

**Азимов Аваз Аббасович**

АН РУз центр геномики и биоинформатики м.н.с.

**Хусенов Наим Нутфуллоевич**

АН РУз центр геномики и биоинформатики, начальник отдела PhD

**Усманов Дилшод Эркинбаевич**

АН РУз центр геномики и биоинформатики зав. лаб. PhD

**Абдукаримов Шарофиддин Сайфидинович**

АН РУз центр геномики и биоинформатики м.н.с.

**Салахутдинов Ильхом Бахтиярович**

АН РУз центр геномики и биоинформатики зав.лаб. с.н.с

**Раджапов Фарход Суннатуллаевич**

АН РУз центр геномики и биоинформатики м.н.с

**Собиров Ботиржон Машариф угли**

АН РУз центр геномики и биоинформатики м.н.с.

**Буриев Забардаст Таджибоевич**

АН РУз центр геномики и биоинформатики д.б.н. проф.

**e-mail: [googlazimov@gmail.com](mailto:googlazimov@gmail.com)**

## **МОДЕЛЬ ПРОГНОЗА И ОЦЕНКИ ВЛИЯНИЯ ДИНАМИКИ СТЕПЕНИ ПОРАЖЕНИЯ ХЛОПЧАТНИКА ФУЗАРИОЗНЫМ ВИЛТОМ**

**Аннотация:** В данной статье отражены результаты работы по разработке статистических моделей зависимости количеством баллов максимального поражения растений хлопчатника разных сортов с другими факторами риска развития фузариозного вилта ФВ. Для решения данной задачи нами были использованы программные процедуры множественного регрессионного анализа пакета статистических программ SPSS 21. В начальном этапе была рассчитана описательная статистика. Корреляционный анализ определил показателей факторов риска развития ФВ, которые коррелировали с зависимым переменным - количеством баллов максимального поражения растений в вышесреднем уровне. Далее пошаговым методом исключений переменных найдена оптимальная в смысле структуры и критериев качества, множественная регрессионная модель зависимости количеством баллов максимального поражения растений с факторами риска развития ФВ.

Ключевые слова: фузариозный вилт, математическая модель прогноза, факторы риска, максимального поражения растений хлопчатника, множественный регрессионный анализ, зависимые переменные, независимые переменные.

**Azimov Avaz Abbasovich**

O'zR FA Genomika va bioinformatika markazi k.i.x

**Xusenov Naim Nutfulloevich**

O'zR FA Genomika va bioinformatika markazi bo'lim boshlig'i, PhD

**Usmanov Dilshod Erkinbaevich**

O'zR FA Genomika va bioinformatika markazi, laboratoriya mudiri, PhD

**Abdukarimov Sharofiddin Sayfidinovich**

O'zR FA Genomika va bioinformatika markazi k.i.x

**Salaxutdinov Il'xom Baxtiyarovich**

O'zR FA Genomika va bioinformatika markazi, laboratoriya mudiri, k.i.x.

**Radjapov Farxod Sunnatullaevich**

O'zR FA Genomika va bioinformatika markazi k.i.x

**Sobirov Botirjon Masharif o'g'li**

O'zR FA Genomika va bioinformatika markazi k.i.x

**Buriev Zabardast Tadjiboevich**

O'zR FA Genomika va bioinformatika markazi b.f.d. professor

**e-mail:** [googlazimov@gmail.com](mailto:googlazimov@gmail.com)

## **G'O'ZANING FUZARIOZ VILT KASALI BILAN ZARARLANISH DARAJASI DINAMIKASI TAHLILI VA PROGNOZ MODELI**

**Annotatsiya:** Bu maqolada fuzarioz vilt bilan kasallantirilgan bir necha xil g'o'za navlarning ballar bo'yicha aniqlangan maksimal zararlanish darajasi miqdori va kasallikning rivojlanishiga sababchi bo'lgan boshqa risk faktorlar o'rtasidagi bog'liqlikni aks ettiruvchi statistik modellar yaratilishi natijalari bayon qilingan. Ushbu masalani hal qilishda SPSS 21 statistika paketida mavjud bo'lgan ko'p o'lchamli regression tahlil kompyuter dasturidan foydalanildi. Ishning avvalida dastlabki statistik hisob-kitob amallari bajarildi. Korrelyatsion tahlil kasallikning rivojlanishiga sababchi bo'lgan boshqa risk faktorlar va g'o'zalarning ballar bo'yicha aniqlangan maksimal zararlanish darajasi miqdori o'rtasidagi korrelyatsiya darajasi o'rtachadan yuqoriroq ekanini ko'rsatdi. Shu

