УДК 324.311

ЗАСТОСУВАННЯ КОРЕЛЯЦІЙНО-РЕГРЕСІЙНОГО АНАЛІЗУ В ОЦІНЦІ РЕЗУЛЬТАТИВНОСТІ НАВЧАЛЬНОГО ПРОЦЕСУ

Селезньова Н.П., Українець О.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Запропоноване дослідження присвячене методам освітніх вимірювань, які ґрунтуються на побудові регресійної моделі, дискримінантному та факторному аналізі оцінок, отриманих студентами в результаті складання сесії. Зокрема в роботі досліджено наявність мультиколінеарності та виключено фактор, що «спотворює» адекватність моделі, побудовано «нове» рівняння регресії, та перевірено його на адекватність, розглянуто автокореляцію залишків. Дана інформація дає змогу використовувати її для оптимізації та поліпшення процесу навчання.
Ключові слова: кореляція, регресія, мультиколінеарність, автокореляція, лінія регресії, регресійна модель, коефіцієнт детермінації.

ФІЗИКО-МАТЕМАТИЧНІ НАУКИ

За останні роки система освіти України суттєво реформувалася. Якість вищої освіти сильно погіршилась за рахунок скорочення не тільки аудиторних годин, а й зменшення обсягу матеріалу, що вивчається. В результаті значна частина випускників вищих навчальних закладів стала неконкурентоспроможними на Європейському ринку праці. Конкурентоздатність економіки країни визначається в першу чергу інтелектуальним потенціалом, який плекається у вищих навчальних закладах. Підготовка фахівців у вищих навчальних закладах потребує системної роботи, в якій освітні вимірювання стають тим інструментом, який дає змогу оцінити рівень знань студентів та якість викладання дисциплін і організації навчально-виховної роботи. В наслідок цього постає актуальність створення та дослідження освітніх вимірювань в контексті покращення якості освіти в Україні.

Вивченню проблем освітніх вимірювань присвячено багато праць зарубіжних та вітчизняних учених [5, 6]. Але серед їх мало уваги приділяється освітнім вимірюванням у контексті покращення якості освіти та аналізу якості роботи викладачів.

Регресійний аналіз є основним засобом дослідження залежності між різними випадковими змінними. В тому числі є цікавим дослідження впливу рівня знань учнів чи студентів з одних навчальних предметів на інші. Результати таких досліджень можуть надати достатньо об'єктивну картину навчального процесу, що в подальшому можна буде використати для оптимізації та поліпшення цього процесу. Наскільки нам відомо, дослідження в такому контексті не проводились.

Для дослідження були взяті оцінки студентів II курсу фізико-математичного факультету Національного Технічного Університету України «Київський Політехнічний Інститут ім. І. Сікорського», отримані протягом сесії, з наступних предметів: математичного аналізу (Y), інформатики (X_1), аналітичної геометрії (X_2) та лінійної алгебри (X_3). Всього складало іспит 37 студентів ($n = 37$).

Постановка задачі: знайти регресійну модель успішності студентів (успішність вимірюється оцінками отриманими студентами з даних предметів) з математичного аналізу, інформатики, лінійної алгебри та аналітичної геометрії; дослідити її на адекватність, та вияснити наявність чи відсутність мультиколінеарності та автокореляції. Зробити відповідні висновки на основі отриманих результатів.

Побудова регресійної моделі. Числовими даними дослідження є оцінки студентів, отримані ними в результаті складання сесії, з окремих розділів вищої математики. При цьому можна стверджувати, що оцінки з одного предмета статистично залежать від оцінок з іншого предмета, тобто рівень знань з одного предмета статистично обумовлює рівень знань з іншого. Отже, один із показників формується під впливом трьох інших. Виникає необхідність побудови моделі такого зв'язку, яку також називають багатофакторною моделлю кореляційно-регресійного аналізу. В цьому випадку результативна ознака (Y) пов'язується за допомогою рівняння множинної регресії з трьома факторними ознаками (X_1, X_2, X_3).

