УДК 33.013: 004.8: 519.25 **DOI 10.5281/zenodo.3999192**

Джунь Йосип Володимирович, д.фіз-мат.н., професор, завідувач кафедри математичного моделювання, **Лотюк Юрій Григорович,** к.пед.н, доцент, доцент кафедри математичного моделювання, (Міжнародний економіко-гуманітарний університет імені академіка Степана Дем'янчука, м. Рівне), josif-june@rambler.ru, lotyuk@ukr.net.

РЕГРЕСІЙНИЙ АНАЛІЗ У ПЕДАГОГІЧНИХ ДОСЛІДЖЕННЯХ З МАТРИЧНИМИ РОЗРАХУНКАМИ ПОХИБОК ПАРАМЕТРІВ МОДЕЛІ

Анотація. *У статті розкрито можливість використання регресійного аналізу у педагогічних дослідженнях з матричними розрахунками похибок параметрів моделі. Визначено, що регресійне моделювання можна виконати належним чином тільки при розумінні трьох основних причин, що зумовлюють незадовільну якість та відсутність наукового результату в педагогічному дослідженні. Обґрунтовано, що математична статистика формує доказову базу досліджень більшості наукових робіт, включаючи дисертації. Для забезпечення якості статистичних досліджень у педагогіці рекомендовано проводити їх групою у складі: педагог-дослідник, математик-статистик та програміст.*

Ключові слова: *регресійний аналіз, стандартні похибки регресорів, залишкові похибки математичної моделі, метод найменших квадратів.*

Dzhun Yosyp Volodymyrovych, Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of Mathematical Modeling, **Lotiuk Yurii Heorhiiovych**, Ph.D in Pedagogic Sciences, Associate Professor, Associate Professor of the Department of Mathematical Modeling (Academican Stepan Demianchuk International University of Economics and Humanities, Rivne), josif-june@rambler.ru, lotyuk@ukr.net.

REGRESSION ANALYSIS IN PEDAGOGICAL RESEARCH WITH MATRIX CALCULATIONS OF THE ERRORS OF MODEL PARAMETERS

Abstract.

Introduction. *The main reasons for the low quality of pedagogical research are discussed in the article.*

The first is a very rough idea of teachers about the importance of mathematical statistics in scientific research. The standard definition of mathematical statistics as a science that studies the methods of collecting, systematizing, processing

and interpreting statistical data, their use for scientific conclusions does not fully correspond to its essence, which consists in the fact that mathematical statistics is the main working mathematical tool of a researcher and was created for this. Mathematical statistics form the evidence base for research in most scientific papers, including dissertations.

The second significant reason for the incorrect application of regression analysis is the lack of understanding of the difference between active and passive experiments, which is caused by the unscientific transfer of modelling methods developed for physical phenomena to other industries, in particular, pedagogical research, where research objects are more complex and much less controlled.

The third reason for poor-quality regression modelling in pedagogy is the lack of an estimate of the standard error of the obtained parameters. This allows excluding significant parameters and significantly simplifying the model, it is important in the passive experiments.

***Purpose.** The purpose is to show the main reasons for the incorrect application of regression analysis in pedagogical research and methods for overcoming them. In order to ensure the elimination of the causes of poor-quality modelling, it is recommended to conduct it with the participation of three specialists: a research teacher, mathematician-statistician, and a programmer.*

***Methods.** In order to assess the reliability of regression modelling, a matrix method for calculating the standard errors of the model parameters is proposed. It has a simple algorithm, so it is well understood and easy to use. Recommendations on the application of the method for calculating the marginal errors of the regresses and formulas have been developed to eliminate the factor signs whose action is at the noise level. Such an operation can significantly simplify the regression model in a passive experiment.*

***Results.** The importance of understanding the first, second and third reasons noted in Introduction, the incorrect analysis of statistical data is shown. The main emphasis of the article is aimed at understanding the importance of mathematical statistics as a decisive tool and evidence base for a researcher. A simple algorithm for estimating the parameters of the regression model and its accuracy is proposed. It is recommended to verify the linearity of the model by constructing correlation fields of the effective and factor attributes before using a computer program of regression analysis. The importance of checking the diagonality of the matrix of factor signs is noted. It is recommended to verify the hypothesis of normality of residual model errors after completion of the regression analysis. Confirmation of this hypothesis will indicate a correct and final decision. The rejection of the normality hypothesis of residual errors in the regression model is the evidence that the obtained parameter estimates are not effective. It is necessary to turn to the methods of the non-classical theory of errors with a purpose to obtain effective estimates in this case.*

Originality. *A simple regression analysis algorithm with calculation of standard model errors based on the application of matrix theory is proposed. For the first time, all the necessary mathematical requirements are indicated, the fulfilment of which is mandatory for the proper application of regression analysis. Knowledge of these requirements is very important for the correct application of software products, since programmers, as a rule, are not aware of the presence of these requirements.*

