

УДК 631.421.1:631.524.01:633.11 "321"

СОВЕРШЕНСТВОВАНИЕ БИОМЕТРИЧЕСКОГО МЕТОДА
ОПТИМИЗАЦИИ ВЫБОРКИ РАСТЕНИЙ
ПРИ ЭКСПЕРИМЕНТАЛЬНОМ СРАВНЕНИИ
ГИБРИДНЫХ ПОПУЛЯЦИЙ

А.В. СМИРЯЕВ, Т.И. ХУПАЦАРНЯ, С.С. БАЖЕНОВА, НГУЕН ТХАНЬ ТУАН

(РГАУ-МСХА имени КА. Тимирязева)

Ранее был разработан биометрический метод оценки влияния выборки растений, измеряемых на каждой делянке полевого опыта, на точность сравнения гибридных популяций по среднему значению, а также дисперсии количественного признака. Метод усовершенствован с учетом неравноточности оценок дисперсий, для ситуаций с асимметрией и эксцессом распределения признака. Предложена методика скользящей дисперсии, предназначенная для прогноза влияния протяженности блока в опыте на достаточный объем выборки измеряемых растений. Используются данные о 7 признаках у 17 гибридных популяций F_3 , F_4 и F_5 мягкой яровой пшеницы.

Ключевые слова: биометрия, точность сравнения гибридных популяций, объем выборки растений, дисперсионный анализ, неравноточность, асимметрия, эксцесс, имитационное моделирование, скользящая дисперсия, яровая пшеница.

Проблема точности полевого опыта рассматривалась многими авторами. В литературе изучены различные причины, влияющие на точность сравнения образцов культурных растений.

Краевой эффект. В частности, установлено, что урожайность в крайних рядах некоторых опытов может превышать среднюю по делянке на 30-40% [11]. При стандартном размере делянки в зону краевого эффекта попадает до 40% растений.

Форма и размер делянки. В частности, показано, что квадратные делянки в опытах с зерновыми культурами имеют существенно меньшую вариацию основных признаков и, следовательно, более точные измерения по сравнению с вытянутыми делянками [3, 10, 12].

Влияние конкуренции. Конкуренция растений внутри каждого селекционного образца отрицательно сказывается на точности их сравнения [8]. Установлено, что минимальное влияние конкуренции проявляется при посеве с оптимальным расстоянием между растениями, разным для каждой культуры и признака [9]. Однако в селекционных питомниках необходимо соблюдать принцип типичности.

Выборка измеряемых растений. В ранних звеньях селекционного процесса сложно выполнить требования, предъявляемые к полемому опыту, и основные причины этого — малое количество семян для посева каждого из вариантов и многовариантность. При этом анализ полученного растительного материала и, в частности, измерение количественных признаков растений является довольно трудоемким. По-

этому важной задачей является разработка рекомендаций о достаточном объеме (m) выборки растений, измеренных на каждой делянке полевого опыта, т.е. таком объеме, увеличение которого незначительно повышает точность сравнения образцов.

Начиная с 2006 г. в РГАУ-МСХА имени К.А. Тимирязева разрабатывается соответствующий биометрический метод определения достаточного объема m . В результате получены оценки влияния объема на точность сравнения форм по среднему значению количественного признака [2]. Проверка и разработка рекомендаций проводились на двухлетних данных, полученных по 10 сортам гексаплоидной озимой тритикале.

Проведена модификация этого метода [5], предназначенная для сравнения гибридных популяций по их внутривидовой дисперсии признака, как показателя генетической изменчивости, доступной для отбора. Проверка модифицированного метода и выработка рекомендаций были успешны на данных о семи признаках 17 гибридных популяций мягкой яровой пшеницы (F_2 — 2008 г., F_3 — 2009 г., и F_5 — 2011 г.). Однако в 2010 г. проверка для трех из семи количественных признаков у 17 популяций /•', выявила отклонения в методе, свидетельствующие о существенных нарушениях модельных предположений дисперсионного анализа. Следовало установить причины отклонений и, по возможности, разработать рекомендации для их устранения.

Кроме того, разработанный метод проверялся на опытах с одинаковым по годам числом делянок в рандомизированных блоках. Не учитывалось возможное влияние на полученные рекомендации изменений общей протяженности блока, принимая во внимание градиенты почвенного плодородия. Желательно разработать специальную биометрическую методику, позволяющую без проведения дополнительных опытов прогнозировать достаточное число (m) растений, измеренных на каждой делянке, при изменении протяженности блока.

Цель работы — разработка и проверка усовершенствованного метода с учетом названных недостатков и ограничений.

Материалы и методика

Использовали 17 гибридных популяций F_3 (2009), F_4 (2010), и F_5 (2011) без отбора, полученных после попарного скрещивания восьми сортообразцов яровой мягкой пшеницы из коллекции системы ВИР имени Н.И. Вавилова [5].

Опыты проводили на поле лаборатории селекции и семеноводства полевых культур РГАУ-МСХА имени К.А. Тимирязева. Ежегодно посев 17 гибридных популяций в трех рандомизированных блоках был произведен вручную. Каждая делянка состояла из пяти учетных и двух защитных рядков по 20 зерен в каждом. Размер делянки: 1х1,26 м. Посев проводили поперёк полосы шириной 1 м с междурядьями 18 см, между соседними делянками расстояния 36 см, межполосная дорожка — 50 см. Уборку проводили вручную в фазе полной спелости, защитные и учетные рядки убирали отдельно.

Общее количество растений в учетных рядках из-за негативного влияния погодных условий разных лет колебалось в пределах от 70 до 100. На каждой делянке со всех пяти учетных рядков было использовано для измерений по 30 случайно выбранных растений (опорное число для биометрического анализа). У растений оценивали следующие признаки: 1 — высота растений; 2 — длина колосового стержня главного колоса; 3 — продуктивная кустистость; 4 — число озерненных колосков главного колоса; 5 — число зерен с растения; 6 — масса зерен с растения; 7 — масса 1000 зерен.

В основе вычисления любого доверительного интервала для парного сравнения количественного параметра (среднего значения признака, дисперсии и т.п.) у селекционных образцов в однолетнем полевом опыте лежит нахождение $SI^2_{\text{ош}}$ — среднего квадрата ошибки опыта (MS остатка), получаемого из таблицы обычного дисперсионного анализа [1].

В частности, НСР = $t_T \sqrt{\frac{2SI^2_{\text{ош}}}{n}}$, где n — число повторений опыта, t_T — таб-

личное значение критерия Стьюдента.