*bilan birga o'zgaruvchilarni ketma-ket yo'qotish usuli bo'yicha maksimal zararlanish darajasi miqdorining boshqa risk faktorlarga bog'liqligini aniqlovchi optimal strukturali va xarakteristikali ko'p o'lchamli regression modeli tuzildi.*

**Kalit so'zlar:** *fuzarioz vilt, matematik prognoz modeli, risk faktorlari, g'o'za o'simligining maksimal zararlanishi, ko'p o'lchamli regression tahlil, erksiz o'zgaruvchilar, erkli o'zgaruvchilar.*

**Azimov Avaz Abbasovich**

AS RUz, Center for genomics and bioinformatics, junior researcher

**Khusenov Naim Nutfulloevich**

AS RUz, Center for genomics and bioinformatics, Head of Department, PhD

**Usmanov Dilshod Erkinbaevich**

AS RUz, Center for genomics and bioinformatics, Head. lab. PhD

**Abdukarimov Sharofiddin Sayfidinovich**

AS RUz, Center for genomics and bioinformatics, junior researcher

**Salakhutdinov Ilkhom Bakhtiyarovich**

AS RUz, Center for genomics and bioinformatics, Head of Lab. senior researcher

**Radjapov Farkhod Sunnatullaevich**

AS RUz, Center for genomics and bioinformatics, junior researcher

**Sobirov Botirjon Masharif ugli**

AS RUz, Center for genomics and bioinformatics, junior researcher

**Buriev Zabardast Tadjiboevich**

AS RUz, Center for genomics and bioinformatics, Doctor of Biology. prof.

**e-mail:** [googlazimov@gmail.com](mailto:googlazimov@gmail.com)

## **MODEL FOR PREDICTION AND ASSESSMENT OF THE INFLUENCE OF THE DEGREE OF COTTON DAMAGE BY FUSARIUM WILT**

**Abstract:** *This article reflects the results of work on the development of statistical models of the dependence of the number of points of the maximum damage to cotton plants of different varieties with other risk factors for the development of Fusarium wilt FW. To solve this problem, we used software procedures for multiple regression analysis of the SPSS 21 statistical software package. At the initial stage, descriptive statistics were calculated. Correlation analysis determined the indicators of risk factors for the development of FW, which correlated with the dependent variable - the number of points for the maximum damage to plants at the higher average level. Then, using a step-by-step*

method of excluding variables, an optimal in terms of structure and quality criteria, and multiple regression model of the dependence of the number of points of maximum damage to plants with risk factors for the development of EF was found.

**Keywords:** *fusarium wilt, mathematical prediction model, risk factors, the maximum damage to cotton plants, multiple regression analysis, dependent variables, independent variables.*

### **Введение**

Фузариозный вилт (ФВ) хлопчатника является вредоносным заболеванием. При поражении восприимчивых сортов потери урожая могут достигать 45-50%ов. Заражение хлопчатника с возбудителем болезни происходит в почве через корни. Первые признаки болезни могут обнаруживаться на всходах. *Fusarium oxysporum* f.sp. *vasinfectum* (FOV) развиваясь в сосудистой системе хлопчатника, вызывает трахеомикозное увядание (вилт), что приводит растение к гибели. Вначале болезнь проявляется на семядолях в виде некротических пятен, далее на листьях появляются сетчатые некрозы вдоль жилок, пораженные листья опадают и растения усыхают. Пораженная ткань проводящих сосудов листьев, черешков, стеблей и коробочек становится темно-бурой, что является характерным признаком болезни [1].

Гриб *FOV* растет и развивается при температуре от 10 до 35 °С. Оптимальными условиями для его развития являются температура 18-27 °С, влажность почвы от 40 до 70% и кислотность среды около pH=5.3. Патоген поражает преимущественно тонковолокнистые сорта хлопчатника (*Gossypium barbadense*) но однако сегодняшний день он также вредить средне волокнистых сортов хлопчатника (*Gossypium hirsutum*). Заболевание интенсивнее проявляется в местах с высокой насыщенностью хлопчатником полевых севооборотов и при несвоевременной сортосмене на более устойчивые сорта [2, 3].

