

KANONIK DISKRIMINANT FUNKSIYA YORDAMIDA G'O'ZANING SHO'RGACHIDAMLILIGINI BAHOLASH

Azimov A.A., Usmanov D.E., Bo'riev Z.T.

O'zbekiston Respublikasi Fanlar Akademiyasi Genomika va bioinformatika markazi

Tel: +99893 392-05-96 e-mail: googlazimov@gmail.com

Annotatsiya

Ushbu maqolada fanning obrazlarni o'qituvchi vositasida farqlash sohasiga tegishli bo'lgan kanonik diskriminant taxlil usuli yordamida g'o'zaning sho'rga chidamlilik darajasini prognoz qilish natijalari bayon qilingan. Bu usul asosida sho'rga chidamlilik darajasi hali noma'lum bo'lgan g'o'zani shu darajani prognoz qilish maqsadida sho'rlanish kontsentratsiyasi turlicha bo'lgan tuproqlarga ekilib, sho'rga chidamlilik darajasi bo'yicha avvaldan taqsimlangan uch g'o'za guruhining har biri uchun diskriminant funktsiya-modeli yaratildi. Bu funktsiya-modellar chidamliligi prognoz qilinishi kerak bo'lgan g'o'zaning va tuproqning parametrik belgilari asosida shu g'o'zaning sho'rga chidamlilik darajasini aniqlab beradilar. Bu diskriminant funktsiyalar yuqori sifatli xarakteristikalarga ega bo'lib, kiritilgan ma'lumotlarni guruhlarga to'g'ri prognoz qilish imkoniyati 96,2% ga teng bo'ldi.

Kalit so'zlar: Sho'rga chidamlilik, kanonik diskriminant tahlil, klassifikatsiya, prognoz, tuproq, g'o'za.

ОЦЕНКА СОЛЕУСТОЙЧИВОСТИ ХЛОПЧАТНИКА КАНОНИЧЕСКОЙ ДИСКРИМИНАНТНОЙ ФУНКЦИЕЙ

Азимов А.А., Усманов Д.Э., Буриев З.Т.

Центр геномики и биоинформатики Академии наук Республики Узбекистан

Тел: +998 93 392-05-96 e-mail: googlazimov@gmail.com

Аннотация.

В статье представлены результаты исследований по прогнозированию степени солеустойчивости хлопчатника методом канонического дискриминантного анализа, который является одним из методов области науки распознавания образов с учителем. С помощью

данного метода для заранее классифицированных трех групп хлопчатника, выращенных на почве с разной концентрацией солености и разделанные по степени устойчивости к соли, разработаны три дискриминантные функции-модели для прогнозирования степени устойчивости к засолению вновь представленного образца хлопчатника по измеренным значениям показателей почвы и хлопчатника. Полученные дискриминирующие функции, обладая качественными характеристиками, обеспечивали суммарную точность правильной классификации методом дискриминантных функций на 96,2% исходных сгруппированных наблюдений.

Ключевые слова: Солеустойчивость, канонический дискриминантный анализ, классификация, прогноз, почва, хлопчатник.

ASSESSMENT OF COTTON SALT RESISTANCE CANONICAL DISCRIMINANT FUNCTION

Azimov A.A., Usmanov D.E., Buriev Z.T.

Center of genomics and bioinformatics Academy of Sciences of the Republic of
Uzbekistan

Tel: +998 93 392-05-96 e-mail: googlazimov@gmail.com

The article presents the results of research on predicting the degree of salt tolerance of cotton by the method of canonical discriminant analysis, which is one of the methods in the field of science of pattern recognition with a teacher. Using this method, for pre-classified three groups of cotton grown on soil with different salinity concentrations and divided by the degree of salt tolerance, three discriminant functions-models were developed to predict the degree of salinity tolerance of a newly introduced cotton sample from the measured values of soil and cotton parameters. The resulting discriminant functions, having qualitative characteristics, ensured the total accuracy of the correct classification by the discriminant functions method for 96.2% of the initial grouped observations.