Методика предложенной оценки достаточного (оптимального) объема выборки растений, измеренных на каждой делянке полевого опыта, основана на значении параметра относительной ошибки K (прогнозирует изменение ошибки опыта при любом конкретном значении m растений по сравнению с ошибкой при реальном числе растений, измеренных на каждой делянке в опыте) [5]:

$$K = \sqrt{\frac{SI^2_{\text{ош}}(m)}{SI^2_{\text{ош}}(30)}}$$

Формула дана для случая измерения 30 случайно отобранных растений на каждой делянке опыта (30 — опорное число растений). В этой формуле $SI^2_{\text{ош}}(30)$ — средний квадрат ошибки опыта (MS остатка) из дисперсионного анализа экспериментальных данных в трех повторениях. $SI^2_{\text{ош}}(m)$ — также средний квадрат ошибки опыта, но прогнозируемый по формулам математической статистики для любого числа (m) растений, измеренных на делянке. При сравнении гибридных популяций по среднему значению признака параметр K далее имеет обозначение $K_{\text{ср}}$ для сравнения дисперсии признака у гибридных популяций — K_D .

В формуле параметра K используется квадратный корень, так как в НСР для парного сравнения вариантов опыта входит именно корень из $SI^2_{\text{ош}}$. Увеличение числа m вызывает снижение $SI^2_{\text{ош}}(m)$ и как следствие снижение параметра ошибки K , что означает увеличение точности сравнения в опыте и наоборот.

Предлагается следующий 10%-й критерий подбора достаточного числа растений, измеренных на делянке при шаге прогноза в 10 растений. Если оценка параметра относительной ошибки (K) при изменении объема выборки, например, на 10 растений ($m = 30 \pm 10$) осталась в пределах $0.9 \leq K \leq 1.1$, то считается, что это изменение по отношению к опорному числу 30 не влияет существенно на относительную точность сравнительной оценки образцов в опыте. Тогда m можно даже сократить до 20 растений на делянке. Если при снижении объема до $m = 30 - 10 = 20$ растений параметр $K > 1,1$, то это снижение нежелательно: НСР, как ожидается, увеличится минимум на 10% по сравнению с НСР при 30 измеренных растений. Если же при увеличении m на 10 (до $w = 30 + 10 = 40$ измеренных растений) параметр $K < 0,9$, точность сравнения образцов должна повыситься существенно: НСР сократится минимум на 10%. Такое увеличение m желательно.

Методика оценки параметра K на основе экспериментальных данных состоит в следующем. Получив из дисперсионного анализа реальных данных опыта значение $SI^2_{\text{ош}}(30)$ — среднего квадрата ошибки для опорного числа растений, можно, как отмечалось выше, оценить $SI^2_{\text{ош}}(m)$ — средний квадрат ошибки для любого числа рас-

тений (m), измеренных на делянке. Для этого учитываем, что на $\Pi^2_{\text{о||}}(30)$ влияют две независимые составляющие.

Первая составляющая — это S^2_m — дисперсия ошибки, которая останется даже для очень большого числа растений, измеренных на каждой делянке. Она обусловлена градиентами — тенденциями изменений почвенного плодородия, т.е. трендами характеристик почвы вдоль каждого блока, а также случайными, бессистемными изменениями параметров почвы по делянкам («пятнистость» плодородия на поле).

Вторая составляющая, оказывающая влияние на $SF_{\text{ian}}(30)$. это $\Pi(30)$ — дополнительная дисперсия ошибки выборочности, возникающая из-за ограниченного числа растений, измеренных на делянке (в данном случае 30).

Таким образом, получаем

$$SF^2_{\text{ou}}(30) = S^2_{\infty} + S^2(30). \quad (1)$$

Если известно значение $S^2(30)$. то из этой формулы можно оценить S_{-T} :

$$S^2_{\infty} = SF^2_{\text{ou}}(30) - S^2(30). \quad (2)$$

Для определения $S^2(30)$ обозначим через D_i — оценку дисперсии признака на i -й делянке опыта по 30 растениям. Как известно, для любого числа m измеренных растений

$$D_i = \sum_j (x_j - x_i)^2 / (m - 1),$$

где x_j — величина изучаемого признака на j -м растении, x_i — среднее значение по m растениям на i -и делянке опыта.

Напомним, что целью сравнительных экспериментов с n повторениями считалась оценка значимости различий изучаемых популяций либо по среднему значению признака ($\Sigma x_k / n$), либо по дисперсии признака ($\Sigma D_k / n$). Здесь суммирование ведется по номерам повторений ($k = 1 \dots n$) для каждой популяции, где n — общее число повторений-блоков опыта.

Рассмотрим ситуацию сравнения средних значений количественного признака у изучаемых популяций. Дисперсия ошибки среднего на i -и делянке, как известно, равна D_i / m . В частности, при $m = 30$ получаем $D_i / 30$. Вычисляем параметр $S^2(30)$ для всего опыта, усредняя эти дисперсии ошибки выборочности растений по всем делянкам (17 популяций \times 3 повторения = 51 делянка):

$$S^2(30) = \frac{\sum_1^{17} \sum_1^3 (D_i / 30)}{17 \times 3}. \quad (3)$$

В модификации метода для сравнения гибридных популяций по значениям дисперсии количественного признака используем формулу дисперсии ошибки выборочной оценки D_i из справочника Ллойда и Ледермана [4]: $2D_i^2(m-1)$. Поэтому для усреднения этих дисперсий ошибки выборочности растений по 51 (17 \times 3) делянке с опытными популяциями F_3 , F_4 или F_5 использована формула

$$S^2(30) = \frac{\sum_1^{17} \sum_1^3 (2D_i^2 / 29)}{17 \times 3}. \quad (4)$$

В обоих случаях сравнения, оценив $S^2(30)$, легко получить оценку прогнозируемого параметра $S^2(m)$ для любого заданного числа измеряемых на делянке растений (m) по формуле

$$S^2(m) \approx S^2(30) \times 30/m. \quad (5)$$

Тогда $SI_{\text{ош}}^2(m)$ — прогнозируемый средний квадрат ошибки для любого m вычисляется по формуле

$$SI_{\text{ош}}^2(m) = S_{\infty}^2 + S^2(m) \quad (6)$$

где значение S_{∞}^2 предварительно получаем из формулы (2) при $m = 30$.

