НИЖЕГОРОДСКАЯ ГОСУДАРСТВЕННАЯ СЕЛЬСКОХОЗЯЙСТВЕННАЯ АКАДЕМИЯ

ДИСПЕРСИОННЫЙ АНАЛИЗ МНОГОУРОВНЕВЫХ ИЕРАРХИЧЕСКИХ КОМПЛЕКСОВ

Методические указания для студентов и аспирантов очного и заочного отделений факультета лесного хозяйства
по специальности 25020165 – лесное хозяйство

Нижний Новгород - 2012

Составители: профессор Бессчетнов В.П., доцент Бессчетнова Н.Н., доцент Храмова О.Ю., доцент Орнатский А.Н., Горелов Н.И.

УДК 630.

Дисперсионный анализ многоуровневых иерархических комплексов: Методические указания для студентов и аспирантов очного и заочного отделений факультета лесного хозяйства по специальности 25.02.01.65 – Лесное хозяйство / Нижегородская государственная сельскохозяйственная академия. – Нижний Новгород, 2012.

В методических указаниях приведены основы дисперсионного анализа и перспективы его применения в лесоводственных исследованиях, принципы организации многоуровневых комплексов и их образование при создании селекционно-семеноводческих объектов, особенности и порядок вычисления итоговых величин двухфакторного и трехфакторного иерархических дисперсионных анализов. Методические указания адресованы студентам, аспирантам очного и заочного отделений сельскохозяйственных вузов, интересны также всем кто занимается научными исследованиями в сельскохозяйственной и лесохозяйственной отраслях.

Печатается по решению редакционно-издательского совета НГСХА

Рецензент: заведующий кафедрой плодоовощеводства, селекции и семеноводства НГСХА, доктор сельскохозяйственных наук, профессор Лебедев В.М.

Оглавление

Введение

Дисперсионный анализ, предложенный в 1925 году Р.А. Фишером, позволяет решать очень важные задачи научных исследований, в том числе и в лесном хозяйстве. Он позволяет рассматривать общую фиксируемую в опыте изменчивость показателей изучаемого объекта, возникающую под действием каких-либо факторов, как изменчивость, складывающуюся из нескольких отдельных её составляющих, каждая из которых обусловлена действием того или иного фактора или их комплекса.

В настоящее время хорошо известно и является общепризнанным положением то, что варьирование признаков у растений обусловлено действием многообразных факторов. Часть из них, оказывая влияние на признак, изменяет его в сторону увеличения, а другие факторы приводят к уменьшению его значений. Совместное действие факторов также вносит какое-то дополнительное влияние на формирование и развитие признака и, следовательно, создает дополнительный источник его изменчивости.

В результате такого многообразного воздействия на признак со стороны комплекса факторов внешней и внутренней среды (а точнее – факторов внешней среды и внутренних биологических особенностей организма) возникает варьирование признака у членов исследуемой совокупности (насаждения, популяции, лесосеменной плантации и т.п.). В этом общем варьировании отражается интегрированное (суммарное) влияние всех раздельно и совместно действующих факторов. С этой конечной общей (результирующей) изменчивостью мы и имеем дело при изучении любой совокупности объектов, в которой влияние каждого отдельного фактора нивелируется действием всех остальных факторов. Для познания закономерностей изменчивости объектов какой-либо совокупности важно вычленить долю влияния отдельных факторов, изучение действия которых на данные объекты может служить целью специальных опытов.

Обрабатывая опытный материал дисперсионным методом, можно получить математическое выражение и определить величину изменчивости, обусловленную воздействием учтенных в опыте факторов, и изменчивость остаточную, возникшую под влиянием всех других, не учтенных в опыте, факторов. В этом состоит первое и наиболее важное свойство дисперсионного анализа. Второе свойство дисперсионного анализа заключается в том, что он позволяет определить статистическую достоверность доли влияния изучаемых факторов.

Ключевые слова: *дисперсионный комплекс, совокупность, подсовокупность, группы, классы, дисперсия, варианса, сумма квадратов отклонений, средняя сумма квадратов отклонений, иерархия связей, организованные факторы, неорганизованные факторы, факториальная дисперсия, остаточная дисперсия, дисперсионное отношение, сила влияния фактора, критерий Фишера.*

1. Принципиальные основы дисперсионного анализа и перспективы его применения в лесоводственных исследованиях

Основное назначение дисперсионного анализа – это разлагать общую изменчивость признака на изменчивость частную, возникающую у членов совокупности (насаждения или популяции) под влиянием многообразных факторов. В ходе анализа устанавливают:
- долю изменчивости, обусловленную каждым учитываемым в опыте фактором;
- долю изменчивости, вызванную совместным действием этих факторов;
- а также выявляют ту долю изменчивости, которая является результатом воздействия многих не учтенных в опыте факторов, что образует так называемую случайную, или остаточную, изменчивость признака.

Указанное свойство дисперсионного анализа имеет большое значение при анализе изменчивости, наблюдаемой у биологических объектов, в том числе и у древесных растений и кустарников.

Дисперсионный анализ является методом анализа количественной информации (преимущественно). Его важной особенностью является то, что он может применяться на разных типах выборок (больших и малых) и, что особенно важно, он позволяет обрабатывать разнообразные совокупности, включающие в себя как однородный, так и неоднородный материал:

- разнополые особи у двудомных растений (тополя, ивы, облепиха и др.);
- совокупности, состоящие из групп растений разного генетического происхождения (особи из естественных насаждений, клоны плюсовых деревьев, гибриды, сорта и т.п.).

При проведении дисперсионного анализа исходят из предположения о том, что некоторая совокупность объектов под действием какого либо фактора (или нескольких факторов) разделяется на некоторое число групп (две и более). При этом каждая из групп объектов отличается от других групп величиной среднего группового значения признака и характером его изменчивости – дисперсией или вариансой (вариансный анализ) в пределах группы. Учитывается также и то, что изменчивость признака у объектов возможна как внутри таких групп, так и между ними.

Если в имеющейся совокупности объектов изменчивость (рассеянье, дисперсия) признака в любой из её частей одинакова, то совокупность не разделена на группы, между которыми можно провести отчетливые границы. В такой ситуации совокупность признается однородной. Это единая совокупность – одно и то же. Если же разные части анализируемой совокупности характеризуются разными дисперсиями – разным характером рассеянья признака – то совокупность не рассматривается как единая, и в её составе признается наличие некоторых внутренних групп. Разделение совокупности на части происходит под действием какого-либо фактора. Например, мы рукой можем разделить равномерно рассыпанные на столе шишки сосны или ели на группы (части совокупности). Выявлению его влияния и подчинен дисперсионный анализ. На разные по своим характеристикам группы могут быть разделены сеянцы, выращиваемые на разном субстрате (фактор разделения на группы – разница в свойствах субстратов), при разном режиме внесения удобрений (действующий фактор – различия в режимах внесения удобрений) или орошения (фактор – режимы орошения) и т.п. На различающиеся между собой группы объектов могут быть разделены деревья в лесу, произрастающие в разных типах лесорастительных условий (ТЛУ), разных типах леса и т.п. Разными группами объектов могут явиться семена, заготовленные с разных по происхождению (генотипу) деревьев. Таких примеров можно привести много.

Если уровень изменчивости признака, оцениваемый величиной дисперсии (среднего квадрата отклонений), между группами больше, чем его изменчивость внутри групп, то наличие различий между группами принимается установленным, а сами различия между группами признаются существенными. Если же уровень различий между группами соответствует уровню различий внутри групп (случайное варьирование, не вызванное действием анализируемого фактора), то действие фактора признается неэффективным, не приводящим к разделению общей рассматриваемой совокупности (дисперсионного комплекса) на реально существующие группы (классы).

Организованными факторами в дисперсионном анализе выступают те факторы, действие которых признается (принимается) доступным для изучения, собственно те факторы, которые вызывают разделение совокупности на части, которые организованы в ту или иную систему (факторы той или иной системной организации). Частным случаем такой организации групп является иерархическая система. В ней одни группы имеют более высокий ранг организации, и сами содержат в своем составе меньшие по своей численности группы, относящиеся к группам низшего ранга организации.

В этой связи необходимым условием эффективного использования метода является правильная организация дисперсионных комплексов.

Следует учитывать логику и структуру естественной группировки членов выборки (деревьев в лесу или на опытном участке) при выделении градаций и классов в дисперсионных комплексах, что имеет место в отношении количественных признаков.

Следует избегать или сводить к минимуму применение таких схем расчета, при которых дисперсионные комплексы (основа статистической обработки) искусственно строятся в виде чередующихся классов, границы которых устанавливаются самим исследователем.