Key words: Salt tolerance, canonical discriminant analysis, classification, forecast, soil, cotton.

Введение

Хлопок, будучи одним из важных сельхозкультур, кроме обеспечения текстильной промышленности натуральным волокном и является во многих отраслях промышленности необходимым востребованным сырьем. С вегетационного периода до созревания он сталкивается с рядом биотических и абиотических стрессов, среди которых засоленность стала одной из основных угроз для устойчивого производства хлопка во всем мире [1].

С торможением ростовых процессов в условиях засоления в значительной мере изменяется и развитие хлопчатника. Таким образом, высокая концентрация солей в почве влияет на темпы ее развития. Задержка ростовых процессов хлопчатника значительно влияет и на прирост листовой поверхности, который под действием солей резко сокращается с последствием нарушения процессов фотосинтеза [2].

Солеустойчивость культур по критериям ФАО учитывается через электрическую проводимость (концентрацию) насыщенного почвенного раствора. Так, при засолении выше 2 dS/m происходит снижение урожая более чувствительных культур (помидоры, перец, лук и др.), а для солеустойчивых (пшеница, ячмень, хлопок) потери урожая происходят при засолении свыше 7 dS/m [3].

На основе данных почвенного и солевого обследования орошаемых земель институтом «Узерлойиха», и, принятых ими же понижающих коэффициентов для расчета бонитета почв:

- при слабом засолении 0,85 (потери 15 %)
- при среднем засолении 0,6 (потери 40 %)
- при сильном засолении 0,4 (потери 60 %)
- при очень сильно засолении 0,3 (потери 70 %),

можно определить, что из-за засоления потери урожая в Республике (как по хлопчатнику, так и по пшенице) достигают 30 % [4].

Кроме того, неумелое использование поливных вод в период вегетации или неправильная методология промывок земель в зимний и весенний период способствуют интенсивному засолению земель. Все это снижает эффективность

использования водных ресурсов при производстве удельной единицы сельскохозяйственной продукции и создает условия постоянного накопления солей, снижающих продуктивность орошаемых земель.

Во всех странах-поставщиках хлопка-сырца для возделывания устойчивых экстремальным и другим стрессовым факторам линии и сортов хлопчатника стало одной из самых необходимых задач агропромышленных комплексов. В своих поисках селекционеры, и связанные с решением данной задачи ученые биологи, химики, биотехнологи и инженеры-агротехники работают над проблемой с привлечением последних достижений теоретических и прикладных арсеналов не только агрокультуры и биологии, но и интеграцией возможности современных многих точных и естественных наук.

Актуальность указанной проблемы не оставил без внимания и ученые Центра геномики и биоинформатики АН РУз, которые в последние годы занимаются поиском отбора солеустойчивых сортов хлопчатника и на основе их параметров методами молекулярной биологии определением генов, отвечающих на стрессовые факторы, в том числе и на солеустойчивость.

Данное исследование в Центре будет произведено отдельно по этапам от дифференциации и отбора существующих солеустойчивых сортов и молекулярно-генетического анализа для определения генов, отвечающих на устойчивость к соли до роста культуры методом *in-vitro*, получаемой на базе специально созданной векторной конструкции посредством рекомбинантной ДНК с последующим выращиванием солеустойчивого растения хлопчатника.

Материалы и методы.

Материалом для исследования служила группа данных, содержащая следующих показателей: сумма сульфатных и хлоридных солей на 100 г.

почве (г) *Salt_100*, поверхностная плотность (см²/г) *SurfaseDensity* и общая площадь листовой пластинки куста (см²) *LeafSurface* и сухая масса листа хлопчатника (г) *SubstantContent*, полученная из источника [5] и классифицированная нами на 3 класса методом к-средних с количеством наблюдений, равными 6, 12 и 8 соответственно.

В активизации процесса фотосинтеза растений с последующим повышением продуктивности показана особая роль общей листовой поверхности растений в работе [6].

