 \cdot q_1}{\sum q_1} : \frac{\sum z_0 \cdot q_0}{\sum q_0}$$

Статистичне вивчення взаємозв'язку

Розрахунок емпіричного кореляційного відносини заснований на використанні відомої теореми додавання дисперсій. Загальна дисперсія результативної ознаки (σ_0^2) може бути розкладена на дві складові. Перша складова - міжгрупова дисперсія (δ^2), характеризує ту частину коливання результативної ознаки, яка складається під впливом зміни ознаки-фактору, покладеного в основу угруповання:

$$\delta^2 = \frac{\sum_{j=1}^{k} (\bar{y}_j - \bar{y}_0)^2 \cdot n_j}{\sum_{j=1}^{k} n_j}$$

де $\bar{y}_j$ - середнє значення результативної ознаки у відповідних групах;

$\bar{y}_0$ - загальна середня величина результативної ознаки для всієї сукупності;

n_j - число спостережень у відповідній групі;

k - число виділених груп.

Друга складова - середня з внутрішньогрупових дисперсій - ($\overline{\sigma}^2$) оцінює ту частину варіації результативної ознаки, яка обумовлена дією інших, "випадкових" причин

$$\overline{\sigma}^2 = \frac{\sum \sigma_j^2 \cdot n_j}{\sum n_j}$$

де σ_j^2 - дисперсія результативної ознаки у відповідній групі:

$$\sigma_j^2 = \frac{\sum (y_j - \bar{y})^2}{n}$$

Загальна дисперсія

$$\sigma_0^2 = \delta^2 + \overline{\sigma}^2 .$$

Частка міжгрупової дисперсії у загальній дисперсії

$$\eta^2 = \frac{\delta^2}{\delta^2 + \overline{\overline{\sigma}}^2}$$

Вибіркове кореляційне відношення

$$\eta = \sqrt{\eta^2}$$

Міри зв’язку вибірок

Коефіцієнт кореляції
$$r = \frac{n \cdot \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{\sqrt{\left(n \cdot \sum x_i^{\ 2} - \left(\sum x_i\right)^2\right) \cdot \left(n \cdot \sum y_i^{\ 2} - \left(\sum y_i\right)^2\right)}}$$

Коефіцієнтом детермінації називається квадрат коефіцієнта кореляції (r^2).

На практиці для оцінки ступені взаємозв'язку можна керуватись наступними емпіричними правилами:

1) $|r| > 0{,}95$ - існує практично лінійна залежність;
2) $0{,}8 < |r| < 0{,}95$ - сильна ступінь лінійної залежності;
3) $0{,}6 < |r| < 0{,}8$ - належність лінійного зв'язку;
4) $|r| < 0{,}4$

Розрахунок емпіричного кореляційного відносини заснований на використанні відомої теореми додавання дисперсій. Загальна дисперсія результативної ознаки (σ_0^2) може бути розкладена на дві складові. Перша складова - міжгруппова дисперсія (δ^2), характеризує ту частину коливання результативної ознаки, яка складається під впливом зміни ознаки-фактору, покладеного в основу угруповання:

ІНША ФОРМА КОЕФІЦІЄНТА КОРЕЛЯЦІЇ

$$\operatorname{cov}(X,Y) = R_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \ddot{m}_x)(y_i - \ddot{m}_y)$$

$$cor(X,Y) = r_{XY} = \frac{R_{XY}}{\ddot{\sigma}_X \cdot \ddot{\sigma}_Y}$$

Множинна кореляція — метод кореляційного аналізу для вимірювання кореляційного зв'язку одночасно між двома, трьома і більше кореляційними ознаками.

Частковий коефіцієнт кореляції між ознаками у та x_1 без урахування впливу ознаки x_2

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} * r_{x_1x_2}}{\sqrt{(1-r^2_{yx_2})(1-r^2_{x_1x_2})}}$$

Частковий коефіцієнт кореляції між у та x_2 без урахування впливу $x_{1:}$

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} * r_{x_1x_2}}{\sqrt{(1-r^2_{yx_1})(1-r^2_{x_1x_2})}}$$

При лінійному двофакторному зв’язку коефіцієнт лінійної кореляції можна визначити за такою формулою:

$$R_{y/x_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} * r_{yx_2} * r_{x_1x_2}}{1-r^2_{x_1x_2}}}$$

- Статистичні рішення мають імовірнісний характер, тобто завжди існує ймовірність того, що прийняті рішення будуть помилковими.
- Головна цінність прийняття статистичних рішень полягає в тому, що в межах імовірнісних категорій можна об'єктивно виміряти ступінь ризику, що відповідає тому чи іншому рішенню.
- Будь-які статистичні висновки, здобуті на підставі обробки вибірки, називають статистичними гіпотезами.
- Статистичні гіпотези про значення параметрів ознак генеральної сукупності називають параметричними.
- Наприклад, висувається статистична гіпотеза про числові значення генеральної середньої $X_Г$, генеральної дисперсії $D_Г$, генерального середнього квадратичного відхилення $6_Г$ та ін.
- Статистичні гіпотези, що висуваються на підставі обробки вибірки про закон розподілу ознаки генеральної сукупності, називаються непараметричними.
- Наприклад, на підставі обробки вибірки може бути висунута гіпотеза, що ознака генеральної сукупності має нормальний закон розподілу, експоненціальний закон та ін.

Статистичний критерій. Емпіричне значення критерію

Для перевірки правильності висунутої статистичної гіпотези вибирають так званий статистичний критерій, керуючись яким відхиляють або не відхиляють нульову гіпотезу. Статистичний критерій, котрий умовно позначають через K, є випадковою величиною, закон розподілу ймовірностей якої нам заздалегідь відомий.

