КОМП'ЮТЕРНІ ЗАСОБИ, МЕРЕЖІ ТА СИСТЕМИ

O. Kovyrova

METHODS OF CHLOROPHYLL FLUORESCENCE INDUCTION CURVES PROCESSING

In paper methods of fluorescence chlorophyll induction curves, received by device "Floratest", are considered and the results of analysis of variance test for normality are given.
Key words: chlorophyll fluorescence induction, fluorometer, normal distribution.

Рассмотрены методы обработки измерений кривых индукции флуоресценции хлорофилла, полученных с помощью прибора "Флоратест" и представлены результаты анализа на нормальность.
Ключевые слова: индукция флуоресценции хлорофилла, флуориметр, нормальное распределение.

Розглянуті методи обробки вимірів кривих індукції флуоресценції хлорофіла, отриманих за допомогою приладу "Флоратест" і представлені результати аналіза на нормальність.
Ключові слова: індукція флуоресценції хлорофілу, флуориметр, нормальний розподіл.

УДК 578.01+681.7.08+535.3+681.335.2

О.В. КОВИРЬОВА

МЕТОДИ ОБРОБКИ ВИМІРІВ КРИВИХ ІНДУКЦІЇ ФЛУОРЕСЦЕНЦІЇ ХЛОРОФІЛУ

Вступ. Портативний прилад "Флоратест" для експрес-діагностики стану рослин (спектральний діапазон вимірювання інтенсивності флуоресценції від 670 до 770 нм) [1, 2] створено і поставлено на серійне виробництво в Інституті кібернетики імені В.М. Глушкова НАН України. Прилад дає змогу швидко оцінити рівень впливу природного оточуючого середовища і забруднень на живі рослини. Робота приладу базується на вимірюванні в реальному часі кривої індукції флуоресценції хлорофілу (ІФХ) без пошкодження рослин.

Побудова графіків ІФХ. Розглянемо детальніше криві ІФХ рослини дурман, отримані за допомогою приладу "Флоратест" (рис. 1). Кожна крива містить 90 відліків по осі *X*. По осі *Y* наводяться значення ІФХ для кожного відліку, які виражені у градаціях шкали приладу.

Прилад "Флоратест" побудований на базі мікроконтролеру ADuC 842 [2], який містить 12-бітний аналого-цифровий перетворювач (АЦП). 12-бітний АЦП здатний сформувати 2^{12} = 4096 комбінацій коду. В подальшому при побудові кривих пропонується використовувати значення у відносних одиницях, яке отримуємо наступним чином:

$$\mathrm{iph}_i(r.u.) = \frac{k_1 \cdot iph_i}{k_2} \quad i = \overline{1,90}\,, \qquad (1)$$

де $k_1 = 2{,}5\,\text{В}$ – напруга вихідного сигналу АЦП, $k_2 = 4096$ – кількість комбінацій коду, iph_i – значення ІФХ в *i*-точці, $\mathrm{iph}_i(r.u.)$ – значення ІФХ в *i*-точці в відносних одиницях. По осі *x* на рис. 1 показано номер відліку