Необходимо сравнивать и сопоставлять между собой такие группы объектов в составе какой-либо изучаемой совокупности, которые сформировались вследствие воздействия на неё (совокупность) известного исследователю фактора или множества факторов, действие которых и подвергается анализу. Здесь в задачу исследователя входит установление существенности различий между сравниваемыми группами совокупности, возникшими под действием фактора, но не (произвольное) субъективное определение границ таких групп и произвольное утверждение собственно факта наличия таких групп.

Целесообразно принимать исходную посылку дисперсионного анализа как предположение (как выдвижение рабочей гипотезы) о наличии в составе некоторой исходной совокупности групп объектов с границами, оформленными в той или иной степени. Само возникновение таких групп рассматривается как результат воздействия на совокупность некоторых факторов. Выдвижение так называемой «нулевой гипотезы» предполагает отсутствие между сравниваемыми группами различий, уровень которых превышал бы уровень случайного разброса значений, возникающего в совокупности независимо от действия (или без такового) какого-либо направленного фактора. То есть различий между группами не существует, и, следовательно, не существует собственно групп в составе совокупности – совокупность однородна по своему составу и строению и не дифференцирована на внутренние части.

Оценка изменчивости дается по величине дисперсий. По этой причине в совокупности формально предполагается наличие как минимум двух групп объектов, а численность каждой из групп – два и более объектов. Кроме того, предполагается различие в значениях признаков, как в группах, так и между ними. Именно в такой ситуации возможен расчет дисперсий как оценок степени разброса или рассеянья значений признака объектов относительно среднего. По соотношению величин дисперсий судят о том: превышает ли уровень изменчивости признаков между видимыми группами объектов уровень изменчивости, соответствующий уровню изменчивости объектов в пределах групп, или соответствует такому уровню случайной изменчивости.

Сами дисперсии рассчитываются с помощью следующих сумм квадратов отклонений по ниже приведенным формулам:

$C_y = \sum_{1}^{m} (V_i - M_o)^2$ – сумма квадратов отклонений отдельных вариант всего комплекса наблюдений (V_i) от их общей средней (M_o) - для расчета общей дисперсии (SS_y – sum square);

$C_x = \sum_{i=1}^{a} n_i \cdot (M_{xi} - M_o)^2$ – сумма квадратов отклонений групповых или частных средних (M_{xi}) от общей для всего комплекса средней (M_o) – для расчета частной или факториальной дисперсии (SS_x – sum square);

$C_z = \sum_{1}^{a} \sum_{i=1}^{na} (V_i - M_{xi})^2$ – сумма по всем группам сумм квадратов отклонений отдельных вариант (V_i) в пределах группы от их групповых средних (M_{xi}) для расчета случайной или остаточной дисперсии (SS_z – sum square).

При этом справедливо равенство: $C_y = C_x + C_z$.

В несколько иной форме это равенство можно представить в виде:

$$\sum_{i=1}^{N}(x_i-\bar{x})^2=\sum_{j=1}^{a}n_j(\bar{x}_j-\bar{x})^2+\sum_{j=1}^{a}\left[\sum_{i=1}^{n}(x_i-\bar{x}_j)^2\right],$$

где:

x_i - отдельный вариант (дата);

$\bar{x}$ - общее среднее для всего комплекса;

$\bar{x}_j$ - частное (групповое) среднее;

N - общее число отдельных вариант в пределах всего комплекса, $N = n_1 + n_2 + n_3 + \ldots + n_a$; а в случае равенства численностей групп $N = n \times a$;

n_j - частное число вариант в пределах отдельной *a*-той группы, в случае равенства численностей всех групп (равномерные комплексы) $n_1 = n_2 = n_3 = \ldots = n_a = n$;

a - число групп в комплексе.

Это основная формула дисперсионного анализа.

Вместе с тем в ней возможно разложение факториальной дисперсии на ряд дисперсий каждого из факторов и дисперсии, определяемой взаимодействием учитываемых факторов, в случае если таких факторов много (более одного). Когда учитывается влияние нескольких факторов, они могут быть как независимыми один от другого, так и зависимыми (взаимосвязанными). В последнем случае говорят об иерархическом дисперсионном анализе на том основании, что взаимозависимые факторы образуют некую систему соподчиненности – т.е. иерархию.

Учесть влияние различий в происхождении растений, составляющих какую-либо филогенетическую систему, на проявление их признаков (размеры или энергия прорастания семян) можно следующим образом. Для сосны обыкновенной может быть учтено: принадлежность к какому либо экотипу или подвиду (сосна кулундинская) – первый уровень градации фактора; принадлежность к той или иной популяции – второй уровень градации фактора, который в свою очередь зависим от первого фактора. В пределах второго уровня градации фактора учитываются различия между особями популяции. Возможна и более сложная система иерархии: семейство сосновых, в его составе род сосна сравнивается с другими родами (ель, лиственница, пихта), в составе каждого рода различные виды, в составе видов внутривидовые таксоны и в конечном итоге популяции и особи, составляющие их.

Важным этапом дисперсионного анализа является определение средних квадратов отклонений или, как их иногда называют, дисперсий. Они представляют собой отношение сумм квадратов отклонений (общей, факториальной или частных факториальных и остаточной) к соответствующим числам степеней свободы.

Отношение факториальной дисперсии к остаточной дисперсии при условии, что факториальная дисперсия больше или равна остаточной, называется дисперсионным отношением или критерием Фишера. По его величине соотнесенной с табличным значением судят об эффективности действия фактора (например, фактора **А**). В случае, когда остаточная дисперсия больше факториальной, допускается расчет дисперсионного отношения как отношения остаточной дисперсии к факториальной (Лакин, 1980; стр. 219 – 222). Название статистики – «критерий Фишера» предложено Дж. У. Снедекором в честь своего учителя Р.А. Фишера.

При определении силы влияния фактора (факторов) как отношения «исправленных» дисперсий факторов к общей дисперсии в состав общей дисперсии не включаются дисперсии факторов, не имеющих достоверного влияния. То есть в этом случае общая дисперсия является суммой только дисперсий факторов с достоверным влиянием и остаточной дисперсии.

Сфера применения дисперсионного анализа достаточно широка, что обусловлено его свойствами, но особенно важен он в селекционных и генетических исследованиях. Оценка одного из важнейших показателей в селекции древесных видов – коэффициента наследуемости признака в широком смысле (h^2) – основана на результатах дисперсионного анализа. В генетических работах дисперсионный анализ используется для подтверждения или отрицания нулевой гипотезы о наличии или отсутствии наследования признаков. В процессе дисперсионного анализа можно получить сведения о криволинейности связей и её величине, путем вычисления значений корреляционного отношения.

Применение многофакторного иерархического дисперсионного анализа перспективно в исследованиях популяционной структуры видов древесных и кустарниковых растений: для доказательства популяционной структуры вида облепиха нами применен такой анализ (Бессчетнов , 1994 а, б; Чмыр, Бессчетнов, 1998). Он эффективен в селекционной оценке популяций и составляющих их особей, при оценке достоверности различий по таксационным показателям между различными насаждениями, объединенными в территориальные иерархические системы или по экологическому принципу. Двухфакторный иерархический анализ пригоден при сравнительной оценке клоновых (или иного происхождения) групп на коллекционных участках, архивах клонов лесосеменных плантациях, в опытах с испытательными и географическими культурами и т.п. Такой подход реализован Н.Н. Бессчетновой (2006а, б) при сравнительной оценке клонов и семенных полусибсовых потомств разных плюсовых деревьев сосны обыкновенной на объектах Постоянной лесосеменной базы (ПЛСБ) и единого генетико-селекционного комплекса (ЕГСК) в Нижегородской области. Весьма эффективен двухфакторный дисперсионный иерархический анализ при расчетах оценок общей комбинационной способности (ОКС) и специфической комбинационной способности (СКС) в селекционной работе с лесными древесными породами.

Вместе с тем применение дисперсионного анализа имеет свои ограничения. Следует учитывать то, что результат оценки действия фактора на вариабельность признака зависит от того, как сгруппированы члены выборки по классам, и какова разница в уровнях варьирующего признака между градациями. Кроме того при изучении двух или большего числа организованных факторов к включению в статистический комплекс допускаются только такие факторы, которые не находятся в корреляционной или функциональной связи. Следует также иметь в виду недостаточную теоретическую разработанность этого метода (многофакторного иерархического дисперсионного анализа), о чем сообщается в ряде специальных литературных источников (Меркурьева, 1970).

2. Принципы организации многоуровневых комплексов и их образование при создании селекционно-семеноводческих объектов

В сложно организованных комплексах, примерами которых могут выступать лесосеменные плантации вегетативного происхождения, архивы клонов, испытательные культуры и другие объекты постоянной лесосеменной базы (ПЛСБ) и единого генетико-селекционного комплекса (ЕГСК), важно корректно учитывать действие каждого из факторов. Особенностью организационной структуры дисперсионных комплексов, которые строят на основе селекционно-семеноводческих объектов, является участие в их составе плюсовых деревьев, представленных своим вегетативным или семенным потомством. При этом каждое плюсовое дерево выступает инициальным ядром формирования некоторого комплекса своих вегетативных или семенных потомков. В свою очередь потомки одного плюсового дерева (его прививки, черенки или же заготовленные с него семена или сеянцы, выращенные из них) не могут являться потомками другого плюсового дерева. Это обусловливает иерархическую природу подобных совокупностей, в соответствие с чем, форма их анализа также должна быть иерархической.