Спостережуване значення критерію, який позначають через K*, обчислюють за результатом вибірки.

Сукупність значень статистичного критерію K, за яких нульова гіпотеза не відхиляється, називають областю прийняття нульової гіпотези.

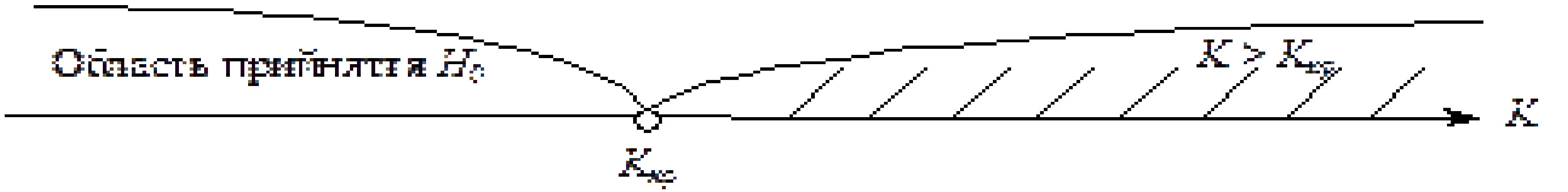
Сукупність значень статистичного критерію K, за яких нульова гіпотеза не приймається, називають критичною областю.

- Нульова й альтернативна гіпотези
- Гіпотезу, що підлягає перевірці, називають основною. Оскільки ця гіпотеза припускає відсутність систематичних розбіжностей (нульові розбіжності) між невідомим параметром генеральної сукупності і величиною, що одержана внаслідок обробки вибірки, то її називають нульовою гіпотезою і позначають H_0.
- Зміст нульової гіпотези записується так:

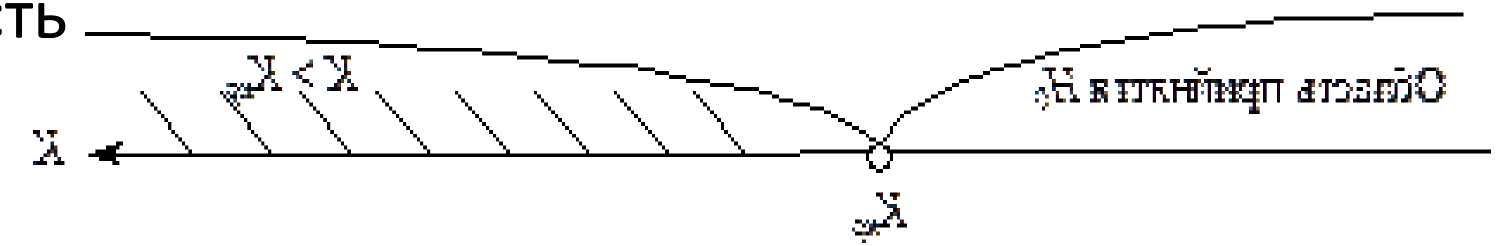
- $H_0: \; X_Г = a;$
- $H_0: \; 6_Г = 2;$
- $H_0: \; r_{xy} = 0{,}95$

Існують три види критичних областей:

- Якщо при $K < K_{кр}$ нульова гіпотеза відхиляється, то в цьому разі ми маємо лівобічну критичну область, яку умовно можна зобразити .



Якщо при $K > K_{кр}$ нульова гіпотеза відхиляється, то в цьому разі маємо правобічну критичну область



Якщо ж при $K < K_{кр1}$ і при $K < K_{кр2}$ нульова гіпотеза відхиляється, то маємо двобічну критичну область

Перевірка статистичних гіпотез щодо середніх величин

З метою перевірки гіпотези про рівність середніх в генеральній сукупності необхідно сформулювати нульову гіпотезу. При цьому, як правило, виходять з того, що обидві вибірки узяті з нормально розподіленої генеральної сукупності з математичним сподіванням, рівним *X* і з дисперсією, рівною c_0.

Тоді задача перевірки гіпотези зводиться до перевірки істотності різниці

$$\left|\tilde{x}_1 - \overline{x}\right| - \left|\tilde{x}_2 - \overline{x}\right| = \left|\tilde{x}_1 - \tilde{x}_2\right| = \varepsilon_{факт}.$$

кожна вибіркова середня має свою помилку*:*

$$\mu_1^2 = \frac{S_1^2}{n_1}; \qquad \mu_2^2 = \frac{S_2^2}{n_2}.$$

де скориговані дисперсії

$$S_1^2 = \frac{\Sigma(x_{1i} - \tilde{x}_1)^2}{n_1 - 1}; \qquad S_2^2 = \frac{\Sigma(x_{2i} - \tilde{x}_2)^2}{n_2 - 1},$$

