

ДОКЛАДЫ

Российской академии
сельскохозяйственных наук

Научно-теоретический журнал

5

сентябрь - октябрь

2006

ОЦЕНКА БИОЭКВИВАЛЕНТНОСТИ ДВУХ УЧАСТКОВ НА СЕЛЬСКОХОЗЯЙСТВЕННОМ ПОЛЕ

Член-корреспондент Россельхозакадемии В.П.Якушев, В.М.Буре

Предложен алгоритм статистической оценки однородности двух участков на сельскохозяйственном поле по имеющимся экспериментальным данным об уровне их средней урожайности за несколько предыдущих лет.

Процесс получения растениеводческой продукции распределен во времени и пространстве на конкретной территории, неоднородной даже в пределах одного поля или его части. Поэтому проводимые технологические операции должны быть дифференцированы не только во времени (с учетом изменчивости погодных условий) и по полям севооборота, но и в пределах одного поля. Появившиеся в последнее время новые технические средства и успехи информатики обеспечивают переход к прецизионным технологиям в сельскохозяйственном производстве. Адаптивно-ландшафтная методология и применение технологий точного земледелия объективно являются во всем мире одними из самых перспективных направлений совершенствования производства растениеводческой продукции [1, 2].

Принципиальное отличие новой концепции состоит в том, что каждое сельскохозяйственное поле рассматривается как неоднородное. В этом случае его можно разделять на некоторое количество новых единиц управления, которые являются однородными (квази-однородными) участками. Однако выде-

ление однородных участков на заданном поле представляет собой объективно сложную задачу. Для ее решения важны сведения об урожайности с точной привязкой к месту (карты урожайности). Такие данные имеют уникальное значение, так как нет более объективного показателя неоднородности сельскохозяйственного поля, чем карта урожайности. Она позволяет судить о степени варьирования почвенно-климатических условий внутри поля и принимать обоснованные технологические решения. Представляется возможным также более дифференцированно проектировать технологии возделывания сельскохозяйственных культур применительно к различным по урожайности участкам. Вместе с тем неясно, каким образом по величине урожайности на тех или иных участках заданного сельскохозяйственного поля можно судить о степени их однородности или неоднородности.

В настоящей работе рассмотрен один из возможных подходов к оценке биоэквивалентности двух участков по урожайности и предложен соответствующий алгоритм, использующий идею статистическо-

го моделирования выборок из нормальных распределений с заданными параметрами математических ожиданий и дисперсий по каждому анализируемому участку. В биометрической литературе имеется большое количество работ, посвященных изучению биоэквивалентности [3-5]. В частности, в работе [3] сформулировано понятие популяционной биоэквивалентности на основе анализа соответствующих экспериментальных данных.

Методика. Предположим, что на некотором поле можно выделить два участка A и B . Требуется принять решение о степени их однородности или неоднородности по уровню средней урожайности, то есть о биологической эквивалентности. Обозначим средние урожайности на выделенных участках как Y_A (на участке A) и Y_B (на участке B). Допустим, что каждый участок состоит из большого числа небольших по площади делянок, тогда средняя урожайность представляет собой суммарный урожай всех делянок, поделенный на их количество. Урожайность различается по делянкам и является реализацией случайной величины, распределение которой неизвестно. Следовательно, средняя урожайность участка – также случайная величина, подчиняющаяся некоторому распределению. Биологическая эквивалентность участков означает совпадение распределений этих случайных величин (средних урожайностей) или достаточную близость (сходство) распределений. В дальнейшем будем считать эти случайные величины взаимно независимыми; подобное допущение оправдано, если участки достаточно большие по площади.

Предположим, имеются две выборки, представляющие собой значения средних урожайностей участков A и B за несколько предыдущих лет: $Y_{A1}, \dots, Y_{An}, \dots, Y_{B1}, \dots, Y_{Bn}$. Допустим, что распределения введенных случайных величин Y_A и Y_B – нормальные с параметрами μ_A, μ_B (математические ожидания) и σ_A^2, σ_B^2 (дисперсии), так как средние значения представляют нормированные суммы урожайностей небольших делянок, составляющих рассматриваемые участки, и, следовательно, при выполнении дополнительных предположений технического характера применима центральная предельная теорема Ляпунова.

