СОВРЕМЕННЫЕ ПРОБЛЕМЫ ОПЫТНОГО ДЕЛА

Материалы международной научно – практической конференции

6 – 9 июня 2000 г.

ТОМ I

Санкт - Петербург
2000 г.

…количеству

…содержания его в суб-

Пр… позволяют сделать следующие основные выводы:

1. Урожайность томатов при выращивании на цеолитах Сокирницкого и Пегас-ского месторождений на 10-17% выше, чем на торфе верховом.
2. Способы внесения удобрений на цеолите (разовое и дробное) при одинаковой дозе азотных удобрений не оказали существенного влияния на урожайность томатов.
3. Минимальное количество нитратного азота (в 10 раз меньше ПДК) содержалось в плодах растений, выращенных на цеолитах при разовом внесении удобрений.
4. Способы размещения цеолитового субстрата (на пленке и без пленки) не оказали существенного влияния на урожайность томатов.

SUMMARY

The affectivity of zeolite substratums for glasshouse tomatoes growing was studied. The experiments were held during 3 years in winter block glasshouse of the joint-stock company "Niva" (Mockow oblast) with zeolite of Sokirnisky (West Ukraine) and Pegassky (Kemerovo oblast) deposist. Thecomparative effectivity of entire and subdivided methods of fertilizers applying was determined. The yield of tomatoes grown on zeolite substratums was found 10-17% higher than on peat (control) within 3 yeas on average. The methods of fertilizers applying didn't effect considerably on plant productivity and fruit guilty of tomatoes.

МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ. НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА

В.П. Якушев, В.М. Буре

Агрофизический научно-исследовательский институт

ВВЕДЕНИЕ

В математической статистике можно выделить два основных подхода к анализу статистических данных. Эти подходы обусловили возникновение двух больших групп методов статистического анализа. Первый подход основан на предположении, что распределение наблюдений, о которых идет речь, принадлежит некоторому параметрическому семейству распределений, например соответствующие случайные величины имеют нормальное, или гамма- , или пуассоновское, или какое-либо другое распределение.

То есть предполагается, что известна форма или семейство распределений, но неизвестны точные значения параметров, выделяющие из семейства распределений истинное распределение.

Чаще всего в практике анализа данных предполагается, что речь идет о нормальном распределении. Именно это предположение имеет принципиальное значение для обоснованного применения регрессионного и дисперсионного анализов, использовании всевозможных стандартных критериев проверки гипотез, например таких, как t – критерий, построения доверительных оценок среднего и дисперсии и т.д. На практике все эти методы анализа наблюдений часто применяют и в тех случаях, когда наблюдения распределены иначе, что иногда вносит серьезную погрешность в результаты статистического исследования. Выяснилось, что нарушения, кажущиеся незначительными и потому трудно обнаружимыми, могут привести к существенному смещению оценок, доверительных границ, коэффициентов доверия. Хорошо известно, что даже такая привычная для первой группы методов оценка, как выборочное среднее, теряет свои оптимальные свойства, если распределение генеральной совокупности заметно отличается от нормального распределения (например, при наличии резко выделяющихся наблюдений [5].), а если генеральная совокупность распределена по закону Коши, то использование среднего арифметического, по-видимому, не имеет смысла.

Следует отметить, что гауссовские модели характерны для таких дисциплин, как астрономия или геодезия, где было достигнуто высокое качество измерений. Ошибки в измерениях имели ясно выраженное нормальное распределение.

С постепенным переходом других наук на количественную основу появилась потребность в обработке данных, не имеющих гауссовского характера. Подобная ситуация характерна для биологии, экономики, социологии и ряда других дисциплин. К подобным данным трудно подобрать подходящее параметрическое семейство. К тому же выводы обычно надо делать по малым выборкам. Если априори закон распределения неизвестен, то по такой выборке установить тип закона невозможно. Использование статистических методов из первой группы в этих условиях представляется малообоснованным и может привести к серьезным ошибкам.