С точки зрения статистического анализа в целом в указанных комплексах градации одного фактора не могут свободно (независимо) группироваться по градациям другого фактора. Глубина иерархической организации (эшелонирование) таких комплексов может быть весьма значительной и может содержать целый ряд уровней, связанных иерархической зависимостью. Комплексы могут быть 2-уровневыми, 3-уровневыми и более глубоко эшелонированными. Для их анализа дисперсионными методами применяют схемы двухфакторных, трехфакторных и более сложных дисперсионных анализов. При этом все они основаны на сопоставлении величин дисперсионных отношений различного ранга. Вполне понятно, что принципом организации иерархических комплексов должна выступать логически понятная общность их системы: плюсовые деревья одной лесосеменной плантации (первый уровень организации), клоны этих плюсовых деревьев (второй уровень организации), побеги на учетных деревьях, образующих клон плюсового дерева в составе одной плюсовой плантации. Здесь критерием обобщения выступает принадлежность всех элементов каждого из уровней к одной ЛСП. Возможны и более сложные комплексы, например, те, в которых критерием обобщения выступает принадлежность к одному лесорастительному району. Тогда на первом уровне будут представлены различные ЛСП одного района, далее плюсовые деревья, их раметы, учетные побеги, позволяющие сформировать совокупность метамеров (хвоинок, листовых пластинок, шишек и т.п.).

Фактор, обеспечивающий возникновение дисперсии значений на каждом из уровней иерархического комплекса, принимается как один из организованных факторов дисперсионного анализа. На селекционно-семеноводческих объектах к числу организованных факторов могут быть отнесены: «различия в генотипах плюсовых деревьев» (различия между ортетами), «различия между клонами одного плюсового дерева» (различия между раметами), «различия между побегами в кроне одного учетного дерева». Разница в значениях признаков метамеров на одном учетном побеге вызывает возникновение так называемой случайной или остаточной дисперсии.

При работе с иерархическими многоуровневыми комплексами следует учитывать, что к так называемой остаточной дисперсии относят ту долю общей дисперсии, которая не связана с действием какого-либо организованного фактора. Принципиально общая дисперсия, фиксируемая в анализируемом комплексе, представлена двумя её категориями: факториальной и остаточной. К факториальной дисперсии относят те её компоненты, которые сформировались под влиянием организованных факторов. Последние рассматривают как факторы, действие каждого из которых может быть объяснено и учтено отдельно. В соответствии с этим на долю случайной дисперсии однофакторного комплекса будет приходиться та её часть, которая останется помимо влияний 1-го фактора, в двухфакторном – помимо влияния 2-х, в трехфакторном – помимо влияния 3-х. Иными словами, если один и тот же комплекс объектов последовательно подвергается дисперсионному анализу разного иерархического уровня, доля остаточной дисперсии не может оставаться неизменной и будет уменьшаться по мере увеличения числа организованных факторов. Это справедливо, если они детализируют сведения о так называемой «неучтенной дисперсии».

3. Особенности двухфакторного иерархического дисперсионного анализа

В иерархических комплексах вследствие наличия соподчиненности между действующими факторами, существует такая особенность: градации одного фактора не могут свободно комбинироваться по градациям другого фактора. Например, градации между филогенетическими или родственными группами растений имеют строгую последовательность и соподчиненность (семейство, род, вид, подвид, экотип, популяция, форма, особь). Сосна обыкновенная никак не может быть отнесена к роду ель, а род лиственница – к роду пихта и т.п.

В иерархических комплексах имеется ряд других особенностей. Так, например, для них вычисляют общую (s_y^2) и случайную (s_z^2) дисперсии, дисперсию по фактору **A**, (т.е. s_A^2), дисперсию по фактору **B** (т.е. s_B^2), но у них нет вычлененной дисперсии от совместного действия **A** и **B** (т.е. s^2_{AB}). Тем самым иерархические комплексы дают уменьшенную информацию о характере и структуре изменчивости, так как не учитывается дисперсия от совместного действия организованных факторов дисперсионного анализа. В некоторых публикациях и в электронных таблицах Excel вместо принятого нами символа s^2 применяют символ ms с соответствующими метками и индексами.

Для иерархических комплексов существует различие в определении достоверности с помощью критерия **F**. В частности установления $\mathbf{F_A}$ средний квадрат по фактору **A** (фактору первого – высшего – уровня) сопоставляют не со средним квадратом по случайным факторам **Z**, а со средним квадратом, вычисленным для фактора **B** (фактор второго – низшего – уровня):

$$F_A = \frac{s_A^2}{s_B^2} = \frac{ms_A}{ms_B}.$$

Критерий $\mathbf{F_B}$ для фактора **B** вычисляют традиционным способом, также как в обычном двухфакторном анализе: берут соотношение ms_B (или s_A^2), с ms_z (или s_z^2), т.е.

$$F_B = \frac{s_B^2}{s_Z^2} = \frac{ms_B}{ms_Z}.$$

Иначе вычисляют числа степеней свободы для фактора **B**, а именно: не по формуле $\nu_B = l_B - 1$, а по формуле $\nu_B = l_A - l_B$, где

- $\boldsymbol{l_B}$ – число градаций фактора **B**;
- $\boldsymbol{l_A}$ – число градаций фактора **A**.

Иерархические комплексы могут быть равномерными и неравномерными. У равномерных комплексов количество градаций в пределах каждого фактора низшего порядка одинаково (одинаковое количество особей в каждой из обследованных популяций, если фактором высшего порядка выступает принадлежность к той или иной популяции). При этом количество наблюдений и замеров в пределах каждой градации низшего фактора (в пределах особи) также берется одинаковым. Такие комплексы иногда называются ортогональными. Кроме того, среди иерархических комплексов различают двух-, трех- и многофакторные, т.е. их структура и особенности находятся в зависимости от сложности иерархического соподчинения между несколькими факторами.

С учетом особенностей структуры иерархических комплексов используют различные рабочие формулы для их обработки. При этом в основе всех схем и алгоритмов расчетов остается:

- вычисление сумм квадратов отклонений вариантов от своего среднего значения, (D_y, D_x, D_A, D_B, D_z для двухфакторных иерархических комплексов) – в ряде публикаций и в электронных таблицах Excel этот показатель обозначают символом ss (ss_y, ss_x, ss_A, ss_B, ss_z);

- затем вычисление среднеквадратических отклонений или средних квадратов отклонений (C_y, C_x, C_A, C_B, C_z или как их иногда обозначают σ^2_y, σ^2_A, σ^2_B, σ^2_z), – в ряде публикаций и в электронных таблицах Excel этот показатель обозначают символом ms от английского middle square (ms_y, ms_A, ms_B, ms_z);
- далее определение достоверности действия факторов на результативный признак с помощью критерия **F**, и как уже отмечалось:

$$F_A = \frac{ms_A}{ms_B};$$

$$F_B = \frac{ms_B}{ms_z}.$$

Если ставится цель определить только достоверность влияния фактора на результативный признак, то дисперсионный анализ комплекса заканчивается вычислением величины соответствующих F-критериев Фишера. Если же требуется найти долю разнообразия результативного признака, возникшего под влиянием воздействующих на комплекс факторов, то вычисляют величину соответствующего отношения факториальных дисперсий к общей дисперсии. При этом используют алгоритмы, предложенные С.А. Плохинским (1963) либо Дж. У. Снедекором (1961).

Конкретные схемы выполнения расчетов при реализации двухфакторного и трехфакторного иерархического дисперсионного анализа приведены ниже.

4. Порядок вычислений итоговых величин двухфакторного иерархического дисперсионного анализа

Порядок вычисления итоговых величин двухфакторного иерархического дисперсионного анализа рассмотрим на примере анализа двухфакторного иерархического неравномерного (не ортогонального) комплекса. Принципиально, алгоритм расчетов неравномерных комплексов применим и для равномерных комплексов, но не наоборот.

1. Пусть требуется оценить влияние исходных родительских форм на продуктивность межвидовых гибридов тополей, получаемых для создания промышленных плантаций. Напомним, что тополя относятся к двудомным растениям, т.е. их особи являются разнополыми. Учитываемым признаком будет объем ствола к установленному возрасту (20 лет). Планом гибридизации было предусмотрено использование в качестве отцовских особей 3-х перспективных форм тополя белого, а в качестве материнских – 8 особей тополя Болле. При этом схема скрещивания и количество оставшихся к моменту учета гибридов по каждой комбинации скрещивания будет отражена в таблице 1.