Целью настоящего исследования является изучение влияния динамики степени поражения растения хлопчатника фузариозным вилтом на гибели растений и установления роли предыдущих состояний, измеренных в разных этапах, посредством множественных регрессионных моделей.

### **Материалы и методы**

Объектами исследования являлись квалифицированные по морфо биологическим и хозяйственным качествам как относительно устойчивыми сорта Las Brenas 347, Л-237025N517, Cokers -124, Mebane B-1, Tamcot sp, PD 648, DPZ 554085, Л-4112-1, Rex, Meade 14-2, Deltopine 14 и как стандартный сорт Ташкент-1. В качестве контроля использована линия неустойчивого сорта 108-Ф. Вместе с этими, подобраны два высокоурожайные, скороспелые и обладающие высококачественными показателями волокна сорта Равнак и Порлок, созданные в Центре геномики и биоинформатики по технологий маркер ассоциированной селекции и ген накаута [4].

Для оценки устойчивости всех опытных образцов к фузариозному вилту, привезены образцы зараженных фузариозом вилтом почвы и растений из хлопковых полей Бухарской области. Далее методом Kim Y (2005) из них выделены споры грибков – возбудители вилта и в специальной питательной среде они культивированы и размножены.

Учет поражения растений вилтом проводился согласно Ulloa M [5] по шести бальной системе, при которой устойчивая иммунная форма отмечалась 0-баллами, а растениям, погибшим и не достигшим фазы плодоношения, присуждались 5-баллов.

При подготовке файла данных по конечной дате по всем сортам хлопчатника баллы, отражающие соответствующие степени поражения растений, количественно просуммированы соответственно и сформирована таблица, по строкам которой расположены сорта, а по столбцам - количества баллов, причем, число строк 15 и число столбцов 6 (ball\_0 - ball\_5).

Для обработки данных использовали программу множественного регрессионного анализа по требованию подготовки данных, были

произведены некоторые формализации, согласно которых обозначены зависимой и независимые переменные через  $x_1, x_2, \dots, x_6$  и их, кроме  $x_1$ , можно считать факторами риска развития ФВ.

- $x_1 - \text{ball}_0$  – количество растений с нулевым баллом – здоровое растение,
- $x_2 - \text{ball}_1$  – количество растений с единичным баллом,
- $x_3 - \text{ball}_2$  – количество двухбалльных растений,
- $x_4 - \text{ball}_3$  – количество трех балльных растений,
- $x_5 - \text{ball}_4$  – количество четырёх балльных растений,
- $x_6 - \text{ball}_5$  – количество пяти балльных растений – погибающее растение.

Последняя переменная  $x_6$  ( $Y$ ) зависимая переменная и в уравнении регрессионной модели она будет стоять слева от знака равенства.

Для обработки данных и получения необходимых количественных значений оценок для построения модели зависимости показателя погибающего растения  $x_6$  с другими факторами риска развития ФВ, была использована одна из процедур метода статистического моделирования – множественный регрессионный анализ, который содержится в пакете статистических программ SSPS 21 [6, 7, 8].

Регрессионное уравнение имеет форму  $Y = c + b_1*x_1 + b_2*x_2 + \dots$ , где  $Y$  является зависимым членом,  $b_1, b_2, \dots, b_m$  коэффициентами регрессии соответствующих  $x_1, x_2, \dots, x_m$  независимых членов и  $c$  константой, которая включает в себя и член ошибки  $e$ .

### **Результаты и обсуждение.**

После обработки группы данных, программой множественного линейного регрессионного анализа в начальном этапе выдаётся информация о действительных и пропущенных значениях, приводятся средние значения, стандартные отклонения, количество наблюдений в отдельности и суммарные показатели.

Корреляционный анализ показал, что зависимая переменная  $x_6$  коррелировала в достаточно высоком уровне и обратном направлении только со следующими двумя показателями. Это переменные  $x_1$  и  $x_5$ , коэффициенты корреляций, которых соответственно были равны  $-0,793$  и  $-0,611$ , а с остальными переменными  $x_6$  коррелировал очень слабо.

Программный пакет SPSS 21 допускает возможность ввода независимых переменных в модель несколькими способами. Например, пошаговые методы исключения и включения переменных, принудительный ввод переменных и другие. Каждый из этих методов имеет свои особенности, и используются в зависимости от требований решаемых задач.