Результаты и обсуждение. Будем считать, что рассматриваемые участки биологически эквивалентны тогда, когда распределения случайных величин Y_A, Y_B одинаковы или достаточно близки. Даже при одинаковых распределениях значения случайных величин, представляющих собой среднюю урожайность участков, в каждом конкретном испытании окажутся различными в силу предполагаемой независимости. В качестве меры схождения случайных величин Y_A, Y_B естественно выбрать вероятность

$$F_{A,B}(t) = P\{|Y_A - Y_B| \leq t\}. \quad (1)$$

Если участки A и B биологически эквивалентны, то, как уже отмечалось, распределения средних урожайностей должны быть одинаковыми или достаточно близкими, поэтому вероятность (1) следует сравнивать с вероятностью

$$F_A(t) = P\{|Y_A - Y_{\bar{A}}| \leq t\}, \quad (2)$$

где случайная величина $Y_{\bar{A}}$ независима от Y_A и имеет такое же распределение. В качестве величины t в формулах (1) и (2), как это было предложено в работе [3], выберем $t\sigma_A\sqrt{2}$. Тогда вероятности (1) и (2) можно записать с использованием функции стандартного нормального распределения $\Phi(t)$:

$$F_{A,B}(t) = P\{|Y_A - Y_B| \leq t\sigma_A\sqrt{2}\} = \Phi\left(\frac{t\sigma_A\sqrt{2} + \mu_B - \mu_A}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right) - \Phi\left(\frac{-t\sigma_A\sqrt{2} + \mu_B - \mu_A}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right),$$

соответственно

$$F_A(t) = P\{|Y_A - Y_{\bar{A}}| \leq t\sigma_A\sqrt{2}\} = \Phi(t) - \Phi(-t).$$

Близость функций $F_{A,B}(t)$ и $F_A(t)$ показывает, что участки A и B можно считать биоэквивалентными. Более строго биоэквивалентность можно определить в терминах разности или отношения введенных функций, для чего необходимо ввести пороговые значения f_d для разности и f_r для отношения.

Будем считать участки A и B биоэквивалентными, если

$$F_{A,B}(t) - F_A(t) \geq f_d \text{ или } F_{A,B}(t) / F_A(t) \geq f_r, \quad (3)$$

где $f_d < 0, f_r < 1$.

Очевидно, большее значение вероятности означает более высокую степень «близости» между средними урожайностями, что свидетельствует о более высокой биологической эквивалентности участков.

С использованием имеющихся выборок построим оценки неизвестных параметров $\mu_A^*, \mu_B^*, \sigma_A^{2*}, \sigma_B^{2*}$, после чего подставив их в функцию $F_{A,B}(t)$, получим статистическую оценку искомой вероятности: $F_{A,B}^*(t)$, далее можно сравнить найденную оценку с $F_A(t)$. Заметим, что функция $F_A(t)$ полностью известна. Проблема заключается в том, что вместо истинной вероятности $F_{A,B}(t)$ в сравнении используется статистическая оценка $F_{A,B}^*(t)$. В связи с этим можно предложить следующий алгоритм, использующий идею статистического моделирования выборок из нормальных распределений с параметрами $\mu_A^*, \mu_B^*, \sigma_A^{2*}, \sigma_B^{2*}$ (отдельно для участка A и участка B).

1. По полученным в результате многолетних наблюдений опытным данным строятся оценки $\mu_A^*, \mu_B^*, \sigma_A^{2*}, \sigma_B^{2*}$.

2. Моделируются на компьютере две выборки: Y_{A1}, \dots, Y_{An} из нормального распределения с параметрами μ_A, σ_A^2 и Y_{B1}, \dots, Y_{Bn} из нормального распределения с параметрами μ_B, σ_B^2 .

3. По полученным выборкам строится новая оценка $F_{A,B}^*(t)$, после чего для ранее выбранного порогового значения f_d (или f) проверяются условия биоэквивалентности (3) в форме разности или отношения.

4. Пункты 2 и 3 многократно повторяются и подсчитывается относительная частота выполнения условия биоэквивалентности (3).

5. Принимается решение о правомерности гипотезы биоэквивалентности участков или об отклонении этой гипотезы.

Рассмотрим применение предложенного алгоритма на двух примерах.