Далее для оценки K при любом m используется формула

$$K = \frac{\sqrt{SI_{\text{ош}}^2(m)}}{\sqrt{SI_{\text{ош}}^2(30)}} = \frac{\sqrt{S_{\infty}^2 + S^2(m)}}{\sqrt{SI_{\text{ош}}^2(30)}}. \quad (7)$$

В частности, для очень большого числа m растений (по отношению к опорному числу 30) величина $S^2(m)$ близка к 0, что следует из формулы (5). В этом случае из формулы (7) получаем оценку $K(m \rightarrow \infty)$:

$$K(m \rightarrow \infty) = \sqrt{\frac{S_{\infty}^2}{SI_{\text{ош}}^2(30)}}. \quad (8)$$

Преобразование экспериментальных данных, стабилизирующее дисперсию ошибки опыта.

Основным предположением модели дисперсионного анализа является нормальное распределение ошибок опыта в повторениях, причем с одинаковой по вариантам дисперсией (равноточность оценок) [7]. Однако при сравнении популяций по дисперсии (D) признака ошибка ее оценки на любой делянке, как отмечалось выше, равна $2D^2/(m - 1)$, т.е. дисперсия ошибки не просто зависит от самого значения основного оцениваемого показателя D , но пропорциональна его квадрату. Эта связь может вызвать существенную неравноточность оценок D на делянках. Возможно, именно неравноточность вызвала отклонения при анализе данных 2010 г., экстремального по условиям выращивания.

Для снятия влияния связи оценок D с дисперсиями их ошибок воспользовались известным методом поиска преобразования случайной величины, стабилизирующей дисперсию [4]. Если случайная величина — оценка \hat{D} имеет математическое ожидание D и дисперсию, приблизительно равную $2\hat{D}^2/(m - 1)$, то следует подобрать такое преобразование $Y = h(D)$, чтобы было выполнено:

$$\sigma^2(Y) \approx [2\hat{D}^2/(m-1)][h'(\hat{D})]^2 = c, \quad (9)$$

где $h'(\hat{D})$ — производная функции Y по D при $D = \hat{D}$, а c — постоянная величина.

Следовательно $\sigma^2(Y)$ — дисперсия ошибки преобразованной случайной величины $Y = h(\hat{D})$ уже не будет зависеть от D .

Из (9) получаем $h'(\hat{D}) = c^{0.5} [2\hat{D}^2/(m-1)]^{-0.5}$.

Следовательно, постоянство дисперсии (9) будет приблизительно достигнуто, если

$$h(\hat{D}) = c^{0.5} \int [2\hat{D}^2/(m-1)]^{-0.5} d\hat{D}, \quad (10)$$

где \int — знак интеграла.

Решение (10) дает преобразование $Y = \ln(\hat{D})$, т.е. «логарифм \hat{D} » обеспечит постоянство дисперсии ошибки новой случайной величины Y . Эта дисперсия зависит только от m и для любого признака приблизительно равна $2/(m-1)$. В частности, при $m = 30$ получаем постоянное значение $2/(m-1) \approx 0,07$.

Поскольку $\ln(\hat{D})$ — монотонно возрастает по \hat{D} , то преобразованные значения $\ln(\hat{D})$ можно вместо самих значений \hat{D} использовать для сравнения популяций по их генетическому разнообразию (D). В частности, можно проводить дисперсионный анализ значений $\ln(\hat{D}_{ij})$, где i — номер популяции, j — номер повторения в опыте.

Имитационное моделирование для оценки влияния асимметрии и эксцесса

В качестве известных причин нарушения модельных предположений дисперсионного анализа в работе также рассмотрено влияние отклонений от нормального распределения исходного количественного признака, измеряемого в опытах. В частности, влияние варьирования по популяциям асимметрии (рис. 1) и эксцесса (рис. 2) распределения.

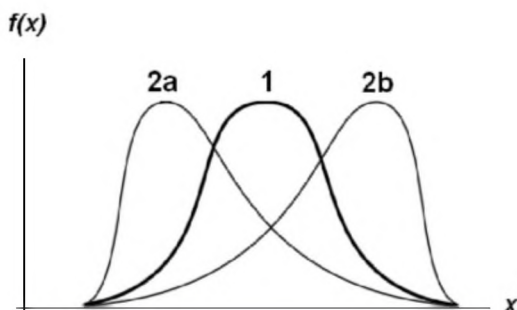


Рис. 1. Кривая 1 — нормальное распределение; 2 — плотности распределения с асимметрией: 2a — положительная асимметрия, 2b — отрицательная асимметрия

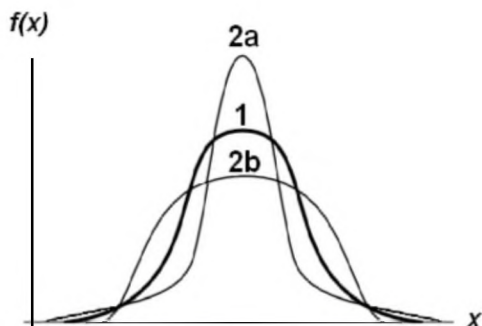


Рис. 2. Кривая 1 — нормальное распределение; 2a — положительный эксцесс (крутовершинное распределение), 2b — отрицательный эксцесс (плосковершинное распределение)

Сложное влияние асимметрии и эксцесса исходных измеренных признаков 30 растений на результаты сравнения генетического разнообразия (D) селекционных образцов практически невозможно оценить аналитически. Поэтому в работе использовали метод имитационного моделирования, для чего была разработана специальная компьютерная программа на языке BASIC. В основе — генерация методом Монте-Карло «данных измерения признака у 30 растений на делянке».

В программе 500 раз генерируется выборка 30 случайных величин x , подчиняющихся распределению с заданным генеральными параметрами: математическим ожиданием, дисперсией (D), асимметрией (As) и эксцессом (Ex). Рассчитываются выборочная оценка D для каждой выборки из 30 значений x . К этой оценке применяется преобразование, выравнивающее дисперсии ошибки оценки D : $Y = \ln(D)$. Вычислялась σ_y — дисперсия ошибки полученных значений $Y = \ln(D)$ по 500 повторам.

Затем меняют один из генеральных параметров и повторяют весь расчет. Таким способом удается оценить влияние варьирования генеральных параметров D , As , Ex , на σ_y — дисперсию ошибки преобразованного показателя $Y = \ln(D)$.

Методика скользящего дисперсии

Желательно без проведения дополнительных экспериментов корректировать рекомендации по объему m выборки измеряемых растений при изменении числа сравниваемых селекционных образцов (делянок в блоке) на том же поле, точнее при тех же градиентах — трендах почвенного плодородия. Так, очевидно, что увеличение числа аналогичных делянок (например, в опыте следующего года) приведет к «удлинению» каждого блока-повторения в опыте. Это, как предполагается, не повлияет закономерно на показатель $S^2(30)$ в уравнении (1) и на вклад в ошибку «пятнистости» почвенного плодородия. Но удлинение участка под опытом может привести к увеличению S^2_∞ за счет усиления влияния трендов плодородия. Последнее вызовет сокращение вклада параметра $S^2(30)$ в относительную ошибку K . Следовательно, в опыте с «удлиненными» блоками рекомендации относительно m могут измениться в сторону снижения, и наоборот, в сторону увеличения — при уменьшении числа делянок в блоке. Для количественной корректировки рекомендаций на m необходимо построить прогноз зависимости параметра S^2_∞ от длины блока или от w — общего числа стандартных делянок в нем.