Досліджувану залежність будемо шукати у вигляді [2, с. 40]:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + e,$$

де Y – вектор спостережень за незалежною змінною; $Y = \{y_1, y_2, ..., y_n\}$;

X – вектор спостережень за незалежною змінною; $X = \{x_1, x_2, ..., x_n\}$;

a_0, a_1, a_2, a_3 – невідомі параметри регресійної моделі;

e – вектор випадкових величин (похибок); $e = \{e_1, e_2, ..., e_n\}$.

Введемо такі позначення: X – матриця, складена з додаткового стовпця одиниць та X_1, X_2, X_3.

$$Y = \begin{pmatrix} 40 \\ 95 \\ ... \\ 97 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 85 & 62 & 70 \\ 1 & 95 & 61 & 62 \\ ... & ... & ... & ... \\ 1 & 98 & 97 & 98 \end{pmatrix}.$$

Використовуючи матричний метод побудови рівняння лінії регресії, отримаємо:

$$Y = -15{,}627 + 0{,}532X_1 + 0{,}085X_2 + 0{,}46X_3 + e. \quad (1)$$

Дане рівняння показує, що при покращенні успішності студентів з будь-кого предмета (X_1, X_2, X_3), буде підвищуватись успішність студентів з математичного аналізу Y При цьому, найбільший вклад дає предмет $X1$ – інформатика, трохи менший – X_3 – лінійна алгебра і найменший вплив має аналітична геометрія X_2. Хоча, виходячи з логічних міркувань, геометрія мала б більш суттєво впливати на рівень успішності з математичного аналізу ніж інформатика, адже раніше студенти досить добре засвоювали математичний аналіз, взагалі не вивчаючи інформатику. Отже можна зробити висновок, що геометрія вивчається студентами на недостатньому рівні.

Для перевірки адекватності функції регресії обчислимо наступні показники (див. табл. 1).

Для контролю правильності обчислень використаємо основну тотожність дисперсійного аналізу [2, с. 446] $s_\gamma^2 = s_f^2 + s_\varepsilon^2$: $96{,}61 + 238{,}61 \approx 324{,}6411$.

Мірою адекватності функції регресії служить величина [2, с. 448]:

$$R^2 = \frac{s_f^2}{s_\gamma^2}, \quad (2)$$

яка називається *коефіцієнтом детермінації.* Даний коефіцієнт набуває значень від до і показує, наскільки є великим загальне відхилення значень функції регресії від фактичних значень Y. Отже, чим ближчим є коефіцієнт до , тим точніше вибрана функція регресії відповідає даним. У нашому випадку:

$$R^2 = 0{,}297, \; R = 0{,}546.$$

Результат вказує на помірний кореляційний зв'язок.

Для визначення значущості знайденого рівняння регресії застосуємо критерій Фішера-Снедекора [2, с. 536]:

$$F = \frac{R^2(n-k-1)}{(1-R^2)k} > F_{\alpha;k;n-k-1}; \quad (3)$$

$$F = 6{,}99; \; F_{a;k;n-k-1} = 2{,}86,$$

де $a = 0{,}05$, $k = 3$, $n = 37$.

Можна зробити наступний висновок: $F > F_{a;k;n-k-1}$, і за згаданим вище критерієм, модель є адекватною.

Дослідження побудованої моделі (1) на наявність мультиколінеарності.

Найважливішими умовами побудови багатофакторної моделі зв'язку є достатня кількість елементів у вибірці та відсутність мультиколінеарності факторів. У тому випадку, коли два факторних показники є мультиколінеарними, один із них повинен бути виключений з моделі.

Термін *«мультиколінеарність»* означає, що в багатофакторній регресійній моделі дві або біль-

Таблиця 1

Компоненти дисперсії	Сума квадратів	Число степенів свободи	Дисперсія
Регресія	$SS_1 = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 = 3081{,}106$	$k = 2$	$s_f^2 = \frac{SS_1}{n} = 83{,}273$
Залишки	$SS_2 = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = 8605{,}975$	$n - k - 1 = 34$	$s_\varepsilon^2 = \frac{SS_2}{n-k-1} = 260{,}787$
Загальна варіація	$SS = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = 11687{,}081$	$n - 1 = 36$	$s_\gamma^2 = \frac{SS}{n} = 315{,}867$

ше незалежних змінних (факторів) пов'язані між собою лінійною залежністю або, іншими словами, мають високий ступінь кореляції [1, с. 228].