Пример 1. Моделировали разность средней урожайности каждого участка за вычетом многолетней средней урожайности всего поля. Предположим, что $\mu_A = 0, \sigma_A^2 = 1, \mu_B = 0,5, \sigma_B^2 = 1$. Смоделирована выборка опытных данных за 10 лет ($n=10$). Найдены оценки: $\mu_A^* = -0,49, \sigma_A^{*2} = 0,46, \mu_B^* = 0,19, \sigma_B^{*2} = 0,93$. В качестве порогового значения для разности вероятностей принято значение $f_d = -0,3$. Проведено 50 испытаний (пункты 2 и 3 алгоритма), в которых 40 раз выполнено условие биоэквивалентности и 10 раз оно нарушено. Приведен график (рис.а) разности функций вероятностей для одного из испытаний, когда условие биоэквивалентности выполнено. Выборки содержали по 10 наблюдений. Результаты примера 1 показывают, что, как правило, различия между участками А и В незначительны и, следовательно, их можно рассматривать как биоэквивалентные. Все вычисления проводили в программе Maple 7:

$y_A := \text{stats}[\text{random}, \text{normald}] (10)$ – опытные данные для участка А;

$y_B := \text{stats}[\text{random}, \text{normald} [0,5,1]] (10)$ – опытные данные для участка В;

$f := (t) \rightarrow -0,3$ – пороговое значение для разности вероятностей;

$\text{plot}([g-z, f], 0..8, \text{color}=[\text{blue}, \text{red}], \text{style}=[\text{line}, \text{point}])$.

Пример 2. Моделировали разность средней урожайности каждого участка за вычетом многолетней средней урожайности всего поля. Предположим, что $\mu_A = 0, \sigma_A^2 = 1, \mu_B = 2,5, \sigma_B^2 = 1$. Смоделирована выборка опытных данных за 10 лет ($n=10$). Найдены оценки: $\mu_A^* = 0,54, \sigma_A^{*2} = 0,84, \mu_B^* = 2,65, \sigma_B^{*2} = 0,38$. В

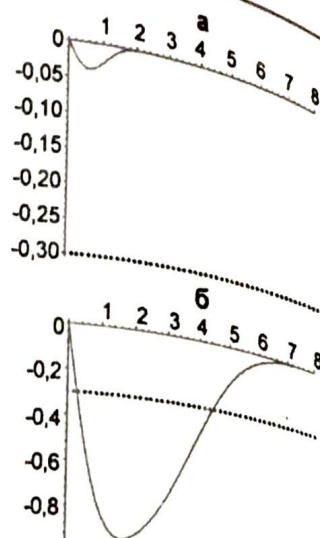


График разности вероятностей с линией порогового значения; а, б – пример 1 и 2.

качестве порогового значения для разности вероятностей принято значение $f_d = -0,3$. Проведено 50 испытаний (пункты 2 и 3 алгоритма), в которых 50 раз условие биоэквивалентности не выполнено. Приведен график (рис.б) разности функций вероятностей для одного из испытаний. Выборки содержали по 10 наблюдений. Результаты примера 2 показывают, что различия между участками А и В значительны. Все вычисления проводили в программе Maple 7:

$y_A := \text{stats}[\text{random}, \text{normald}] (10)$ – опытные данные для участка А;

$y_B := \text{stats}[\text{random}, \text{normald} [2,5,1]] (10)$ – опытные данные для участка В;

$f := (t) \rightarrow -0,3$ – пороговое значение для разности вероятностей;

$\text{plot}([g-z, f], 0..8, \text{color}=[\text{blue}, \text{red}], \text{style}=[\text{line}, \text{point}])$.

Рассмотренные примеры показывают работоспособность предлагаемого алгоритма. Он может быть реализован и в других программных продуктах, включая Excel, и позволяет при задаваемых различных уровнях урожайности проводить компьютерный анализ и оценку степени однородности тех или иных попарно сравниваемых участков на заданном сельскохозяйственном поле.

Литература. 1. *Агроэкологическая оценка земель, проектирование адаптивно-ландшафтных систем земледелия и агротехнологий. Методическое руководство.* – М.: Росинформагротех, 2005. 2. *Якушев В.П.* На пути к точному земледелию. – СПб.: Изд-во ПРИФ РАН, 2002. 3. *Shall R.* // *Biometrics*–1995.–V.51.–P.615-626. 4. *Westlake, W.J.* // *Biometrics.* – 1976. – V.32. – P. 741-744. 5. *Berger R.L., Hsu J.C.* // *Statistical science.* – 1996. – V.11. – N 4. – P.283-319.

Агрофизический научно-исследовательский институт, 195220, Санкт-Петербург
Поступила в редакцию 26.12.05

Yakushev V.P., Bure V.M. Approach to evaluating bioequivalence of two plots on an agricultural field

Suggested algorithm of statistical evaluating the homogeneousness of the two plots on an agricultural field is based on available experimental data about the level of average crop yield in the course of several previous years.