Другой подход не предполагает использования какого-либо определенного параметрического семейства распределений. Статистические методы, относящиеся к этому подходу, для своего обоснования не требуют наличия какого-либо определенного параметрического семейства распределений в качестве распределения генеральной совокупности, поэтому их называют непараметрическими или свободными от распределения. Единственное предположение, которое делается в большинстве непараметрических процедур, - это непрерывность распределения случайных величин. Непараметрические методы обладают рядом преимуществ перед гауссовскими. Основные из них - более широкое поле приложений, отсутствие требования о нормальном распределении генеральных совокупностей, для многих непараметрических методов требуются не действительные значения наблюдений, а только их ранги. Хотя на первый взгляд может показаться, что, применяя ранговые методы, мы теряем много существенной информации, содержащейся в выборке, теоретические исследования показали, что это не так. Обычно непараметрические процедуры лишь немного менее эффективны, чем их конкуренты из первой группы, если рассматриваемые генеральные совокупности нормальны, зато они оказываются значительно эффективнее методов из первой группы, если распределения случайных величин отличны от нормального [1].

Несмотря на то, что непараметрические методы начали развиваться значительно позднее, чем гауссовские, их возможности интенсивно расширялись и сегодня многие методы из первой группы имеют непараметрические эквиваленты. Так, например, часто применяемый в практике статистических исследований однофакторный дисперсионный анализ имеет непараметрический эквивалент (ранговые критерии Краскела-Уоллиса, Джонкхиера, оценка различия между эффектами обработки двух выборок посредством нахождения медианы Ходжеса-Лемана и т.д.). Имеется непараметрический аналог двухфакторного дисперсионного анализа.

В непараметрической статистике разработана корреляционная теория, основанная на ранговых коэффициентах корреляции.

Из литературы посвященной непараметрической статистике можно отметить следующие руководства [1,2,3,4].

Следует так же отметить, что большое количество работ по непараметрической статистике опубликовано в зарубежных биометрических журналах, что, по-видимому, отражает важность методов непараметрической статистики именно для экспериментальных исследований в биологии.

В каком-то смысле промежуточное положение между параметрическими и непараметрическими моделями занимают модели так называемой робастной статистики [5,6]..

ПРОВЕРКА ОДНОРОДНОСТИ ДВУХ ВЫБОРОК. КРИТЕРИЙ МАННА-УИТНИ

Назначение критерия – проверка гипотезы о статистической однородности двух выборок [7]. Иногда эту гипотезу называют гипотезой об отсутствии эффекта обработки (имея в виду, что одна из выборок содержит характеристики объектов, подвергшихся некоему воздействию, а другая – характеристики контрольных объектов).

Данные. Рассматриваются две выборки $x_1, x_2, \ldots, x_m$ (выборка x) и $y_1, \ldots, y_n$ (выборка y), объемов m и n соответственно. Обозначим закон распределения первой выборки через F, а второй - через G

Предположения. Выборки независимы, функции распределения непрерывны.

Проверяемая гипотеза. Утверждение об однородности выборок можно записать в виде H: F=G.

Альтернативы. Критерий Манна-Уитни предназначен для проверки H против альтернативы $F \ge G$ (т.е. для любого x выполняется $F(x) \ge G(x)$, правосторонняя альтернатива) или альтернативы $F \le G$. (левосторонняя альтернатива). Можно рассматривать и двусторонние альтернативы.

Статистика критерия. Условимся, что всякое событие $x_i < y_j$ обозначает «успех», а всякое событие $x_i > y_j$ - «неудачу». Изменяя i от 1 до m и j от 1 до n, получаем mn парных сравнений элементов выборок x и y. Обозначим число успехов в этих парных сравнениях через U. Очевидно, что U может принимать значения от 0 до mn. Статистика Манна-Уитни не имеет биномиального распределения, так как слагаемые являются зависимыми. При выполнении гипотезы H распределение статистики U не зависит от закона распределения выборок (если эти распределения непрерывны). Распределение U при гипотезе H зависит только от объемов выборок. В справочниках [1,2] приводятся таблицы, по которым можно найти вероятность $P(U \ge k)$ для различных k при небольших m и n.

При справедливости гипотезы H выполняется $P(x > y_j) = P(x < y_j) = 0.5$. Поэтому при H количества успехов и неудач должны быть приблизительно равны, т.е. не должны значительно отклоняться от mn/2.

При нарушении гипотезы H для правосторонних альтернатив значение U, как правило, должно превосходить mn/2, для левосторонних альтернатив соотношение обратное.

Отмеченное свойство позволяет построить критерий для проверки гипотезы H против односторонних альтернатив. Либо по заданному уровню значимости α по таблицам для правосторонних альтернатив находим критическое значение $U_1(m,n)$, что $P(U \ge U_1(m,n)) = \alpha$ при этом критическая область для гипотезы H против правосторонней альтернативы будет иметь вид:

$$U \ge U_1(m,n),$$