Таблица 1 - Двухфакторный иерархический дисперсионный анализ (Реконструкция по Г.Ф. Лакину, 1980)

№	Показатели	Отцы (a=3)								Сумма
		A1		A2			A3			
		Матери (b=8)								
		B1	B2	B3	B4	B5	B6	B7	B8	
1	2	3	4	5	6	7	8	9	10	11
1.	Запас древесины, $м^3$	1,0 0,8 0,6 0,8	0,9 0,7 0,8 0,7 0,5	1,0 1,1 0,9 1,0	1,2 1,0 1,0 0,9 1,0 1,1 0,8	0,9 0,9 1,0 0,8	1,1 1,2 1,0 0,9 1,0 0,8	1,0 1,1 1,1 0,8 0,9	0,9 0,9 0,8 1,1 0,6	
2.	n_B	4	5	4	7	4	6	5	5	40 Σn=N
3.	$\sum x_i$	3,2	3,6	4,0	7,0	3,6	6,0	4,9	4,3	36,6
4.	$\left(\sum x_i\right)^2$	10,24	12,96	16,0	49,0	12,96	36,0	24,01	18,49	1339,56
5.	$\frac{\left(\sum x_i\right)^2}{n}$	2,56	2,59	4,0	7,0	3,28	6,0	4,80	3,70	33,93
6.	$\sum x_i^2$	2,64	2,68	4,02	7,10	3,26	6,10	4,87	3,83	34,50
7.	n_A	9		15			16			40
8.	$\sum x_A$	6,8		14,6			15,2			36,6
9.	$\left(\sum x_A\right)^2$	46,24		213,16			231,04			-
10.	$\frac{\left(\sum x_A\right)^2}{n}$	5,14		14,21			14,44			33,79

2. Используя материалы первичных наблюдений, приведенные в таблице 1, вычисляем промежуточные вспомогательные величины.

2.1. Средний квадрат суммы вариант:

$$H = \frac{\left(\sum x_i\right)^2}{\sum n_B} = \frac{\left(\sum x_i\right)^2}{N} = \frac{36{,}6^2}{40} = 33{,}49.$$

3. Затем находим суммы квадратов отклонений.

3.1. Общая сумма квадратов отклонений:

$$D_y = \sum x_i^2 - H = 34{,}50 - 33{,}49 = 1{,}01.$$

3.2. Межгрупповая сумма квадратов отклонений:

$$D_x = \sum \frac{\left(\sum x_i\right)^2}{n_B} - H = 33{,}93 - 33{,}49 = 0{,}44\,.$$

3.3. Остаточная сумма квадратов отклонений:

$$D_z = D_y - D_x = 1{,}01 - 0{,}44 = 0{,}57\,.$$

3.4. Сумма квадратов отклонений по фактору А:

$$D_A = \sum \frac{\left(\sum x_A\right)^2}{n_A} - H = 33{,}79 - 33{,}49 = 0{,}30\,.$$

3.5. Сумма квадратов отклонений по фактору В:

$$D_B = \sum \frac{\left(\sum x_i\right)^2}{n_B} - \sum \frac{\left(\sum x_A\right)^2}{n_A} = 33{,}93 - 33{,}79 = 0{,}14\,.$$

4. Определяем числа степеней свободы.

4.1. Общее число степеней свободы (для общей дисперсии):

$$k_y = \sum b_B - 1 = \sum n_A - 1 = \sum n_i - 1 = N - 1 = 40 - 1 = 39\,.$$

4.2. Число степеней свободы по фактору А (для дисперсии по фактору А):

$$k_A = a - 1 = 3 - 1 = 2\,.$$

4.3. Число степеней свободы по фактору В (для дисперсии по фактору В):

$$k_B = b - a = 8 - 3 = 5\,.$$

4.4. Число степеней свободы по случайным факторам (для остаточной дисперсии):

$$k_z = N - b = 40 - 8 = 32\,.$$

4.5. Выполняем проверку правильности расчетов, исходя из представлений о равенстве числа общего степеней свободы сумме чисел степеней свободы всех факторов дисперсионного комплекса:

$$k_y = k_A + k_B + k_z = 2 + 5 + 32 = 39\,.$$

Равенство подтверждено, следовательно, вычисления выполнены корректно.

5. Находим средние квадраты отклонений по организованным факторам, используя полученные величины соответствующих сумм квадратов отклонений и числа степеней свободы.

5.1. По фактору А – фактору высшего уровня иерархии:

$$S_1^2 = \frac{D_A}{k_A} = \frac{0{,}30}{2} = 0{,}15$$

5.2. По фактору В – фактору низшего уровня иерархии:

$$S_2^2 = \frac{D_B}{k_B} = \frac{0{,}14}{5} = 0{,}028$$

5.3. По остаточной вариации – по случайным факторам – Z:

$$S_Z^2 = \frac{D_z}{k_z} = \frac{0{,}57}{32} = 0{,}018$$

5.4. Результат записываем в таблицу 2.

Таблица 2 - Результаты дисперсионного анализа гибридов тополей

Вариация	Степени свободы	Суммы квадратов	Средние квадраты	Дисперсионные отношения		
				$F_{факт}$	$F_{ст}$	
					5%	1%
1	2	3	4	5	6	7
По фактору А	2	0,30	0,150	$F_A = S_A^2 / S_B^2$=0,150/0,028 = 5,4	3,3	5,3
По фактору В	5	0,14	0,028	$F_A = S_B^2 / S_Z^2$=0,028/0,018 = 1,6	2,7	3,7
Остаточная	32	0,57	0,018	-	-	-
Общая	39	1,01	-	-	-	-

Статистически доказанным оказалось влияние фактора **А**. Действие фактора **В** статистически не достоверно, что будет учтено в дальнейших вычислениях.

6. Определяем усредненные значения степеней свободы для неравномерных комплексов, позволяющие учесть различия в числе степеней свободы по градациям организованных факторов. Необходимость такого преобразования вызвана тем, что в расчетах факториальных дисперсий при неодинаковой численности вариант в градациях комплекса используются усредненные значения *bn* и *n*.

6.1. Усредненное значение для расчета дисперсии по первому фактору – фактору **А**:

$$(bn)_0 = \frac{1}{a-1}\left(N - \frac{\sum n_A^2}{N}\right).$$

$$(bn)_0 = \frac{1}{3-1}\left(40 - \frac{9^2 + 15^2 + 16^2}{40}\right) = \frac{1}{2}\left(40 - \frac{526}{40}\right) = 12{,}975 \cong 13$$

6.2. Усредненное значение для расчета дисперсии по второму фактору – фактору **B**:

$$(n)_0 = \frac{(n_A)_0 + (n_B)_0}{2}$$

6.3. Усредненное значение числа наблюдений по фактору **A** (*a - 1* – адекватно числу степеней свободы по фактору **A**) – коррекция неравномерности градаций фактора высшей иерархи. Примечание: в учебнике Г.Ф. Лакина (1980, стр. 251) приведена некорректная запись алгоритма вычислений, которая в предложенном виде нами исправлена:

$$(n_A)_0 = \frac{1}{a-1} \times \left(\sum \frac{\sum n_{iA}^2}{n_A} - \frac{\sum n_i^2}{N} \right);$$

$$(n_A)_0 = \frac{1}{3-1} \times \left[\left(\frac{4^2+5^2}{9} + \frac{4^2+7^2+4^2}{15} + \frac{6^2+5^2+5^2}{16} \right) - \left(\frac{4^2+5^2+4^2+7^2+4^2+6^2+5^2+5^2}{40} \right) \right] =$$

$$= \frac{1}{2} \times \left(15{,}33 - \frac{208}{40} \right) = \frac{15{,}33 - 5{,}20}{2} = 5{,}065$$

6.4. Усредненное значение числа наблюдений по фактору **B** (*b – a* – адекватно числу степеней свободы по фактору **B**) – коррекция неравномерности градаций фактора низшей иерархи. Примечание: в учебнике Г.Ф. Лакина (1980, стр. 251) приведена некорректная запись алгоритма вычислений, которая нами исправлена:

$$(n_B)_0 = \frac{1}{b-a} \left(N - \sum \frac{\sum n_{iA}^2}{n_A} \right).$$

$$(n_B)_0 = \frac{1}{8-3} \left(40 - \frac{4^2+5^2}{9} + \frac{4^2+7^2+4^2}{15} + \frac{6^2+5^2+5^2}{16} \right) = \frac{1}{5} (40 - 15{,}33) = 8{,}000 - 3{,}066 = 4{,}934$$

6.5. После чего:

$$(n)_0 = \frac{(n_A)_0 + (n_B)_0}{2} = \frac{5{,}065 + 4{,}934}{2} = 4{,}9995 \cong 5 \ .$$

7. На следующем этапе рассчитываем факториальные и общую дисперсии по алгоритму Снедекора с учетом поправок на влияние остаточной дисперсии.