С целью оценки солеустойчивости хлопчатника канонической дискриминантной функцией, являющиеся как прогнозная модель отнесения представленных данных в свой класс с возможно большей вероятностью, нами была использована программная процедура дискриминантный анализ статистического пакета IBM Statistica SPSS 21 [7, 8, 9, 10].

Результаты и обсуждение.

В начальном этапе программных расчетов получены результаты описательной статистики исследуемых показателей, таких как средние, стандартные отклонения и количество наблюдений, отдельно по группам и общие, которые расположены в таблице 1.

Таблица 1.

Группы	Показатели	Среднее	Стд. Отклонение	Колич. наблюд.
Устойчивые	<i>Salt_100</i>	0,3843	0,0636	6
	<i>SurfaseDensity</i>	0,0325	0,0039	6
	<i>LeafSurface</i>	3074,1667	563,6564	6
	<i>SubstantContent</i>	0,1957	0,0214	6
Средне устойчивые	<i>Salt_100</i>	0,3551	0,1315	8
	<i>SurfaseDensity</i>	0,0319	0,0056	8
	<i>LeafSurface</i>	2101,6250	103,8941	8
	<i>SubstantContent</i>	0,2300	0,0364	8
Слабо устойчивые	<i>Salt_100</i>	0,5168	0,1895	12
	<i>SurfaseDensity</i>	0,0388	0,0071	12
	<i>LeafSurface</i>	1415,0833	182,4507	12
	<i>SubstantContent</i>	0,2268	0,0254	12

Итого	<i>Salt_100</i>	0,4365	0,1653	26
	<i>SurfaseDensity</i>	0,0352	0,0068	26
	<i>LeafSurface</i>	2009,1923	724,9615	26
	<i>SubstantContent</i>	0,2206	0,0307	26

На рисунке 1 представлена зависимость стандартизованных значений средних показателей хлопчатника от степени устойчивости к засолению. Стандартизация осуществлена путем деления среднего значения каждого показателя на его средне квадратического отклонения, и таким образом получены безразмерные величины, способствующие удобству при выполнении сравнительных процедур.

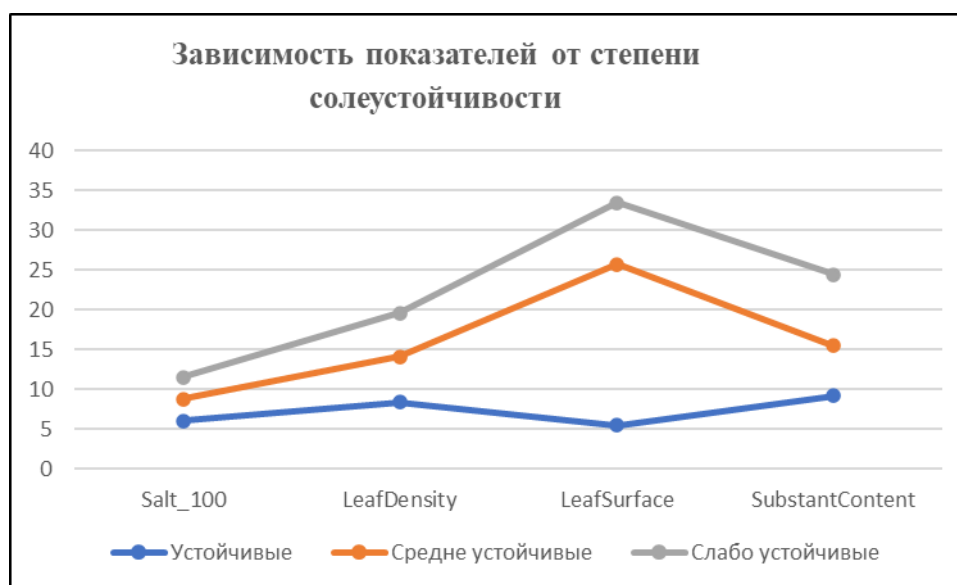


Рис.1. Зависимость стандартизованных значений средних показателей от степени засоления

На графике видно явное отличие Устойчивой группы от двух других групп по изменениям стандартных средних показателей почвы и хлопчатника. Здесь группы по показателю *Salt_100* не слишком отличались, а по показателю *LeafSurface* отличие групп довольно ощутимое. Отличие групп по показателям *SurfaseDensity* и *SubstantContent* оказалось среднее и выше среднего соответственно.