Для этого по аналогии с названием скользящей средней [6] была разработана методика скользящей дисперсии. В зависимости от целей сравнения популяций в данном методе используется один из двух индикаторов влияния градиентов почвенного плодородия в опыте на параметр S^2_∞ . Если стоит цель — сравнение по среднему значению количественного признака, то используется $d_i(X_{cp})$ — дисперсия оценок средних значений (X_{cp}) селекционного признака группы популяций на i соседних делянках блока. Если же поставлена цель — сравнение популяций по дисперсии признака, то $d_i(D)$ — дисперсия оценок показателя D для группы из i соседних делянок.

Так, по данным каждого однолетнего опыта с гибридными популяциями яровой пшеницы, где в каждом блоке было 17 делянок, число соседних делянок варьировали в расчетах от $i_{\min} = 7$ до $i_{\max} = 17$ и оценивали 11 значений скользящей дисперсии. Выбор группы смежных делянок для получения элементов при оценке каждого значения d_i из 11 поясняет рисунок 3.

Элементы скользящей дисперсии d_i определяют в группе из семи делянок: с 1-й по 7-ю и получают дисперсию по ним — первый элемент $d_{7-1} = \sum_{j=1}^7 (P_j - P_{7-1})^2 / (7-1)$, где P_j — значение изучаемого параметра (X_{cp} или D) у популяции на j -й делянке, P_{7-1} — среднее значение на первых семи смежных делянках блока. Затем оценивают второй элемент $d_{7-2} = \sum_{j=2}^8 (P_j - P_{7-2})^2 / (7-1)$, где P_{7-2} — среднее значение признака со 2-й по 8-ю делянку, затем третий — с 3-й по 9-ю (d_{7-3}), ... с 11-й по 17-ю (d_{7-11}). Сред-

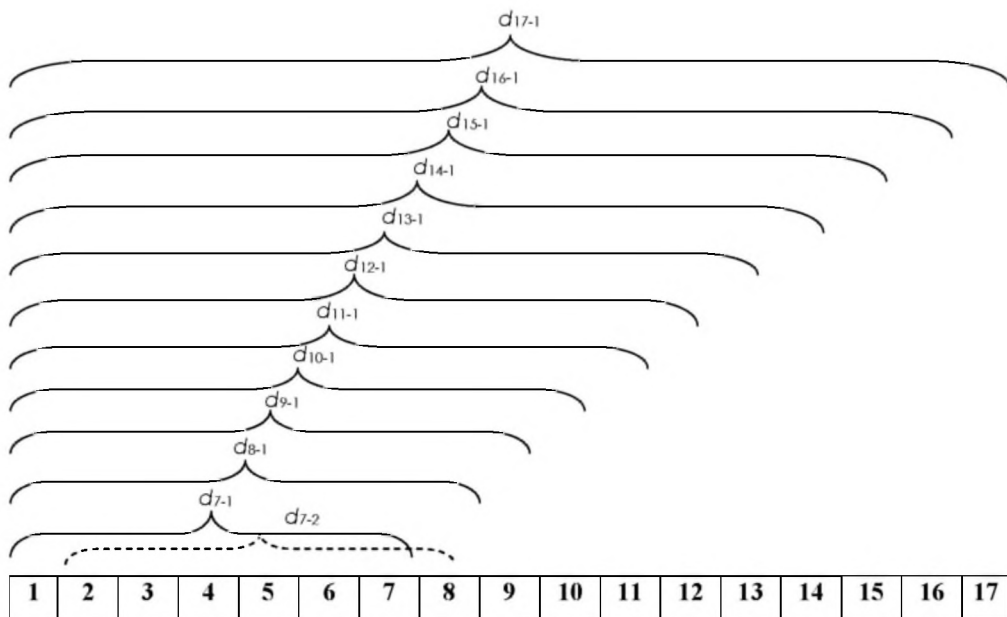


Рис. 3. Группировка соседних делянок в блоке для нахождения элементов каждой из 11 оценок скользящей дисперсии d ,

нее по 11 полученным элементам d_{7-k} — оценкам дисперсий ($k = 1 \dots 11$) и будет d_7 — первой оценкой скользящей дисперсии.

Аналогично d_8 будет оценкой скользящей дисперсии — среднее значение из дисперсий, полученных по 10 группам из 8 смежных делянок, d_9 — из 9 ... d_{16} (из 16), а d_{17} — одна дисперсия 17 чисел определенная для всей группы делянок блока: с 1-й по 17-ю. Всего таких индикаторных оценок 11 (d_7, \dots, d_{17}) в каждом блоке-повторении полевого опыта.

Модель изменчивости параметра P_j (X_{cp} или D) у популяции с номером j в любом блоке опыта

$$P_j = \mu + G_j + e_j,$$

где μ — среднее значение признака по всему блоку, G_j — генотипическое отклонение параметра у популяции на j -й делянке блока, e_j — ошибка оценки параметра этой популяции, имеющая структуру среднего квадрата в соответствии с уравнением (1): $SI^2_{\text{ош}}(30) = S^2_{\infty} + S^2(30)$.

Поэтому ожидаемая структура индикаторного значения d_{17} — для всего блока из 17 делянок:

$$d_{17} = \sum_{i=1}^{17} (P_i - P_{17-1})^2 / (17-1) = S_{G17} + S^2_{\infty 17} + S^2(30), \quad (11)$$

где S_{G17} — генотипическая дисперсия параметра (μ или u) всех 17 популяции опыта, $S^2_{\infty 17} = S^2_{\infty}$ — влияние градиентов почвенного плодородия и «пятнистости» во всем блоке из 17 делянок.