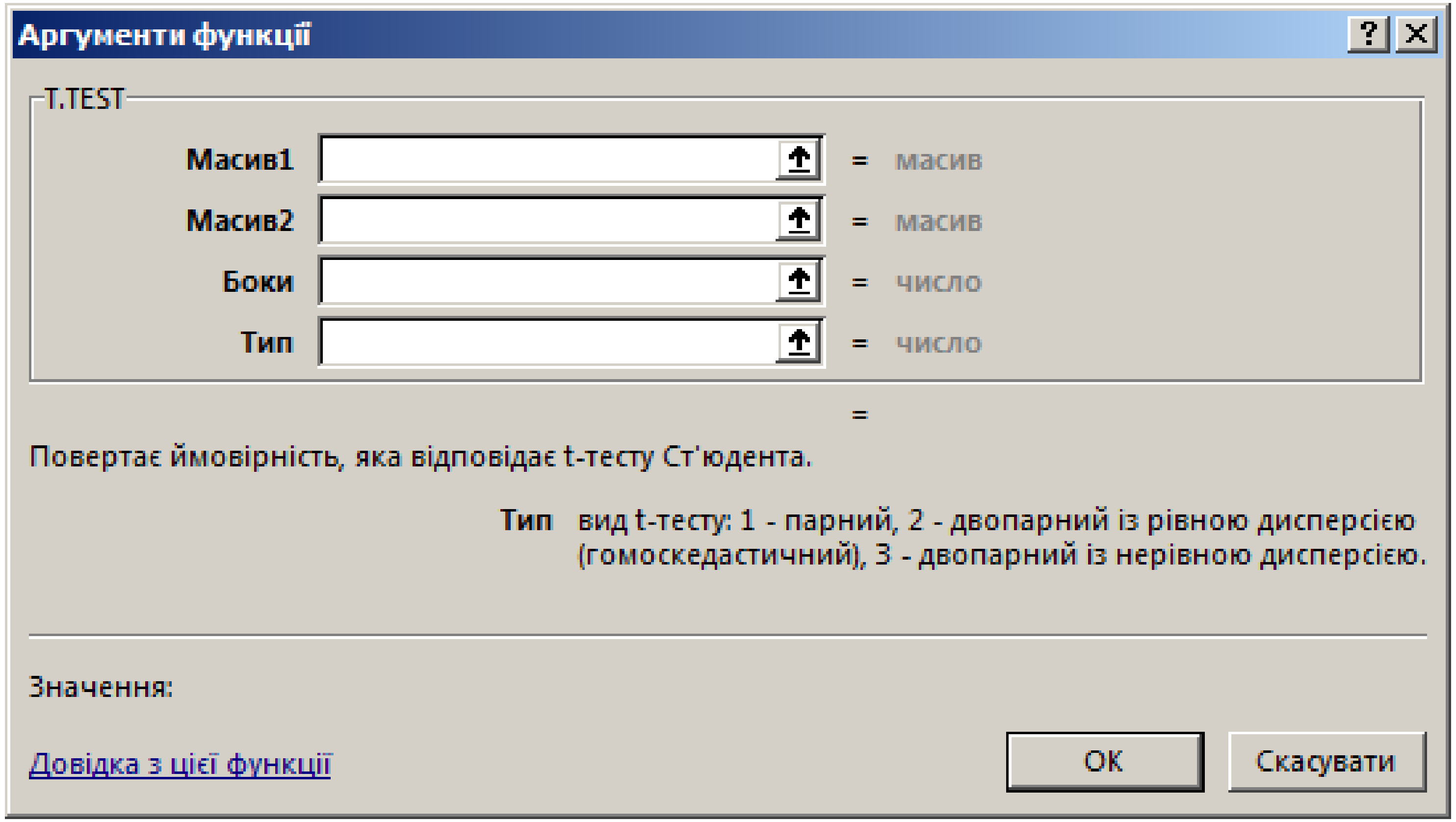
або за іншими формулами

$$S_1^2 = \frac{\Sigma x_{1i}^{\ 2} - n_1\tilde{x}_1^{\ 2}}{n_1 - 1}; \qquad S_2^2 = \frac{\Sigma x_{2i}^{\ 2} - n_2\tilde{x}_2^{\ 2}}{n_2 - 1}.$$

Узагальнена середня помилка двох вибіркових середніх

$$\overline{\mu}_{1-2} = \sqrt{\mu_1^2 + \mu_2^2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

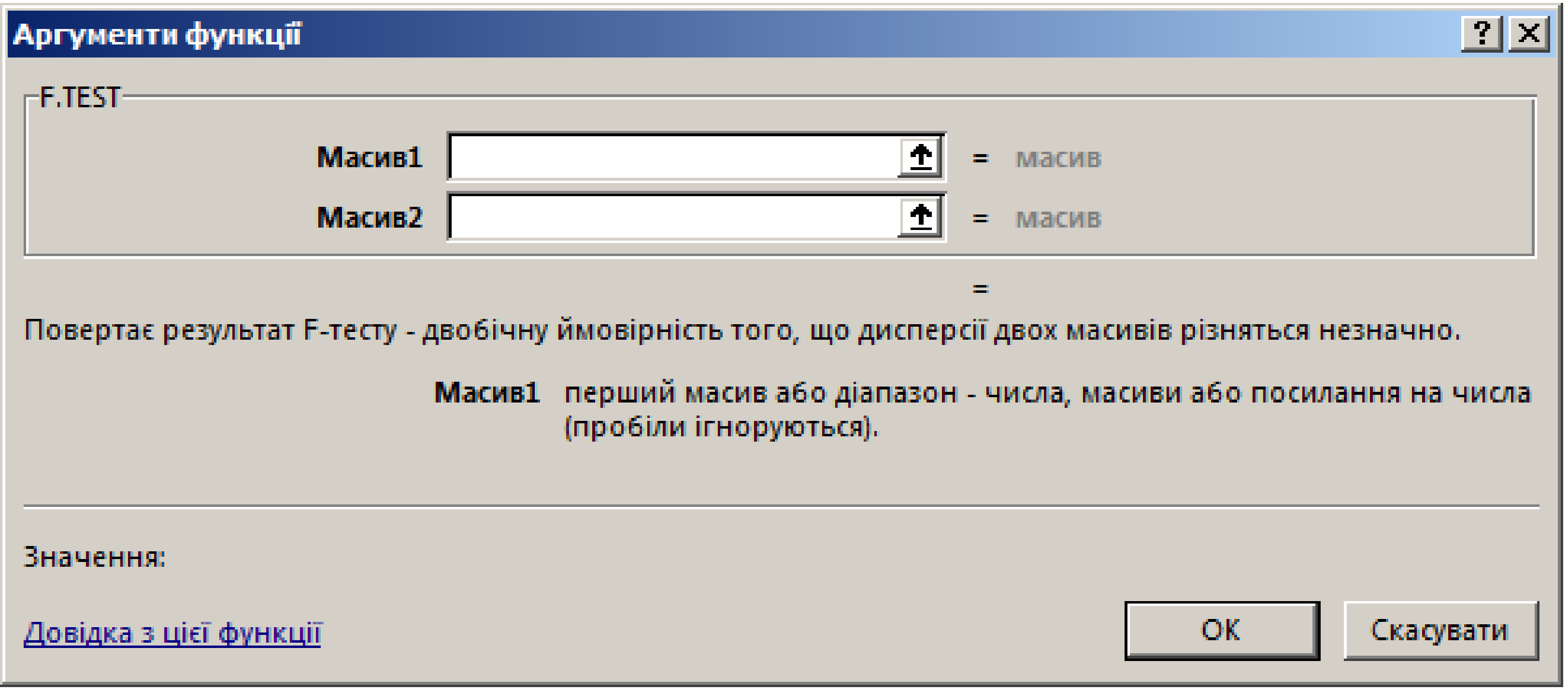
Визначивши дисперсії і середню помилку вибіркових середніх, можна обчислити фактичне значення І-критерію і порівняти його з критичним (табличним) значенням при відповідному рівні значущості і числі ступенів свободи варіації (для вибірок з чисельністю *п* > 30 використовується t-критерій нормального розподілу, а для вибірок з чисельністю *п* < 30 – і-критерій Стьюдента).