При последовательном исключении программа начинает с введения всех независимых показателей в модель и рассчитывает долю, вносимую каждой из них в точность прогноза модели с проверкой по t-тесту для каждого показателя. Если доля вклада показателя для повышения точности предсказания значения зависимой переменной не значительной, то данный показатель исключается из модели, и последняя переоценивается. Также подвергаются повторной оценке, оставшиеся в модели предикторы.

Пошаговые методы предпочтительнее использовать в тех случаях, когда цель построения множественной регрессионной модели – предсказание значения зависимой переменной по значениям нескольких независимых переменных-предикторов, т.е. когда необходимо выбрать по возможности наименьшее количество предикторов, которые обеспечат максимальную точность предсказания. Из всех пошаговых методов метод последовательного исключения предпочтителен, т.к. он несет наименьший риск ошибки второго типа (наименьшую вероятность исключить из модели предиктор, который на самом деле оказывает влияние на зависимую переменную) [9].

Для построения множественных регрессионных моделей нами был использован пошаговый метод исключения переменных и в результате после двух итераций программой найдена самой оптимальной структуры модель и

поиск завершен. Ниже представлен анализ регрессионной модели, сформированной в последней итерации со соответствующей структурой.

Обычно регрессионные модели характеризуются некоторыми критериями качества. К ним относятся так называемые коэффициент множественной корреляции  $R$ , коэффициент детерминации  $R^2$ , стандартная ошибка оценки  $S$ , значение  $F$  критерия Фишера и его значимость. Кроме того, проверяется по  $t$ -критерию Стьюдента значимости всех коэффициентов, стоящих при независимых переменных-предикторах, вошедших в модель [10].

Коэффициент множественной корреляции  $R$  указывает совокупного влияния независимых показателей – предикторов на зависимой переменной и принимает значения от нуля - когда сила влияния предикторов отсутствует, до единицы – если сила совокупного влияния предикторов очень велика. Этот коэффициент в нашей регрессионной модели имел значения 0,981, который считается довольно высоким.

Коэффициент детерминации  $R^2$  есть квадрат коэффициента множественной корреляции. Поскольку в наших расчетах модели  $R$  имел высокие значения, то  $R^2$  тоже как критерий качества моделей, имел соответственно близкие к единице значения и указывал, на то, что больше 90 процентов вариации в двух наших моделях объясняются предикторами, вошедшими в регрессионные уравнения.

Стандартная ошибка оценки  $S$  представляет собой стандартное отклонение остатков и в больших выборках  $S$  приближается к стандартной ошибке прогнозируемого значения. Кроме того, в качественных моделях стандартная ошибка оценки будет заметно ниже, чем стандартное отклонение зависимой переменной. Данное предположение подтверждается в полученных результатах по нашим расчетам значения  $S$ , равное 0,640 против 2,883 в самом последнем уравнении.

Критерий Фишера  $F$  вычисляется методом дисперсионного анализа ANOVA и служит для проверки значимости модели. Если значение



коэффициента  $F$  больше 1, то доля вариабельности, объясняемая моделью, больше той, которая не может быть ею объяснена, и чем больше это значение по сравнению с единицей, тем выше значимость модели. Если полученное значение  $F$  превышает критическое, определяемое в соответствующих таблицах, это означает, что регрессионная модель статистически значима при соответствующем уровне значимости (чаще всего это 5% уровень). В наших расчетах для последней модели значение  $F$  равно 63,98 и значение уровня значимости  $P < 0,0001$ .

Роли коэффициентов регрессии,  $b$  в моделях особенные - они характеризуют среднее увеличение зависимой переменной, когда независимая переменная увеличивается на одну единицу, а другие независимые переменные остаются неизменными. Коэффициент  $b$  отражает влияние изменения зависимой переменной на независимую переменную. Чем больше коэффициент у независимой переменной, тем больше зависимая переменная изменяется за каждую единицу изменения в независимой переменной.

Проверка нулевой гипотезы о равенстве нулю коэффициента регрессии и для оценки значимости отдельных коэффициентов  $b$  используются  $t$ -тесты критерия Стьюдента и в результате удаляются из уравнения модели независимые переменные, с уровнем значимости  $P > 0,05$ .