Рассмотрим параметр $S^2_{\infty 17}$ — влияние градиентов и «пятнистости» в группе из $i \leq 7$ смежных делянок. При увеличении d_i по i от d_7 до d_{17} ожидаемое значение $S^2(30)$

не меняется. То же относится к вкладу «пятнистости» почвенного плодородия. При условии рандомизированного размещения популяций в блоке значение S_{G17} также не имеет какой-либо тенденции при изменении d_i по i от d_- до d_{17} . Но $S^2_{\infty i}$ при существенном влиянии градиентов имеет тенденцию к увеличению по i из-за все большего «охвата» участков поля — числа смежных делянок, попадающих под усреднение в процессе вычисления от d_- до d_+ . Из уравнения (11):

$$S^2_{\infty i} = d_i - S_{G17} - S^2(30). \quad (12)$$

Оценку генотипической дисперсии S_{G17} получаем из обычного дисперсионного анализа изучаемого параметра (X_{cp} или D) по формуле

$$S_{G17} = (ms_l - ms_e)/n.$$

Для уточнения оценок $S^2_{\infty i}$ из уравнения (12) каждую из 11 оценок ($d_7...d_{17}$) следует предварительно усреднить по n повторениям-блокам.

Итак, при существенных градиентах почвенного плодородия на поле проявится достоверная положительная регрессионная связь между числами i и оценками d_i скользящей дисперсий для этих групп соседних делянок. С помощью такой регрессионной связи появится возможность прогнозировать параметр S^2_{∞} для любого w — общего числа делянок, предполагаемых в блоке нового опыта на данном поле.

В результате на основе параметра K удастся заранее, не проводя дополнительных полевых опытов, подбирать новое достаточное (оптимальное) значение m для любого w .

Результаты и их обсуждение

Сравнение популяций по среднему значению признака

При сравнении 17 популяций по среднему значению признака увеличение числа растений, измеренных на делянке с опорного числа 30 до 40, не принесет значимого (на 10%) выигрыша в точности сравнения ни по одному из семи признаков (табл. 1). Снижение объема выборки до 20 растений также не приведет к значимому увеличению НСР. Снижение объема выборки с 30 до 10 растений приведет к значимому увеличению НСР для одного признака из семи. Снижение объема выборки с 30 до пяти приводит к снижению точности сравнения более чем на 10% для четырех признаков из семи.

Таким образом, результаты 2011 г. подтвердили рекомендации, полученные в 2009-2010 гг. [5]: при сравнении средних достаточно 20 растений, измеренных на

Т а б л и ц а 1

Зависимость относительной ошибки сравнения (K_{cp}) от объема выборки растений, измеренных на делянке (m), 2011 г.

| Признак | $m = 5$ | $m = 10$ | $m = 20$ | $m = 30$ | $m = 40$ | $m \rightarrow \infty$ |
|---------------------------|---------|----------|----------|----------|----------|------------------------|
| Высота растения | 1,01 | 1 | 1 | 1 | 1 | 1 |
| Длина колосового стержня | 1,49 | 1,22 | 1,06 | 1 | 0,97 | 0,87 |
| Кустистость | 1,21 | 1,09 | 1,02 | 1 | 0,99 | 0,95 |
| Число озерненных колосков | 1,10 | 1,04 | 1,01 | 1 | 0,99 | 0,98 |
| Число зерен с растения | 1 | 1 | 1 | 1 | 1 | 1 |
| Масса зерен с растения | 1,17 | 1,07 | 1,02 | 1 | 0,99 | 0,96 |
| Масса 1000 зерен | 1,03 | 1,01 | 1 | 1 | 1 | 0,99 |

каждой делянке. Отметим, что для признаков «высота», «число зерен с растения» и «масса 1000 зерен» варьирование числа m практически не влияет на K_{cp} . Следовательно, внутриделяночная изменчивость этих признаков мала по сравнению с влияниями «пятнистости» и градиентов почвенного плодородия, отраженных в параметре S^2_{∞} (см. формулу (1) из раздела «Материалы и методика»). Вывод о трех признаках также соответствует результатам биометрического анализа данных 2009 и 2010 гг. [5].

Сравнение популяций по дисперсии признака

Снижение с 30 до 20 числа m — объема выборки растений, измеренных на делянке, нежелательно (табл. 2). При увеличении же выборки до 40 растений точность сравнения повысится на 11% только для двух признаков из семи. Лишь при очень значительном увеличении объема выборки растений ($m \rightarrow \infty$) можно добиться существенного увеличения точности сравнения популяций по дисперсии (D). Практически это трудно реализовать: резко увеличатся трудозатраты.

Т а б л и ц а 2

Зависимость относительной ошибки сравнения популяций по дисперсии (K_D) от объема выборки растений (m), 2011 г.

| Признак | $m = 5$ | $m = 10$ | $m = 20$ | $m = 30$ | $m = 40$ | $m \rightarrow \infty$ |
|--------------------------|---------|----------|----------|----------|----------|------------------------|
| Высота растения | 2,25 | 1,62 | 1,18 | 1 | 0,89 | 0,44 |
| Длина колосового стержня | 2,02 | 1,49 | 1,14 | 1 | 0,92 | 0,62 |
| Кустистость | 1,75 | 1,35 | 1,10 | 1 | 0,95 | 0,77 |
| Число озеренных колосков | 1,76 | 1,36 | 1,10 | 1 | 0,95 | 0,76 |
| Число зерен растения | 1,80 | 1,38 | 1,11 | 1 | 0,94 | 0,74 |
| Масса зерен растения | 1,72 | 1,34 | 1,09 | 1 | 0,95 | 0,78 |
| Масса 1000 зерен | 2,25 | 1,62 | 1,19 | 1 | 0,89 | 0,43 |

Представленные в таблице 2 результаты анализа, проведенного в 2011 г. по семи исследуемым признакам подтвердили выводы, сделанные на основании исследований 2009-2010 гг. [5]: при сравнении по дисперсиям D (оценкам генетической дивергентности) популяций яровой пшеницы оптимальный объем выборки составляет 30 растений на делянке.

Преобразование данных для дисперсионного анализа

Как отмечалось выше, по данным 2010 г., для трех из семи количественных признаков дисперсионный анализ параметра D у 17 популяций выявил отклонения от ожидаемых соотношений (табл. 3). А именно, для признаков 1, 5 и 7 значения $S^2_{\text{ош}}(30)$ оказались меньше $S^2(30)$ (отмечено жирным шрифтом). Это невозможно, исходя из формулы (1), приведенной в разделе «Материалы и методика».