Мультиколінеарність пояснюючих змінних призводить до зміщення оцінок параметрів моделі, через що з їх допомогою не можна зробити коректні (адекватні) висновки про результати взаємозв'язку залежної і пояснюючої змінної. Наслідком наявності мультиколінеарності є те, що не можна точно визначити параметри регресії для сукупності.

Для дослідження мультиколінеарності використаємо алгоритм *Фаррара-Глобера*. Даний алгоритм включає в себе три основні статистичні критерії перевірки:

а) перевірка всього масиву незалежних змінних за допомогою критерію χ^2-квадрат;

б) перевірка кожної незалежної змінної за допомогою F-статистики;

в) перевірка кожної пари незалежних змінних за допомогою t-критерію.

Результати всіх трьох критеріїв дають нам змогу зробити висновки щодо наявності чи відсутності мультиколінеарності пояснюючих змінних.

Реалізуємо алгоритм Фаррара-Глобера для дослідження мультиколінеарності між змінними X1, X2, X3.

Маємо наступні вхідні дані (див. табл. 2).

Крок 1: Нормалізація значень факторів за допомогою формули:

$$x_{ij}^* = \frac{x_{ij} - \overline{x_j}}{\sqrt{n\sigma_{x_j}^2}},$$

де $\overline{x}_j$ – середнє значення j-ї пояснюючої змінної; x_{ij} – індивідуальне значення j-ї пояснюючої змінної; σ_j – стандартне відхилення j-ї пояснюючої змінної; n – число спостережень.

Нормалізована матриця має вигляд:

$$X^* = \begin{pmatrix} 0,061 & -0,132 & 0,034 \\ 0,207 & -0,150 & -0,089 \\ 0,076 & -0,113 & 0,110 \\ 0,091 & -0,150 & 0,003 \\ -0,157 & -0,039 & -0,089 \\ \dots & \dots & \dots \\ 0,251 & 0,514 & 0,462 \end{pmatrix}.$$

Крок 2: Знаходження кореляційної матриці, елементами якої є парні коефіцієнти кореляції $r_{ij} = r_{x_i x_j}$.

$$R = \left(X^*\right)^T X^* = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0,3 & 0,478 \\ 0,3 & 1 & 0,597 \\ 0,478 & 0,597 & 1 \end{pmatrix}$$

де $(X^*)^T$ – транспонована матриця X^*. Елементи матриці R характеризують щільність зв'язку однієї незалежної змінної з іншою.

Крок 3: Визначення критерію χ^2-квадрат:

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m+5)\right] \cdot \ln|R|,$$

де $|R|$ – визначник кореляційної матриці R.

$$|R| = 0,497,\ \ln|R| = -0,699.$$

Тоді $\chi^2 = -\left[37 - 1 - \frac{1}{6}(2 \cdot 3 + 5)\right] \cdot (-0,699) = 23,91.$

Порівняємо це значення з табличним значенням при $\frac{1}{2}m(m-1) = 3$ ступенях свободи, де m – кількість факторів, і рівні значущості $\alpha = 0,05$. Оскільки $\chi^2 = 23,91 > \chi^2_{табл} = 7,815$ [2, с. 535], то в масиві значень існує мультиколінеарність.

Крок 4: Знаходження оберненої матриці до кореляційної матриці

$$C = R^{-1} = \left(\left(X^*\right)^T X^*\right)^{-1} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m1} & \dots & c_{mm} \end{pmatrix} =$$

$$= \begin{pmatrix} 1,297 & -0,030 & -0,602 \\ -0,030 & 1,553 & -0,912 \\ -0,602 & -0,912 & 1,832 \end{pmatrix}.$$

Крок 5: Обчислення F-критеріїв:

$$F_k = \frac{(c_{kk} - 1)(n - m)}{(m-1)},$$

де c_{kk} – діагональні елементи матриці C; n – число спостережень; m – кількість факторів, $k = 1, 2, 3$.