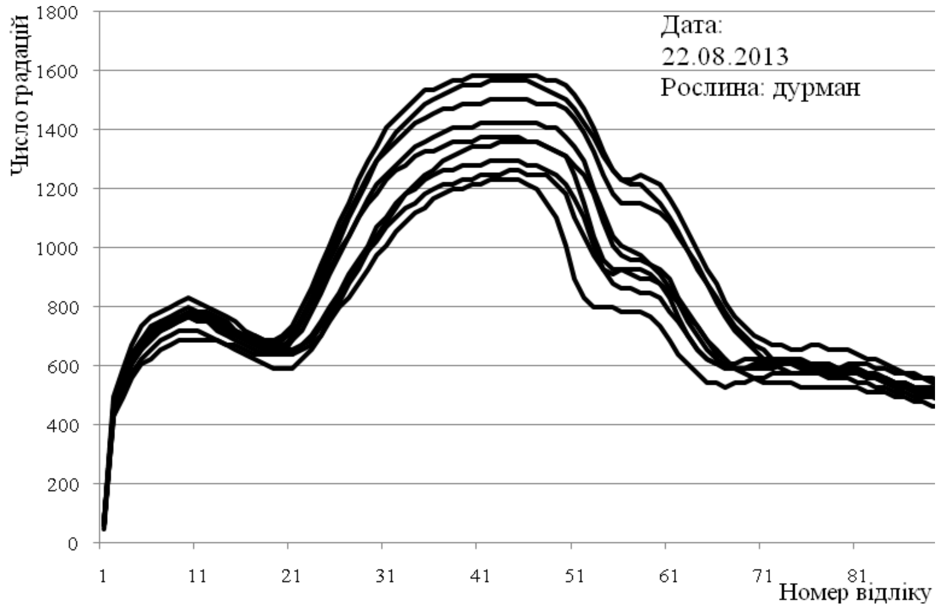

РИС. 1. Криві ІФХ рослини дурману

а саме значення від 1 до 90. Однак крива ІФХ вимірюється не рівномірно, а відповідно до експоненціальної шкали [2]. Тому графік ІФХ у відносних одиницях має наступний вигляд (рис. 2).

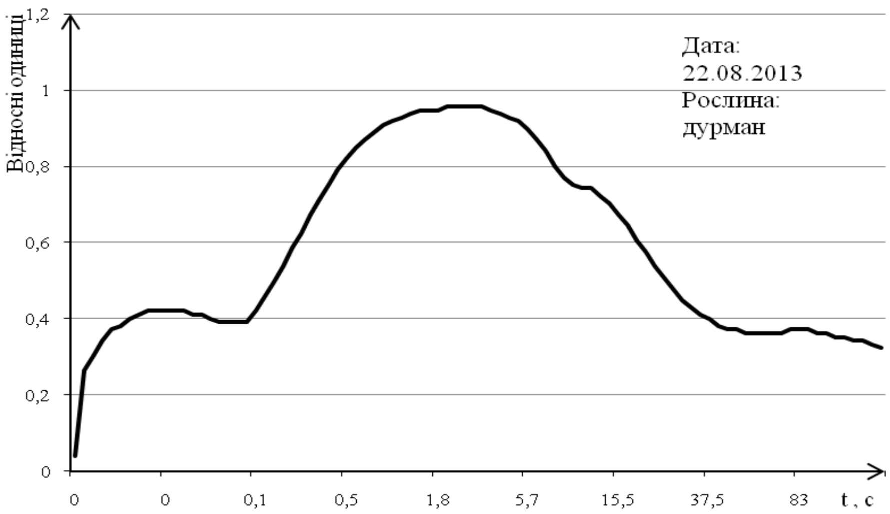

РИС. 2. Результати вимірювання кривих ІФХ рослини дурману у контрольній групі за допомогою портативного флуориметра "Флоратест" у відносних одиницях

Протягом 2012 – 2013 рр. в Інституті кібернетики проведено ряд натурних експериментів [3, 4] з метою оцінки чутливості флуориметра "Флоратест" до стресових факторів, які впливають на стан рослин. Під час експериментів отримано близько 1160 кривих ІФХ. Серед них 460 в контрольних групах (рослини не піддавалися негативному впливу).

Під час експерименту з рослинами дурману [3, 4] в контрольній групі розкид кривих по амплітуді складав до 50 %. Для перевірки достовірності отриманих результатів статистичними методами, як правило, робиться припущення, що вибірки мають нормальний розподіл. Але для цього попередньо слід визначити, чи підпорядковуються результати вимірювань нормальному розподілу.

Для перевірки на нормальність існує більше двадцяти різноманітних методів. Найбільш поширеними є критерій Андерсона – Дарлінга, критерій типу Колмогорова – Смірнова, критерій χ^2 Пірсона, критерій Шапіро – Вілка [5] та його модифікація критерій Шапіро – Франсія.

При використанні вищезгаданих методів слід враховувати об’єм вибірки, на якому даний метод може працювати. Більшість згаданих критеріїв розглянуто в роботі [6]. В роботах [7–9] проведено порівняльний аналіз критеріїв перевірки відхилення розподілу від нормального закону, виконано дослідження особливостей і потужностей деяких критеріїв нормальності.