7.1. Дисперсия по фактору **A**:

$$s_A^2 = \frac{s_1^2 - s_2^2}{(bn)_0} = \frac{0{,}150 - 0{,}028}{13} = 0{,}010.$$

7.2. Дисперсия по фактору **B**:

$$s_B^2 = \frac{s_2^2 - s_z^2}{(n)_0} = \frac{0{,}028 - 0{,}018}{5} = 0{,}002.$$

7.3. Общая дисперсия:

$$s_y^2 = s_A^2 + s_B^2 + s_z^2 = 0{,}010 + 0{,}002 + 0{,}018 = 0{,}030.$$

8. Далее рассчитываем силу влияния факторов.

8.1. Сила влияния фактора **A** (отцовский эффект):

$$h_A^2 = \frac{s_A^2}{s_y^2} = \frac{0{,}010}{0{,}030} = 0{,}33333.$$

8.2. Сила влияния фактора **B**, включая совместное влияние **AB** (материнский эффект совместно с эффектом взаимодействия материнской и отцовской компоненты):

$$h_B^2 = \frac{s_B^2}{s_y^2} = \frac{0{,}002}{0{,}030} = 0{,}06667.$$

8.3. Сила влияния неорганизованных факторов

$$h_z^2 = \frac{s_z^2}{s_y^2} = \frac{0{,}0180}{0{,}030} = 0{,}60000.$$

8.4. Проверяем правильность расчета показателей силы влияния факторов, исходя из того, что сумма оценок доли влияния каждого из факторов должна быть равна «единице»:

$$h_y^2 = h_A^2 + h_B^2 + h_Z^2 = 0{,}33333 + 0{,}06667 + 0{,}60000 = 1{,}0 \; .$$

9. В данной ситуации более корректным является расчет по схеме, исключающей расчет силы влияния по фактору **B** и предусматривающей расчет общей «исправленной дисперсии» без учета величины или доли дисперсии по фактору **B**. Это обусловлено тем, что выполненная ранее оценка эффективности факторов иерархического дисперсионного комплекса не подтвердила достоверность влияния фактора **B**.

9.1. Рассчитываем общую «исправленную дисперсию» как сумму «исправленной дисперсии» по фактору **A** и остаточной дисперсии:

$$s_y^2 = s_A^2 + s_B^2 + s_z^2 = 0{,}010 + 0{,}018 = 0{,}028.$$

9.2. Сила влияния фактора **A** (отцовский эффект) рассчитывается как:

$$h_A^2 = \frac{s_A^2}{s_y^2} = \frac{0{,}010}{0{,}028} = 0{,}35714.$$

9.3. Сила влияния фактора **B**, включая и совместное влияние **AB** (материнский эффект совместно с эффектом взаимодействия отцовского и материнского факторов) не рассчитывается. Эффект от совместного влияния факторов **A** и **B** не учитывается на том основании, что один из компонентов совместного влияния, а именно фактор **B**, не имеет достоверного влияния.

9.4. Сила влияния неорганизованных факторов будет вычислена как:

$$h_z^2 = \frac{s_z^2}{s_y^2} = \frac{0{,}0180}{0{,}0280} = 0{,}64286.$$

9.5. Проверяем правильность расчета показателей силы влияния факторов, исходя из того, что сумма оценок доли влияния каждого из факторов должна быть равна «единице»:

$$h_y^2 = h_A^2 + h_Z^2 = 0{,}35714 + 0{,}64286 = 1{,}0\ .$$

Исходное условие удовлетворяется, следовательно, анализ выполнен верно.

Полученные оценки силы влияния факторов можно выразить в процентном отношении, для чего каждую из оценок умножают на 100%. В этом случае при интерпретации полученных результатов можно говорить о том, на сколько процентов изменчивость результирующего признака обусловлена влиянием каждого из учтенных факторов. Таким образом, удается разложить общую дисперсию комплекса на составляющие её компоненты и выяснить силу влияния каждого компонента на общую изменчивость результативного признака.

5. Трехфакторный иерархический дисперсионный анализ

Источниками дисперсии в трехфакторных комплексах могут выступать различия между ортетами, различия между раметами, различия между частями рамет (их побегами), на которых формируются элементарные совокупности метамеров. Различия между метамерами (хвоя, листья, побеги) создают (позволяют оценить) так называемую остаточную дисперсию. Именно метамеры в рассматриваемом примере выступают предметом учета, измерений (оценок в общем случае). Мы определяем параметры хвои вначале на разных побегах одного учетного дерева, затем на побегах других учетных деревьев, являющихся клонами одного плюсового дерева, затем – на других плюсовых деревьях. Однако во всех случаях происходит оценка параметров хвои. Подобных примеров можно привести много (шишки, их части, семена и пр.).

Распределение субъектов случайной дисперсии по уровням иерархического комплекса определяет дисперсию всех остальных уровней или дисперсию по всем остальным источникам в дисперсионном комплексе.

В селекционных исследованиях иерархическая организация и степень близости объектов в иерархической системе (место в системе) соответствует степени генетической близости объектов (степени эволюционной близости).

Расчет дисперсий в неравномерных дисперсионных иерархических комплексах носит черты общие для всех дисперсионных анализов. Они (общие черты) сводятся к определению дисперсии как отношения соответствующего квадрата отклонений к соответствующему ему числу степеней свободы (числу степеней свободы каждого организованного фактора, общему числу степеней свободы и числу степеней свободы для остаточной дисперсии). Вместе с тем имеются и свои особенности проведения анализа, которые состоят в особенностях организации вычисления его итоговых значений.

6. Порядок вычисления итоговых величин трехфакторного иерархического дисперсионного анализа

1. На первом этапе вычисляют значения

- общей суммы квадратов отклонений (D_y или ss_y),
- факториальной суммы квадратов отклонений (D_x или ss_x), в которую входят все организованные факторы, и
- суммы квадратов отклонений остаточной дисперсии (D_z или ss_z).

Этот этап работы проводят так же, как в любом однофакторном дисперсионном анализе.

2. Используя материалы первичных наблюдений, приведенные в таблице, вычисляем промежуточные вспомогательные величины.

2.1. Средний квадрат суммы вариант (**H**):

$$H = \frac{\left(\sum x_i\right)^2}{\sum n_i} = \frac{\left(\sum x_i\right)^2}{N}.$$

$$H = \frac{(26{,}96)^2}{538} = \frac{726{,}57}{538} = 1{,}3505056.$$

3. Затем находим суммы квадратов отклонений.

3.1. Общая сумма квадратов отклонений:

$$D_y = \sum x_i^2 - H.$$

$$D_y = 1{,}474 - 1{,}351 = 0{,}123.$$

3.2. Межгрупповая сумма квадратов отклонений:

$$D_x = \sum \frac{\left(\sum x_i\right)^2}{n_i} - H.$$

$$D_x = 1{,}448 - 1{,}351 = 0{,}097.$$

3.3. Остаточная сумма квадратов отклонений:

$D_z = D_y - D_x$.

$D_z = 0{,}123 - 0{,}097 = 0{,}026$.

На следующем этапе находят значения сумм квадратов отклонений для каждого организованного фактора отдельно.

3.4. Сумма квадратов отклонений по фактору высшей иерархии **A** (здесь используют сумму средних квадратов значений для <u>*именно*</u> фактора высшей иерархии **A** и средний квадрат общей суммы вариант **H**):

$$D_A = \sum \frac{\left(\sum x_{iA}\right)^2}{n_A} - H.$$

$D_A = 1{,}356 - 1{,}351 = 0{,}006$ (с учетом всех округлений).

3.5. Сумма квадратов отклонений по фактору промежуточной иерархии **B** (здесь используют сумму средних квадратов значений для фактора промежуточной иерархии **B** и сумму средних квадратов значений для фактора высшей иерархии **A**):

$$D_B = \sum \frac{\left(\sum x_{iB}\right)^2}{n_B} - \sum \frac{\left(\sum x_{iA}\right)^2}{n_A}.$$

$D_B = 1{,}412 - 1{,}356 = 0{,}056$.

3.6. Сумма квадратов отклонений по фактору низшей иерархии **C** (здесь используют сумму средних квадратов значений для фактора низшей иерархии **C** и сумму средних квадратов значений для фактора промежуточной иерархии **B**):

$$D_C = \sum \frac{\left(\sum x_{iC}\right)^2}{n_C} - \sum \frac{\left(\sum x_{iB}\right)^2}{n_B}.$$

$D_C = 1{,}448 - 1{,}412 = 0{,}036$.