Гіпотези про дисперсії виникають достатньо часто, оскільки дисперсія характеризує такі виключно важливі показники, як точність машин, технологічних процесів, пристроїв, степінь однорідності сукупностей, ризик, пов'язаний із відхиленням дохідності активів від очікуваного рівня тощо.

F-розподіл Фішера-Снедекора

$$F = \frac{\dfrac{1}{n_1 - 1}\left[(n_1 - 1)\dfrac{\hat{s}_1^2}{\sigma^2}\right]}{\dfrac{1}{n_2 - 1}\left[(n_2 - 1)\dfrac{\hat{s}_2^2}{\sigma^2}\right]} = \frac{\hat{s}_1^2}{\hat{s}_2^2}$$



Довірчий інтервал

– для математичного сподівання або середнього

$$\varepsilon_m = \ddot{\sigma}_x \Phi^{-1}(\beta);$$

– для дисперсії

$$\varepsilon_D = \ddot{D}_x \Phi^{-1}(\beta)\sqrt{\frac{0,8N+1,2}{N(N-1)}};$$

– для відносної частоти в інтервалі гістограми

$$\varepsilon_{p_i} = \Phi^{-1}(\beta)\sqrt{\frac{k_i(1-k_i)}{N}},$$

де, $\ddot{\sigma}_m = \sqrt{\frac{D_x}{N}}$, а $Л(\beta)$– зворотне значення функції Лапласа, тобто таке значення аргументу (квантиля z), при якому функція Лапласа дорівнює β.

РЕГРЕСІЙНИЙ АНАЛІЗ

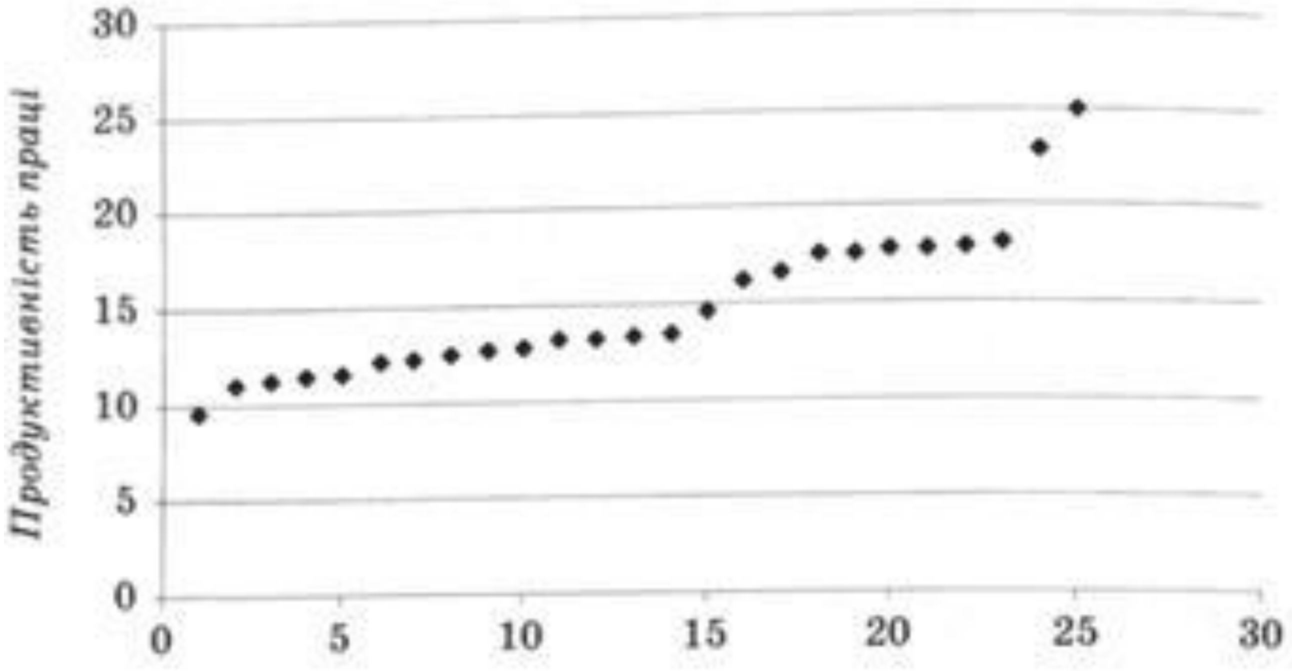
(англ. regression analysis) – це метод визначення відокремленого і спільного впливу факторів на результативну ознаку та кількісної оцінки цього впливу шляхом використання відповідних критеріїв.