Фактичні значення критеріїв порівнюють з табличним значенням при $n - m$ і $m - 1$ ступенях свободи та рівні значущості $\alpha = 0,05$. Якщо $F_{крит.} > F_{табл.}$, то відповідна k-змінна є мультиколінеарною іншим.

$$F_1 = 5,041,\ F_2 = 9,404,\ F_3 = 14,14.$$

Обчислені критерії порівнюємо з табличним значенням $F_{кр} = 19,500$. Отримуємо $F_1 < F_2 < F_3 < F_{крит.}$. Це означає, що в сукупності всі три змінні X_1, X_2, X_3 є мультиколінеарними за критерієм χ^2-квадрат.

Таблиця 2

№	X1	X2	X3	$(X_{1i} - \overline{X}_1)^2$	$(X_{2i} - \overline{X}_2)^2$	$(X_{3i} - \overline{X}_3)^2$
1.	85	62	70	17,78	50,91	4,79
2.	95	61	62	202,1	66,18	33,777
3.	86	63	75	27,21	37,64	51,68
4.	87	61	68	38,64	66,18	0,04
5.	70	67	62	116,29	4,56	33,77
...	...	...	...	...	...	...
37.	98	97	98	296,4	776,45	911,39
Сума				4714,27	2942,32	4267,68
Середнє значення	80,78	69,14	67,82			
Дисперсія	11,29	8,92	10,74			
Середньоквадратичне відхилення	11,44	9,04	10,89			

Уточнимо наявність мультиколінеарності факторів за допомогою T-критерію (крок 7).

Крок 6: Обчислимо частинні коефіцієнти кореляції, які характеризують тісноту зв'язку між двома змінними за умови, що інші змінні не впливають на цей зв'язок (існування парної мультиколінеарності):

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk}c_{jj}}},$$

де c_{kj} – елементи матриці C (2), що розміщені в k-ому та j-ому стовпці; c_{kk} та c_{jj} – діагональні елементи матриці C.

$$r_{12,3} = 0{,}021,\ r_{13,2} = 0{,}39,\ r_{23,1} = 0{,}541.$$

Спираючись на отримані значення частинних коефіцієнтів кореляції, можна відмітити, що:

→ зв'язок між успішністю студентів з інформатики та лінійної алгебри – помірний;

→ зв'язок між успішністю з лінійної алгебри та аналітичної геометрії помірний;

→ зв'язок між успішністю з інформатики та аналітичної геометрії – слабкий.

Крок 7: Обчислення t-критеріїв:

$$t_{kj} = \left|r_{kj}\right| \cdot \sqrt{\frac{n-m}{1-r_{kj}^2}}.$$

Фактичні значення t_{kj} порівнюють із табличним значенням при $n - m$ ступенях свободи і рівні значущості $\alpha = 0{,}05$. Якщо $t_{kj} > t_{табл.}$, то між незалежними змінними існує мультиколінеарність.

$$t_{12} = 0{,}123,\ t_{13} = 2{,}473,\ t_{23} = 3{,}75.$$

Обчислені t-критерії порівнюють із табличним значенням $t_{табл} = 2{,}031$ [2, с. 534], коли маємо $n - m = 37 - 3 = 34$ ступенів свободи, та при рівні значущості $\alpha = 0{,}05$.

Оскільки $t_{13} > t_{табл}$ та $t_{23} > t_{табл}$, то можемо зробити висновок, що мультиколінеарність існує між рівнем успішності студентів з інформатики та лінійної алгебри, і між рівнем успішності з аналітичної геометрії та лінійної алгебри. Оскільки $t_{12} > t_{табл}$, то мультиколінеарність між рівнем успішності з інформатики та аналітичної геометрії не існує.

Дослідження, проведені за алгоритмом Фаррара-Глобера показали, що мультиколінеарність між пояснюючими змінними даного прикладу існує. Отже, для того щоб можна було застосувати метод найменших квадратів для оцінювання параметрів регресійної моделі за цією інформацією, необхідно, в першу чергу, звільнитись від мультиколінеарності. Для цього виключимо змінну X_3 з подальших досліджень.