Дослідження показали, що на вибірках невеликого розміру об’ємом 10–20 спостережень та при невеликих рівнях значущості α (ймовірність помилки першого роду) критерії Шапіро – Вілка та Еппса – Паллі є зміщеними. Водночас потужність цих критеріїв на вибірках об’ємом $n \leq 50$ вище потужності непараметричних критеріїв (згоди) типу Колмогорова, типу ω^2 Крамера – Мізеса – Смірнова та типу $\boldsymbol{\Omega}^2$ Андерсона – Дарлінга, які в такій ситуації потужніші критеріїв типу χ^2. Водночас використання критеріїв згоди для перевірки відхилень від нормального закону при невеликих об’ємах вибірок є безперспективним внаслідок їх низької потужності за відношенням до близьких альтернатив. При невеликій кількості елементів автори рекомендують віддавати перевагу спеціальним критеріям перевірки відхилень від нормальності. Результатом аналізу отриманих у роботі оцінок потужності критеріїв [9] є висновок авторів, що слід віддати перевагу наступним критеріям: Шпилельхальтера, Херази – Грина Т2, Еппса – Паллі, Шапіро – Вілка, Гірі.

В результаті проведених експериментів над рослинами дурману в вересні 2013 року отримано дві вибірки об’ємом 35 елементів. Для перевірки на нормальність кожну таку вибірку розділили на групи відповідно до номеру відліку. Тобто, аналізувалось 90 груп ординат, які мають 35 значень. За результатами розрахунків визначено кількість груп, які підпорядковуються нормальному розподілу.

Перевірка на нормальність здійснювалася методом Саркаді [10], який за словами авторів роботи [6] є більш досконалим ніж інші. Перевірка показала, що всі 90 груп підпорядковуються нормальному розподілу. Однак цей метод має низький ранг [6], тому отримані дані не можна вважати достовірними і в подальшому не враховуються.

З урахуванням вище наведеного додатково для аналізу обрано метод Шапіро – Вілка [5] (використовується на вибірках об’єму 8–50 елементів). Об’єм вибірки більше 20, тому недолік, зазначений вище, не повинен впливати на результат обчислень.

Критерій Шапіро – Вілка базується на відношенні оптимальної лінійної незсуненої оцінки дисперсії до її звичайної оцінки, отриманої методом максимальної вірогідності. Статистика критерію має вигляд:

$$W = \frac{1}{s^2}\left[\sum_{i=1}^{k} a_{n-i+1}(x_{n-i+1} - x_i)\right]^2, \tag{2}$$

де $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Коефіцієнти a_{n-i+1} та критичні значення статистики $W(\alpha)$ наведені в роботі [5]. Якщо $W < W(\alpha)$, то гіпотеза нормальності розподілу відкидається на рівні значущості α.

Результати розрахунків наведені у таблиці.

ТАБЛИЦЯ. Результати перевірки нормальності методом Шапіро – Вілка

№	Вибірка	Кількість кривих	Кількість груп серед 90, які підпорядковуються нормальному розподілу з рівнем значущості α, %		
			α = 0,01	α = 0,05	α = 0,1
1	1	35	25,5	14	7,7
2	2	35	13,3	1	1
Середнє значення			19,4	7,5	4,35

Як видно з таблиці, в середньому лише 20 % груп підпорядковуються нормальному розподілу з рівнем значущості α = 0,01, ще менша кількість груп з рівнем значущості α = 0,05 та α = 0,1.

Метод Шапіро – Вілка реалізовано за допомогою мови Visual Basic в середовищі Microsoft Excel. Для перевірки роботи програми використано тестові дані, які згенеровано за допомогою функції rnorm () пакету nortest мови програмування R [11].

Отже, в результаті аналізу отриманих нами експериментальних даних було виявлено, що вони не підпорядковуються нормальному розподілу. Тому стандартні статистичні методи обробки даних не доцільно використовувати для аналізу груп кривих. Нами рекомендується замість середніх значень груп ординат використати значення медіани, 25 і 75 процентиля [12].

Побудова регресійної моделі кривої ІФХ за допомогою методу крокової регресії. Одним із шляхів подальшого аналізу кривих ІФХ є представлення отриманої кривої за допомогою математичного рівняння. На даному етапі як математичний апарат обрано регресійний аналіз, а саме побудова поліноміальної моделі восьмого порядку методом крокової регресії. Тип моделі визначено в результаті аналізу моделей різного типу, а саме моделі з «оберненим» перетворенням відносно предиктора, моделі з логарифмічним перетворенням відносно предиктора, моделі з перетворенням типу квадратного кореня відносно предиктора, мультиплікативної моделі, експоненціальної та поліноміальної моделей. Стандартна похибка прогнозованого значення і значення залишкової суми квадратів є найменшими, а значення коефіцієнту кореляції є найбільшим (0,99 або 99 %) для поліноміальної моделі.