В первой полученной нами модели, хотя основные критерии как  $R^2$ ,  $S$ ,  $F$  и  $P$ , характеризующие качества модели довольно высокие, данная модель содержит статистически незначимого коэффициента, определенного  $t$ -тестом критерия Стьюдента и с уровнем значимости  $P > 0,05$ . И поэтому, в практических целях не рекомендуется пользоваться этой моделью.

В ходе итераций изменением структуры уравнений методом пошагового исключения не информативных переменных, достигнуть упрощение структуры и улучшение критерии качеств регрессионной модели.

$$Y = x_6 = 13,668 - 0,950 \cdot x_1 - 1,192 \cdot x_2 - 0,992 \cdot x_3 - 1,086 \cdot x_5 \quad (1)$$

$t$ : 21,22; -11,44; -6,99; -6,23; -4,56;

$$p: 0,0001; 0,0001; 0,0001; 0,0001; 0,01;$$

$$R^2 = 0,962; S = 0,640; F = 63,984; P < 0,0001.$$

Последняя модель (1) имеет оптимальную структуру и обладает высокими значениями критериев, характеризующих данной модели. В эту модель вошли четыре показателя степени поражения как независимые переменные: количество нулевых баллов, соответствующее здоровому растению  $x_1$ , количество единичных баллов  $x_2$ , количество двухбалльных поражений  $x_3$  и количество поражений в четырёх балльной степени  $x_5$ . Коэффициенты всех переменных модели по  $t$ -тесту критерия Стьюдента имели высокие значения соответственно - 0,950; - 1,192; - 0,992; - 1,086 и оказались статистически значимыми на уровне  $P < 0,0001$ , а у  $x_5$   $P < 0,01$ .

Как уже выше отмечалось, что величина и знаки коэффициентов предикторов модели непосредственно повлияют на изменения значений зависимой переменной. Так и в нашей последней модели коэффициент -0,950 при предикторе количество нулевых баллов  $x_1$  означает, что если количества нулевых баллов повысится на одну единицу, количества степени максимального поражения, соответствующее гибели растения  $x_6$  уменьшает на 0,950 единицу при фиксированных значениях других независимых переменных модели. Аналогичным образом, при изменении предиктора количество единичных баллов  $x_2$  на одну единицу измерения, количества степени максимального поражения  $x_6$  ( $Y$ ) изменится на величину коэффициента, равную 1,192 единицу только в противоположную сторону, поскольку данный коэффициент имеет отрицательный знак. Роль следующих коэффициентов тоже аналогично интерпретируются.

Константа  $c = 13,668$  в модели является оценкой среднего  $x_6$  при условии, что все коэффициенты, стоящие при независимых переменных должны быть равны нулю. Так как, такие условия не выполнимы, в данной ситуации величину  $c$  невозможно интерпретировать.

На рисунке 1 приводится график сопоставления табличных данных  $x_6$  с прогнозными значениями  $x_6$ , рассчитанными с помощью моделью. Этот

график показывает, что кривые, соответствующие табличным и прогнозным значениям, почти совпали друг с другом. Данный факт подтвердил высокой степени адекватности разработанной нами регрессионной модели.