Побудуємо рівняння регресійної моделі, виключивши змінну X_3.

Рівняння лінії регресії будемо шукати у вигляді: $Y_x = a + bX_1 + cX_2$ [2, с. 40]. Позначимо: X – матриця, складена з додаткового стовпця одиниць та X_1, X_2.

$$Y = \begin{pmatrix} 40 \\ 95 \\ \dots \\ 97 \end{pmatrix},\quad X = \begin{pmatrix} 1 & 85 & 62 \\ 1 & 95 & 61 \\ \dots & \dots & \dots \\ 1 & 98 & 97 \end{pmatrix}.$$

Тоді

$$X'X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 85 & 95 & \dots & 98 \\ 62 & 61 & \dots & 97 \end{pmatrix}\begin{pmatrix} 1 & 85 & 62 \\ \dots & \dots & \dots \\ 1 & 98 & 97 \end{pmatrix} = \begin{pmatrix} 37 & 2989 & 2558 \\ 2989 & 246177 & 207762 \\ 2558 & 207762 & 179790 \end{pmatrix},$$

де X' – транспонована матриця до матриці X.

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 85 & 95 & \dots & 98 \\ 62 & 61 & \dots & 97 \end{pmatrix}\begin{pmatrix} 40 \\ 95 \\ \dots \\ 97 \end{pmatrix} = \begin{pmatrix} 2384 \\ 196177 \\ 166636 \end{pmatrix}$$

Знайдемо матрицю $A^{-1} = (X'X)^{-1}$.

$$A^{-1} = \begin{pmatrix} 2{,}34473 & -0{,}01271 & -0{,}01867 \\ -0{,}01271 & 0{,}00023 & -0{,}00009 \\ -0{,}01867 & -0{,}00009 & 0{,}00037 \end{pmatrix}.$$

$$\vec{b}(a;b;c) = A^{-1}(X'Y) =$$

$$= \begin{pmatrix} 2{,}34473 & -0{,}01271 & -0{,}01867 \\ -0{,}01271 & 0{,}00023 & -0{,}00009 \\ -0{,}01867 & -0{,}00009 & 0{,}00037 \end{pmatrix}\begin{pmatrix} 2384 \\ 196177 \\ 166636 \end{pmatrix} = \begin{pmatrix} -15{,}124 \\ 0{,}676 \\ 0{,}361 \end{pmatrix}.$$

$$a = -15{,}124,\ b = 0{,}676,\ c = 0{,}361.$$

Отже, маємо таке рівняння регресії:

$$Y_x = -15{,}124 + 0{,}676X_1 + 0{,}361X_2. \qquad (4)$$

Отже, з цього рівняння, випливає, що рівень знань студентів з інформатики статистично впливає на рівень знань з математичного аналізу. Відповідні коефіцієнти у рівнянні регресії зросли.

Перевіримо адекватність функції регресії за допомогою коефіцієнта детермінації, обчисленого за формулою (2): $R^2 = 0{,}264$, $R = 0{,}513$. Результат вказує на помірний кореляційний зв'язок між випадковими змінними Y та (X_1, X_2).

Визначимо значущість знайденого рівняння за допомогою формули (3):

$$F = 6{,}09;\ F_{\alpha;k;n-k-1} = 3{,}28,$$
$$\text{де } \alpha = 0{,}05,\ k = 2,\ n = 37.$$

Можна зробити наступний висновок: $F > F_{\alpha;k;n-k-1}$, і за згаданим вище критерієм, модель є адекватною.

Дослідження моделі (4) на наявність автокореляції.