4. Определяем числа степеней свободы. При этом исходим из того, что в нашем примере в формировании дисперсионного комплекса участвовало вегетативное потомство 3-х плюсовых деревьев (ортетов), следовательно, число градаций фактора **A** (фактор высшей иерархии) равно 3 (a=3). Каждое плюсовое дерево было представлено некоторым количеством своих клонов (рамет): в нашем примере по 3. Тогда общее количество градаций фактора **B** (фактор второго или промежуточного уровня иерархии) составит 9 (b=9). С каждого из 9-ти клонов было заготовлено по 3 учетных побега, которые представляли фактор **C** (фактор низшего уровня иерархии). Число градаций этого фактора составило 27 (c=27). Каждый учетный побег служил местом

заготовки первичных единиц выборки (отдельных пучков хвои), различия в параметрах которых формирует случайную или остаточную дисперсию. По условия нашего примера число заготовленных с каждого учетного побега хвоинок оказалось неодинаковым и в сумме составило 538 – это общая численность дисперсионного комплекса. Принципиально числа степеней свободы вычисляются в соответствии с общепринятыми представлениями.

4.1. Общее число степеней свободы (для общей дисперсии):

$k_y = \sum n_C - 1 = \sum n_B - 1 = \sum n_A - 1 = \sum n_i - 1 = N - 1$.

$k_y = 538 - 1 = 537$.

4.2. Число степеней свободы по фактору **A** (для дисперсии по фактору **A**):

$k_A = a - 1$.

$k_A = 3 - 1 = 2$.

4.3. Число степеней свободы по фактору **B** (для дисперсии по фактору **B**):

$k_B = b - a$.

$k_B = 9 - 3 = 6$.

4.4. Число степеней свободы по фактору **C** (для дисперсии по фактору **C**):

$k_C = c - b$.

$k_C = 27 - 9 = 18$.

4.5. Число степеней свободы по случайным факторам (для остаточной дисперсии), оно вычисляется как разность значений общей численности дисперсионного комплекса (общее число наблюдений) и числа градаций фактора низшей иерархии (в нашем примере число учетных побегов):

$k_z = N - c$.

$k_z = 538 - 27 = 511$.

5. На основе установленных выше значений находим средние квадраты отклонений по факторам. Для этого используем полученные величины сумм квадратов отклонений по каждому действующему фактору и соответствующие им числа степеней свободы.

5.1. По фактору **A** – фактору высшего уровня иерархии (на данном этапе обозначим её $S^2{}_1$):

$$S_1^2 = \frac{D_A}{k_A}.$$

$$S_1^2 = \frac{0{,}006}{2} = 0{,}003.$$

5.2. По фактору **B** – фактору промежуточного уровня иерархии (на данном этапе обозначим её $S^2{}_2$):

$$S_2^2 = \frac{D_B}{k_B}.$$

$$S_2^2 = \frac{0{,}056}{6} = 0{,}009.$$

5.3. По фактору **C** – фактору низшего уровня иерархии (на данном этапе обозначим её $S^2{}_3$):

$$S_3^2 = \frac{D_C}{k_C}.$$

$$S_3^2 = \frac{0{,}036}{18} = 0{,}002.$$

5.3. По остаточной вариации – для дисперсии по случайным факторам – **Z**:

$$S_Z^2 = \frac{D_z}{k_z}$$

$$S_Z^2 = \frac{0{,}026}{511} = 0{,}000051$$

5.4. Результат записываем в таблицу (табл. 3).

Таблица 3 - Результаты дисперсионного анализа массы хвои плюсовых деревьев сосны обыкновенной в составе лесосеменной плантации № 24 ГБУ НО «Семеновский спецсемлесхоз» Нижегородской области

Вариация	Степени свободы (k)	Суммы квадратов (D)	Средние квадраты (S)	Дисперсионные отношения		
				$F_{факт}$	$F_{ст}$	
					5%	1%
1	2	3	4	5	6	7
По фактору А	k_A=2	D_A=0,006	$S^2{}_1$=0,003	$F_A = S^2{}_1/S^2{}_2$ = 0,315	1,57	1,89
По фактору В	k_B=6	D_B=0,056	$S^2{}_2$=0,009	$F_B = S^2{}_2/S^2{}_3$ = 4,719	1,32	1,48
По фактору С	k_C=18	D_C=0,036	$S^2{}_3$=0,002	$F_C = S^2{}_3/S^2{}_Z$ = 38,888	1,24	1,36
Остаточная (Z)	k_z=511	D_Z=0,026	$S^2{}_z$=0,00005	-	-	-
Общая (Y)	k_y=538	D_y=0,123	-	-	-	-

Статистически доказанным оказалось влияние факторов **B** и **C**; влияние фактора **A** статистически недостоверно.

Для получения исправленных значений дисперсий, используемых в расчетах по методу Снедекора, в неравномерных комплексах необходимо вычислить усредненные значения численностей по каждому организованному фактору.

6. На следующем этапе переходим к расчету усредненных численностей иерархических групп дисперсионного комплекса. Мы с вами уже знаем, что усредненные численности групп используются в расчетах исправленных значений дисперсий организованных факторов при вычислении коэффициента наследуемости (эффективности действия фактора) по методу Снедекора в неравномерных дисперсионных комплексах. В приведенном ниже примере использованы значения, предоставляемые студентам в учебном расчетном модуле электронных таблиц Excel, файл: «Хвоя-3-факторный».

В трехфакторном анализе учитывают величины дисперсий, вызванных отдельным действием каждого из факторов, имеющих иерархическую соподчиненность. Эффект взаимодействия факторов в иерархических комплексах, в том числе и в трехфакторных, не вычисляют. В соответствии с этим находят:

- дисперсию по фактору **A** – различия между плюсовыми деревьями (ортетами);

- дисперсию по фактору **B** – различия между учетными деревьями одного плюсового дерева (раметами);

- дисперсию по фактору **C** – различия между учетными побегами одного учетного дерева.

В вычислениях используют усредненные численности групп соответствующих иерархий (иерархических уровней):

- усредненное число наблюдений в пределах одного плюсового дерева (численность всех учетов на всех раметах одного плюсового дерева) для расчета по фактору **A**;

- усредненное число наблюдений в пределах одного учетного дерева (численность всех учетов на всех побегах одного учетного дерева) для расчета по фактору **B**;

- усредненное число наблюдений в пределах одного учетного побега (численность всех учетов на одном учетном побеге) для расчета по фактору **C**.

6.1. Усредненное значение численности градаций фактора высшей иерархии (градации фактора **A**) неравномерного иерархического комплекса для расчета дисперсии по первому фактору – фактору **A** (усредненное значение численности одной градации высшего – первого – порядка):

$$(bn)_0 = \frac{1}{a-1} \times \left(N - \frac{\sum n_A^2}{N} \right) = \frac{1}{k_a} \times \left(N - \frac{\sum n_A^2}{N} \right).$$

$$(bn)_0 = \frac{1}{3-1}\left(538 - \frac{180^2 + 180^2 + 178^2}{538}\right) = \frac{1}{2}\left(538 - \frac{96484}{538}\right) = 179{,}3383 \cong 179$$

6.2. Усредненное значение численности градаций фактора промежуточной иерархии (градации фактора **В**) неравномерного иерархического комплекса для расчета дисперсии по второму фактору – фактору **В** (усредненное значение численности одной градации промежуточного – второго – порядка):

$$(n)_0 = \frac{(n_A)_0 + (n_B)_0}{2}, \text{ где}$$

6.3. Усредненное значение числа наблюдений в градациях фактора **В** по градациям фактора **А** – коррекция неравномерности градаций фактора высшей иерархи (*a - 1* – адекватно числу степеней свободы по фактору **А**). Примечание: в учебнике Г.Ф. Лакина (1980, стр. 251) приведена некорректная запись алгоритма вычислений. Представленная в настоящем виде формула корректна (Бессчетнов В.П.):

$$(n_A)_0 = \frac{1}{a-1} \times \left(\sum \frac{\sum nb_{iA}^2}{n_A} - \frac{\sum nb_i^2}{N} \right) = \frac{1}{k_a} \times \left(\sum \frac{\sum nb_{iA}^2}{n_A} - \frac{\sum nb_i^2}{N} \right);$$

Усредненная численность группы (комплекса столбцов, принадлежащих одному учетному дереву – усредненная численность одной градации фактора **В**) с поправкой на неравномерность численностей групп по плюсовым деревьям – учитывает как неодинаковое число учетных деревьев в каждом плюсовом дереве (в составе клонов одного плюсового дерева), так и неодинаковую численность учетов (дат) в каждом частном столбце, относящемся к одному плюсовому дереву.