С помощью ДА представляется возможным построение оптимальных поверхностей дискриминантных функций (ДФ) Y в пространстве признаков, разделяющих все пространство на области, соответствующие объектам разных классов. Эти поверхности (функции) служат границами между областями и обеспечивают оптимальное разделение объектов, относящихся к разным классам. Правило классификации в этом случае заключается в определении по величине дискриминантной функции принадлежности объекта к той или иной из выделенных областей. Принадлежность к соответствующей области означает принадлежность соответствующему классу [11, 12, 13].

Количество дискриминантных функций определяется как $k-1$ или $m-1$, где k – число групп, а m – число переменных. Как правило, для определения количества необходимых дискриминантных функций выбирают наименьшее из этих чисел [14, 15, 16].

Дискриминантная функция – выведенная посредством дискриминантного анализа есть линейная комбинация независимых переменных, с помощью которой можно наилучшим образом различить (дискриминировать) категории зависимой переменной [17].

$$Y_i = C_i + w_{i1}X_1 + w_{i2}X_2 + \dots + w_{im}X_m$$

где, $i=1... m$ - номер группы,

m – количество переменных,

Y_i - результат показателя классификации,

C_i - константа для i -ой группы,

w_{ij} - веса для j -ой переменной i -ой группы,

X_j - наблюдаемое значение j -ой переменной.

Согласно вышеизложенным, функции классификации позволяют для прямого вычисления показателя классификации Y_i для новых наблюдений.

В рассматриваемом нашем анализе используется 4 переменных и 3 групп, т.е. число дискриминантных функций должно быть две.

Для каждой совокупности в выборке можно определить положение точки, представляющей средние для всех переменных в многомерном пространстве, определенном значениями переменных рассматриваемой модели. Эти точки называются центроидами группы. В ходе работы программы ДА в зависимости от числа дискриминантных функций образуется ось, плоскость или пространство, разделяющие исследуемые группы. Оценка степени сходства групп осуществляется посредством расчета расстояния между их центроидами. В качестве меры сходства, из множества существующих расстояний, чаще всего применяется квадрат евклидового или Махаланобиса расстояние; затем, используя для классификации вычисленное расстояние, можно получить вероятность того, что образец принадлежит к конкретной совокупности [18].

Эффективность дискриминации в ДА оценивается для каждой ДФ критерием λ -Уилкса, посредством который устанавливается остаточная дискриминационная способность различать группы, в исключении информации из предыдущей функции и тем самым останавливается поиск следующей функции при малости остаточной дискриминации. Поскольку данный критерий имеет распределения

Таблица 2.

Лямбда Уилкса				
Проверка функции(й)	Лямбда Уилкса	Хи-квадрат	ст.св.	Знч.
от 1 до 2	,128	44,123	8	,000
2	,835	3,889	3	,274

χ^2 - Пирсона, межгрупповые различия для каждой из полученных функций оценивается достоверно вычислением статистики χ^2 . Значения статистики λ -Уилкса близкие к нулю, свидетельствуют о том, что центроиды групп сильно отличаются друг от друга по отношению к степени разброса наблюдений внутри групп и тем самым показывает о высоком качестве дискриминации

[19, 20]. В наших расчетах, приведенных в таблице 1, для первой канонической дискриминантной функции значение критерия λ -Уилкса равно 0,128, статистики χ^2 -Пирсона - 44,123 и уровень значимости довольно высокая ($P < 0,0001$), а у второй ДФ указанные характеристики имеют относительно низкие значения и уровень значимости ($P > 0,05$).