Conclusion. *The first thing that is necessary for the correct conduct of regression analysis and ensuring the correct conclusions is the knowledge of three reasons for poor-quality modelling mentioned above. It is also recommended to abandon the Gaussian least squares schemes, which are rather cumbersome and specific, and move on to simple, understandable, and universal matrix calculations. It is recommended to observe all of the specified mathematical conditions for the proper application of regression analysis.*

Key words: *regression analysis, standards errors of the regresses, observations-calculations errors of the mathematical models, least squares method.*

Існують три основні причини, які обумовлюють незадовільну якість наукового результату педагогічного статистичного дослідження. Перша з них – нечітке уявлення про значення математичної статистики в педагогічних та психологічних дослідженнях. В [1] сказано, що «математична статистика це розділ математики, присвячений методам збору, систематизації, обробки і інтерпретації статистичних даних, а також їх використання для наукових висновків. Процедури математичної статистики опираються на теорію ймовірностей, яка дозволяє оцінити точність і надійність висновків, отриманих в кожній задачі на основі наявного статистичного матеріалу». Ця цитата стисла і чітка, проте сказана не цілком по суті. А суть полягає в тому, що математична статистика є головним робочим інструментом дослідника і створена для цього. А це міняє акценти. Головне в тому, що саме математична статистика, а не щось інше, формує, як правило, доказову базу дослідження наукових робіт, включаючи дисертації.

Роль математичної статистики в проведенні наукових досліджень вивчали відомі зарубіжні і вітчизняні науковці. Так, Л. О. Атраментова наголошує, що без математичної статистики, як правило, дисертація не має доказового обґрунтування [2]. Загальний аналіз проблем математичного моделювання в педагогічних дослідженнях здійснено Й. В. Джунем [3].

М. Б. Малютов і А. Ю. Заіграєв розглянули сучасні завдання оптимального планування регресивних експериментів у педагогіці [4].

Метою нашої статті є розкриття можливості використання регресійного аналізу у педагогічних дослідженнях з матричними розрахунками похибок параметрів моделі.

Наукові дослідження вимагають наявності достовірних статистичних даних, без яких неможливо отримати об'єктивні результати. Другою причиною незадовільної якості наукового результату при проведенні статистичних педагогічних досліджень є нерозуміння різниці між активним і пасивним експериментами при проведенні регресійного аналізу. Це питання детально розглянуто в роботі [3]. Пасивний характер педагогічних експериментів означає довільну трактовку факторних границь. На практиці це означає, що дослідник може не врахувати дію певних факторів, а неможливість змінювати силу дії тих факторів, які враховані, зводить дію деяких з них до рівня шумів. Третьою важливою причиною неякісного моделювання є невміння оцінювати стандартні похибки отриманих регресорів. Ця операція дозволяє відсіяти незначимі регресійні коефіцієнти і істотно спростити саму модель, особливо при реалізації пасивних експериментів. Є також інші, менш значні причини непрофесійного проведення досліджень в педагогіці, наприклад, невміння їх оптимізувати, як це рекомендовано в [4].

Існує одне дієве правило, яке дозволяє уникнути помилок в регресійному моделюванні і провести його на належному рівні. Це правило «трійці» яке використовують завжди при проведенні міжнародних експериментів, наприклад, при проведенні лазерної локації штучних супутників Землі по проекту MERIT (Monitoring Earth Rotation and Intercom paring the Techniques of observation and analysis), у якому прийняли участь 22 країни, включаючи УРСР [5; 6, с. 279]. Моделювання є професійним тільки в тому разі, коли воно виконується «трійцею»: педагог-дослідник, програміст, математик-статистик (який обов'язково мусить мати досвід дослідницької роботи).

Слід зауважити, що на перших етапах впровадження комп'ютерних технологій у наукові дослідження не було розділення праці між математиками і програмістами. Згадаємо знамениті механіко-математичні факультети, де кожен програміст об'єктивно був і математиком, а кожен математик – професіоналом-програмістом. Адже тоді програмування було спрощенням реалізації тих чи інших математичних ідей, які є завжди суттю досліджень. Потім відбулося розділення праці: програмісти стали механічно і масово програмувати різноманітні математичні процедури, іноді без уявлення про те, в яких випадках ці процедури можна застосовувати, не знаючи, коли ці процедури є адекватними, а коли ні, виконуючи чисто технічну роботу.