В представленной выше формуле приняты следующие обозначения:

- nb – численность одной градации фактора В (численность учетов в одной рамете);

- Σnb^2_{iA} – сумма квадратов численностей одной градации фактора В (численность учетов в одной рамете) в пределах одной градации фактора А;

- Σnb^2_i – общая сума квадратов численностей градаций фактора В по всему дисперсионному комплексу;

- n_A – численность соответствующей градации фактора А (число учетов в отдельной градации фактора А – в одном ортете);

- N – общая численность дисперсионного комплекса;

- k_a – число степеней свободы фактора А.

$$(n_A)_0 = \frac{1}{3-1} \times \left[\left(\frac{60^2+60^2+60^2}{180} + \frac{60^2+60^2+60^2}{180} + \frac{58^2+60^2+60^2}{178} \right) - \right.$$

$$\left. - \left(\frac{60^2+60^2+60^2+60^2+60^2+60^2+58^2+60^2+60^2}{538} \right) \right] =$$

$$= \frac{1}{2} \times \left(179{,}3483 - \frac{32164}{538} \right) = \frac{179{,}3483 - 59{,}7844}{2} = 59{,}7819639$$

6.4. Усредненное значение числа наблюдений в градациях по фактору **В** (b – a – адекватно числу степеней свободы по фактору **В**) – коррекция неравномерности градаций фактора промежуточной иерархии. Примечание: в учебнике Г.Ф. Лакина (1980, стр. 251) приведена некорректная запись алгоритма вычислений. Представленная в настоящем виде формула корректна (Бессчетнов В.П.):

$$(n_B)_0 = \frac{1}{b-a} \times \left(N - \sum \frac{\sum n_{iA}^2}{n_A} \right) = \frac{1}{k_b} \times \left(N - \sum \frac{\sum n_{iA}^2}{n_A} \right).$$

Усредненная численность группы (комплекса столбцов, принадлежащих одному учетному дереву – усредненная численность одной градации фактора **В**) с поправкой на неравномерность численностей групп по учетным деревьям – учитывает как неодинаковое число учетных деревьев в каждом плюсовом дереве (в составе клонов одного плюсового дерева), так и неодинаковую численность учетов (дат) в каждом частном столбце, относящемся к одному плюсовому дереву.

$$(n_B)_0 = \frac{1}{9-3}\left(538 - \frac{60^2+60^2+60^2}{180} + \frac{60^2+60^2+60^2}{180} + \frac{58^2+60^2+60^2}{178} \right) =$$

$$\frac{1}{6}(538 - 179{,}3448) = 59{,}77528$$

6.5. После чего вычисляется общая усредненная численность отдельной градации фактора **В** (усредненное общее число учетов по всей группе столбцов, принадлежащих одному учетному дереву) с поправкой на неравномерность численностей групп по плюсовым деревьям.

$$(n)_0 = \frac{(n_A)_0 + (n_B)_0}{2}.$$

В представленной выше формуле приняты следующие обозначения:

- $(n)_0$ – общая усредненная численность одной полной градации фактора промежуточной иерархии **В** (среднее число учетов – дат – в одной градации фактора **В**) для всего дисперсионного комплекса с учетом неравномерности общих численностей градаций фактора **А**.

- $(n_A)_0$ – усредненная общая численность группы (комплекса столбцов, принадлежащих одному учетному дереву) с поправкой на неравномерность численностей групп по плюсовым деревьям – по фактору **А**;

- $(n_B)_0$ – усредненная численность группы (комплекса столбцов, принадлежащих одному учетному дереву) с поправкой на неравномерность численностей групп по учетным деревьям – по фактору **В**;

$$(n)_0 = \frac{59{,}781964 + 59{,}77528}{2} = 59{,}77862245 \cong 60$$

6.6. Усредненное значение численности градаций фактора низшей иерархии (градации фактора **С**) неравномерного иерархического комплекса для расчета дисперсии по третьему фактору – фактору **С** (усредненное значение численности одной градации низшего – третьего – порядка):

$$(n)_n = \frac{(n_A)_n + (n_B)_n + (n_C)_n}{3}, \text{ где}$$

6.7. Усредненное значение числа наблюдений в градациях фактора **С** с поправкой на неравномерность фактора **А**

$$(n_A)_n = \frac{1}{a-1} \times \left(\sum \frac{\sum nc_{iA}^2}{n_{Ai}} - \frac{\sum na_i^2}{N} \right) = \frac{1}{k_a} \times \left(\sum \frac{\sum nc_{iA}^2}{n_{Ai}} - \frac{\sum na_i^2}{N} \right), \text{ где:}$$

В представленной выше формуле приняты следующие обозначения:

- nc_{iA} – численность одной градации фактора низшей иерархии **С** в пределах одной градации фактора высшей иерархии **А** (численность частных учетов в одном ортете);
- Σnc^2_{iA} – сумма квадратов численностей столбцов в одной градации фактора **А** (численность учетов в каждой из градаций фактора низшей иерархии **С** в одном ортете) в пределах одной градации фактора **А**;
- Σna^2_i – общая сума квадратов численностей градаций фактора **А** по всему дисперсионному комплексу;
- n_{Ai} – численность соответствующей градации фактора **А** (число учетов в отдельной градации фактора А – в одном конкретном ортете);
- N – общая численность дисперсионного комплекса;
- a – число градаций фактора А;
- k_a – число степеней свободы фактора А.

$$(n_A)_n = \frac{1}{3-1} \times \left(\frac{20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2}{180} + \right.$$

$$+ \frac{20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2}{180} +$$

$$+ \frac{20^2+20^2+20^2+20^2+20^2+20^2+19^2+20^2+19^2}{178} -$$

$$- \frac{20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2}{538} -$$

$$- \frac{20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2}{538} -$$

$$\left. - \frac{20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2}{538} \right) = 19{,}9285744\!\!\!1$$

6.8. Усредненное значение числа наблюдений в градациях фактора **С** с поправкой на неравномерность фактора **В**

$$(n_B)_n = \frac{1}{(b-a)+(c-b)} \times \left(N - \sum \frac{\sum nc_{iA}^2}{n_{Ai}} \right) = \frac{1}{k_b + k_c} \times \left(N - \sum \frac{\sum nc_{iA}^2}{n_{Ai}} \right),$$ где:

В представленной выше формуле приняты следующие обозначения:

- nc_{iA} – численность одной градации фактора низшей иерархии **С** в пределах одной градации фактора высшей иерархии **А** (численность частных учетов в одном ортете);

- Σnc^2_{iA} – сумма квадратов численностей столбцов в одной градации фактора **А** (численность учетов в каждой из градаций фактора низшей иерархии **С** в одном ортете) в пределах одной градации фактора **А**;

- $\Sigma nca^2_i\ /n_{Ai}$ – усредненное значение численностей градации фактора низшей иерархии **С** в пределах одной градации фактора **А**;

- n_{Ai} – численность соответствующей градации фактора **А** (число учетов в отдельной градации фактора **А** – в одном конкретном ортете);

- N – общая численность дисперсионного комплекса;

- a – число градаций фактора **А**;

- b – число градаций фактора **В**;

- c – число градаций фактора **С**;

- k_b – число степеней свободы фактора **В**;

- k_c – число степеней свободы фактора **С**.

$$(n_B)_n = \frac{1}{(9-3)+(27-9)} \times \left(538 - \sum \frac{\sum nc_{iA}^2}{n_{Ai}} \right) = \frac{1}{6+18} \times \left[538 - \right.$$
$$- \left(\frac{20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2}{180} + \right.$$
$$+ \frac{20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2+20^2}{180} +$$
$$\left. \left. + \frac{20^2+20^2+20^2+20^2+20^2+20^2+19^2+20^2+19^2}{178} \right) \right] = 19{,}9255618$$

6.9. Усредненное значение числа наблюдений в градациях фактора **С** с поправкой на неравномерность градаций фактора **С**

$$(n_C)_n = \frac{1}{(c-b)} \times \left(N - \sum \frac{\sum nc_{iB}^2}{n_{Bi}} \right) = \frac{1}{k_c} \times \left(N - \sum \frac{\sum nc_{iB}^2}{n_{Bi}} \right),$$ где:

В представленной выше формуле приняты следующие обозначения:

- nc_{iB} – численность одной градации фактора низшей иерархии **С** в пределах одной градации фактора промежуточной иерархии **В** (численность частных учетов в одной рамете);

- Σnc^2_{iB} – сумма квадратов численностей столбцов в одной градации фактора **В** (численность учетов в каждой из градаций фактора низшей иерархии **С** в одной рамете) в пределах одной градации фактора **В**;

- $\Sigma nc^2_{iB} / n_{Bi}$ – усредненное значение численностей градации фактора низшей иерархии **С** в пределах одной градации фактора **В**;

- n_{Bi} – численность соответствующей градации фактора **В** (число учетов в отдельной градации фактора **В** – в одной конкретной рамете);

- N – общая численность дисперсионного комплекса;

- b – число градаций фактора **В**;

- c – число градаций фактора **С**;

- k_c – число степеней свободы фактора **С**.

$$(n_C)_n = \frac{1}{(27-9)} \times \left(538 - \sum \frac{\sum nc^2_{iB}}{n_{Bi}} \right) = \frac{1}{18} \times \Bigg[538 - \left(\frac{20^2+20^2+20^2}{60} + \frac{20^2+20^2+20^2}{60} + \right.$$
$$+ \frac{20^2+20^2+20^2}{60} + \frac{20^2+20^2+20^2}{60} + \frac{20^2+20^2+20^2}{60} + \frac{20^2+20^2+20^2}{60} +$$
$$\left. + \frac{19^2+20^2+19^2}{60} + \frac{20^2+20^2+20^2}{60} + \frac{20^2+20^2+20^2}{60} \right) \Bigg] = 19{,}92528736$$

6.10. После чего вычисляют усредненную численность с учетом поправки на неравномерность численностей всех категорий градаций трехфакторного дисперсионного комплекса.

$$(n)_n = \frac{(n_A)_n + (n_B)_n + (n_C)_n}{3},$$

где:

- $(n)_n$ – общая усредненная численность одной градации фактора низшей иерархии **С** для всего дисперсионного комплекса с учетом неравномерностей всех организованных факторов.