Регресійний аналіз проводиться на основі побудованого рівняння регресії і визначає внесок кожної незалежної змінної у варіацію досліджуваної (прогнозованої) залежної змінної величини.

Основним завданням регресійного аналізу є визначення впливу факторів на результативний показник (в абсолютних показниках). Передусім для цього необхідно підібрати та обґрунтувати рівняння зв'язку, що відповідає характеру аналітичної стохастичної залежності між досліджуваними ознаками. Рівняння регресії показує як в середньому змінюється результативна ознака (Y) під впливом зміни факторних ознак (x_i).

$$Y_x = f(x_1, x_2, ..., x_n)$$

Паралельне зіставлення рядів значень рівнів озброєності праці основними засобами та її продуктивності, а також крапковий графік "кореляційного поля" свідчать про наявність і напрям зв'язку (прямий) між наведеними показниками. Причому зміна озброєності праці (факторної ознаки х) приводить до відносно рівномірної зміни продуктивності праці (результативної ознаки у), що видно із графіка.



Нормальна крива

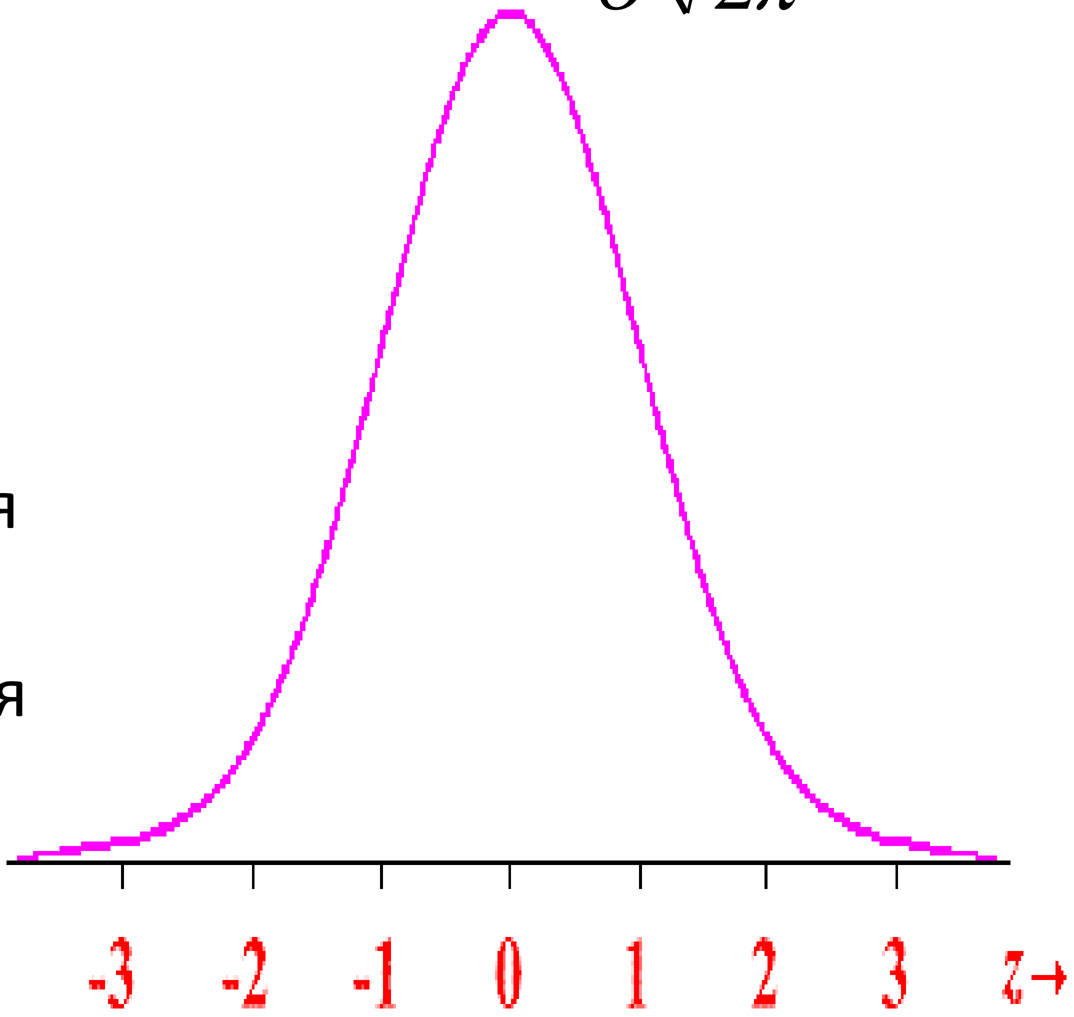
Квантиль кривої

$$z = \frac{x_i - \bar{x}}{\sigma}$$

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

Для розрахунку z-оцінок за формулою, необхідно від оцінки x_i, відняти середнє значення вибірки $\bar{x}$, а потім поділити на стандартне відхилення σ.