Таке розділення праці фактично стало головною причиною двох фінансових катастроф, яскраво описаних в знаменитій книзі Н. Н. Талеба [7]. В ній автор описує причину того, чому не спрацювала методика прогнозування вартості опціонів, яка розроблена нобелівськими лауреатами Р. Мертоном і М. Шоулзом [8] та яка призвела до краху інвестиційну корпорацію в США LTKM влітку 1998 року з обвалом у 19 мільярдів доларів, що створило тяжку ситуацію для фінансової системи країни. Викликає здивування, чому ці лауреати, опираючись на закон похибок Гауса, який дійсно в багатьох випадках себе блискуче зарекомендував, не побажали проконсультуватися з цього приводу із знаменитим математиком-статистиком США Джоном Тьюки, який у своїх роботах [9–11] застерігав від уявлень про всезагальність нормального закону, а в роботах [8; 9] цитує слова А. К. Джері: «нормальність це міф, в реальному світі ніколи не було і ніколи не буде нормального розподілу» [12]. Математики, відокремившись від програмістів, створюють нові математичні методи обробки статистичної інформації, часто не здогадуючись чітко визначити і жорстко підкреслити важливість виконання тих фундаментальних принципів, які саме і покладені в основу і які мають бути адекватні реальності. Програмісти, алгоритмізуючи ці нові методи, часто не здогадувались зазначити умови, за яких їх можна застосовувати і практично реалізувати. Сучасні університети готують ІТ фахівців, які не знають математику і мають просто, так би мовити «технічно-кнопочну» підготовку. Навіть комп'ютерні «генії» нині не мають найменшого уявлення про адекватність деяких математичних процедур в кожному конкретному випадку, які вони програмують, видаючи замовникові безліч статистичних показників за масою програм. А замовник не знає, що з цими показниками робити і використовує деякі з них лише тому, що на багатьох відповідальних наукових працівників, від яких залежить доля дисертації, і які не мають фізико-математичної освіти, слова: «застосовано сучасний програмний продукт», справляють магічну дію, що також відмічає експерт ДАК в роботі [2].

Мова про математичну адекватність застосованої програми часто не йде, хоча саме це визначає бездоганність проведення дослідження. Для педагогів пропонується такий тест для перевірки професійності програміста. Спитайте його, «які математичні умови адекватного застосування найпростішої програми по критерію Стьюдента?». Правильна відповідь: «нормальність розподілу і рівність дисперсій двох груп». Якщо програміст дасть звичайну відповідь, що про цю вимогу не сказано нічого в програмі, то програмний продукт такого автора може «підтвердити» будь-яку наукову нісенітницю.

Зазначимо, що усі відмічені вище проблеми є досить актуальними в педагогічних дослідженнях, особливо при реалізації регресійного аналізу, який ретельно розглянутий у вітчизняних [12] та фундаментальних зарубіжних працях [14–17]. Проте, ці праці спеціально не призначені для педагогів чи програмістів, що і викликає певні труднощі їх використання. Після знакової фундаментальні роботи [18] пройшло вже 44 роки і до цього часу для педагогів і психологів ще не випущено нового україномовного підручника такого ж рівня, настільки ж виразного і з врахуванням усіх досягнень в математичній статистиці на той час і таким же рівнем ерудиції. В роботі [18] є посилання на 263 джерела із статистичними процедурами та описом основних з них – праця неймовірна за обсягом навіть в наш час.

Перейдемо тепер до суті нашого дослідження. Як відомо, загальна математична формула регресійного аналізу має вигляд:

$$y_i = a + bx_{1i} + cx_{2i} + ... + zx_{pi} + e_i \quad , \tag{1}$$

де y_i – експериментальні значення результативної ознаки; $a,b,c,...,z$ – параметри регресійної моделі, визначення яких і є метою дослідження; $x_{1i}, x_{2i}, ..., x_{pi}, (j = 1,2,..., p)$ – значення факторних ознак $x_{j,i}$; e_i залишкові похибки регресійної моделі:

$$e_i = y_i - Y_i \ , \tag{2}$$

де Y_i – теоретичні значення результативної (досліджуваної) ознаки:

$$Y_i = a + bx_{1i} + cx_{2i} + ... + zx_{pi} \tag{3}$$

Регресори $a,b,c,...,z$ визначаються методом найменших квадратів за умови:

$$f = \sum_{i=1}^{n} \left(y_i - Y_i \right)^2 = \sum_{i=1}^{n} e_i^2 = min \tag{4}$$

Вимога (4) дає можливість отримати найкращі класичні оцінки регресорів, але тільки за умови нормальності значень e_i .

Підставимо в (4) значення Y_i з формули (3):

$$f = \sum_{i=1}^{n} \left(y_i - a - bx_{1i} - cx_{2i} - ... - zx_{pi} \right)^2 = min \tag{5}$$

Щоб знайти мінімум функції продиференціюємо її за невідомими аргументами:

$$\left. \begin{aligned} \frac{df}{da} &= 2\sum_{i=1}^{n} \left(y_i - a - bx_{1i} - cx_{2i} - ... - zx_{pi} \right)(-1) = 0; \\ \frac{df}{db} &= 2\sum_{i=1}^{n} \left(y_i - a - bx_{1i} - cx_{2i} - ... - zx_{pi} \right)(-x_{1i}) = 0; \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \frac{df}{dz} &= 2\sum_{i=1}^{n} \left(y_i - a - bx_{1i} - cx_{2i} - ... - zx_{pi} \right)(-x_{pi}) = 0. \end{aligned} \right\} \tag{6}$$