В таблице 3 приведены собственные значения для каждой дискриминантной функции и кумулятивная доля объясненной дисперсии, накопленной каждой функцией. Как видно из таблицы, функция 1 ответственна за 96,5% объясненной дисперсии, т.е. 96,5% всей дискриминирующей мощности определяется этой функцией. Поэтому эта функция наиболее значима, а функция 2 ответственна только за 3,5% объясненной дисперсии;

Таблица 3.

Собственные значения				
Функция	Собственное значение	% объясненной дисперсии	Кумулятивный %	Каноническая корреляция
1	5,497 ^a	96,5	96,5	,920
2	,198 ^a	3,5	100,0	,407

Для оценки информативности ДФ еще одной важной характеристикой является коэффициент канонической корреляции (R), роль который заключается в оценке степени тесноты связи между двумя переменными. При этом, чем ближе значение R к единице, тем и выше разделительная способность дискриминантной функции. В наших вычислениях для первой ДФ значение R равно 0,920, которое означает, что данная ДФ обладает превосходной способностью разделить данных в свои классы, а вторая функция обладает этой способностью в два раза меньше ($R = 0,407$), чем первой.

Таблица 4.

Нормированные коэффициенты канонической дискриминантной функции	
	Функция

	1	2
<i>Salt_100</i>	,021	,398
<i>SurfaseDensity</i>	,019	,439
<i>LeafSurface</i>	,999	,047
<i>SubstantContent</i>	-,056	-,572

В таблице 4 приведены стандартизованные коэффициенты дискриминантных функций. Эти коэффициенты, основанные на стандартизованных переменных, принадлежат к одной и той же шкале измерений, поэтому их можно сравнивать, чтобы определить величины и направления вкладов переменных в каждую каноническую функцию. Из таблицы видно, что наибольший вклад:

- в дискриминантную функцию 1 вносят переменные: *LeafSurface*; *SubstantContent*, *Salt_100*; *SurfaseDensity*

- в дискриминантную функцию 2 – *SubstantContent*, *SurfaseDensity*, *Salt_100*, *LeafSurface*.

На диаграмме, показанной на рисунке 2 видно, что наблюдения образцов хлопчатника, принадлежащие одинаковым группам, локализованы в определенных областях плоскости, при этом расстояния между центроидами групп Слабо устойчивые, Среднеустойчивые и Устойчивые не одинаковые. Это показывает на то, что «соседние» группы отличны друг от друга. Самое большее расстояние имеется между группами Слабоустойчивые и Устойчивые, вследствие чего они наиболее различны между собой.