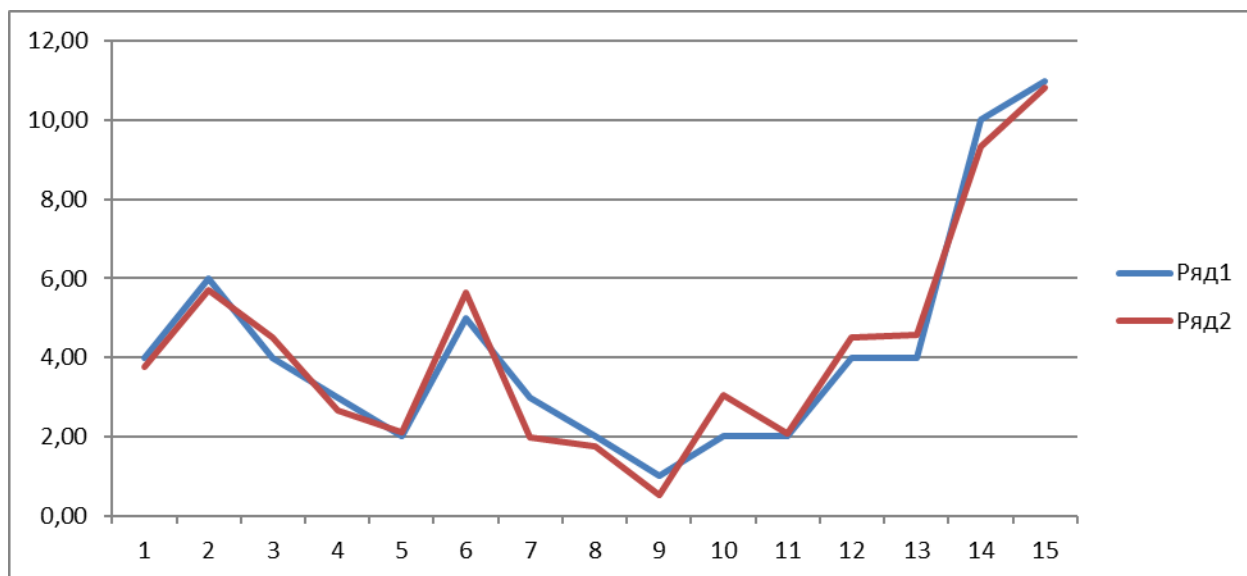


Рис.1. График табличных и прогнозированных моделью значений  $x_6$  (ball\_5).

### Заключение

Таким образом, полученная математическая модель зависимости показателя состояния гибели растения от четырёх факторов риска развития ФВ хлопчатника в виде регрессионного уравнения предоставляет исследователю возможность косвенно определить самую высокую степени поражения  $x_6$  по значениям этих факторов. И с учётом их относительных вкладов по коэффициентам будет возможность управлять зависимым переменным и более целенаправленно планировать проведения агрокультурную тактику для защиты урожая от фузариозного вилта.

### Использованная литература

1. [http://www.agroatlas.ru/ru/content/diseases/Gossypii/Gossypii\\_Fusarium\\_oxysporum/index.html](http://www.agroatlas.ru/ru/content/diseases/Gossypii/Gossypii_Fusarium_oxysporum/index.html).
2. Ахмеджанов А. Н. и др. Важная веха в селекции вилтоустойчивости хлопчатника: Достижения и перспективы экспериментальной биологии растений. Ташкент, 2013. С. 195-197.

3. Ahmedzhanov A., Mamaruziev A., Akkuzhin D., Sherimbetov A. Problems of modern science and education (2016). Creation of viltresistant perspective cotton varieties wilt a complex of economic–valuable traits by using intva and inter species hubbrisations [№ 18 \(60\). P. \(17 – 22\).](#)
4. Abdurakhmonov IY, Abdullaev AA, Buriev ZT, Shermatov SS, Kushanov F, Makamov A, Shapulatov U, Egamberdiev S, Salakhutdinov IB, Ayubov M, Darmonov M, Adylova AT, Rizaeva SM, Abdullaev F, Namazov S, Khalikova M, Saydaliev H, Avtonomov VA, Sanamyan M, Duiesenov TK, Musaev JA, Abdullaev A, Abdukarimov A. (2014). Cotton germplasm collecion of Uzbekistan. World Cotton Germplasm Resources, Ibrokhim Y. Abdurakhmonov (Ed.), ISBN 978-953-51-1622-6, InTech, Croatia.
5. Ulloa M, Hutmacher RB, Roberts PA, Wright SD, Nichols RL, Davis RM (2013). Inheritance and QTL mapping of Fusarium wilt race 4 resistance in cotton. Theor Appl Genet 126:1405–1418
6. Мелник М. Основы прикладной статистики. М. Энергоатомиздат (1983). Глава 14. Корреляционный и регрессионный анализ. – 416 с.
7. Бююль А., Цеффель П. SPSS: искусство обработки информации. – М., (2005). Глава 18. Множественный регрессионный анализ.
8. Наследов, А. Д. SPSS 19. Профессиональный статистический анализ данных [Текст] / А. Д. Наследов. – СПб.: Питер, 2011. – 400 с.
9. Шарашова Е.Е., Холматова К.К., Горбатова М.А., Гржибовский А.М. ([2017](#)). Применение множественного линейного регрессионного анализа в здравоохранении с использованием пакета статистических программ SPSS Наука и Здравоохранение. №3. С. 5-31.
10. Дрейпер Н , Смит Г. Прикладной регрессионный анализ. В 2-кн. - М : Финансы и статистикаю. 1987. -351с.