Виконуючи в (6) множення і розкриваючи знак суми, отримуємо таку систему рівнянь для отримання невідомих значень регресорів

$$\left.\begin{array}{llllllllllll} an & + & b\sum_{i=1}^{n} x_{1i} & + & c\sum_{i=1}^{n} x_{2i} & + & ... & + & z\sum_{i=1}^{n} x_{pi} & = & \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_{1i} & + & b\sum_{i=1}^{n} x_{1i}^2 & + & c\sum_{i=1}^{n} x_{1i}x_{2i} & + & ... & + & z\sum_{i=1}^{n} x_{1i}x_{pi} & = & \sum_{i=1}^{n} y_i x_{1i} \\ .... & ... & .... & ... & .... & ... & ... & ... & .... & ... & .... \\ \sum_{i=1}^{n} x_{pi} & + & b\sum_{i=1}^{n} x_{1i}x_{pi} & + & c\sum_{i=1}^{n} x_{2i}x_{pi} & + & ... & + & z\sum_{i=1}^{n} x_{pi}^2 & = & \sum_{i=1}^{n} y_i x_{pi} \end{array}\right\} \quad (7)$$

Рівняння (7) називаються системою нормальних рівнянь методу найменших квадратів для визначення невідомих $a,b,c,...,z.$ Запишемо розв’язок цих рівнянь в матричній формі. Позначимо матриці невідомі через $D_a, D_b, D_c, ..., D_z$. Тоді:

$$D_a = \begin{vmatrix} \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & ... & \sum_{i=1}^{n} x_{pi} \\ \sum_{i=1}^{n} y_i x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{2i}x_{1i} & ... & \sum_{i=1}^{n} x_{pi}x_{1i} \\ .... & .... & .... & .. & .... \\ \sum_{i=1}^{n} y_i x_{pi} & \sum_{i=1}^{n} x_{1i}x_{pi} & \sum_{i=1}^{n} x_{2i}x_{pi} & ... & \sum_{i=1}^{n} x_{pi}^2 \end{vmatrix} \quad (8)$$

$$D_b = \begin{vmatrix} n & \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_{2i} & ... & \sum_{i=1}^{n} x_{pi} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} y_i x_{1i} & \sum_{i=1}^{n} x_{2i}x_{1i} & ... & \sum_{i=1}^{n} x_{pi}x_{1i} \\ .... & .... & .... & ... & .... \\ \sum_{i=1}^{n} x_{pi} & \sum_{i=1}^{n} y_i x_{pi} & \sum_{i=1}^{n} x_{2i}x_{pi} & ... & \sum_{i=1}^{n} x_{pi}^2 \end{vmatrix} \quad (9)$$

$$D_c = \begin{vmatrix} n & \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_{2i} & ... & \sum_{i=1}^{n} x_{pi} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{2i}x_{1i} & ... & \sum_{i=1}^{n} x_{pi}x_{1i} \\ .... & .... & .... & ... & .... \\ \sum_{i=1}^{n} x_{pi} & \sum_{i=1}^{n} y_i x_{pi} & \sum_{i=1}^{n} x_{2i}x_{pi} ... & ... & \sum_{i=1}^{n} x_{pi}^2 \end{vmatrix} \quad (10)$$

$$
D_z = \begin{vmatrix}
n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} y_i \\
\sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{2i} x_{1i} & \cdots & \sum_{i=1}^{n} y_i x_{1i} \\
\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\
\sum_{i=1}^{n} x_{pi} & \sum_{i=1}^{n} x_{1i} x_{pi} & \sum_{i=1}^{n} x_{2i} x_{pi} \cdots & \cdots & \sum_{i=1}^{n} y_i x_{pi}^2
\end{vmatrix} \qquad (11)
$$

Детермінант системи рівнянь (7):

$$
D = \begin{vmatrix}
n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} x_{pi} \\
\sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{2i} x_{1i} & \cdots & \sum_{i=1}^{n} x_{pi} x_{1i} \\
\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\
\sum_{i=1}^{n} x_{pi} & \sum_{i=1}^{n} x_{1i} x_{pi} & \sum_{i=1}^{n} x_{2i} x_{pi} \cdots & \cdots & \sum_{i=1}^{n} x_{pi}^2
\end{vmatrix} \qquad (12)
$$

Розв’язок системи нормальних рівнянь (7) в матричній формі є таким:

$$
a = \frac{D_a}{D}; \quad b = \frac{D_b}{D}; \quad c = \frac{D_c}{D}; \quad \ldots; \quad z = \frac{D_z}{D}. \qquad (13)
$$

Матричний підхід в регресійному моделюванні в належному обсязі розглянуто в підручнику [19] Большакова В. Д., колишнього голови Всесоюзної акредитаційної комісії ВАК СРСР.