- $(n_A)_n$ – усредненная численность одной градации фактора низшей иерархии **С** с учетом поправки на неравномерность градаций фактора высшей иерархии **А**;

- $(n_B)_n$ – усредненная численность одной градации фактора низшей иерархии **С** с учетом поправки на неравномерность градаций фактора промежуточной иерархии **В**;

- $(n_C)_n$ – усредненная численность одной градации фактора низшей иерархии **С** с учетом поправки на неравномерность численности градаций фактора низшей иерархии **С**.

$$(n)_n = \frac{19{,}92857 + 19{,}92556 + 19{,}92529}{3} = 19{,}92647 \cong 20.$$

7. На следующем этапе рассчитываем факториальные и общую дисперсии. Используем алгоритм Снедекора, учитывающий поправку на влияние остаточной дисперсии.

7.1. Дисперсия по фактору **А**:

$$\hat{s}_A^2 = \frac{s_A^2 - s_B^2}{(bn)_0}.$$

$$\hat{s}_A^2 = \frac{0{,}003 - 0{,}009}{179} = -0{,}00004.$$

Возникновение отрицательного значения оценок дисперсии вызвано, тем, что в рассматриваемом примере дисперсия по фактору **А** не получила подтверждения статистической достоверности (см. табл. 1) и оказалась меньше дисперсии по фактору **В**. В подобных случаях значения дисперсии не вычисляют, а её долю в общей дисперсии признака не учитывают.

7.2. Дисперсия по фактору **В**:

$$\hat{s}_B^2 = \frac{s_B^2 - s_C^2}{(n)_0}.$$

$$\hat{s}_B^2 = \frac{0{,}009 - 0{,}002}{60} = 0{,}000123$$

7.3. Дисперсия по фактору **С**:

$$\hat{s}_C^2 = \frac{s_C^2 - s_z^2}{(n)_n}.$$

$$\hat{s}_C^2 = \frac{0{,}002 - 0{,}000051}{20} = 0{,}000097.$$

7.4. Общая дисперсия в трехфакторных иерархических дисперсионных комплексах вычисляется по формуле:

$$\hat{s}_y^2 = \hat{s}_A^2 + \hat{s}_B^2 + \hat{s}_B^2 + s_z^2.$$

Однако, поскольку в нашем примере действие фактора **А** оказалось статистически недостоверным, долю соответствующей ему дисперсии не учитывают в общем расчете. Тогда вышеприведенная формула примет вид:

$$\hat{s}_y^2 = \hat{s}_B^2 + \hat{s}_B^2 + s_z^2.$$

$$\hat{s}_y^2 = 0{,}000123 + 0{,}000097 + 0{,}000051 = 0{,}000271.$$

8. Далее рассчитываем силу влияния факторов.

8.1. Сила влияния фактора **A** (эффект влияния ортетов):

$$h_A^2 = \frac{s_A^2}{s_y^2}.$$

Для нашего приме эта величина не рассчитывается, поскольку действие фактора **A** статистически недостоверно, или принимается равной 0.

8.2. Сила влияния фактора **B**, включая совместное влияние **AB** (эффект влияния рамет совместно с эффектом взаимодействия компонентов дисперсии ортетов и рамет):

$$h_B^2 = \frac{s_B^2}{s_y^2}.$$

$$h_B^2 = \frac{0{,}000123}{0{,}000271} = 0{,}4539.$$

8.3. Сила влияния фактора **C**, включая совместное влияние **BC** (эффект влияния рамет совместно с эффектом взаимодействия компонентов дисперсии рамет и учетных побегов):

$$h_C^2 = \frac{s_C^2}{s_y^2}.$$

$$h_C^2 = \frac{0{,}000097}{0{,}000271} = 0{,}3579.$$

8.4. Сила влияния неорганизованных факторов:

$$h_z^2 = \frac{s_z^2}{s_y^2}.$$

$$h_z^2 = \frac{0{,}000051}{0{,}000271} = 0{,}1882.$$

8.5. Проверяем правильность расчета показателей силы влияния факторов, исходя из того, что сумма оценок доли влияния каждого из факторов должна быть равна «единице»:

$$h_y^2 = h_A^2 + h_B^2 + h_C^2 + h_Z^2 = 1{,}0$$

$$h_y^2 = 0 + 0{,}4539 + 0{,}3579 + 0{,}1882 = 1{,}0.$$

Используя представленный выше алгоритм вычислений, студенты самостоятельно находят значения всех итоговых величин трехфакторного иерархического дисперсионного анализа. Расчеты ведут в предложенном преподавателем исходном файле электронных таблиц Excel: «Хвоя-3-факторный».

Библиографический список

1. Гатаулин А.М. Система прикладных статистико-математических методов обработки экспериментальных данных в сельском хозяйстве. Ч. 2 / А.М. Гатаулин. – М.: Изд-во МСХА, 1992. – 192 с.
2. Доспехов Б.А. Методика полевого опыта / Б.А. Доспехов. - М.: Колос, 1985. – 416 с.
3. Зайцев Г.Н. Математическая статистика в экспериментальной ботанике / Г.Н. Зайцев. - М.: Наука, 1984. – 424 с.
4. Лакин Г.Ф. Биометрия. Учебное пособие для биологич. спец. Вузов / Г.Ф. Лакин – 3-е изд., перераб. и доп. – М.: Высшая школа, 1980. – 293с.
5. Моисейченко В.Ф. Основы научных исследований в плодоводстве, овощеводстве и виноградарстве / В.Ф. Моисейченко, А.Х. Заверюха, М.Ф.Трифонова. – М.: Колос, 1994. – 383 с.
6. Меркурьева Е.К. Биометрия в селекции и генетике сельскохозяйственных животных. Учебное пособие для зоотехнических вузов и факультетов /Е.К. Меркурьева. - М.: Колос, 1970. – 424 с.
7. Никитин К.Е. Методы и техника обработки лесоводственной информации / К.Е. Никитин, А.З. Швиденко. – М.: Лесная промышленность, 1978. – 272 с.
8. Плохинский, Н. А. Биометрия / Н.А. Плохинский. – Новосибирск: Наука. Сиб. отд., 1961. – 364 с.
9. Плохинский, Н. А. Наследуемость / Н.А. Плохинский. – Новосибирск: Наука. – Сиб. отд-ние, 1964. – 195с.
10. Плохинский, Н.А. Алгоритмы биометрии / Н.А. Плохинский. – М.: Издательство МГУ, 1967. – 82 с.
11. Рокицкий, П.Ф. Введение в статистическую генетику / П.Ф. Рокицкий. – Минск: Высшая школа, 1978. – 448 с.
12. Свалов, Н.Н. Вариационная статистика. Учебное пособие для вузов / Н.Н. Свалов. – М.: Лесная промышленность, 1977, - 176 с.
13. Снедекор, Дж. У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. Пер. с англ. / Дж. У. Снедекор. – М.: Сельхозиздат, 1961. – 502 с.

ДИСПЕРСИОННЫЙ АНАЛИЗ МНОГОУРОВНЕВЫХ ИЕРАРХИЧЕСКИХ КОМПЛЕКСОВ

Методические указания для студентов и аспирантов очного и заочного отделений факультета лесного хозяйства по специальности 25.02.01.65 – Лесное хозяйство

Составители: Бессчетнов Владимир Петрович
Бессчетнова Наталья Николаевна
Храмова Ольга Юрьевна
Орнатский Александр Николаевич
Горелов Николай Иванович

Редактор О.Ф. Костина

Подписано в печать 14.05.2012 г.
Формат 60x84 1/16. Печать офсетная
Печ. л. 2,1. Тираж 100 зкз. Заказ

Нижегородская государственная сельскохозяйственная академия
603107, г. Нижний Новгород, проспект Гагарина, 97

Типография НГСХА