Відповідно до [13] пари випадкових змінних мають двомірний розподіл ймовірностей деякого типу. Якщо існує зв'язок між залежною ймовірною величиною *Y* і змінною не випадковою величиною *X*, то рівняння *Y* відносно *X* буде називатися рівнянням регресії. Змінна *X* називається предиктором.

Нехай рівняння регресії змінної *Y* від змінної *X* має вигляд $\alpha_0 + \alpha_1 X$. Тоді можна записати лінійну модель регресії (модель регресії першого порядку) у наступному вигляді:

$$Y = \alpha_0 + \alpha_1 X + \varepsilon, \tag{3}$$

де α_i – параметри моделі, ε – випадкова величина. Рівняння (3) – це модель, яка постулюється. Вважаємо, що вона встановлена. На наступних кроках необхідно перевірити, чи відповідає вона реальності. Оцінки значень параметрів у рівнянні (1) мають вигляд:

$$\hat{Y} = a_0 + a_1 X \,, \tag{4}$$

де $\hat{Y}$ – прогнозоване значення для даного X , у випадку, коли коефіцієнти a_i визначені. Як процедуру оцінювання параметрів α_i найчастіше використовують метод найменших квадратів (МНК).

В загальному випадку тип лінійної моделі зі змінними X_1, X_2, …, X_k можна представити у вигляді:

$$Y = \alpha_0 Z_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \ldots + \alpha_p Z_p + \varepsilon, \tag{5}$$

де $Z_0 = 1$ – це фіктивна змінна, яка завжди рівна 1 і зазвичай не записується.

Кожна Z_j , $j = 1, 2, \ldots, p$ – це відома функція від X_1, X_2, …, X_k :

$$Z_j = Z_j (X_1, X_2, \ldots, X_k), \tag{6}$$

яка може мати будь-яку форму. Інколи кожна функція Z_j включає лише одну змінну X . Будь-яку таку модель можна записати після перетворення у вигляді (5) і аналізувати її загальними методами. Більш узагальненим є представлення Y , X , α_i у вигляді векторів.

Фактори, які використовуються в регресійних задачах, зазвичай можуть приймати значення з певного неперервного інтервалу. Інколи можна вводити фактор, який має два або більше різноманітних рівня. Існує можливість приписати цим факторам деякі рівні по порядку. Змінні такого типу називають фіктивними змінними. У більшості випадків вони не пов'язані з фізичними рівнями, які можуть існувати у факторів самі по собі.

Для будь-якої регресійної задачі існує безліч способів вибору фіктивних змінних. Одним із варіантів представлення є матриця Вандермонда [14, 15]. В нашому випадку використано найпростіший випадок: як фіктивні змінні використано змінні наступного вигляду з максимальною ступінню рівною 8 ($k = 8$, $n = 90$):

$$X = \begin{pmatrix} 1 & 1 & 1^2 & \dots & 1^8 \\ 1 & 2 & 2^2 & \dots & 2^8 \\ 1 & 3 & 3^2 & \dots & 3^8 \\ 1 & \vdots & \vdots & \dots & \vdots \\ 1 & n-1 & (n-1)^2 & \dots & (n-1)^8 \\ 1 & n & n^2 & \dots & n^8 \end{pmatrix}$$

$$\det(X'X) = 1{,}7878 \cdot 10^{116} \neq 0. \tag{7}$$

Визначник матриці не рівний нулю. Отже, фіктивні змінні можна використовувати при побудові моделі.

Задача вибору «найкращого» рівняння регресії досить складна, оскільки не існує однозначної статистичної процедури. До того ж використання різних методів для однієї і тієї ж задачі не обов'язково приводить до отримання однакового розв'язку.

Для побудови моделі нами обрано метод крокової регресії. Під час виконання методу крокової регресії відбувається включення змінних за чергою у рівняння до тих пір, поки рівняння не стане задовільним. Порядок включення визначається за допомогою часткового коефіцієнта кореляції, як міри значущості змінних, ще не включених у рівняння. На кожному кроці відбувається перевірка значущості змінної, яка вводиться в модель, за допомогою F-критерію та перевірка гіпотези про рівність нулю кожного коефіцієнта α_i за допомогою часткового F-критерію. Після отримання моделі розраховують довірчі інтервали для параметрів моделі та стандартну похибку для $\hat{Y}$.