Після того, як обчислені регресори (13), регресійна модель (3) набуває робочого вигляду, оскільки відомі усі її параметри. На цьому найчастіше, і на превеликий жаль, закінчується регресійний аналіз. Але такий аналіз є спрощеним, неповним, оскільки в пасивному експерименті дію значної частини регресорів можна віднести до шумового поля.

Ці регресори значимого впливу не мають і лише ускладнюють модель. Тому важливо її відсіяти. Але для того, щоб це зробити такий відсів необхідно обчислити стандартні похибки визначених нами по формулі (13) регресорів. Обчислення цих похибок відбувається в три етапи.

На *першому етапі* обчислюють залишкові похибки e_i за формулою (2), де y_i – результати спостережень, а Y_i – отримують за формулою (3) використовуючи параметри моделі (13).

Використовуючи значення e_i вираховують стандартну похибку вимірів по формулі:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-k}}\,, \tag{14}$$

де n – кількість вимірів; k – кількість регресорів.

На *другому етапі* обчислюють вагові коефіцієнти регресорів:

$$Q_{11} = \frac{A_{11}}{D}; \quad Q_{22} = \frac{A_{22}}{D}; \quad \ldots; \quad Q_{pp} = \frac{A_{pp}}{D}\,, \tag{15}$$

де A мінори детермінанту D.

На останньому етапі вираховують стандартні похибки регресорів:

$$\sigma_a = \sigma\sqrt{Q_{11}}; \quad \sigma_b = \sigma\sqrt{Q_{22}}; \quad \ldots; \quad \sigma_p = \sigma\sqrt{Q_{pp}}\,. \tag{16}$$

Гранична похибка Δ_j регресора j при ризику α визначається так:

$$\Delta_j = \sigma_j t_\alpha\,, \tag{17}$$

де квантиль t_α – аргумент функції Лапласа $\Phi(t_\alpha)$, який визначають для заданого α по формулі:

$$\Phi(t_\alpha) = \frac{1-\alpha}{2} \tag{18}$$

Наприклад, при ризику $\alpha = 0{,}05$ маємо:

$$\Phi(t_\alpha) = \frac{1-0{,}05}{2} = \frac{0{,}95}{2} = 0{,}47500 \tag{19}$$

По таблицям функції Лапласа знаходимо, що значення 0,47500 функція $\Phi(t_\alpha)$ набуває при $t = 1{,}96$.

При рівні ризику $\alpha = 0{,}05$ гранична похибка (17) набуває вигляду:

$$\Delta_j = \sigma_j * 1{,}96\,. \tag{20}$$

У тому разі, коли якийсь регресор, наприклад:

$$(b) \le \Delta_b, \qquad (21)$$

то ним можна знехтувати для спрощення моделі. Діючим можна вважати лише той регресор, у якого, наприклад:

$$|b| > \Delta_b. \qquad (22)$$

Після перевірки усіх регресорів на значущість, вилучаємо з моделі (3) ті з них, у яких $\Delta_j \le \sigma_j t_\alpha$, після чого модель розраховуємо заново.

Розглянемо приклад реалізації регресійної моделі з оцінкою стандартних похибок регресорів. В табл. 1 наведені такі вихідні дані для обчислення: x_i – коефіцієнт фондоозброєності комп'ютерної школи для менеджерів; y_i – середня заробітна плата викладача школи. Діаграма розсіювання залежностей пар $x_j \to y_i$ підтвердила їх лінійний характер.

Для дослідження взято кількість спостережень $n = 32$, оскільки регресійні оцінки параметрів моделі набувають статистичного характеру за умови $n > 30$. Зауважуємо, що приклади регресійного моделювання в підручниках, коли $n > 30$, не є свідченням високої їх якості.

Отже регресійна модель для вихідних даних має такий вигляд:

$$Y_i = a + bx_i \qquad (23)$$

У відповідності з формулами (8), (9), (12), і (13), знаходимо невідомі параметри a і b по формулам:

$$a = \frac{D_a}{D} = \frac{\begin{pmatrix} \sum y_i & \sum x_i \\ \sum y_i x_i & \sum x_i^2 \end{pmatrix}}{\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}} = \frac{\sum y_i \sum x_i^2 - \sum (y_i x_i) \sum x_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} \qquad (24)$$

Підставивши в (24) значення елементів матриць, обчислені в таблиці, отримаємо:

$$a = \frac{456{,}83 * 78358{,}2579 - 23081{,}6544 * 1400{,}53}{32 * 78358{,}2579 - 1400{,}53^2} = 6{,}355276\,; \qquad (25)$$

$$b = \frac{D_b}{D} = \frac{\begin{pmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{pmatrix}}{D} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} \qquad (26)$$

Таблиця 1

Приклад реалізації двовимірної регресійної моделі з матричною оцінкою стандартних похибок регресорів