Автокореляцією називається залежність між значеннями однієї вибірки з запізненням в один лаг. Наприклад, якщо між значеннями однієї вибірки $e_1, e_2, \dots, e_p$ та $e_2, e_3, \dots, e_{p+1}$ є залежність, то маємо справу з автокореляцією, якщо така залежність є між значеннями двох різних вибірок $e_1, e_2, \dots, e_p$ та $w_2, w_3, \dots, w_{p+1}$, то це свідчить про

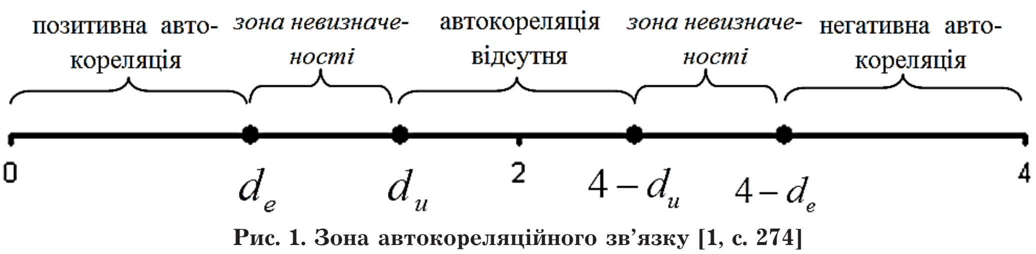

Рис. 1. Зона автокореляційного зв'язку [1, с. 274]

наявність серійної кореляції [1, с. 273]. Для дослідження наявності автокореляції використовують *d-тест Дарбіна-Уотсона.*

Крок 1. Обчислимо значення d-статистики [1, с. 274]:

$$dw = d = \frac{\sum_{i=2}^{n}\left(e_i - e_{i-1}\right)^2}{\sum_{i=1}^{n} e_t^2}, \text{ де } 0 \le dw \le 4.$$

Крок 2. За таблицею Дарбіна-Уотсона, при заданому рівні значущості $a = 0{,}05$, кількості факторів $m = 2$ та кількості спостережень $n = 37$, знайдемо значення dw_1 та dw_2.

→ якщо $0 < d < dw_1$, то є присутньою додатна автокореляція;

→ якщо $dw_1 < d < dw_2$ або $4 - dw_2 < d < 4 - dw_2$, то значення d потрапляє в зону невизначеності, тобто ми не можемо зробити висновки а ні про її наявність, а ні про відсутність;

→ якщо $4 - dw_1 < d < 4$, то маємо від'ємну автокореляцію;

→ якщо $dw_1 < d < 4 - dw_2$, то автокореляція відсутня (див. рис. 1).

Дослідимо нашу вибірку на наявність автокореляції. Знайдемо значення d-статистики: $d = 2{,}43$. За таблицею Дарбіна-Уотсона: $dw_1 = 1{,}27$, $dw_2=1{,}49$ [1, с. 485]. Отримали $1{,}27 < 2{,}43 < 4 - 1{,}49$, тобто $dw_1 < d < 4 - dw_2$. Отже, з похибкою щонайбільше 5%, можна стверджувати, що автокореляція відсутня.

Відсутність автокореляції свідчить про те, що випадкові величини не корелюють із своїми попередніми значеннями, іншими словами, можемо з впевненістю сказати, що успішність студентів з інформатики статистично не впливає на успішність з аналітичної геометрії, та навпаки. Отже, оцінки отримані при обчисленні t- та F-критерію є ефективними, а сам метод найменших квадратів у дослідженні застосований коректно.

Висновки. Багатофакторну лінійну регресію доцільно будувати тоді, коли на змінну Y впливає два та більше фактори (X_1, X_2, X_n), а у нашому дослідженні їх три (X_1, X_2, X_3). Тому нами була побудована відповідна залежність:

$$Y = -15{,}627 + 0{,}532X_1 + 0{,}085X_2 + 0{,}46X_3 + e. \quad (1)$$

Рівняння (1) є адекватним, про це свідчать як виконання основної дисперсійної тотожності так і значення коефіцієнту детермінації. За допомогою критерію Фішера-Снедекора встановлено значущість знайденого рівняння регресії. Отже, знайдену модель (1) можна вважати адекватною.

Встановлено, що має місце той факт, що при зміні порядку трьох факторів рівняння залишається незмінним з відповідними коефіцієнтами, проте при зміні досліджуваної змінної, отримаємо дещо інше рівняння.