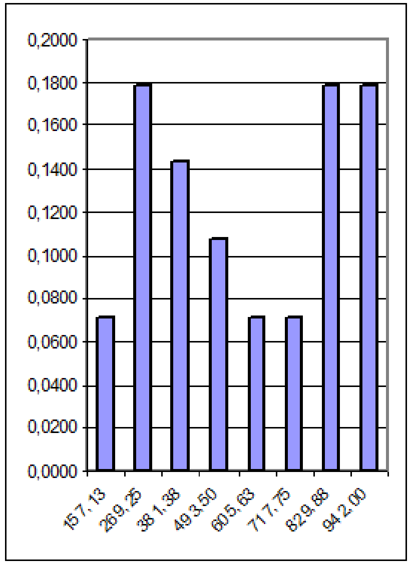
- Додатна z-оцінка означає, що значення вище середнього.
- Від’ємна z-оцінка означає, що значення знаходиться нижче середнього.

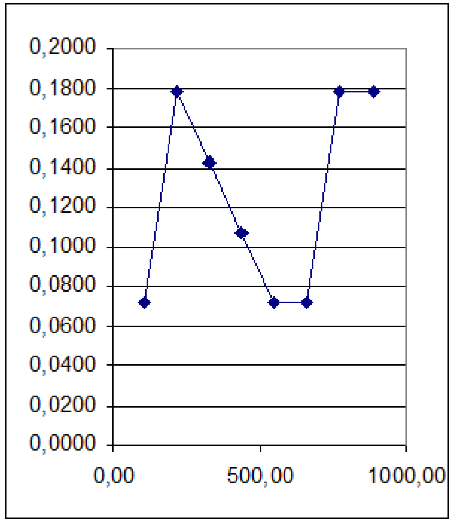


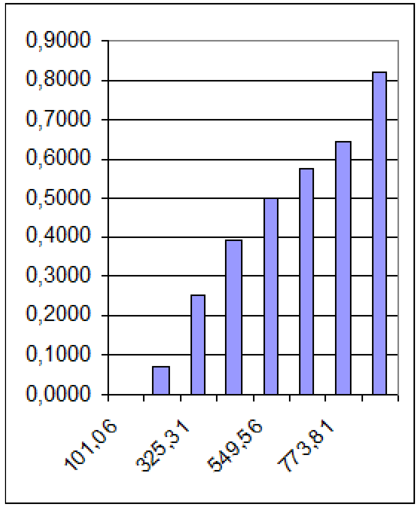
ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

$$d_{max}(i) = x_{min} + \frac{(x_{max} - x_{min}) \cdot i}{d} \qquad d_{op} = \frac{x_{max} - x_{min}}{1 + 3{,}332 \cdot \ln N}$$

$$F(d_i) = \sum_{l=0}^{i-1} k_l$$







Гістограма, Полігон Кумулята

Віднесення випадкової величини до певного закону розподілу

$$P(x_i<x<x_{i+1}) = F(x_{i+1}) - F(x_i)$$

$$\chi^2 = n\sum_{i=1}^{d}\frac{(p_i - k_i)^2}{p_i}$$

$$r = d - s - 1$$

Синтез статистичних лінійних та квазілінійних моделей

Можливі перетворення

$$y = a_0 + \sum_{i=1}^{K} a_i x_i \qquad y = a_0 \lfloor^{a_i x_i} \qquad y = a_0 \log_n x$$