Вихідні дані			Регресійні обчислення					
i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	Y_i	$e_i = y_i - Y_i$	e_i^2
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	5,00	6,00	30,0000	25,0000	36,0000	7,2602	-1,2602	1,588104
2	7,47	7,50	56,0250	55,8009	56,2500	7,7072	-0,2072	0,042932
3	9,90	5,90	58,4100	98,0100	34,8100	8,1469	-2,2469	5,048560
4	12,60	9,10	114,6600	158,7600	82,8100	8,6356	0,4644	0,215667
5	15,15	6,14	93,0210	229,5225	37,6996	9,0971	-2,9571	8,744440
6	17,55	11,60	203,5800	308,002	134,5600	9,5314	2,0686	4,279106
7	19,99	8,08	161,5192	399,6001	65,2864	9,9730	-1,8930	3,583449
8	22,39	11,33	253,6787	501,3121	128,3689	10,4073	0,9227	0,851375
9	25,15	9,95	250,2425	632,5225	99,0025	10,9068	-0,9727	0,946145
10	27,50	16,00	440,0000	756,2500	256,0000	11,3321	4,6678	21,789290
11	29,89	9,00	269,0100	893,4121	81,0000	11,7646	-2,7646	7,643013
12	32,40	13,83	448,0920	1049,7600	191,1689	12,2189	1,6111	2,598643
13	35,05	10,99	385,1995	1228,5025	120,7801	12,6985	-1,7085	2,918972
14	37,38	19,00	710,2200	1397,2644	361,0000	13,1201	5,8799	34,573224
15	39,92	14,07	561,6744	1593,6064	197,9649	13,5798	0,4902	0,240296
16	42,50	10,90	463,2500	1806,2500	118,8100	14,0467	-3,1467	9,901721
17	45,12	20,00	902,4000	2035,8144	400,0000	14,5209	5,4791	30,020537
18	47,47	11,00	522,1700	2253,4009	121,0000	14,9462	-3,9462	15,572494
19	50,51	16,09	812,7059	2551,2601	258,8881	15,4963	0,5937	0,352480
20	52,52	14,05	737,9060	2758,3504	197,4025	15,8601	-1,8101	3,276462
21	54,87	16,06	881,2122	3010,7169	257,9236	16,2854	-0,2254	0,050805
22	57,59	21,00	1209,3900	3316,6091	441,0000	16,7776	4,2224	17,828662
23	60,00	15,00	900,0000	3600,0000	225,0000	17,2138	-2,2138	4,900910
24	62,60	20,88	1307,0880	3918,7600	435,9744	17,6843	3,1957	10,212498
25	65,05	18,21	1184,5605	4231,5025	331,6041	18,1277	0,0823	0,006773
26	67,65	20,00	1353,0000	4576,5225	400,0000	18,5982	1,4018	1,965043
27	69,69	18,00	1254,4200	4856,6961	324,0000	18,9674	0,9674	0,935863
28	72,35	20,00	1447,0000	5234,5225	400,000	19,4488	0,5512	0,303821
29	74,93	21,15	1584,7695	5614,5049	447,3225	19,9157	1,2343	1,523496
30	77,76	19,00	1477,4400	6046,6176	361,0000	20,4279	-1,4279	2,038898
31	80,01	18,00	1440,1800	6401,6001	324,0000	20,8351	-2,8351	8,037792
32	82,57	19,00	1568,8300	6817,8049	361,0000	21,2984	-2,2984	5,282643
$\sum 1400,53$ $\overline{x_j} = 43,7665625$		$\sum 456,83$ $\overline{y_i} = 43,76656$	$\sum x_i * y_i = 23081,65$	$\sum x_i^2 = 78358,257$	$\sum y_i^2 = 7287,7265$		-32,8812 32,9352	207,27114

$$b = \frac{32 * 23081,6544 - 1400,53 * 456,83}{32 * 78358,2579 - 1400,53^2} = 0,180975 \qquad (27)$$

Обчислення стандартних похибок регресорів a і b виконується в три етапи.

Етап А. Обчислюємо по формулі (14) стандартну похибку вимірів, використовуючи дані в таблиці:

$$\sigma = \sqrt{\frac{\sum e_i^2}{n-k}} = \sqrt{\frac{207{,}271114}{32-2}} = 2{,}628505\,. \qquad (28)$$

Етап B. По формулам (15) знаходимо вагові коефіцієнти регресорів:

$$Q_{11} = \frac{\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}}{\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}} = \frac{A_{11}}{D} = \frac{\sum x_i^2}{D} = \frac{78358{,}2579}{545979{,}9721} = 0{,}143519 \qquad (29)$$

$$Q_{22} = \frac{\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}}{\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}} = \frac{A_{22}}{D} = \frac{n}{D} = \frac{32}{545979{,}9721} = 0{,}00005861 \qquad (30)$$

Етап C. Обчислюємо стандартні похибки регресорів a і b за формулами (16):

$$\sigma_a = \sigma * \sqrt{Q_{11}} = 2{,}628505 * \sqrt{0{,}143519} = 0{,}99578\ ; \qquad (31)$$

$$\sigma_b = \sigma * \sqrt{Q_{22}} = 2{,}628505 * \sqrt{0{,}00005861} = 0{,}0120123\,. \qquad (32)$$