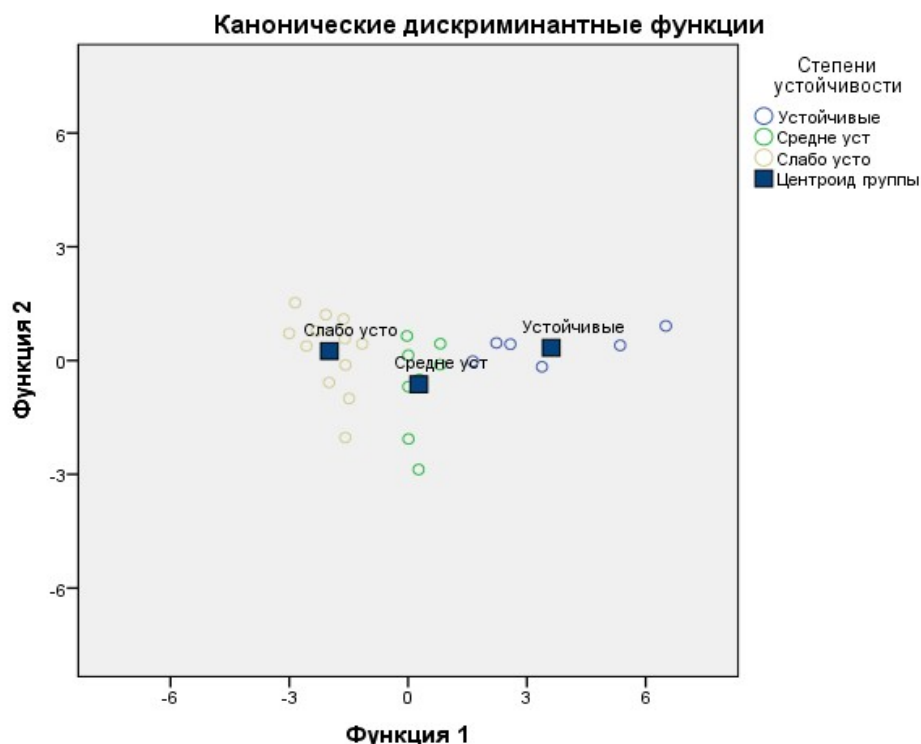


Рис.2. Диаграмма рассеяния канонических значений для пар значений дискриминантных функций 1 и 2.

Из диаграммы видно, что группы Слабоустойчивые и Устойчивые наиболее отдалены друг от друга, т.е. показателям в этих группах соответствуют большие значения Функции 1. Поэтому Функция 1 главным образом дискриминирует образцов между этими группами и Среднеустойчивую группу. Дискриминантная функция 2, дискриминирует образцов тоже как в случае Функция 1, и всех групп - друг с другом.

Классифицирующие функции, восстановленные на базе специальных математических методов ДА, предназначены отдельно для каждой группы и могут быть использованы для классификации новых представленных образцов растений хлопчатника с целью прогноза их солеустойчивости. Новый образец растения относится к той группе, для которой классификационная функция имеет наибольшее значение. В таблице 5, приведены коэффициенты и свободные члены при переменных восстановленных линейных функций.

Таблица 5.

Коэффициенты классифицирующей функции	
	Степени устойчивости

	Устойчивые	Средне устойчивые	Слабо устойчивые
<i>Salt_100</i>	-48,363	-51,339	-49,375
<i>LeafThick</i>	2494,683	2414,868	2470,870
<i>LeafSurface</i>	,046	,034	,027
<i>SubstantContent</i>	315,534	341,406	328,232
(Константа)	-133,871	-105,914	-92,219

Линейные дискриминантные функции Фишера

В конечном этапе работы программы ДА производит расчет классификационной матрицы (Таблица 6), которая является еще одним методом проверки качества дискриминации. Расчет этой матрицы осуществляется на базе переменных и коэффициентов по каждой в таблице 5 определенной функции классификации. Результатом правильной классификации объектов является процент отнесения их в свою группу. В данной таблице представлена информация

Таблица 6.

Результаты классификации ^а						
		Степени устойчивости	Предсказанная принадлежность к группе			Итого
			Устойчивые	Средне устойчивые	Слабо устойчивые	
Исходные	%	Устойчивые	83,3	16,7	,0	100,0
		Среднеустойчивые	,0	100,0	,0	100,0
		Слабо устойчивые	,0	,0	100,0	100,0

а. 96,2% исходных сгруппированных наблюдений классифицировано правильно.

о проценте корректно классифицированных наблюдений в каждой группе.

Строки матрицы — исходные классы, столбцы – предсказанные классы. Как видно из таблицы, исходные и предсказанные Среднеустойчивые и Слабо устойчивые классы полностью, на 100%, совпали, а Устойчивые – на 83,3%. Но в конечном счете суммарная точность правильной классификации составила 96,2% исходных сгруппированных наблюдений.

Заключение

Таким образом, на основе выявленных результатов программных расчетов и представленных исходных данных можно сделать следующие выводы:

1. Изменение количественных признаков листовых пластинок и других биометрических признаков листа может быть связано, вместе с засоленностью почвы и с разнообразием других локальных условий среды;
2. Результаты дисперсионного анализа показали, что межгрупповые различия по всем признакам достоверны на уровне $P < 0,05$;
3. Показатель сумма сульфатных и хлоридных солей на 100 г. почве вносит наибольший вклад в изменчивость признаки листа;
4. Показатель сумма сульфатных и хлоридных солей на 100 г. почве с показателем поверхностной плотности листовой пластинки коррелировал в очень высокой степени ($r = 0,89$).
5. Восстановлены две канонические дискриминантные функции с качественными характеристиками (λ -Уилкса = 0,128 и каноническая корреляция $R=0,92$) дифференцирующие исходных данных в свой класс с высокой вероятностью;
6. Разработаны для прогнозирования три модельные дискриминантные функции, позволяющие по вновь поступающим данным прогнозировать степень устойчивости хлопчатника;
7. Суммарная точность правильной классификации методом дискриминантных функций составила 96,2% исходных сгруппированных наблюдений.
8. В перспективе предполагается повышение разнообразия исследуемых групп и целенаправленное планирование опытно - полевых и лабораторных экспериментов с последующим привлечением разных методов анализа, в том числе и биохимических и молекулярно-биологических.

Использованная литература.

1. Saddiqe Z, Javeria S, Khalid H, Farooq A. Effect of salt stress on growth and

antioxidant enzymes in two cultivars of maize (*Zea mays* L.) *Pak J Bot.* 2016; 48:1361–1370.

2. Шерматов Ермат и др. Экспресс-метод оценки засоленности мелиорированных почв // EESJ East European Scientific Journal.

3. D. Hillel, “Salinity Management for Sustainable Irrigation: Integrating Science, Environment, and Economics,” World Bank Publications, Washington DC, 2000.

4. Ю.И. Широкова, А.К. Чернышев, Н.Ш. Шарафутдинова Опыт, задачи и перспективы улучшения мелиоративного контроля орошаемых земель. // Мелиорация: этапы и перспективы развития. Материалы международной научно-производственной конференции. – М., 2006. – С. 35-46.

5. Радкевич М.В., Юлчиев Д.Г., Арипов И. Использование хлопчатника для биоиндикации засоления почвы // *Universum: технические науки.* 2022. 2(95).

6. Мансуров Н.И. Физиолого-генетический анализ признаков фотосинтеза и продуктивности у хлопчатника: Автореф. Дис. ... канд. Биол. Наук / Н.И. Мансуров. - Душанбе ИФИБР АН РТ.- 1991.- с. 156.

7. Мелник М. Основы прикладной статистики. М. Энергоатомиздат. 1983. Глава 13. Регрессионный и корреляционный анализ. – 416 с.

8. Бьюль А., Цеффель П. SPSS: искусство обработки информации. – М., 2005. Глава 14. Корреляционный анализ.

9. Наследов, А.Д. SPSS 19. Профессиональный статистический анализ данных [Текст] // А. Д. Наследов. – СПб. Питер, 2011. – 400 с.

10. Гржибовский А. М., Иванов С. В., Горбатова М. А. Сравнение количественных данных двух независимых выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и Здравоохранение. 2016. №2. С. 5-28.

11. Гланц С. Медико-биологическая статистика. М. : Практика, 1998. 459 с.

12. Лакин Г.Ф. Биометрия. М. : Высшая школа, 1990. 351 с
13. Davis J.C. Statistics and data analysis in geology. – 3rd ed. – John Wiley & Sons, 2002.– 656 p.
14. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974. – 240 с.
15. Machine learning, neural and statistical classification / ed. D. Mitchie [et al.]. – Ellis Horwood, Chichester, 1994. – 304 p.
16. Li C.C. Introduction to experimental statistics. – New York: McGraw-Hill, Inc., 1964. – 460 p.
17. Cacoullos T. Discriminant analysis and applications. – London: Academic Press, Inc. ltd, 1972. – 434 p.
18. Ефимов В.М., Ковалева В.Ю. Многомерный анализ биологических данных: учеб. пособие. Горно-Алтайск: ГорноАлтайский гос. ун-т, 2007. 75 с.
19. Клекка У.Р. Дискриминантный анализ // Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика, 1989. С. 78–137.
20. Дж.Ту, Р. Гонсалес Принципы распознавания образов. М.:Мир, 1978.- 412 с.