Такое соотношение свидетельствует о сильных нарушениях модельных предположений дисперсионного анализа в 2010 г. и может привести к искажениям рекомендаций для подбора m — числа растений, измеряемых на каждой делянке. В 2009 г. (табл. 4) и в 2011 г. (табл. 5) подобных отклонений не наблюдалось, как и при дисперсионном анализе для сравнения популяций по среднему значению для каждого из семи признаков в 2009-2011 гг. Последнее, по-видимому, объясняется

Таблица 3

Значения параметров $S^2_{\text{ош}}(30)$ и $S^2(30)$ при сравнении популяций по дисперсии (2010 г)

| Параметр | Признаки* | | | | | | |
|-----------------------|---------------|------|------|------|----------------------------|------|--------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $S^2_{\text{ош}}(30)$ | 428,12 | 0,22 | 0,18 | 3,19 | 2,06×10⁵ | 0,29 | 31,42 |
| $S^2(30)$ | 504,5 | 0,14 | 0,18 | 1,56 | 2,83×10⁵ | 0,27 | 54,21 |

* В таблицах 3-8 номера признаков — из раздела «Материалы и методика».

Таблица 4

Значения $S^2_{\text{ош}}(30)$ и $S^2(30)$ при сравнении популяций по дисперсии признака (2009 г)

| Параметр | Признаки | | | | | | |
|-----------------------|----------|------|------|------|----------------------|------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $S^2_{\text{ош}}(30)$ | 1102,49 | 0,16 | 0,58 | 2,12 | 6,16×10 ⁵ | 2,01 | 95,64 |
| $S^2(30)$ | 623,10 | 0,13 | 0,31 | 1,58 | 4,81×10 ⁵ | 1,92 | 77,67 |

Таблица 5

Значения $S^2_{\text{ош}}(30)$ и $S^2(30)$ при сравнении популяций по дисперсии признака (2011 г)

| Параметр | Признаки | | | | | | |
|-----------------------|----------|------|------|------|----------------------|------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $S^2_{\text{ош}}(30)$ | 677,17 | 0,14 | 0,65 | 3,35 | 5,45×10 ⁵ | 0,96 | 65,57 |
| $S^2(30)$ | 547,29 | 0,09 | 0,27 | 1,41 | 2,44×10 ⁵ | 0,38 | 53,49 |

отсутствием достоверной статистической связи оценок средних значений признака на делянках с дисперсиями их ошибок, которые, как отмечалось выше, равны D/m .

Как отмечалось в разделе «Материалы и методика», дисперсия ошибки основного оцениваемого показателя D не просто зависит от самого значения D , но пропорциональна его квадрату. В 2010 г., экстремальном по условиям выращивания, различия D по делянкам опыта весьма значительны, что видно из таблицы 6. Несмотря на высокую робастность дисперсионного анализа, такие различия могли грубо нарушить его предположения. Для проверки этой гипотезы в соответствии с рекомендованным

Таблица 6

Диапазоны (*min-max*) разброса значений показателя D и дисперсии его ошибки — $2D^2/(30-1)$, оцененные по 51 делянке в опыте 2010 г.

| Параметр | Признаки | | | | | | |
|-------------------------------|-------------------|---------------|---------------|----------------|--|---------------|------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $D(\text{min-max})$ | 21,90– 172,66 | 0,56– 2,13 | 0,66– 3,35 | 1,00– 14,29 | 832,32– 4464,91 | 0,69– 2,96 | 14,06– 46,65 |
| $2D^2/(30-1)(\text{min-max})$ | 33,09– 2056,99 | 0,02– 0,31 | 0,03– 0,77 | 0,07– 15,31 | 4,78×10 ⁴ – 1,37×10 ⁶ | 0,03– 0,60 | 13,64– 150,16 |

преобразованием (раздел «Материалы и методика») было проведено преобразование оценок дисперсий на каждой делянке по формуле $\ln(\hat{D})$ и пересчет (табл. 7).

Т а б л и ц а 7

Пересчитанные значения θ и $S^2(30)$ при анализе величин $\ln[D]$, преобразованных значениям дисперсий на каждой делянке в опыте 2010 г.

| Параметр | Признаки | | | | | | |
|-----------------------|----------|------|------|------|------|------|-------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $S^2_{\text{ош}}(30)$ | 0,07 | 0,12 | 0,08 | 0,12 | 0,07 | 0,10 | 0,05 |
| $S^2(30)$ | 0,07 | 0,07 | 0,07 | 0,07 | 0,07 | 0,07 | 0,07 |

Для преобразованных данных недопустимые отклонения проявляются при анализе только последнего признака (масса 1000 зерен) из трех, отмеченных в таблице 3. Следовательно, устраненная зависимость дисперсий ошибок параметра \hat{D}_k от значений самого этого параметра — не единственная причина, нарушающая модельные предположения дисперсионного анализа.

Результаты применения имитационного моделирования для оценки влияния асимметрии и эксцесса

Возможно, дополнительной причиной, вызывающей существенные нарушения модельных предположений дисперсионного анализа, является значительный разброс значений асимметрии (As) и особенно эксцесса (Ex) по исходным данным на 51 делянке в опыте 2010 г. (табл. 8).

Т а б л и ц а 8

Диапазоны (min-max) разброса значений асимметрии (As) и эксцесса (Ex), оцененных для 51 делянки по выборкам объема 30 растений (2010 г.)

| Параметр | Признаки | | | | | | |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| $As (min-max)$ | -1,41– 0,88 | -1,97– 1,13 | -0,44– 2,11 | -1,85– 0,86 | -0,27– 2,00 | -0,2– 2,32 | -0,48– 2,71 |
| $Ex (min-max)$ | -1,44– 3,86 | -1,31– 6,36 | -1,31– 7,68 | -1,31– 6,41 | -1,34– 1,99 | -1,41– 6,48 | -1,31– 11,26 |

Полученные результаты имитационного моделирования, описанного в разделе «Материалы и методы», использовали прежде всего для оценки влияния параметров D , As , Ex , на σ^2_Y — дисперсию ошибки преобразованного показателя D . Анализ влияния параметров позволил сделать следующие выводы:

1. При фиксированных значениях показателей D , As , Ex варьирование математического ожидания признака на делянке, как и предполагалось, не влияет на σ^2_Y ;
2. Проверка выравнивания дисперсии ошибки параметра D через преобразование $Y = \ln(D)$, если изменчивость количественного признака подчиняется закону нормального распределения ($As = 0$, $Ex = 0$), подтвердила высокое качество этого пре-

образования. Дисперсия ошибки преобразованных значений $Y = \ln(D)$, независимо от D , составляла приблизительно 0,07, что соответствует оценке $2/(30-1)$, полученной при аналитическом подборе преобразования (см. раздел «Материалы и методы»);

3. При фиксированных показателях As и Ex значение дисперсии ошибки преобразованных значений $Y = \ln(D)$ мало зависит от величины генерального параметра D у исходных выборок «данных по 30 растениям». Распределение самих преобразованных значений $Y = \ln(D)$ близко к нормальному;

4. Дисперсия ошибки преобразованных значений $Y = \ln(D)$ существенно меняется (по сравнению с 0,07) при варьировании показателя Ex у распределения количественного признака и в меньшей степени — при варьировании As .