Однією з класичним умов лінійної регресії є відсутність мультиколінеарності між пояснюючими змінними. Тому було зроблено перевірку на мультиколінеарність за допомогою критерію Фаррара-Глобера. Вона показала, що змінна X_3 (успішність студентів з лінійної алгебри) є мультиколінеарною з іншими. Отже, доцільно було виключити її з подальшого дослідження та побудувати нове регресійне рівняння (4): $Y_x = -15{,}124 + 0{,}676X_1 + 0{,}361X_2$. Побудована модель є адекватною і між змінними X_1 та X_2 відсутня лінійна залежність.

Для дослідження наявності автокореляції між пояснюючими змінними X_1 та X_2 було застосовано тест Дарбіна-Уотсона. Він показав, що між змінними автокореляція відсутня. Отже, можна сказати, що успішність студентів з інформатики статистично не впливає на успішність з аналітичної геометрії, та навпаки. А також оцінки отримані при обчисленнях t- та F-критерію можна вважати ефективними і метод найменших квадратів був застосований коректно.

Для моніторингу якості освіти, маючи рівняння регресії, надалі необхідно побудувати контрольні карти Шухарта і, таким чином, отримати статистичні діаграми управління процесом.

Список літератури:

1. Лук'яненко І., Краснікова Л., Економетрика / І. Лук'яненко, Л. Краснікова // Київ «Знання», 1998 р. – С. 228-278.
2. Кремер Н.Ш. Теория вероятностей и математическая статистика / Н.Ш. Кремер // Москва «ЮНИТИ», 2010 р. – С. 395-474.
3. Селезньова Н.П., Українець О.В. Моніторинг якості освіти через призму кореляційного аналізу. У кн. Матеріали Сімнадцятої міжнародної наукової конференції ім. академіка Михайла Кравчука (Т. 3. Теорія ймовірностей та математична статистика. Історія та методик математики) / Н.П. Селезньова, О.В. Українець // Київ: НТУУ «КПІ», 2016 р. – С. 146-149.
4. Селезньова Н.П., Українець О.В. Перевірка адекватності функції регресії. У кн. «Математика в сучасному технічному університеті» Матеріали V Міжнародної науково-практичної конференції 29-30 грудня 2016 р. (С. 4. Сучасні освітні технології у вищій школі) / Н.П. Селезньова, О.В. Українець // Київ: НТУУ «КПІ», 2017 р. – С. 252-256.
5. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стэнли // Москва: Прогресс, 1976. – С. 494.
6. Сидоренко Е.В. Методы математической обработки в психологи / Е.В. Сидоренко // Санкт-Петербург: ООО «Речь», 2003. – С. 350.

Селезнева Н.П., Украинец О.В.
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»

ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА В ОЦЕНКЕ РЕЗУЛЬТАТИВНОСТИ УЧЕБНОГО ПРОЦЕССА

Аннотация
Предложенное исследование посвящено методам образовательных измерений, основанных на построении регрессионной модели, дискриминантном и факторном анализе оценок, полученных студентами в результате сдачи сессии. В работе исследовано наличие мультиколлинеарности и исключено фактор, который «искажает» адекватность модели, построено «новое» уравнение регрессии, проверено его на адекватность и автокорреляцию остатков. Данная информация позволяет использовать ее для оптимизации и улучшения процесса обучения.
Ключевые слова: корреляция, регрессия, мультиколлинеарность, автокорреляция, линия регрессии, регрессионная модель, коэффициент детерминации.

Selezneva N.P., Ukrainets O.V.
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»

THE USING OF CORRELATION AND REGRESSION ANALYSIS DURING THE EFFECTIVENESS OF THE EDUCATIONAL PROCESS ASSESSING

Summary
The research suggested is devoted to the methods of educational measurements which are based on the regressive model construction, discriminant and factor points assessment, received by the students as the result of passing end-of-semester exams. In particular, the presence of multicollinearity was investigated, the factor, which distorts the model validity was excluded, the new regression equation was created and checked for adequacy, and the residues autocorrelation was considered in this research. The following information gives the opportunity to use it for the educative process optimization and improving.
Keywords: correlation, regression, multicollinearity, autocorrelation, regression line, regression model, the coefficient of determination.