Використовуючи значення (31) і (32) записуємо регресійну модель (23) у кінцевому вигляді:

$$\left| \begin{array}{ccccccccc} Y_i & = & a & + & bx_i & = & 6{,}35528 & + & 0{,}180975x_i \\ & & \pm\sigma_a & & \pm\sigma_b & & \pm 0{,}99578 & & \pm 0{,}020123 \end{array} \right. \qquad (33)$$

Так подають результати регресійного аналізу у ядерній фізиці, квантовій фізиці, космічних дослідженнях тощо. Таке подання є професійним.

Висновок. Усвідомлення першої, другої і третьої із зазначених вище причин некоректного аналізу, сприяють належній якості регресійного моделювання в педагогічних дослідженнях. Запропонований нами математичний апарат, а головне, розуміння значення математичної статистики, як вирішального інструмента і доказової бази дослідження є головними складовими успішного захисту дисертації. Рекомендуємо також перед застосуванням регресійного аналізу будувати кореляційні поля пар значень $y_i \to x_{ji},\ \ j = 1,2,...,p$ з метою перевірки лінійності моделі. Бажано, щоб кореляційні поля пар залежностей $x_{1i} \to y_i, x_{2i} \to y_i, ..., x_{pi} \to y_i$ мали вигляд еліпсів. Важливо перед застосуванням регресійного аналізу перевірити діагональність кореляційної матриці факторних ознак. Недіагональні елементи цієї матриці мають дорівнювати нулю, що необхідно перевірити за допомогою z-перетворення Фішера для коефіцієнтів кореляції. Рекомендуємо також, після завершення регресійного моделювання, тестувати на нормальність залишкові похибки e_i . Якщо результат такої перевірки буде позитивним, то можна вважати отримані параметри кінцевими. В тому разі, коли гіпотеза нормальності для залишкових похибок e_i буде відхилена, то результати регресійного аналізу не можна вважати остаточними. Щоб зрозуміти чому це так необхідно звернутися до фундаментальної праці знаменитого професора Кембріджського університету Г. Джеффріса [20], яка витримала у Великобританії 9 перевидань. Для отримання остаточних результатів регресійного аналізу у випадку негаусового характеру залишкових похибок e_i , необхідно застосувати методи некласичної теорії похибок, розглянутих в розробках кафедри математичного моделювання Міжнародного економіко-гуманітарного університету імені академіка С. Дем'янчука [21; 22].

Наведені дані засвідчують, що регресійний аналіз має дуже важливі умови свого застосування, які забезпечують його якість і надійність, про що часто не повідомляють підручники з математичної статистики. Ні дослідники-педагоги, ні програмісти цими умовами зазвичай не цікавляться, і як правило, на превеликий жаль, їх не знають.

Список використаних літературних джерел

1. Вероятность и математическая статистика: Энциклопедия. Гл. ред. Ю. В. Прохоров. М.: Большая Российская энциклопедия, 1999. 910 с.

2. Атраментова Л. О. Наукове дослідження і статистика. *Науковий світ.* 2006. № 4, С. 6–7.

3. Джунь Й. В. Загальний аналіз проблем математичного моделювання в педагогічних дослідженнях. *Психолого-педагогічні основи гуманізації навчально-виховного процесу в школі та ВНЗ: збірник наукових праць,* № 1 (23). Рівне: РВЦ МЕГУ ім. акад. С. Дем'янчука, 2020. С. 28–39.

4. Малютов М. Б., Заиграев А. Ю. Современные задачи оптимального планирования регрессионных экспериментов. Киев: Выща школа, 1989. 64 с.

5. Dzhun I. V. Pearson's Distribution of type VII of the Errors of Satellite Laser Ranging Data. *Kinematic and Physics of Celestial Bodies.* 1991, vol. 7, pp. 74–84. Allenton Press, Inc., New York.

6. Астронометричний енциклопедичний словник. За заг. редак. І. А. Климішина та А. О. Корсунь. Львів, 2003. 548 с.

7. Taleb N. N. The Black Swan: The Impact of the Highly Improbable. New York: Random House. 2007. 440 p.

8. Blach F., Scholes M. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 1973, № 81, pp. 637–654.

9. Tukey J. W. A Survay of Sampling from Contaminated Distribution. Paper 39. In «Contribution to Probability and Statist» (ed. Olkin I. et al.). Stanford Univ, Press, 1960, pp. 448–485.

10. Tukey J. W. The future of Data Analysis. *Ann Math. Stat, 1962*, № 33, № 1, pp. 1–67.

11. Tukey J. W. Data Analysis and the Frontiers of Geophysics. *Science*. 1965. Vol. 148, № 3675, pp. 1283–1289.

12. Geary R. C. Testing for normality. *Biometrics,* 1947, № 34, pp. 209–242.

13. Радченко С. Г. Методология регрессионного анализа. К.: «Корнийчук», 2011. 376 с.

14. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Пер. с англ., 2 изд., кн. 1–2, М., 1986.

15. Draper N. R. Smith H. Applied Regression Analysis, 3rd Edition. By John Wiley & Sons, Inc, 1998. 736 p.

16. Rao C. R. Linear Statistical inference and its Application. 2nd ed., New York: John Wiley & Sons, Inc. 1973.

17. Hocking R. R. Methods and Application of Linear Models: Regression and Analysis of Variance. Third Edition. Ishpeming, Michigan: John Wiley & Sons, Inc., 2013. 720 p.

18. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. М.: Изд. «Прогресс», 1976. 478 с.

19. Большаков В. Д. Теория ошибок наблюдений. Учебник для вузов. 2-е изд., перераб. и доп. М.: Недра, 1983. 223 с.

20. Jeffreys H. Theory of Probability. Sec. Edition. Oxford, 1940. 468 p.

21. Джунь Й. В. Неклассическая теория погрешностей измерений. Ровно: Естеро, 2015. 168 с.

22. Dzhun I. V. Non-Classical Theory Measurements Error. Amazon, 2020. 200 p.

References

1. Veroyatnost i matematicheskaya statistika: Entsiklopediya. (1999). Gl. red. Yu. V. Prohorov. M.: Bolshaya Rossiyskaya entsiklopediya (In Russian).

2. Atramentova L. O. (2006). Naukove doslidzhennia i statystyka. *Naukovij svit (Scientific world),* 4, 6–7 (In Ukrainian).

3. Dzhun Y. V. (2020). Zahalnyi analiz problem matematychnoho modeliuvannia v pedahohichnykh doslidzhenniakh. *Psykholoho-pedahohichni osnovy humanizatsii navchalno-vykhovnoho protsesu v shkoli ta VNZ: zbirnyk naukovykh prats (Psychological and pedagogical bases of humanization of education and educational process at school and high school: collection of scientific works),* 1 (23), 28–39 (In Ukrainian).

4. Malyutov M. B., Zaigraev A. Yu. (1989). Sovremennyie zadachi optimalnogo planirovaniya regressionnyih eksperimentov. Kiev: Vyischa shkola (In Russian).

5. Dzhun I. V. (1991). Pearson’s Distribution of type VII of the Errors of Satellite Laser Ranging Data. *Kinematic and Physics of Celestial Bodies.* 7, 74–84. Allenton Press, Inc., New York.

6. Astronometrychnyi entsyklopedychnyi slovnyk. (2003). Za zah. redak. I. A. Klymishyna ta A. O. Korsun. Lviv (In Ukrainian).

7. Taleb N. N. (2007). The Black Swan: The Impact of the Highly Improbable. New York: Random House.

8. Blach F., Scholes M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81, 637–654.

9. Tukey J. W. (1960). A Survay of Sampling from Contaminated Distribution. Paper 39. In «Contribution to Probability and Statist» (ed. Olkin I. et al.). Stanford Univ, Press, 448–485.

10. Tukey J. W. (1962). The future of Data Analysis. *Ann Math. Stat*, 33, 1, 1–67.

11. Tukey J. W. (1965). Data Analysis and the Frontiers of Geophysics. *Science*, 148, 3675, 1283–1289.

12. Geary R. C. (1947). Testing for normality. *Biometrics*, 34, 209–242.

13. Radchenko S. G. (2011). Metodologiya regressionnogo analiza. K.: «Korniychuk» (In Russian).

14. Dreyper N., Smit G. (1986). Prikladnoy regressionnyiy analiz. Per. s angl., 2 izd., kn. 1–2, M. (In Russian).

15. Draper N. R. Smith H. (1998). Applied Regression Analysis, 3rd Edition. By John Wiley & Sons, Inc.

16. Rao C. R. (1973). Linear Statistical inference and its Application. 2nd ed., New York: John Wiley & Sons, Inc.

17. Hocking R. R. (2013). Methods and Application of Linear Models: Regression and Analysis of Variance. Third Edition. Ishpeming, Michigan: John Wiley & Sons, Inc.

18. Glass Dzh., Stenli Dzh. (1976). Statisticheskie metodyi v pedagogike i psihologii. M.: Izd. «Progress».

19. Bolshakov V. D. (1983). Teoriya oshibok nablyudeniy. Uchebnik dlya vuzov. 2-e izd., pererab. i dop. M.: Nedra (In Russian).

20. (1940). Jeffreys H. Theory of Probability. Sec. Edition. Oxford,

21. Dzhun Y. V. (2015). Neklassicheskaya teoriya pogreshnostey izmereniy. Rovno: Estero (In Russian).

22. Dzhun I. V. (2020). Non-Classical Theory Measurements Error. Amazon.

Стаття поступила в редакцію 13.06.2020 р.