Для иллюстрации 4-го вывода на рисунке 4 представлен характер зависимости σ^2_Y — дисперсии ошибки преобразованных значений $Y = \ln(D)$ от Ex — генерального значения показателя эксцесса распределения признака x у 30 растений при отсутствии асимметрии этого распределения ($As = 0$).

Пересчет при дополнительном внесении асимметрии ($As \neq 0$) показывает, что независимо от знака величины As при варьировании степени асимметрии от -2 до 2 показатель σ^2_Y увеличивается на 5-80% (числа на поле рисунка 4 обозначают величины As).

Расчеты варьирования опытных оценок Ex и As (см. табл. 8) при сопоставлении с графиком на рисунке 4 показывают, что параметр Ex в 2010 г. для признака масса 1000 зерен мог колебаться в диапазоне от -1,44 до 11,26. Следовательно, дисперсия σ^2_Y могла колебаться по делянкам опыта в 2010 г. приблизительно от $0,07 \times 0,3$ при $Ex = -1,44$ до $0,07 \times 18$ при $Ex = 11,26$ (см. рис. 4). По-видимому, превосходство параметра $S^2(30)$ над параметром $SI^2_{ou}(30)$, получаемым из дисперсионного анализа оценок $Y = \ln(D)$ для признака масса 1000 зерен (см. табл. 7), вызвано именно такой неравноточностью этих оценок.

Добиваться устранения такого недостатка преобразованных значений $Y = \ln(D)$ проблематично и, по-видимому, нецелесообразно, так как при анализе семи признаков за четыре года (28 расчетов) подобный недостаток проявился только один раз (7-й признак в таблице 7). Такие редко встречающиеся результаты просто не следует учитывать при разработке рекомендаций по числу измеряемых растений (m) для сравнения показателя D у селекционных образцов по элементам структуры урожая любой культуры.

Можно сделать вывод, что преобразование $Y = \ln(D)$ весьма чувствительно к различиям по делянкам эксцессов и асимметрий исходных распределений признаков. Поэтому в 2009 и 2011 гг., когда преобразование данных не потребовалось,

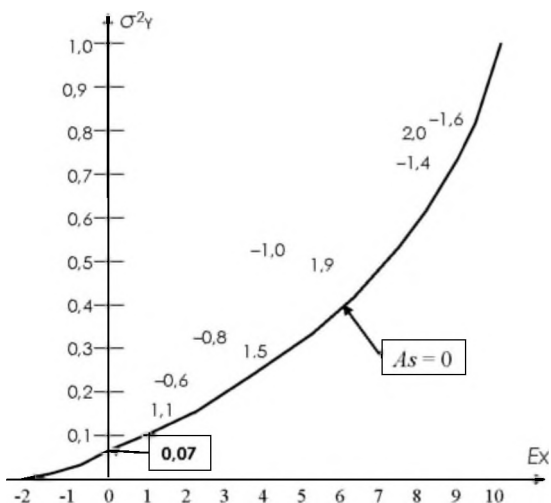


Рис. 4. Характер зависимости σ^2_Y от Ex при $As = 0$ (жирная линия) и увеличение σ^2_Y при $As \neq 0$

дисперсионный анализ проявил большую робастность к эксцессам и асимметриям исходных данных.

Таким образом, для анализа и сравнения параметра D (генетического разнообразия гибридных семей) при возникновении проблем с дисперсионным анализом следует использовать не сами оценки параметра D , а их преобразованные значения $Y = \ln(D)$, выравнивающие дисперсии ошибки.

Применение методики скользящей дисперсии

Для построения наглядных графиков на оси абсцисс можно отложить целые числа ($i = 7 \dots 17$) соседних делянок, используемых при подсчете скользящих дисперсий. На оси ординат — оценки $d_7 \dots d_{17}$, предварительно усредненные по n повторениям-блокам. В результате возникает возможность выявить тенденцию — связь размеров участков поля с градиентами (трендами) почвенного плодородия. Однако при этом следует учитывать, что оценки $d_7 \dots d_{17}$ получены с ошибками выборочности — на график следует нанести их доверительные интервалы. В качестве «плеча» такого интервала при построении графиков можно использовать параметр u ошибки выборочно сти:

$$\sigma = \sqrt{\frac{2d_i^2}{n(w-1)}},$$

где n — число повторений-блоков (в наших опытах $n = 3$). w — число делянок в блоке ($w = 17$).

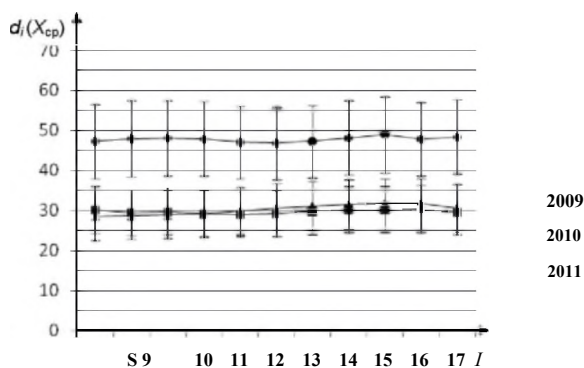


Рис. 5. Зависимость скользящих дисперсий $d_i(X_{ср})$ при сравнении популяций по среднему значению признака высота растений от длины опытного блока (l' — количество делянок в блоке)

При анализе гибридных популяций по средним значениям признака методика скользящей дисперсии не выявила достоверных тенденций для большинства признаков и опытов 2009-2011 гг., т.е. не проявилась достоверная (с учетом доверительных интервалов оценок $d_i(X_{ср})$) регрессионная зависимость ошибки, связанной с пестротой почвенного плодородия (S^2_{∞}), от длины опытного блока (количества делянок i). Для иллюстрации на рисунке 5 приведен результат анализа признака высота растения; графики для каждого из трех лет усреднены по трем повторениям (блокам) опыта.

Напомним, что если бы проявилась тенденция роста S^2_{∞} при увеличении длины опытного блока, то влияние ошибки, связанной с ограниченным числом растений, измеряемых на каждой делянке (S^2_{30}), на относительную ошибку (K) сравнения средних снизилось бы. Тогда при значительном удлинении блока, возможно, удалось бы сократить число растений, измеряемых на делянке без существенного снижения точности сравнения популяций ($K \leq 1, 1$).

Аналогично при сравнении семей по дисперсии (D) признака за 2009-2011 гг. методика скользящей дисперсии не выявила достоверных тенденций $d_i(D)$ по большинству признаков. Для иллюстрации на рисунке 6 показан признак высота растений. Каждый график усреднен по трем повторениям (блокам) опыта.