$$y = a_0 + \sum_{i=-n}^{-1} a_i x^i$$

$x_1\ x_2$, $x_1/\ x_2$, x_1-x_2, log $x_1\ x_2$,

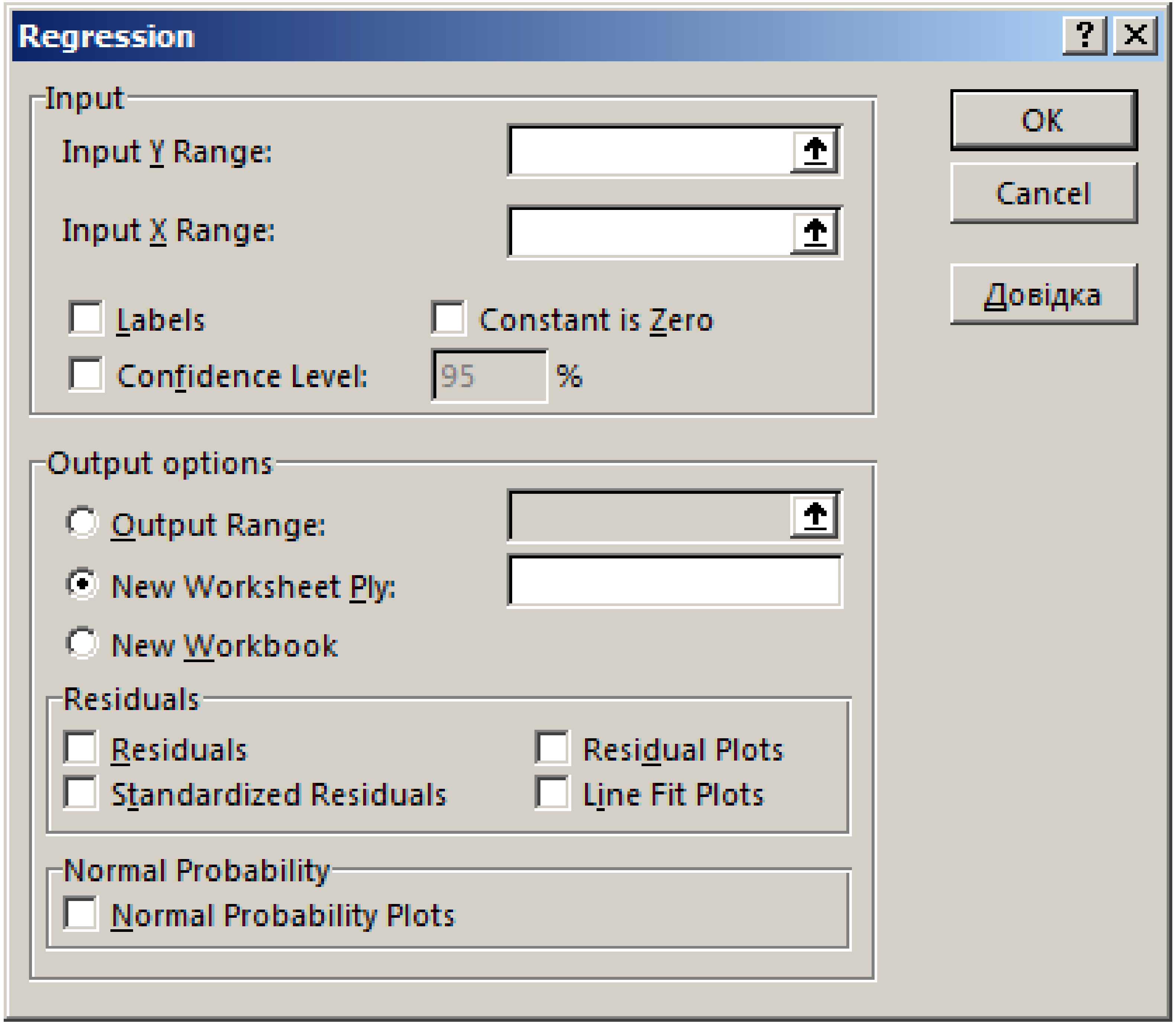
Приклад нормування-денормування

$y = a_0 + a_1 x + a_2 x^2$

$$y = \left[M_y + \sigma_y \left(a_0 - \frac{a_1 M_x}{\sigma_x} - \frac{a_2 M_{x2}}{\sigma_{x2}} \right) \right] + \frac{a_1 \sigma_y}{\sigma_x} x + \frac{a_2 \sigma_y}{\sigma_{x2}} x^2$$

Ln$y = a_0 + a_1$*Ln$x_1 + a_2$*Lnx_2

$$\mathrm{Ln}y = \left[M_y + \sigma_y \left(a_0 - \frac{a_1 M_x}{\sigma_x} - \frac{a_2 M_{x2}}{\sigma_{x2}} \right) \right] + \frac{a_1 \sigma_y}{\sigma_x} \mathrm{Ln}x_1 + \frac{a_2 \sigma_y}{\sigma_{x2}} \mathrm{Ln}x_2$$


Regression
?
Input
Input Y Range:
Input X Range:
Labels
Constant is Zero
Confidence Level:
95
%
OK
Cancel
Довідка
Output options
Output Range:
New Worksheet Ply:
New Workbook
Residuals
Residuals
Residual Plots
Standardized Residuals
Line Fit Plots
Normal Probability
Normal Probability Plots

Для оцінки значущості рівняння регресії (2) використовується ***F-тест***. Для цього виконується зрівняння фактичного $\boldsymbol{F_{факт}}$ та критичного (табличного) $\boldsymbol{F_{табл}}$ значення ***F-критерію Фішера***.

$$F_{факт} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}(n-2) = \frac{r^2}{1-r^2}(n-2)$$

Для оцінки значущості коефіцієнтів регресії b_0, b_1 використовується t-тест. Для цього зрівнюються фактичне $t_{факт}$ та критичне (табличне) $t_{табл}$ значення t-критерія Стьюдента. tфакт для коефіцієнтів b_0, b_1 визначається за наступними формулами:

$$t_{b_0} = \frac{|b_0|}{S}\sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{\overline{x^2}}}$$

Оцінка адекватності регресійної моделі робиться на підставі коефіцієнта детермінації:

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Значущість коефіцієнта кореляції

за статистикою Стьюдента. :

$$|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > t_{1-\alpha, n-2}$$

КОЕФІЦІЄНТ СПІРМЕНА

часто використовується тоді, коли дані не мають нормального розподілу або коли змінні вимірюються в порядковій шкалі.

Цей коефіцієнт обчислюється шляхом порівняння рангів відповідних спостережень для обох змінних. Він може приймати значення від -1 до 1, де:

* 1 вказує на ідеальну пряму монотонну залежність між змінними (одна збільшується, коли інша також збільшується),

* -1 вказує на ідеальну зворотну монотонну залежність (одна збільшується, коли інша зменшується),

- 0 вказує на відсутність монотонної залежності.

Коефіцієнт Спірмена є корисним інструментом в аналізі залежностей між змінними, особливо, коли точні величини вимірюються в порядковій шкалі або коли ви хочете оцінити взаємозв'язок безпосередньо за рангами.

$$\rho_{x/y} = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

$$t_\rho = \rho_{x/x}\sqrt{\frac{n-2}{1-\rho_{x/y}^2}}$$

Значення коефіцієнта кореляції вважається статистично істотним, якщо $t_\rho > t_{кр}$

ЗГЛАДЖУВАННЯ ЧАСОВОГО РЯДУ МЕТОДОМ ЗМІННОГО СЕРЕДНЬОГО

Часовий ряд (англ. time series) — це ряд точок даних, проіндексованих (або перелічених, або відкладених на графіку) в хронологічному порядку. Найчастіше часовий ряд є послідовністю, взятою на рівновіддалених точках в часі, які йдуть одна за одною. Таким чином, він є послідовністю даних дискретного часу.