Процес припиняється, в тому випадку, якщо жодну зі змінних, які містяться в поточному рівнянні, не вдається виключити з нього, а найближчий найкращий предиктор не може зайняти місце. Недоліком даного методу є те, що важлива змінна може ніколи не включитися в модель, а другорядні будуть включені.

Розглянемо результати побудови регресійних моделей восьмого порядку кривих ІФХ методом крокової регресії:

$$Y = \beta_0 + \beta_1 q + \beta_2 q^2 + \beta_3 q^3 + \beta_4 q^4 + \beta_5 q^5 + \beta_6 q^6 + \beta_7 q^7 + \beta_8 q^8 + \varepsilon, \tag{8}$$

де q – вектор довжиною n.

Всі розрахунки виконувалися за допомогою функції ЛИНЕЙН програми Microsoft Office Excel для визначення статистики ряду з використанням методу найменших квадратів.

На рис. 3 показано графік ІФХ та графік моделі 5-го порядку липи широколистної (дата виміру – 14.06.2011), стандартна похибка для оцінки $\hat{Y}$ рівна 0,047. На рис. 4 показано графік ІФХ та графік моделі 6-го порядку рослини дурману (дата виміру – 22.08.2013), стандартна похибка для оцінки $\hat{Y}$ рівна 0,06. Аналогічним чином можна побудувати моделі для різних типів рослин.

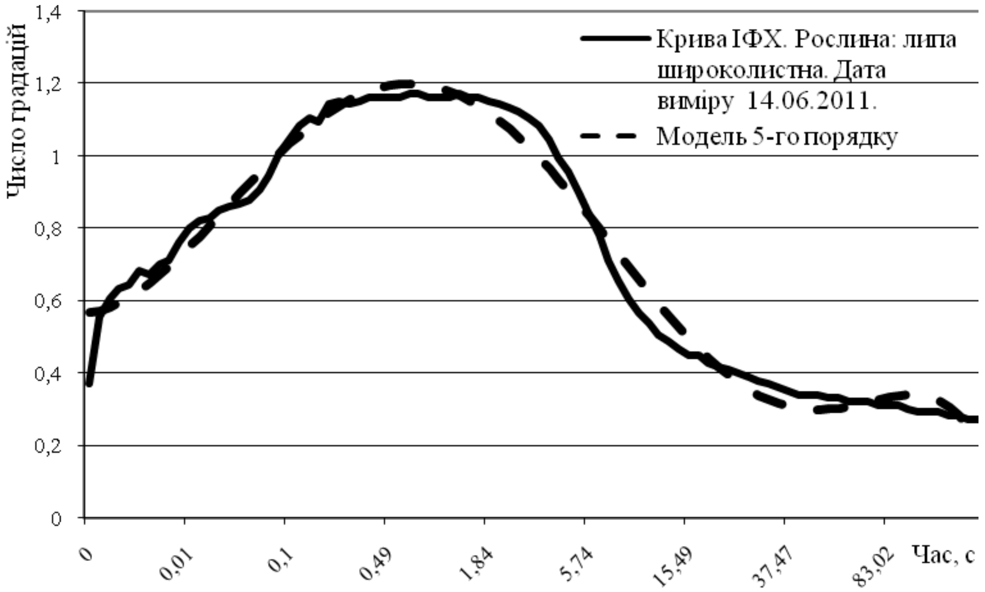

РИС 3. Графік ІФХ рослини липи широколистної та побудована поліноміальна модель 5-го порядку