Причина отсутствия достоверных регрессионных зависимостей d_i от i — малая общая длина каждого блока (17 делянок) в экспериментах 2009-2011 гг. Градиенты почвенного плодородия в сравнительно коротких блоках не проявились в достаточной степени. Основное влияние на параметр S^2_{∞} оказывала «пятнистость».

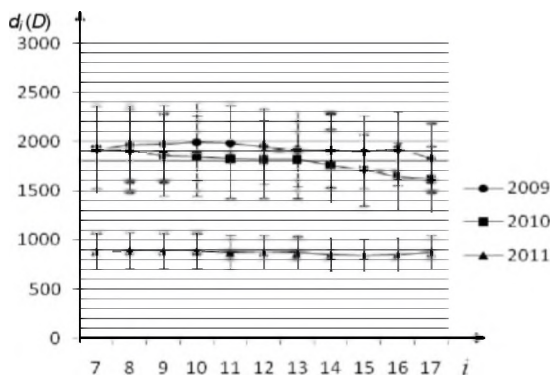


Рис. 6. Зависимость скользящих дисперсий $d_i(D)$ при сравнении гибридных популяций по дисперсии признака высота растений от длины опытного блока (i — количество делянок в блоке)

Выводы

1. В целом анализ данных 2009-2011 гг. с помощью усовершенствованного биометрического метода оптимизации числа растений, измеренных на каждой делянке, указывает, что 20 растений достаточно для сравнения популяций по среднему значению количественного признака. Если стоит цель — сравнить показатель генетического разнообразия (D) популяций, то следует измерять по 30 растений. Вывод относится к опытам с популяциями мягкой яровой пшеницы, но сам биометрический метод применим для оптимизации полевых сравнительных испытаний любых форм культурных растений.

2. При возникновении проблем с дисперсионным анализом результатов сравнительных испытаний по показателю D — генетическому разнообразию популяций — следует анализировать не сами оценки D , а их преобразованные значения $Y = \ln(D)$, выравнивающие дисперсии ошибки D . Установлено, что преобразование $Y = \ln(D)$ чувствительно к различиям по делянкам эксцессов и асимметрий исходных распределений признаков. Однако, судя по 4-летним результатам анализа семи количественных признаков, это влияние проявляется сравнительно редко и не снижает возможностей разработанного биометрического метода оптимизации числа растений, измеряемых на каждой делянке.

3. Предложенную методику скользящей дисперсии желательно проверить на данных полевых сравнительных испытаний с более длинными блоками. Другой подход — применить имитационное моделирование, накладывая с помощью метода Монте-Карло дополнительный градиент почвенного плодородия на реальные или моделируемые данные эксперимента.

Работа выполнена при частичной финансовой поддержке Министерства сельского хозяйства РФ по теме: Разработка прогрессивных методов селекции и технологий возделывания новых сортов тритикале и пшеницы на основе математического моделирования.

Библиографический список

1. Доспехов Б.А. Методика полевого опыта. М.: Агропромиздат, 1985.
2. Комарова Е.А., Смирязев А.В., Пыльное В.В. Влияние объема выборки растений на точность сравнения генотипов в полевом опыте // Известия ТСХА, 2007. Вып. 2.
3. Коновалов Ю.Б. Особенности полевого опыта в ранних звеньях селекционного процесса. М.: ТСХА, 1982. 28 с.
4. Ллойд Э., Ледерман У. Справочник по прикладной статистике. Т. 1. М., 1989.

5. Нгуен Тхань Туан., Смиряев А.В., Баженова С.С. Влияние объема выборки растений на точность сравнения гибридных форм яровой пшеницы // Известия ТСХА, 2010. Вып. 3.

6. Смиряев А.В., Кильчевский А.В. Генетика популяций и количественных признаков. М.: «КолосС», 2007.

7. Шеффе Г. Дисперсионный анализ. М.: Физматгиз, 1963.

8. Dionysia A. Fasoula. Correlations between auto-, alio- and nil-competition and their implications in plant breeding//Euphytica, 1990. Vol. 50. № 1. P. 57-62.

9. Iliadis G.C., Roupakias D.G., Goulas C.K. Effectiveness of honeycomb selection for yield superiority at three interplant distances: a field simulation study using chickpea (*Cicer arietinum* L.) inbred lines // Euphytica, 2003. Vol. 133. № 3. P. 299-311.

10. Pouliney R., Riley J. and Webster R. Optimizing plot size and shape for field experiments on terraces//Experimental Agriculture, 1997. P. 51-64.

11. Romani M., Borghi B., Aiberici R., Delogu G., Hesselbach J. and Salamini F. Intergenic competition and border effect in bread wheat and barley // Euphytica, 1993. Vol. 69. № 1-2. P. 19-31.

12. Zhang R., Warrick A. W, Myers D. E. Heterogeneity, plot shape effect and optimum plot size // Geodenna, 1994. Vol. 62. P. 183-197.

Рецензент — к. с.-х. н. Р.Р. Усманов

SUMMARY

Biometrical method for estimating the influence of plants number, measured on each plot of field experiment, upon accuracy comparison the hybrid populations by the averages and variances of a quantitative trait, had been worked out before. The method has been improved taking into account not equal accuracy of variances estimate and for situations with skewness and kurtosis of trait distribution. The technique of moving variance is used to forecast the influence of block length in experiment on the enough sample size of measured plants on each plot. Data of seven traits in 17 hybrid populations F3, F4 and F5 of soft spring wheat are used.

Key words: biometry, accuracy of hybrid population comparison plants sampling volume, dispersion analysis, irregularity, asymmetry, excess, imitation modeling, sliding dispersion, spring wheat.

Смиряев Анатолий Владимирович — д. б. н., проф. каф. генетики и биотехнологии РГАУ-МСХА имени Тимирязева (127550, г. Москва, ул. Тимирязевская, 49; тел. (499) 976-08-94; e-mail: genetics'rttimacad.ru).

Хунацария Титико Ипполитович — к. б. н., проф. каф. селекции и семеноводства полевых культур РГАУ-МСХА имени Тимирязева (тел. (499) 976-12-72).

Баженова Светлана Сергеевна — к. с.-х. н., доц. каф. селекции и семеноводства полевых культур РГАУ-МСХА имени Тимирязева (e-mail: selection@timacad.ru).

Нгуен Тхань Туан (Вьетнам) — к. б. н., преподаватель каф. селекции и генетики Ханойского аграрного университета (e-mail: thanglongmos@yahoo.com).