Використовується для розрахунку значень у прогнозованому періоді на основі значення змінної для вказаної кількості попередніх періодів. Змінне середнє, на відміну від простого середнього для всієї вибірки, містить відомості про тенденції змінення даних. Цей метод може використовуватися для прогнозування збуту, запасів та інших процесів. Розрахунок прогнозованих значень виконується за такою формулою.

$$F_{(t+1)} = \frac{1}{N}\sum_{j=1}^{N} A_{t-j+1}$$

де:

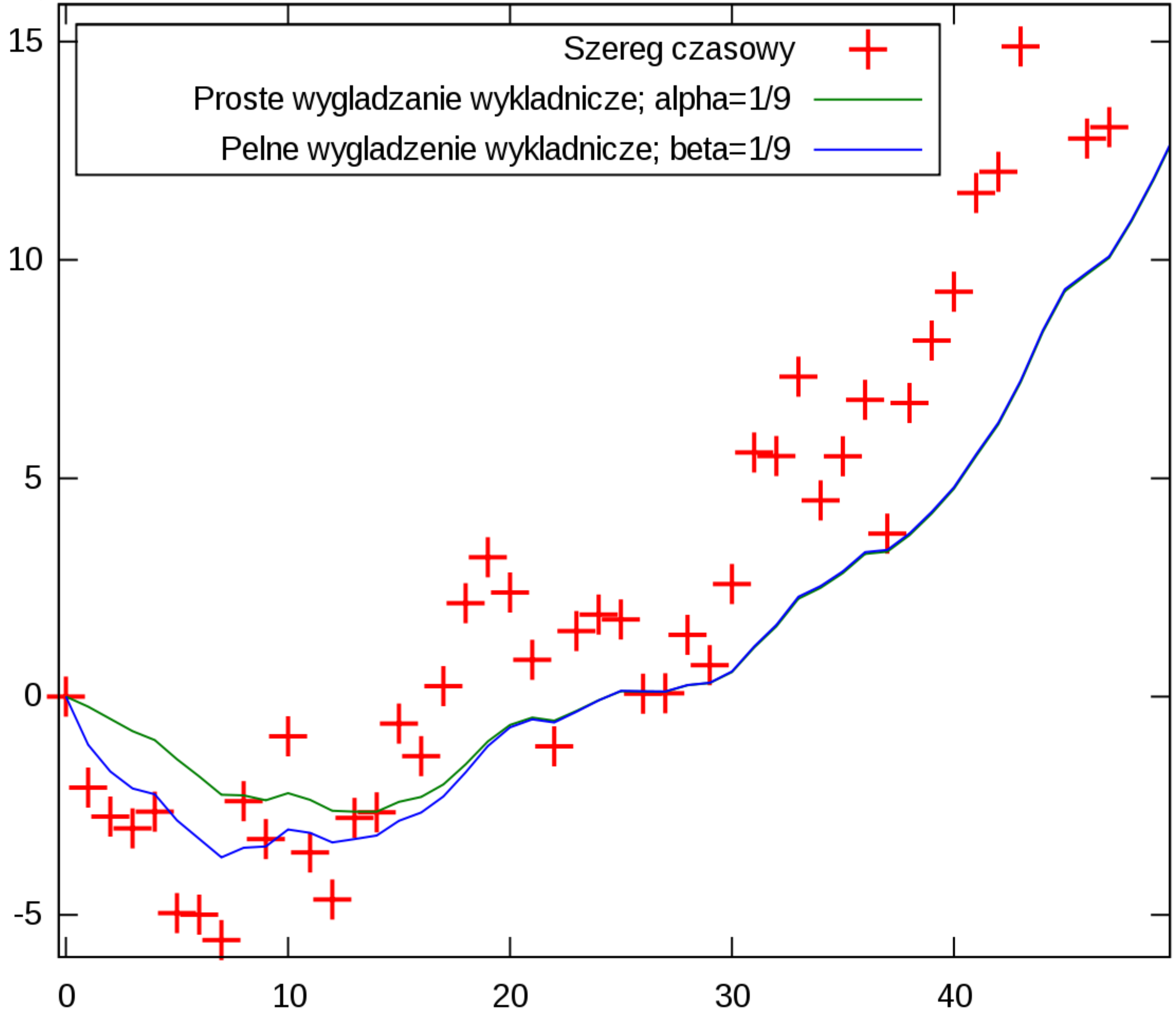
- *N* – кількість попередніх періодів, які потрібно долучити до змінного середнього;
- *A j* фактичне значення на момент часу *j*
- *F j* прогнозоване значення на момент часу *j*


Simple Moving Average (SMA)
115.0000
110.0000
105.0000
100.0000
SMA 10
SMA 30
SMA 60
08
Feb
Mar
Apr
May

ЕКСПОНЕНЦІЙНЕ ЗГЛАДЖУВАННЯ

$$y_t^* = \alpha y_t + (1-\alpha) y_{t-1}^*,$$

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



ЛІТЕРАТУРА

- Пістунов І.М. Фінансово-економічні розрахунки на персональному комп'ютері. [Електронний ресурс]: Навч. посібник/ І.М. Пістунов, І.С. Попова/ М-во освіти і науки України; Нац. Гірн. ун-т. – Д.: НГУ, 2017. – 124 с. Режим доступу: http://pistunovi.inf.ua/ FERPC.pdf (дата звернення: 17.06.2017). – Назва з екрана.
- Економічна інформатика: Навч. посібник/ Є.В. Кочура, Р.І. Лєпа, І.М. Пістунов, Т.В. Борщ, Ю.І. Рогов. - Дніпропетровськ: Освіта і наука, 2008.- 324 с.
- Моторин Р.М. Економічна статистика. КНЕУ-2004
- Мармоза А.Т. Економічна статистика. Підручник. — К.: Центр учбової літератури, 2017. — 600 с.