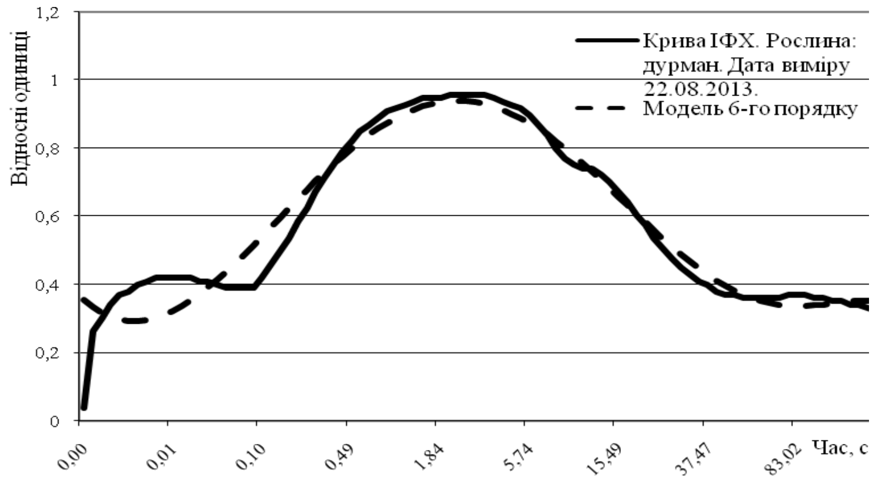

РИС 4. Графік ІФХ рослини дурману та побудована поліноміальна модель 6-го порядку

Очевидно, що модель для липи більш відповідає реальності, у порівнянні з моделлю для дурману. Це пов'язано з тим, що криві ІФХ відрізняються для різних типів рослин. Виходячи з цього, метою наступних досліджень є пошук математичних методів, які будуть враховувати особливості різних кривих ІФХ при побудові математичної моделі.

Висновки. 1. В результаті аналізу отриманих нами експериментальних даних виявлено, що вони не підпорядковуються нормальному розподілу. Тому стандартні статистичні методи обробки даних не доцільно використовувати для аналізу груп кривих. Нами рекомендується замість середніх значень груп ординат використати значення медіани, 25 і 75 процентиля. 2. Побудована поліноміальна модель кривої індукції флуоресценції хлорофілу методом крокової регресії для заміни реальних досліджень комп'ютерним моделюванням.

1. *Palagin O., Romanov V., Galelyuka I., Voronenko et al.*Computer devices and mobile information technology for precision farming // Proceeding of the 7th IEEE International conference on "Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications", IDAACS'2013. – Berlin, Germany. – 2013, September 12–14. – P. 47–51.
2. *http*://www.dasd.com.ua/
3. *Груша В.М., Ковирьова О.В.* Дослідження чутливості флуориметра "Флоратест" до дії стресових факторів на стан рослин // Комп'ютерні засоби, мережі та системи. – 2012. – № 11. – С. 119 – 126.
4. *Груша В.М., Ковирьова О.В.* Особливості обробки кривих індукції флуоресценції хлорофілу // Материалы 18-го Юбилейного Международного молодежного форума «Радиоэлектроника и молодежь в XXI веке». Сб. материалов форума. Т. 1. – Харьков: ХНУРЭ. – 2014. – С. 138–139.
5. *Shapiro S.S., Wilk M.B.* An analysis of variance test for normality // Biometrika. – 1965. – 52, N 3. – P. 591–611.
6. *Кобзарь А. И.* Прикладная математическая статистика. – М.: Физматлит, 2006. – 238 с.
7. *Лемешко Б.Ю., Лемешко С.Б.* Сравнительный анализ критериев проверки отклонения распредления от нормального закона // Метрология. – 2005. – № 2. – С. 3–24.
8. *Лемешко Б.Ю., Рогожников А.П.* Исследование особенностей и мощности некоторых критериев нормальности // Метрология. – 2009. – № 4. – С. 3–24.
9. *Лемешко Б.Ю., Рогожников А.П.* О нормальности погрешностей измерений в классических экспериментах и мощности критериев, применяемых для проверки отклонения от нормального закона // Метрология. – 2012. – № 5. – С. 3–26.
10. *http*://www.r-project.org/.
11. *Sarkadi Karoly.* On testing for normality. Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., Vol. 1 (Univ. of Calif. Press, 1967), P. 373–387.
12. *Стентон Гланц.* Медико-биологическая статистика. Пер. с англ. – М.: Практика, 1998. – 459 с.
13. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ (книга 2). – М.: Финансы и статистика, 1986. – 351 с.
14. *Стрижов В.В., Крымова Е.А.* Методы выбора регрессионных моделей. – М.: ВЦ РАН, 2010. – 60 с.
15. *Ильин В.А., Позняк Э.Г.* Линейная алгебра. – М.: Наука. – Физматлит, 1974. – С. 35.

Одержано 15.09.2014